UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Common independent components of the P3b, N400, and P600 ERP components to deviant linguistic events

Permalink

https://escholarship.org/uc/item/4fn7c2kc

Author

Groppe, David Martin

Publication Date

2007

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Common Independent Components of the P3b, N400, and P600 ERP Components to Deviant Linguistic Events

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Cognitive Science

by

David Martin Groppe

Committee in charge:

Marta Kutas, Chair
Anders M. Dale
Virginia R. de Sa
Steven A. Hillyard
Scott Makeig
Martin I. Sereno

2007

The dissertation of David Martin Groppe is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2007

DEDICATION

To my mother, Rose Marie Groppe, who taught me how to solve for x, and

to my father, John Daniel Groppe, who taught me how to write.

TABLE OF CONTENTS

Signature Page....iii

Dedication....iv

Table of Contents....v

List of Tables....x

List of Figures....xii

Acknowledgements....xxiii

Curriculum Vitae....xxviii

Abstract of the Dissertation....xxxi

Chapter I: Introduction
I.A: Comprehending the Predictable and the Unexpected....1
I.B: Better ERP Componentry through Independent Component Analysis....5
I.C: Thesis Goals and Organization....6
I.D: References....10

Chapter II: Three Components of the Unexpected
II.A: Introduction....18
II.A.1: Components and Subcomponents Defined....19
II.B: Features and Functional Correlates of the N400....21
II.C: Features and Functional Correlates of the P600....25
II.D: Features and Functional Correlates of the P3b....29
II.E: Evidence for N400, P600, and P3b Subcomponents....35
II.E.1: Types of Evidence....35
II.E.2: N400 Subcomponents....39
II.E.2.a: Evidence for N400 Subcomponents....39
II.E.2.b: Possible N400 Subcomponents....43
II.E.3: P600 Subcomponents....44
II.E.3.a: Evidence for P600 Subcomponents....44
II.E.3.b: Possible P600 Subcomponents....48
II.E.4: P3b Subcomponents....49
II.E.4.a: Evidence for and Identify of Possible P3b Subcomponents...49
II.F: Subcomponentry and Problems with Conventional ERP Analysis....53
II.G: A Contribution of the P3b to the N400 and P600....56
II.G.1: Evidence for a Contribution of the P3b to the N400....56

II.G.2: Evidence for a Contribution of the P3b to the P600.....61
II.G.3: On the Significance of Extracting a P3b Contribution to the N400 and P600.....63
II.G.4: On the Likelihood of Extracting a P3b Contribution to the N400 and P600.....68
II.H: References.....71

Chapter III: Independent Component Analysis for ERP Subcomponentry
III.A: Introduction.....95
III.B: Sources of Scalp Potentials.....96
III.B.1: Neural Sources.....96
III.B.2: EEG Artifacts.....99
III.C: Linear Spatial Filters.....100
III.C.1: LSF Limitations and Difficulties.....102
III.D: Independent Component Analysis for Deriving LSFs.....105
III.D.1: An Example of ICA.....107
III.D.2: ICA Ambiguities.....111
III.D.2.a: ICA Scaling Ambiguity.....111
III.D.2.b: ICA Component Order Ambiguity.....113
III.D.3: Physiological Motivation for ICA.....113
III.D.4: Past ICA Findings.....116
III.D.5: ICA Caveats.....119
III.D.5.a: Component Reliability.....119
III.D.5.b: Component Accuracy.....120
III.E: PCA for Deriving LSFs.....124
III.F: Source Localization for Deriving LSFs.....127
III.G: Temporal PCA for Defining ERP Subcomponents.....132
III.H: Summary.....133
III.I: References.....161

Chapter IV: Electrode Level Analysis of the P3b, N400, and P600
IV.A: Introduction.....169
IV.B: Methods: Oddball Tasks (Experiment 1) 172
IV.B.1: Overview and Logic of Experimental Design: Oddball Tasks.....172
IV.B.2: Participants Oddball Tasks.....173
IV.B.3: Stimuli: Oddball Tasks.....173
IV.B.4: Procedure: Oddball Tasks.....174
IV.B.5: EEG Recording Parameters: Oddball Tasks.....176
IV.C: Methods: Sentence Acceptability (Experiment 2) 177
IV.C.1: Overview and Logic of Experimental Design: Sentence Acceptability.....177
IV.C.2: Participants: Sentence Acceptability.....178
IV.C.3: Stimuli: Sentence Acceptability.....178
IV.C.3.a: Cloze Probability.....180

IV.C.3.b: Word Length & Frequency....182
IV.C.3.c: Lexical Association....183
IV.C.3.d: Concreteness....184
IV.C.3.e: Repetition....185
IV.C.3.f: Sentence Stem Constraint/Entropy....186
IV.C.3.g: Summary: Sentence Stimuli Properties....188
IV.C.4: Procedure & EEG Recording Parameters: Sentence Acceptability...189
IV.D: EEG Data Preprocessing: Both Experiments....191
IV.E: Results: Oddball Tasks (Experiment 1) 193
IV.E.1: Behavioral Accuracy: Oddball Tasks....193
IV.E.2: Event-Related Potentials: Oddball Tasks....194
IV.E.3: ERP Analyses: Oddball Tasks....197
IV.E.3.a: P3b Effect Peak Latency....197
IV.E.3.b: Mean Amplitude Analysis....197
IV.E.3.c: Mean Amplitude Analysis: Sequence Effects....199
IV.E.3.d: Anatomical Complexity....200
IV.E.3.e: P3b Effect Topographies....202
IV.E.4: Single Trial Analysis: Oddball Tasks....204
IV.E.5: Results Summary: Oddball Tasks....205
IV.F: Results: Sentence Acceptability (Experiment 2) 207
IV.F.1: Analyses of Behavior: Sentence Acceptability....207
IV.F.1.a: Behavioral Accuracy....207
IV.F.1.b: Reaction Time....209
IV.F.2: Event-Related Potentials: Sentence Acceptability....210
IV.F.3: ERP Analyses: Sentence Acceptability....214
IV.F.3.a: N400 and P600 Effect Peak Latency....214
IV.F.3.b: Mean Amplitude Analysis: Congruity Effects....214
IV.F.3.c: Mean Amplitude Analysis: Grammaticality Effects....216
IV.F.3.d: Mean Amplitude Analysis: Congruity vs. Grammaticality Effects....218
IV.F.3.e: Anatomical Complexity....219
IV.F.3.f: N400 and P600 Effect Topographies....220
IV.F.3.g: Effects of Cloze Probability....221
IV.F.3.h: Effects of Entropy/Constraint....225
IV.F.4: Results Summary: Sentence Acceptability....229
IV.G: Results: P3b, N400, P600 and LPC Comparisons....231
IV.G.1: Peak Latency and Amplitude Correlation....231
IV.G.2: Least-Squares Source Model of the N400....233
IV.G.3: Topography Similarity....235
IV.G.4: Results Summary: P3b, N400, P600 and LPC Comparisons....238
IV.H: Discussion....239
IV.H.1: Replication of the Canonical P3b, N400, and P600....239
IV.H.2: Between-Participant and Single Trial Variation....241
IV.H.3: Evidence for Subcomponents of the P3b, N400, and P600....243

IV.H.4: On the Relationship Between the P3b and LPC.....244
IV.H.5: On the Relationship Between the N400, P3b, and LPC.....245
IV.H.6: On the Relationship Between the P600, N400, P3b, and LPC.....247
IV.H.7: Other Effects and Null Effect of Note.....248
IV.I: Appendix 1: 64 Channel Electrode Labels and Locations.....320
IV.J: References.....324

Chapter V: Identifying Reliable Independent Components
V.A: Introduction.....337
V.B: Assessing Component Reliability with 50% Splits.....339
V.B.1: How to Split the Data.....341
V.B.2: How to Pair Independent Components.....341
V.B.3: How to Test for Significant Similarity.....343
V.B.3.a: Activation and Topography Similarity Metrics.....343
V.B.3.b: How Similar is Homologous? 346
V.B.3.c: Choosing a Critical Region.....348
V.B.4: An Example: Visual Oddball Data.....349
V.C: Comparison with Bootstrap Resampling.....352
V.D: Reliability vs. Amount of Data.....354
V.E: Types of Reliable ICs.....356
V.E.1: Large ICs.....356
V.E.2: Dipolar ICs.....358
V.F: Improving IC Reliability.....360
V.G: References.....390

Chapter VI: Independent Component Analysis of the P3b, N400, and P600
VI.A: Introduction.....394
VI.B: Independent Component Derivation and Complexity of Deviance Effects.....396
VI.B.1: ICA Parameters & IC Reliability.....396
VI.B.2: IC Complexity of the P3b, N400, & P600 Effects.....398
VI.C: Independent Components Common to Experiments 1 and 2.....400
VI.C.1: Identifying Common ICs.....400
VI.C.2: Contribution of Common ICs to Grand Average ERPs.....403
VI.C.2.a: Interim Summary I: Degree of Commonality and Distinctiveness.....405
VI.C.3: The Class of "Best P3b" ICs.....408
VI.C.3.a: Common Best P3b IC Oddball Task Activity.....409
VI.C.3.b: Common Best P3b IC Sentence Acceptability Task Activity.....413
VI.C.3.c: Common Best P3b IC Variability.....423
VI.C.3.d: Interim Summary II: Common Best P3b ICs.....427
VI.C.4: Cluster Analysis of Common ICs.....436
VI.C.4.a: IC Source Modeling.....437

VI.C.4.b: Clustering Algorithm and Features.....438
VI.C.4.c: Cluster Validation.....440
VI.C.4.d: Cluster Mean Activation Analysis.....442
VI.C.4.e: Clusters of Common ICs.....443
VI.C.4.e.1: Left and Right Frontal Clusters (LFr & RFr).....445
VI.C.4.e.2: Left Fronto-Temporal Cluster (LFTe).....447
VI.C.4.e.3: Left and Right Parietal Clusters (LPa & RPa).....448
VI.C.4.e.4: Left and Right Occipital Clusters (LOc & ROc).....449
VI.C.4.e.5: Midline Central Cluster (MCe) 451
VI.C.4.e.6: Midline Parietal Cluster (MPa) 455
VI.C.4.e.7: Midline Occipital Cluster (MOc) 456
VI.C.4.f: Interim Summary III: Clusters of Common ICs.....457
VI.D: Independent Components Unique to One Experiment.....472
VI.D.1: Interim Summary IV: Independent Components Unique to One Experiment.....476
VI.E: Summary.....478
VI.F: References.....641

Chapter VII: Conclusions and Closing Thoughts
VII.A: Introduction.....652
VII.B: A Possible P3b-like Subcomponent of the N400 effect.....653
VII.C: On the Relationship between the P3b and P600.....656
VII.D: Future Work and Testable Predictions.....657
VII.E: Potential Implications for our Understanding of Language Learning and Comprehension.....662
VII.F: On the utility of ICA for Defining ERP Subcomponents.....667
VII.G: On the Potential Importance of Individual Variability in Studying the Electroencephalogram and the Cognitive Neuroscience of Language.....675
VII.H: References.....682

LIST OF TABLES

Chapter III
Table III.1: Hypothetical EEG source correlations.....137
Table III.2: Source gain vector similarity.....138
Table III.3: Similarity between ICs and sources.....139
Table III.4: Similarity between ICs and sources (without sensor noise).....140

Chapter IV
Table IV.1: Summary statistics for stem completion and stem properties.....254
Table IV.2: Summary statistics describing EEG artifact independent components...255
Table IV.3: Similarity of grand average ERP effect and LPC topographies.....256
Table IV.4: Similarity of ERP effect and LPC topographies.....257
Table IV.5: Peak latency and amplitude rank correlations for ERP effects and the LPC.....258
Table IV.6: Summary statistics for high and low cloze probability nonanomalous nouns and verbs.....259
Table IV.7: Summary statistics for high and low entropy sentence stems.....260

Chapter VI
Table VI.1: Variance of grand average P3b effects accounted for by each cluster of ICs.....481
Table VI.2: Variance of grand average congruity effects accounted for by each cluster of ICs.....482
Table VI.3: Variance of grand average grammaticality effects accounted for by each cluster of ICs.....483
Table VI.4: Variance of the grand average late positive component accounted for by each cluster of ICs.....484
Table VI.5: Variance of various grand average ERP effects and components accounted for by the class of 12 common best P3b IC.....485
Table VI.6: Median one-tailed independent samples *t*-test *p*-values of the three primary deviance effects for the 12 common best P3b ICs and their comparable best P3b electrode.....486
Table VI.7: Oddball task *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests.....487
Table VI.8: Congruity effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (N400 window).....488
Table VI.9: Congruity effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (P600 window).....489
Table VI.10: Grammaticality effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (N400 window).....490
Table VI.11: Grammaticality effect *p*-values from cluster ANOVAs and follow

up one-tailed *t*-tests (P600 window).....491
Table VI.12: Congruity vs. grammaticality effect *p*-values from one-tailed *t*-tests.....492
Table VI.13: Cloze probability effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests.....493
Table VI.14: Entropy effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (incongruous endings).....494
Table VI.15: Entropy effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (ungrammatical endings).....495
Table VI.16: Concreteness effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (N400 window).....496
Table VI.17: Concreteness effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (P600 window).....497
Table VI.18: Frequency effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (N400 window).....498
Table VI.19: Frequency effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (P600 window).....499
Table VI.20: Word length effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (N400 window).....500
Table VI.21: Word length effect *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (P600 window).....501
Table VI.22: Central tendencies and dispersion of stimulus features used to create median split ERAs).....502
Table VI.23: Mean onset and peak latency of the late positivity in the ERAs of ICs in the Midline Central cluster).....503

LIST OF FIGURES

Chapter II
Figure II.1: Hypothetical ERPs that illustrate how the N400 effect could be caused by a delayed P3b....70

Chapter III
Figure III.1: A single EEG source and its projection to two electrodes....141
Figure III.2: Hypothetical EEG and EEG source topographies....142
Figure III.3: Three EEG sources and their combined projections to two electrodes....143
Figure III.4: A hypothetical linear spatial filter....144
Figure III.5: A simple example of ICA....145
Figure III.6: Examples of Gaussian, supergaussian, and subguassian distributions...146
Figure III.7: Hypothetical source ERPs....147
Figure III.8: Scatter plots of the activity of two sources....148
Figure III.9: ERPimages of the activity of two sources....149
Figure III.10: ERP and sensor noise source topographies....150
Figure III.11: Hypothetical electrode ERPs....151
Figure III.12: Hypothetical electrode difference waves....152
Figure III.13: Independent component ERPs....153
Figure III.14: Independent component topographies....154
Figure III.15: Contribution of Component 2 to difference waves....155
Figure III.16: Contribution of Component 5 to difference waves....156
Figure III.17: Representative independent component topographies....157
Figure III.18: Representative principal component topographies....158
Figure III.19: Random principal component topographies....159
Figure III.20: Principal square root of the covariance matrix topographies....160

Chapter IV
Figure IV.1: An example series of tasks that a participant would perform in the experimental session....261
Figure IV.2: Electrode locations....262
Figure IV.3: A hypothetical examples to illustrate the difference between constraint and entropy....263
Figure IV.4: Scatter plot of the 312 sentence stem constraint and entropies....264
Figure IV.5: Properties of an independent component that primarily represents EEG blink artifact....265
Figure IV.6: Properties of an independent component that primarily represents EEG eye movement artifact....266
Figure IV.7: Properties of an independent component that primarily represents EEG heart beat artifact....267

Figure IV.8: Properties of an independent component that primarily represents EEG drift artifact....268
Figure IV.9: Properties of an independent component that primarily represents 60 Hz line noise....269
Figure IV.10: Properties of an independent component that primarily represents muscle activity....270
Figure IV.11: Properties of an independent component that primarily represents a faulty electrode....271
Figure IV.12: Grand average ERPs from the letter case oddball task....272
Figure IV.13: Grand average ERPs from the letter case oddball task at representative electrodes....273
Figure IV.14: Grand average ERPs from the semantic category oddball task....274
Figure IV.15: Grand average ERPs from the semantic category oddball task at representative electrodes....275
Figure IV.16: Grand average difference waves from both oddball tasks....276
Figure IV.17: Grand average difference waves from both oddball tasks at select electrodes....277
Figure IV.18: Grand average ERPs to targets preceded by standards or targets at select electrodes....278
Figure IV.19: Grand average P3b sequence effect at select electrodes....279
Figure IV.20: Grand average P3b sequence effect topography....280
Figure IV.21: Grand average P3b, N400, and P600 effect topographies....281
Figure IV.22: Mean cumulative percentage variance accounted for each participant's largest five principal components....282
Figure IV.23: Mean P3b, N400, P600 effect topographies of the first eight participants....283
Figure IV.24: Mean P3b, N400, P600 effect topographies of the latter eight participants....284
Figure IV.25: Mean P3b, N400, P600 effect and LPC topographies of the first eight participants....285
Figure IV.26: Mean P3b, N400, P600 effect and LPC topographies of the latter eight participants....286
Figure IV.27: Single trials of the EEG response to targets at electrode MiPa in the letter case oddball task....287
Figure IV.28: Single trials of the EEG response to targets at electrode MiPa in the semantic category oddball task....288
Figure IV.29: Mean accuracy and standard error bars in the sentence acceptability task. 289
Figure IV.30: Mean individual participant median reaction times in the sentence evaluation task....290
Figure IV.31: Grand average ERPs to nonanomalous sentence final words....291
Figure IV.32: Grand average ERPs to nonanomalous sentence final words at representative electrodes....292
Figure IV.33: Grand average ERPs to sentence final nouns....293

Figure IV.34: Grand average ERPs to sentence final nouns at representative electrodes.. 294
Figure IV.35: Grand average ERPs to sentence final verbs.....295
Figure IV.36: Grand average ERPs to sentence final verbs at representative electrodes.....296
Figure IV.37: Grand average ERPs to sentence final pronouns.....297
Figure IV.38: Grand average ERPs to sentence final pronouns at representative electrodes.....298
Figure IV.39: Grand average difference waves of semantic congruity effects.....299
Figure IV.40: Grand average difference waves of semantic congruity effects at representative electrodes.....300
Figure IV.41: Grand average difference waves of grammaticality effects.....301
Figure IV.42: Grand average difference waves of grammaticality effects at representative electrodes.....302
Figure IV.43: Grand average difference waves of verb congruity and grammaticality effects.....303
Figure IV.44: Grand average difference waves of verb congruity and grammaticality effects at representative electrodes.....304
Figure IV.45: Grand average topographies of semantic congruity and grammaticality effects for each part of speech.....305
Figure IV.46: Grand average LPC and P3b, N400, and P600 effect topographies.....306
Figure IV.47: Grand average ERPs to sentence final nonanomalous nouns and verbs.....307
Figure IV.48: Grand average ERPs to sentence final nonanomalous nouns and verbs at representative electrodes.....308
Figure IV.49: Single trials of the EEG response to nonanomalous noun and verb endings at electrode MiPa.....309
Figure IV.50: Single trials of the EEG response to nonanomalous noun and verb endings at electrode LLPf.....310
Figure IV.51: Grand average ERPs to incongruous verbs.....311
Figure IV.52: Grand average ERPs to incongruous verbs at representative electrodes.....312
Figure IV.53: Grand average ERPs to ungrammatical verbs.....313
Figure IV.54: Grand average ERPs to ungrammatical verbs at representative electrodes.....314
Figure IV.55: Single trials of the EEG response to incongruous verbs at electrode RDPP.....315
Figure IV.56: Single trials of the EEG response to ungrammatical verbs at electrode RDPP.....316
Figure IV.57: Single trials of the EEG response to incongruous verbs at electrode RDPf.....317
Figure IV.58: Single trials of the EEG response to ungrammatical verbs at electrode RDPf.....318

Figure IV.59: Least-squares N400 effect source model of ERPs to all incongruous and congruous endings.....319

Chapter V
Figure V.1: Hypothetical scalp potential sources that ICA algorithms cannot reliably decompose.....363
Figure V.2: The topographies of seven ICs that are significantly similar to the topography of an IC that primarily represent blink activity.....364
Figure V.3: Histograms of topography and activation similarity for all possible IC pairs.....365
Figure V.4: Joint distribution of IC pair topography and activation similarities from Figure V.5: Topographies of the top 16 ICs from Participant 1's visual oddball task data and their matches from both 50% splits of the data.....366
Figure V.6: Topographies of ICs 17-32 from Participant 1's visual oddball task data and their matches from both 50% splits of the data.....367
Figure V.7: Topographies of ICs 33-48 from Participant 1's visual oddball task data and their matches from both 50% splits of the data.....368
Figure V.8: Topographies of ICs 49-64 from Participant 1's visual oddball task data and their matches from both 50% splits of the data.....369
Figure V.9: Features of IC 4 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....370
Figure V.10: Features of IC 10 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....371
Figure V.11: Features of IC 14 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....372
Figure V.12: Features of IC 7 from Participant 1's visual oddball data and the two ICs with similar topographies to IC 7's from each 50% split of the data.....373
Figure V.13: Features of IC 3 from Participant 1's visual oddball data and the two ICs with the most similar topographies to IC 3's from each 50% split of the data.....374
Figure V.14: Features of IC 30 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....375
Figure V.15: Features of IC 8 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....376
Figure V.16: IC topography and activation reproducibility estimates using 50% splits and bootstrap resamples are highly correlated.....377
Figure V.17: Mean percentages of reliable ICs as a function of data set size.....378
Figure V.18: Median percentage variance accounted for by reliable and unreliable ICs.....379
Figure V.19: Features of IC 32 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data.....380
Figure V.20: Percentage of dipolar ICs that were found to be reliable.....381
Figure V.21: Topographies of the ICs 1-16 of Participant 7's visual data from

Experiment 3 and their matches from both 50% splits of the data (100 ms baseline).....382
Figure V.22: Topographies of the ICs 17-30 of Participant 7's visual data from Experiment 3 and their matches from both 50% splits of the data (100 ms baseline).....383
Figure V.23: Topographies of the ICs 1-16 of Participant 7's visual data from Experiment 3 and their matches from both 50% splits of the data (full epoch baseline).....384
Figure V.24: Topographies of the ICs 17-30 of Participant 7's visual data from Experiment 3 and their matches from both 50% splits of the data (full epoch baseline).....385
Figure V.25: Topographies of the 18 most similar pairs of ICs from Figures V.21-24.....386
Figure V.26: Topographies of the 12 least similar pairs of ICs from Figures V.21-24.....387
Figure V.27: Effect of data baseline on IC reliability.....388

Chapter VI
Figure VI.1: Envtopo plots of Participant 16's letter case oddball task ERPs and difference wave.....504
Figure VI.2: Envtopo plots of Participant 16's ERPs to congruous and incongruous sentence endings and their difference wave.....505
Figure VI.3: Envtopo plots of Participant 16's ERPs to grammatical and ungrammatical sentence endings and their difference wave.....506
Figure VI.4: Mean individual participant P3b, N400, and P600 effect variance accounted for by increasing numbers of "best" independent components.....507
Figure VI.5: Features of an IC derived from both experiments and its homologue derived from only the oddball and sentence acceptability experiments.....508
Figure VI.6: Features of an IC derived from both experiments and its automatically selected pair from only the oddball and sentence acceptability experiments.....509
Figure VI.7: Features of an IC derived from both experiments and its automatically selected pair from only the oddball and sentence acceptability experiments.....510
Figure VI.8: Hypothetical ideal ICA unmixing and mixing vectors of two data sets from two different experiments.....511
Figure VI.9: Features of an IC derived from both experiments and its homologue derived from only the oddball and sentence acceptability experiments.....512
Figure VI.10: Features of an IC derived from both experiments and its homologue derived from only the oddball and sentence acceptability experiments.....513
Figure VI.11: Features of an IC derived from both experiments and its

homologue derived from only the oddball and sentence acceptability experiments....514
Figure VI.12: Example grand average ERPs to standards in the letter case oddball task....515
Figure VI.13: Example grand average ERPs to targets in the letter case oddball task....516
Figure VI.14: Example grand average ERPs to standards in the semantic category oddball task....517
Figure VI.15: Example grand average ERPs to targets in the semantic category oddball task....518
Figure VI.16: Example grand average P3b effect difference waves in the letter case oddball task....519
Figure VI.17: Example grand average P3b effect difference waves in the semantic category oddball task....520
Figure VI.18: Example grand average ERPs to congruous sentence endings in the sentence acceptability task....521
Figure VI.19: Example grand average ERPs to incongruous sentence endings in the sentence acceptability task....522
Figure VI.20: Example grand average ERPs to grammatical sentence endings in the sentence acceptability task....523
Figure VI.21: Example grand average ERPs to ungrammatical sentence endings in the sentence acceptability task....524
Figure VI.22: Example grand average N400 effect difference waves....525
Figure VI.23: Example grand average P600 effect difference waves....526
Figure VI.24: Topographies of the grand average P3b, N400, and P600 effects captured by ICs common to both experiments and by reliable ICs....527
Figure VI.25: Median ERP effect or component PVA by the 10 ICs that individually accounted for the most variance of that effect of component....528
Figure VI.26: Topographies of each participant's hand-picked single most P3b-like IC....529
Figure VI.27: Average oddball task activations of each participant's "best P3b" IC....530
Figure VI.28: Mean features of common best P3b ICs....531
Figure VI.29: Mean peak latency of the late positivity of the common best P3b ICs....532
Figure VI.30: Average responses to targets from the first and second half of both oddball tasks of the common best P3b ICs....533
Figure VI.31: Average responses to standards from the first and second half of both oddball tasks of the common best P3b ICs....534
Figure VI.32: ERPimage of the habituation of the response of common best P3b ICs in the oddball tasks....535
Figure VI.33: Average responses to sentence final words of the common best P3b ICs....536
Figure VI.34: Number of common best P3b ICs whose ERAs to incongruous

words are of greater absolute magnitude than their ERAs to congruous endings.....537
Figure VI.35: Average responses to low and high cloze nonanomalous noun and verb endings of the common best P3b ICs.....538
Figure VI.36: ERPimage of the effect of cloze probability on the response to nonanomalous noun and verb endings of the common best P3b ICs.....539
Figure VI.37: Average responses to low and high entropy incongruous endings of the common best P3b ICs.....540
Figure VI.38: ERPimage of the effect of entropy on the response to incongruous endings of the common best P3b ICs.....541
Figure VI.39: Average responses to low and high entropy ungrammatical endings of the common best P3b ICs.....542
Figure VI.40: ERPimage of the effect of entropy on the response to ungrammatical endings of the common best P3b ICs.....543
Figure VI.41: Average responses to congruous and incongruous sentence endings from the first and second half of the sentence acceptability task of the common best P3b ICs.....544
Figure VI.42: Average responses to grammatical and ungrammatical sentence endings from the first and second half of the sentence acceptability task of the common best P3b ICs.....545
Figure VI.43: ERPimage of the habituation of the response of common best P3b ICs to sentence endings.....546
Figure VI.44: Median topography and ERA/ERP between-participant similarity of the common best P3b ICs and their comparable electrode data.....547
Figure VI.45: Oddball task ERPs from each participant's "best P3b" electrode.....548
Figure VI.46: ERPs from each participant's best P3b electrode to sentence final words.....549
Figure VI.47: Mean classification error from cluster validation procedure and cluster dendrogram.....550
Figure VI.48: Mean features of left frontal IC cluster.....551
Figure VI.49: Topographies of left frontal IC cluster.....552
Figure VI.50: Mean features of right frontal IC cluster.....553
Figure VI.51: Topographies of right frontal IC cluster.....554
Figure VI.52: Mean features of left fronto-temporal IC cluster.....555
Figure VI.53: Topographies of left fronto-temporal IC cluster.....556
Figure VI.54: Mean features of left parietal IC cluster.....557
Figure VI.55: Topographies of left parietal IC cluster.....558
Figure VI.56: Mean features of right parietal IC cluster.....559
Figure VI.57: Topographies of right parietal IC cluster.....560
Figure VI.58: Mean features of left occipital IC cluster.....561
Figure VI.59: Topographies of left occipital IC cluster.....562
Figure VI.60: Mean features of right occipital IC cluster.....563
Figure VI.61: Topographies of right occipital IC cluster.....564
Figure VI.62: Mean features of midline central IC cluster.....565
Figure VI.63: Topographies of midline central IC cluster.....566

Figure VI.64: Mean features of midline parietal IC cluster.....567
Figure VI.65: Topographies of midline parietal IC cluster.....568
Figure VI.66: Mean features of midline occipital IC cluster.....569
Figure VI.67: Topographies of midline occipital IC cluster.....570
Figure VI.68: Mean normalized absolute *t*-scores for each IC cluster.....571
Figure VI.69: Example grand average scalp-projected oddball ERAs of the left frontal cluster.....572
Figure VI.70: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the left frontal cluster.....573
Figure VI.71: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the left frontal cluster.....574
Figure VI.72: Example grand average scalp-projected oddball ERAs of the right frontal cluster.....575
Figure VI.73: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the right frontal cluster.....576
Figure VI.74: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the right frontal cluster.....577
Figure VI.75: Example grand average scalp-projected oddball ERAs of the left fronto-temporal cluster.....578
Figure VI.76: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the left fronto-temporal cluster.....579
Figure VI.77: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the left fronto-temporal cluster.....580
Figure VI.78: Example grand average scalp-projected oddball ERAs of the left parietal cluster.....581
Figure VI.79: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the left parietal cluster.....582
Figure VI.80: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the left parietal cluster.....583
Figure VI.81: Example grand average scalp-projected oddball ERAs of the right parietal cluster.....584
Figure VI.82: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the right parietal cluster.....585
Figure VI.83: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the right parietal cluster.....586
Figure VI.84: Example grand average scalp-projected oddball ERAs of the left occipital cluster.....587
Figure VI.85: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the left occipital cluster.....588
Figure VI.86: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the left occipital cluster.....589
Figure VI.87: Example grand average scalp-projected oddball ERAs of the right occipital cluster.....590
Figure VI.88: Example grand average scalp-projected ERAs to congruous and

incongruous sentence endings of the right occipital cluster.....591
Figure VI.89: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the right occipital cluster.....592
Figure VI.90: Example grand average scalp-projected oddball ERAs of the midline central cluster.....593
Figure VI.91: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the midline central cluster.....594
Figure VI.92: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the midline central cluster.....595
Figure VI.93: Example grand average scalp-projected oddball ERAs of the midline parietal cluster.....596
Figure VI.94: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the midline parietal cluster.....597
Figure VI.95: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the parietal central cluster.....598
Figure VI.96: Example grand average scalp-projected oddball ERAs of the midline occipital cluster.....599
Figure VI.97: Example grand average scalp-projected ERAs to congruous and incongruous sentence endings of the midline occipital cluster.....600
Figure VI.98: Example grand average scalp-projected ERAs to grammatical and ungrammatical sentence endings of the midline occipital cluster.....601
Figure VI.99: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to targets in the letter case oddball task.....602
Figure VI.100: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to standards in the letter case oddball task.....603
Figure VI.101: Envtopo plots of each common IC cluster's contribution to the grand average P3b effect in the letter case oddball task.....604
Figure VI.102: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to targets in the semantic category oddball task.....605
Figure VI.103: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to standards in the semantic category oddball task.....606
Figure VI.104: Envtopo plots of each common IC cluster's contribution to the grand average P3b effect in the semantic category oddball task.....607
Figure VI.105: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to incongruous sentence endings.....608
Figure VI.106: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to congruous sentence endings.....609
Figure VI.107: Envtopo plots of each common IC cluster's contribution to the grand average N400 effect.....610
Figure VI.108: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to ungrammatical verb sentence endings.....611
Figure VI.109: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to grammatical verb sentence endings.....612
Figure VI.110: Envtopo plots of each common IC cluster's contribution to the

grand average verb P600 effect.....613
Figure VI.111: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to ungrammatical pronoun sentence endings.....614
Figure VI.112: Envtopo plots of each common IC cluster's contribution to the grand average ERPs to grammatical pronoun sentence endings.....615
Figure VI.113: Envtopo plots of each common IC cluster's contribution to the grand average pronoun P600 effect.....616
Figure VI.114: Mean IC cluster target sequence effect ERAs.....617
Figure VI.115: Mean IC cluster ERAs to incongruous and ungrammatical verbs.....618
Figure VI.116: Mean IC cluster ERAs to high and low cloze nonanomalous sentence endings.....619
Figure VI.117: Mean IC cluster ERAs to high and low entropy incongruous sentence endings.....620
Figure VI.118: Mean IC cluster ERAs to high and low entropy ungrammatical sentence endings.....621
Figure VI.119: Mean IC cluster ERAs to high and low concreteness congruous sentence endings.....622
Figure VI.120: Mean IC cluster ERAs to high and low concreteness incongruous sentence endings.....623
Figure VI.121: Mean IC cluster ERAs to frequent and infrequent noun and verb sentence endings.....624
Figure VI.122: Mean IC cluster ERAs to frequent and infrequent incongruous sentence endings.....625
Figure VI.123: Mean IC cluster ERAs to long and short noun and verb sentence endings.....626
Figure VI.124: Mean IC cluster ERAs to long and short verb sentence endings.....627
Figure VI.125: Scalp-projected ERAs of ICs in left fronto-temporal cluster.....628
Figure VI.126: Scalp-projected oddball task ERAs of ICs in midline central cluster.....629
Figure VI.127: Scalp-projected sentence acceptability task ERAs of ICs in midline central cluster.....630
Figure VI.128: Mean peak latency of the late positivity of ICs in the midline central cluster.....631
Figure VI.129: ERPimage of the effect of cloze probability on the response to congruous endings of the ICs in the midline central cluster.....632
Figure VI.130: Topographies of ICs unique to the oddball tasks that contribute greatly the to P3b effect.....633
Figure VI.131: Oddball task ERAs of the ICs in Figure IV.130.....634
Figure VI.132: Topographies of ICs unique to the sentence acceptability task that contribute greatly the to N400 effect.....635
Figure VI.133: ERAs to congruous and incongruous sentence endings of the ICs in Figure IV.132 (Part 1).....636
Figure VI.134: ERAs to congruous and incongruous sentence endings of the ICs in Figure IV.132 (Part 2).....637

Figure VI.135: Topographies of ICs unique to the sentence acceptability task that contribute greatly the to P600 effect.....638
Figure VI.136: ERAs to grammatical and ungrammatical sentence endings of the ICs in Figure IV.135 (Part 1).....639
Figure VI.137: ERAs to grammatical and ungrammatical sentence endings of the ICs in Figure IV.135 (Part 2).....640

ACKNOWLEDGEMENTS

As an undergraduate, when I encountered exciting cognitive science papers or books (e.g., Ed Hutchins' *Cognition in the Wild,* Pat Churchland and Terry Sejnowski's *The Computational Brain,* Jeff Elman's *The Importance of Starting Small*) I made a point of noting the institutions with which the authors were affiliated. Time and time again, UCSD was the home of many of these investigators and I developed the impression of a uniquely diverse and vibrant community of scientists working together at UCSD to understand cognition from the cultural to the molecular level using the tools of computer science, neuroscience, psychology, and anthropology. This was a community that I very much wanted to be a part of.

Soon upon entering graduate school, I learned that making meaningful connections across these different levels was much more difficult than I had naively imagined as an undergraduate. However, the intellectual caliber and personal character of the cognitive scientists at UCSD, if anything, have exceeded my expectations, and I have benefited enormously from my membership in this community over the years. In particular there are several people who directly contributed to this dissertation and to the best parts of my graduate school experience whom I would presently like to acknowledge.

First and foremost is my advisor of the past six years, Marta Kutas. When I began graduate school, I did not know what the acronym "ERP" stood for, but under Marta's tutelage I learned to record EEG, to design experiments, to read and discuss

ERP reports, and to apply conventional ERP analysis techniques. Moreover, I do not think I can express how much I have gained from Marta's copious and readily available feedback (even when an ocean away), her uncanny ability to perceive the essence of an issue, her clear and eloquent ability to communicate, her honest criticism, the great care she exhibits when making inferences from data, her vast knowledge of the language comprehension ERP literature, and her basic sense of compassion and human decency. Finally, had it not been for Marta's unfailing support these past few years, it is quite possible that I would not have finished this dissertation. Frequently, when I hear other graduate students talk about their incommunicative or unsupportive advisors, I am reminded of what a privilege it is to be Marta's student. I cannot thank her enough for all that she has done for me and I will continue to strive to live up to her example as my career continues.

Secondly, I need to thank Scott Makeig, whose pioneering work applying ICA to ERP phenomena was largely the inspiration for this thesis project and for my interest in the application of machine learning techniques to neuroscience data. Furthermore, it is directly because of Scott's patient mentorship and support that I was able to pursue this dissertation project (without the computational resources of the Swartz Center for Computational Neuroscience I would still be waiting for my Matlab scripts to finish running). Thanks to Scott, I have learned a valuable set of analytical tools that have enabled me to produce this dissertation and that leave me in a good position to contribute more in the future. Moreover, I have also learned from Scott to appreciate the richness and variability of the EEG and the importance of explaining

this variability for our understanding of brain function. I will always be grateful for both the skills and perspective I have gained from Scott's tutelage.

This dissertation is very much the product not only of my two research mentors, but also of their two labs, which have been a delight to work with. The first of these, Kutaslab, provided valuable ERP and psycholinguistic expertise and the EEG data in this dissertation. Of the members of Kutaslab (past and present), I am particularly grateful to Laura Kemmer and Kara Federmeier for providing pilot data, the analysis of which motivated this dissertation project. I am also indebted to Esmeralda De Ochoa and Julia Cagle for helping me to record my data, to Katherine DeLong for sharing some of her stimuli and knowledge of the language comprehension literature, and to Tom Urbach, Paul Krewski, and Carrie Joyce for their technical assistance. In addition, I would like to thank those people along with Arielle Borovsky, Michael Kiang, Jenny Staab, and Kim Sweeney Wolanyk for their stimulating thoughts, feedback, and findings over the years.

The other lab that was directly involved in this dissertation project was the Swartz Center for Computational Neuroscience. Of this group, I am particularly grateful to my friend Ying Wu for her proofreading and comments on Chapter II and to Bob Buffington, Julie Onton, and Arnaud Delorme for their technical assistance. In addition, I would like to thank those individuals as well as Rey Ramirez and Jason Palmer for their helpful thoughts and feedback on the ICA portions of this thesis.

In addition to my collaborators, I am indebted to the other members of my committee, Anders Dale, Virginia de Sa, Steven Hillyard, and Marty Sereno, for their

time, questions, and comments. If someone were to ask me why cognitive science is worth doing and funding, I would readily point to some of the work each of my committee members has done as prime examples of important questions and valuable contributions toward answers to those questions. It is an honor to have them as my thesis committee. Moreover, it has been a great pleasure to get Marty's polymath perspective on countless questions at departmental happy hours and Hillcrest shuttle commutes, and most of what I know about probability and machine learning I owe to Virginia, in whose class this dissertation project was begun.

Finally, a host of other investigators, friends, and family members have helped in one way or another to make this dissertation possible and to make my time in graduate school so rewarding. John Batali was my advisor and advocate for my first two years of graduate school and introduced me to information theory. A lab rotation with Javier R. Movellan and Tim Marks was key to helping me to understand the basic mathematics of probability and to appreciate the importance of probabilistic inference and reasoning in cognition. Jeff Elman's connectionist models of language comprehension were also an inspiration for my interest in machine learning, brain dynamics, and probabilistic inference in language comprehension. Moreover, I am indebted to Jeff for graciously mentoring my teaching the first time I led a class at UCSD. Seana Coulson was effectively a seventh member of my thesis committee. She provided some of the sentence stimuli as well as many valuable comments on the dissertation project. I am greatly indebted to my girlfriend, Suzanne Wood, for her patient, constant support, understanding and love. I will do my best to do the same for

her when she is finishing her own dissertation. My dear friend, classmate, and model skeptic, Hsin-Hao Yu, provided assistance on every aspect of this thesis. He was always available to answer my questions about cerebral cortex, mathematics, and apple software, and he proofread Chapter III. My dear friend and inveterate roommate, J. R. Osborn, made the sprawl of San Diego a much more interesting and homier place to live, and he provided valuable technical and printing support. I am also grateful to my good friends John "End the Liberation" Lewis, Morana "Frok" Alac, Maurizio Marchetti, Katrina "Soufflé" Hoch, Carl "You can't Hallucinate more than 50% of the Time" Sachs, Alan "Torreon" Robinson, Jymm Gifford, Jonathan Nelson, Flavia Filimon, Ferruh Yilmaz, Colleen Tuffy, Jingle, Gordon "The History of My Skin" Chang, Jenny "Hand Stand" Collins, Rob "Scienceology" Liebscher, Ian "La-la-la-la-la" Fasel, Ayse Saygin, Devin "Corn Dog" Drew and Shannon Dowdy for eight years of great memories. I also wish to thank my siblings, Elizabeth, Jennifer, Maureen, and John for always looking out for their baby brother and for exposing me to the world outside of Rensselaer, Indiana. Lastly, I can not thank my parents, John and Rose Marie, enough for all the love, support, and guidance that they have given me over the years. They probably had no idea how much of my life I would spend in front of computer when they bought that TI-99/4A for us.

CURRICULUM VITAE

EDUCATION

2003-2007 Ph.D., Dept. of Cognitive Science, University of California, San Diego (Supervisor: Dr. M. Kutas). Dissertation: Common Independent Components of the P3b, N400, and P600 ERP Components to Deviant Linguistic Events.

2001 M.S., Dept. of Cognitive Science, University of California, San Diego

1997 B.A. with honors, Grinnell College, Major: Cognitive Science.

HONORS AND AWARDS

2007-2008 Faculty Fellow, University of California, San Diego

2007 Teaching Excellence Award, Department of Cognitive Science, University of California, San Diego

2006 Summer Graduate Teaching Fellowship, University of California, San Diego (Teaching mentor: Jeff Elman).

2006 Teaching Excellence Award, Department of Cognitive Science, University of California, San Diego

2003-2004 Predoctoral Fellowship, Institute for Neural Computation, University of California, San Diego

2002-2003 Predoctoral Fellowship, Center for Research on Language, University of California, San Diego

2000-2001 Predoctoral Fellowship, Center for Research on Language, University of California, San Diego

2001 Teaching Excellence Award, Department of Cognitive Science, University of California, San Diego

2000 Honorable Mention for Teaching Excellence, Department of Cognitive Science, University of California, San Diego

1997 Inducted into Phi Beta Kappa, Grinnell College, Grinnell, Iowa

CONFERENCE PAPERS AND POSTERS

Groppe, D., Kemmer, L., Makeig, S., & Kutas, M. (2006) A method for assessing the reliability of components derived from independent component analysis. Annual Meeting of the Cognitive Neuroscience Society.

Groppe, D., Kemmer, L., Kutas, M., & Makeig, S. (2004) Within-subject reliability of independent component analysis of the electroencephalogram in two text processing tasks. Annual Meeting of the Cognitive Neuroscience Society.

Liebscher, R. & Groppe, D. (2003) Rethinking context availability for concrete and abstract words: a corpus study. Proceedings of Corpus Linguistics 2003, Lancaster, UK.

PROFESSIONAL ACTIVITIES

Co-editor, *UCSD Cognitive Science Online*, 2002-2006

Co-organized weekly departmental talk series on "Probabilistic Functionalism," UCSD Department of Cognitive Science, 2006

Graduate student representative to faculty, UCSD Department of Cognitive Science, 2002-2003

Head Teaching Assistant, UCSD Department of Cognitive Science, 2001-2002

Organized weekly departmental talk series, UCSD Department of Cognitive Science, 2000-2001

Founder of Grinnell College's Experimental College, Grinnell College, Iowa, 1996-1997

TEACHING EXPERIENCE

2006, 2007 Instructor, University of California, San Diego
COGS 14: Design and Analysis of Experiments

2006 Teaching Assistant, University of California, San Diego
COGS 1: Introduction to Cognitive Science
COGS 14: Design and Analysis of Experiments

2005 Teaching Assistant, University of California, San Diego
COGS 1: Introduction to Cognitive Science

2004 Teaching Assistant, University of California, San Diego
COGS 108D: Programming Methods for Cognitive Science

2002 Teaching Assistant, University of California, San Diego
COGS 108B Artificial Intelligence Modeling

2001 Teaching Assistant, University of California, San Diego
COGS 108C: Neural Network Models of Cognition
COGS 18: Introduction to C Programming for Cognitive Modeling
COGS 108A: Programming Methods for Cognitive Science

2000 Teaching Assistant, University of California, San Diego
COGS 18: Introduction to C Programming for Cognitive Modeling
COGS 174: Drugs: Brain, Mind, and Culture

1997-1998 Teacher, Dilko Language Schools, Istanbul, Turkey
Taught English as a foreign language.

1997 Teacher, Grinnell College's Experimental College, Grinnell, Iowa
Taught and designed "Chomsky, Nader, Moore: A Critique of America"

ABSTRACT OF THE DISSERTATION

Common Independent Components of the P3b, N400, and P600 ERP Components to Deviant Linguistic Events

by

David Martin Groppe

Doctor of Philosophy in Cognitive Science

University of California, San Diego 2007

Professor Marta Kutas, Chair

Two of the most reliable and most studied neural correlates of language comprehension are the N400 and P600 ERP components, canonically elicited by semantically and grammatically anomalous words (respectively). Over two decades of research on these phenomena have yielded insight into the time course of sentence comprehension and language learning, and have established links between language comprehension and non-linguistic domains such as music perception. However, the

ability of the N400 and P600 to inform cognitive neuroscience theory has been limited by the high likelihood that they are composed of multiple, functionally and anatomically distinct subcomponents whose contributions to these scalp potentials overlap in time and space. Methods for identifying and extracting these subcomponents are clearly needed to remove these limitations.

One method for potentially extracting such spatio-temporally overlapping ERP subcomponents is independent component analysis (ICA). In this talk I will present the results of my dissertation relevant to its primary focus, the extraction of one putative subcomponent of the N400 and P600, the P3b ERP component, using ICA. The P3b is popularly believed to index the update of a probabilistic model of the sensory environment or the engagement of attentional resources, and is canonically elicited by rare targets in simple target detection tasks. Identifying a P3b subcomponent was done by applying ICA separately to N400/P600 and P3b datasets to find independent components (ICs) that appeared to replicate across the two decompositions. The results of this procedure suggest considerable though far from total overlap between the sources of the N400, P600, and P3b. In particular, there appears to be a common subcomponent with a predominantly midline central distribution that accounts for around 50% of the variance of the N400, P600, and P3b effects. This midline central (MCe) subcomponent is characterized by a late positivity that is enhanced to rare/anomalous stimuli, habituates over the course of the experiment, and tends to peak later for stimuli that presumably take longer to evaluate. In light of these characteristics, I have tentatively interpreted the MCe subcomponent

as an index of the allocation of attentional resources. As there are several reasons why these results might be misleading, it is critical to validate them and my interpretation of them with additional research. If these results prove valid, they would improve our ability to analyze the brain's response to language via ICA, add much needed subtlety to our functional understanding of this response, and potentially enhance our understanding of the neural mechanics of this response by making it possible to relate language comprehension to animal models of the P3b.

Chapter I: Introduction

I.A Comprehending the Predictable and the Unexpected

You may not be aware of it, but you already know a great deal of what you are going to read in these pages. This is not because you have psychic powers or because you have been painstakingly following this research project, but simply due to your command of the English language.

Natural language is a balance of the predictable (redundant) and the unexpected (informative). Rarely, if ever, do you know exactly what another person is going to say next, but based on even a modicum of context, you typically have some idea of what's to come. Take for example the simple sentence fragment

I take my coffee with cream and ....

Based on your knowledge of English and the preceding sentence context you know a lot about how this utterance may continue. On one level, you know some specific words that are likely to come next (e.g. "sugar" or "artificial sweetener")[1] and even

[1] According to Genzel and Charniak (2002), the entropy of the distribution of written sentences between 3 and 25 words in length is approximately between 7 and 8 bits. Bates (1999) claims that fluent adults know between 20,000 and 40,000 words. If a speaker produced utterances from a set of 20,000 words where each word was equally likely and independent of previous words, the entropy of sentences between 3 and 25 words in length would be between 43 and 357 bits. Similarly, Philip B. Gough (1983) has estimated that readers can predict the 9th *open class* word (e.g., nouns, verbs) of 30% of sentences with greater than 10% accuracy and they can predict the 9th *closed class* word (e.g.,

what those words might look or sound like. On a higher level, you can predict what part of speech the next word will probably be; you know that a noun or a verb is likely to occur, but that another conjunction is highly unlikely. At yet another level, you know what the semantic features of the next word might be (e.g. "sweet", "dissolves in liquid", "edible") and how they might relate to the eventual import of the greater discourse.

These multiple levels of redundancy reduce the efficiency of linguistic communication, but their presence is a powerful aid to language comprehension and is probably fundamental to language learning. Behavioral studies have shown that probable words and word pairs are recognized and comprehended more quickly and with weaker input signals than less probable controls (for review see Jurafsky, 2003). Furthermore, comprehenders prefer the most probable interpretation of semantically or syntactically ambiguous words (ibid.). With respect to learning, unsupervised computational models have shown that a great deal of word meaning (Burgess, 1998; Elman, 1990) and grammatical relationships (Elman, 1993; Lewis & Elman, 2001) can be learned from the probabilities of word co-occurrence or the conditional probability of a word given preceding words. Research on infants and children has shown that they may indeed use such statistics to learn simplified, language-like patterns (Altmann, 2002; Saffran, Senghas, & Trueswell, 2001; Seidenberg, MacDonald, & Saffran, 2002) and by extension, language.

---

pronouns, articles) of 78% of sentences with greater than 10% accuracy. Clearly there is a massive degree of redundancy in natural language (see also Gough et al., 1981).

Over the past quarter of a century, one of the primary focuses of the neuroscience of sentence comprehension has been the study of the neural mechanisms for exploiting these redundancies and coping with their violations. This investment of scientific resources is well motivated for several reasons. Most obviously, behavioral measures clearly show that the brain capitalizes on the predictability of language and any adequate neurological theory of language comprehension must account for this. In addition, the aforementioned computational modeling and developmental research suggests that understanding the development of these mechanisms is key to understanding language acquisition, the source of much contentious debate on the nature of cognition (Elman et al., 1996; Pinker, 1994). A more practical motivation is that violating the brain's expectations tends to produce a strong neural response (Van Petten & Luka, 2006), which facilitates the study of when and where different types of linguistic information become relevant to the brain. Finally, understanding how the brain contends with linguistic uncertainty may reveal deep relationships between language comprehension and non-linguistic domains such as simple perceptual discriminations (Friston, 2005; Knill & Pouget, 2004) and information foraging (Nelson, 2005) that also exploit the imperfect predictability of the sensory world (Chater & Manning, 2006).

Two of the most robust neural correlates of word (un)predictability to come out of this enterprise are the N400 (Kutas & Hillyard, 1980b)and P600 (Osterhout & Holcomb, 1992) event-related potential (ERP) components. The N400 is canonically elicited by semantically unexpected or anomalous words such as the word "bake" in:

*The cats won't **bak**e the food that Mary gives them.*

While the P600 is canonically elicited by syntactically unexpected or anomalous words such as the word "eating" in:

*The cats won't **eating** the food that Mary gives them.*

These two phenomena have realized some of the promise of this research endeavor. They have shown that the brain makes predictions about future linguistic input (DeLong, Urbach, & Kutas, 2005) but that its response to a word is not simply a function of the probability of the word (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Chapter IV). Moreover, they have revealed remarkably rapid effects of second-language learning (McLaughlin, Osterhout, & Kim, 2004), demonstrated that the brain exploits temporally distant and diffuse discourse level information as rapidly as recent sentence level information (van Berkum, Hagoort, & Brown, 1999), and established links between language comprehension and non-linguistic abilities such as picture perception (Ganis, Kutas, & Sereno, 1996) and music perception (Koelsch et al., 2004; Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

However, the informativeness of these phenomena has been limited by the high likelihood that the N400 and P600 are the products of multiple anatomically and

functionally distinct neural processes, which are generally difficult to tease apart (Osterhout, 1997; Osterhout & Hagoort, 1999; Van Petten & Luka, 2006). This possibility confounds our ability to understand the functional correlates of these components (e.g., Coulson, King, & Kutas, 1998; Osterhout & Holcomb, 1999). Furthermore, it hinders our understanding of how these scalp potentials might relate to potentials elicited by non-linguistic stimuli as differences between linguistic and non-linguistic paradigms (e.g., difficulty of stimulus discrimination, modality of presentation) can obscure what is common. This is a critical problem for the neuroscience of language, as nothing truly commensurate to language has been discovered in non-humans (Deacon, 1997; Hauser, Chomsky, & Fitch, 2002), but there may be animal analogs of some (if not all) of the mechanisms used in language comprehension, which are necessary for a detailed understanding of any neural process.

I.B Better ERP Componentry through Independent Component Analysis

This frustrating problem caused by the temporal overlap of distinct ERP generators complicates the ERP methodology in general and is widely recognized. Over the past several years, there has been significant interest in the use of learned linear spatial filters to ameliorate this complication (Makeig et al., 1999; Parra, Spence, Gerson, & Sajda, 2005; Spencer, Dien, & Donchin, 1999). Linear spatial

filters can potentially untangle the contributions of these distinct ERP generators, what I will call "ERP subcomponents," thus allowing their independent identification and measurement. Learned linear spatial filters are related to linear spatial filters commonly used for ERP source localization (Baillet, Mosher, & Leahy, 2001; Dale et al., 2000). However, they differ in that they are derived from the statistics of ERPs or the electroencephalogram (EEG) from which ERPs are obtained with much fewer anatomical assumptions. One family of such algorithms, independent component analysis, ICA (Hyvärinen, Karhunen, & Oja, 2001), appears particularly useful for ERP analyses (Debener et al., 2005; Jung et al., 2000; Tang, Sutherland, & McKinney, 2005).

I.C Thesis Goals and Organization

The primary goal of this thesis is to use ICA to extract a "P3b-like" subcomponent of the N400 and P600 effects, if such a subcomponent exists. The P3b (Sutton, Braren, Zubin, & John, 1965), sometimes called the "P3" or "P300," is canonically elicited by unlikely stimuli in simple signal detection tasks. Although the function and neural underpinnings of the P3b are most likely also complex, it quite possibly reflects an update of a rather abstract model of the stimulus environment (Donchin & Coles, 1988; Friston, 2005) and/or an attentional response mediated by the locus coeruleus/norepinephrine system (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Although the function of the P3b is widely believed to be quite general, its role

(if any) in language comprehension is unclear and has been the source of some debate. Some researchers have argued that the N400 is produced by or partially reflects a delay of the P3b to incongruous words (Chung et al., 1996; Curran, Tucker, Kutas, & Posner, 1993) and others have argued that the P600 is functionally, though not necessarily physiologically, equivalent to the P3b (Coulson, King, & Kutas, 1998b; Gunter, Stowe, & Mulder, 1997). While both of these theories have a degree of empirical support, it is clear that the N400 is not reducible to the P3b and there is enough evidence that the P3b and P600 are distinct for the proposed isomorphism to be controversial (Frisch, Kotz, von Cramon, & Friederici, 2003; Osterhout & Hagoort, 1999).

Given the high likelihood that these three ERP components are the sum of multiple, functionally distinct subcomponents, it is quite possible that these three phenomena do share subcomponents and that the distinctions between them are due to some additional unique subcomponents and/or the differences in the relative timing or activation of the common subcomponents. If this is true, it would greatly enrich our understanding of language comprehension with what we know and will learn about the P3b (for which there are animal models--e.g., Pineda et al., 1989) and it could explain some evidence that is inconsistent with conventional interpretations of these components (Chapter II.G). Most importantly, if we could accurately extract subcomponents of these scalp potentials it would greatly enhance the detail with which we could measure and understand the neural mechanisms of language comprehension.

Towards this goal, I have collected these three ERP effects from the same experimental participants and compared them using both conventional "electrode level" analyses and ICA. My results suggest that there is a P3b-like subcomponent that accounts for about 60% of the variance of the P3b, N400, and P600 effects. This ostensible subcomponent is characterized by a late positivity that tends to be enhanced to low probability events, habituates over the course of the experiment, and generally peaks later to stimuli that are (presumably) more difficult to classify. The topography of this subcomponent typically has a midline central focus, reminiscent of the conventionally defined "P3a" subcomponent of the P3b, which is thought by many to be part of the brain's orienting response to attention-capturing stimuli. In light of this similarity, I tentatively interpret this midline central (MCe) subcomponent as an attentional response. However, this ostensible subcomponent and its putative attentional function should be interpreted only tentatively and require additional validation, as there are several reasons why the procedure used to define this subcomponent could be mistaken and there is considerable variation in the activity of this subcomponent across the 12 participants in whom it was found.

A secondary interwoven contribution of this thesis is a method for assessing the reliability of the independent components (ICs) derived by ICA. Since ICs are obtained from the statistics of EEG/ERP data, which can vary considerably from experimental trial to trial, there is no guarantee that any particular IC is reliable. Consequently, it is important to take steps towards ensuring that the results of ICA are not spurious. I have developed a method for doing this out of a need to identify

homologous ICs derived from different data sets (i.e., a P3b data set and an N400/P600 data set). In some ways the method improves upon existing algorithms for assessing IC reliability and I show how the method can be used to tune the application of ICA to improve the reliability of its results.

Before presenting these results, I first provide a review of relevant literature on the N400, P600, and P3b (Chapter II), and ICA (Chapter III), to more clearly motivate this thesis and to explain its methods. I then turn to the electrode level analyses of these ERP phenomena (Chapter IV), which provide some evidence that the N400 and P600 effects are composed of multiple subcomponents and of a relationship between the N400 effect and the P3b-like "late positive component" elicited by sentence endings. Chapter V presents the algorithm for assessing IC reliability along with some validating evidence and demonstrations of its utility. ICA is then used in Chapter VI to explore the subcomponentry of the P3b, N400, and P600. A modified version of the algorithm presented in Chapter V is used to identity ICs that are common or unique to these different ERP effects and the aforementioned MCe subcomponent is identified and analyzed. In addition to the MCe subcomponent, automated cluster analysis suggests some other common subcomponents of the P3b, N400, and P600 effects, and some attempt is made to explore what might be unique to each of these effects. Finally, I end the thesis with a summary of the evidence for a common P3b-like subcomponent of the P3b, N400, and P600 effects, propose future work to validate the MCe subcomponent, discuss possible implications of such a subcomponent for our understanding of language comprehension, and close with some

thoughts on the use of ICA for ERP subcomponentry and the importance of individual variability for the future of EEG analysis and the cognitive neuroscience of language (Chapter VII).

I.D References

Altmann, G. T. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences of the United States of America, 99*(24), 15250-15251.

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14-30.

Bates, E. A. (1999). On the nature and nurture of language. In R. Levi-Montalcini, D. Baltimore, R. Dulbecco, & F. Jacob (Series Eds.) & E. Bizzi, P. Calissano, & V. Volterra (Vol. Eds.), *Frontiere Della Biologia [Frontiers of Biology]: The Brain of Homo Sapiens.* Rome: Giovanni Trecanni.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers, 30*, 188-198.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335-344.

Chung, G., Tucker, D. M., West, P., Potts, G. F., Liotti, M., Luu, P., et al. (1996). Emotional expectancy: Brain electrical activity associated with an emotional bias in interpreting life events. *Psychophysiology, 33*(3), 218-233.

Coulson, S., King, J. W., & Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13*(62), 653-672.

Coulson, S., King, J. W., & Kutas, M. (1998b). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology, 88*(3), 188-209.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Deacon, T. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York: W.W. Norton.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 112*(3), 357-374.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71-99.

Elman, J. L., Bates, E. A., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469-495.

Frisch, S., Kotz, S. A., von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology, 114*(2), 336-340.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London: Series B, Biological sciences, 360*(1456), 815-836.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89-106.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. *Proceedings of the 40th Annual Meeting of the ACL,* Philadelphia. 199-206.

Gough, P. B. (1983). Context, form, and interaction. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes* (pp. 203-211). New York: Academic Press.

Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981). Words and contexts. In O. J. L. Tzeng, & H. Singer (Eds.), *Perception of Print: Reading Research in Experimental Psychology* (pp. 85-102). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*(6), 660-676.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. New York: J. Wiley.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V. J., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics.* Cambridge, MA: MIT Press.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712-719.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302-307.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Lewis, J. D., & Elman, J. L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the 26th Annual Boston University Conference on Language Development,*

Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience, 19*(7), 2665-2680.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7*(7), 703-704.

Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review, 112*(4), 979-999.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510-532.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*(1), 1-14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage, 28*(2), 326-341.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Pinker, S. (1994). *The Language Instinct*. New York: Morrow.

Saffran, J. R., Senghas, A., & Trueswell, J. C. (2001). The acquisition of language by children. *Proceedings of the National Academy of Sciences of the United States of America, 98*(23), 12874-12875.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does grammar start where statistics stop? *Science, 298*(5593), 553-554.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology, 36*(3), 409-414.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(700), 1187-1188.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage, 25*(2), 539-553.

van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279-293.

Chapter II: Three Components of the Unexpected

II.A Introduction

Over the past quarter of a century, event-related potential (ERP) research on language comprehension has shown that brain's immediate response to a word is strongly affected by the probability of that word. Curiously, this response is not simply a function of that probability (contrary to what one might expect from an engineering perspective on communication--Shannon, 1948, or from some connectionist models of language processing--Elman, 2004). Rather it depends upon how the word differs from what is likely given previous context. For example, words that are unlikely because of their semantic import most reliably elicit what appears to be an enhanced N400 ERP component but words that are unlikely because of their syntactic features most reliably elicit the P600 ERP component.

As discussed in the previous chapter, these two ERP components have provided valuable information about the nature and time course of language comprehension and learning, and have revealed some links between language comprehension and non-linguistic perception. However, their utility has been limited by the high likelihood that each component is actually the sum of spatiotemporally

overlapping, functionally distinct "subcomponents." In particular, it is quite likely another ERP component, the P3b, contributes in some way to both of these effects. A main goal of this thesis is to try to determine how much (if at all) the P3b contributes to these effects and more importantly, to extract a P3b contribution from these effects using independent component analysis (ICA).

In this chapter, I briefly summarize the defining features and antecedent conditions of the N400, P600, and P3b. I review current thinking on their functional correlates and evidence that each component is composed of subcomponents. To motivate attempts to identify and extract these putative subcomponents, I then illustrate how the high likelihood that these phenomena are complex limits their ability to inform cognitive neuroscience theories. Subsequently, I review evidence that the P3b contributes in some way to the N400 and P600 and explain why such shared P3b subcomponents are particularly promising to try to extract and identify using ICA.

However, before going any further, I first need to clarify what I mean by the terms "ERP component" and "ERP subcomponent."

II.A.1 Components and Subcomponents Defined

Kutas and Dale (1997) specify two quite different ways to define an ERP component. According to a "functional definition," an ERP component is simply a feature of the ERP waveform (e.g., a peak or a trough) whose polarity, latency, and

scalp topography are predictable from experimental manipulations. In contrast, a "physiological definition" conceives of an ERP component as the activity of a specific neural generator or set of neural generators. In practice, these two definitions are frequently mixed as ERP researchers often use a functional definition in their analyses with the assumption that these ERP features correspond to specific neural generators, even if the identity of the neural generators is not well established (e.g., Van Petten & Luka, 2006; Friederici & Kotz, 2003; Nieuwenhuis, Aston-Jones, & Cohen, 2005).

In this dissertation, when I use the term "ERP component" I intend the mixed "functional/physiological" definition. More specifically, I refer to the most conventional method for measuring the N400, P600, and P3b, which is the mean ERP amplitude across all electrodes in a particular time window (e.g., Section IV.E.3.b). I assume that this measure reflects the activity of a set of neural generators that is largely identical across individuals and across the different experimental paradigms that elicit these components. Furthermore, I hypothesize that for each of the three components, the set of neural generators consists of multiple, functionally distinct and somewhat independent areas of cortex. A contribution of a particular, functionally unified subset of neural generators to the conventional measure of these ERP phenomena is what I refer to with the term "ERP subcomponent."

II.B Features and Functional Correlates of the N400

The canonical N400 (Kutas & Hillyard, 1980b) is elicited by a semantically incongruous or simply unexpected word in a sentence (for review see Kutas & Federmeier, 2000). For example, consider the final words of the following sentences:

A*: The bill was due at the end of the* ***month****.*

B*: The bill was due at the end of the* ***hour****.*

C*: The bill was due at the end of the* ***dog****.*

When such sentences are presented visually, word-by-word, the resultant ERPs to the expected sentence ending ("month") exhibit a large positivity that is greatest at centro-parietal electrodes, beginning roughly 3000 ms after the onset of the word, and lasting for several hundred milliseconds (e.g., Figures IV.31, IV.45: A). The ERPs to the unlikely ending ("hour") and the semantically incongruous ending ("dog") also show this "late positive component" (LPC). However, the peak of the LPC to these endings is preceded by a negative deflection in the ERP, the N400, which peaks approximately 400 ms after the appearance of the word, is broadly distributed across the scalp but greater at centro-parietal locations and slightly bigger over the right hemisphere (e.g., Figures IV.33, IV.45: B).

While the N400 is most clearly apparent when contrasting the ERPs to unlikely or incongruous words with likely control words, decades of research suggest that the N400 is elicited by all words to varying degrees. To date, the best single predictor of the amplitude of the N400 is the "cloze probability" of the eliciting word. The "cloze probability" of a word is the probability that someone would produce that word, given the linguistic context that precedes that word (Section IV.C.3.a). Cloze probability strongly correlates with mean ERP voltage in the N400 time window (i.e., words with lower cloze probabilities elicit larger N400s) and this correlation is strongest at electrodes where the N400 is typically largest (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984). This correlation is even apparent in individual participants (Kutas & Hillyard, 1984) and is weakly apparent in the raw EEG (Dambacher, Kliegl, Hofmann, & Jacobs, 2006).

While the relationship between cloze probability and N400 amplitude is remarkably robust, they can be disassociated by semantic factors. For example, the final words of the following two sentences have approximately equal cloze probabilities (Kutas & Hillyard, 1984):

C*: Don't touch the wet* ***dog****.*

D*: He liked lemon and sugar in his* ***coffee.***

However, the N400 to “coffee” is much smaller than that to “dog” (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984). This appears to be caused by the fact that “coffee” is semantically related to the expected completion ”tea.” Importantly, this effect of semantic relatedness is greater when the most likely completion is highly expected (Federmeier & Kutas, 1999). Thus the relatively smaller N400 to the related ending is unlikely to be due to an underestimation of the probability of the related ending.

N400 amplitude is also sensitive to other semantic manipulations. The N400 to a word is reduced when it is preceded by a semantically related word (Deacon, Dynowska, Ritter, & Grose-Fifer, 2004), gesture (Wu & Coulson, 2005), or piece of music (Koelsch et al., 2004). In addition, the N400 to concrete words (e.g., “highway”) is larger than the N400 to comparable abstract words (e.g., “freedom”), unless the words are likely given previous context (Holcomb, Anderson, Kounios, & West, 1999). Moreover, non-word meaningful stimuli, such as pictures (Ganis, Kutas, & Sereno, 1996; McPherson & Holcomb, 1999), sign language signs (Kutas, Neville, & Holcomb, 1987), and gestures (Wu & Coulson, 2005), elicit N400-like effects, although the ERP morphology and topographies of these effects can be quite different from that of the canonical N400.

In contrast, the N400 is not generally elicited by non-semantic anomalies. For example, misspelled words (Münte, Heinze, Matzke, Wieringa, & Johannes, 1998), rare uppercase words (Kutas & Hillyard, 1980a; Kutas & Hillyard, 1980b; Osterhout, McKinnon, Bersick, & Corey, 1996), sentence internal ungrammatical words

(Coulson, King, & Kutas, 1998b; Hagoort, 2003b; Kolk, Chwilla, van Herten, & Oor, 2003; Kutas & Hillyard, 1983; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995; Osterhout & Nicol, 1999), and musical anomalies (Besson & Macar, 1987; Patel, Gibson, Ratner, Besson, & Holcomb, 1998) do no trigger N400s. However, the N400 to semantically incongruous words is somewhat modulated by their grammaticality (Friederici, Steinhauer, & Frisch, 1999; Hagoort, 2003a; Osterhout & Nicol, 1999)[2]. Thus the N400 is not strictly sensitive to semantic factors and word probabilities[3].

In light of such findings, there appears to be a general consensus that the N400 primarily reflects some aspect of probabilistic semantic processing that is not specific to language (Hagoort, 2003a; Kutas & Federmeier, 2000--but see Deacon, et al., 2004). Presently, more specific details about the function of the neural activity that generates the N400 are still uncertain. Some have suggested that the N400 indexes the retrieval of information from semantic memory (Kutas & Federmeier, 2000) and/or the integration of incoming semantic information with what has been previously comprehended (Friederici, 2002; Hagoort, Hald, Bastiaansen, & Petersson,

---

[2] Using sentence internal anomalies, Friederici et al. (1999) and Osterhout and Nicol (1999) both found that the N400 to incongruous, ungrammatical words was smaller than an additive effect of congruity and grammaticality would have predicted. Hagoort (2003a) and Gunter et al. (1997) found a similar trend for sentence final anomalies, but the trends were not statistically significant (each study involved 24 participants). In contrast, Hagoort (ibid.) found that the N400 to sentence internal incongruous, ungrammatical words was larger than would have been predicted by additive effects of congruity and grammaticality. However, this effect too failed to reach significance (p=.07).

[3] The final word of ungrammatical sentences is often followed by an N400-like effect (Friederici, Pfeifer, & Hahne, 1993; Hagoort, Brown, & Groothusen, 1993; Hagoort, 2003b; McKinnon & Osterhout, 1996; Münte, Szentkuti, Wieringa, Matzke, & Johannes, 1997; Osterhout & Holcomb, 1992; Osterhout & Holcomb, 1993; Osterhout & Mobley, 1995; Osterhout, 1997). If this effect is indeed related to the N400, it would provide additional evidence that the N400 is affected by non-semantic factors. However, to the best of my knowledge, there is no evidence that this N400-like effect is similar to the N400 in anything more than superficial appearance.

2004; Kutas & Federmeier, 2000). A pure integration hypothesis appears to be contradicted by the recent findings of DeLong and colleagues (2005) who found that the less expected member of a pair of indefinite articles (specifically "a" and "an") elicit a larger N400 than the more expected indefinite article. Since the meaning of "a" and "an" are effectively the same, the greater N400 to the less expected article cannot reflect the greater difficulty of integrating a less expected received word meaning into the current interpretation of the sentence. Thus, the N400, at least to some extent, seems to more generally reflect the update of the expected interpretation of an utterance.

II.C Features and Functional Correlates of the P600

The ERP component most reliably elicited by syntactically unexpected words is the P600. The canonical P600 (Osterhout & Holcomb, 1992) is elicited by ungrammatical words or words that are syntactically legal but dispreferred after a syntactically ambiguous sentence fragment. An example of the former would be the final word of:

E: *Every weekend, we go to the beach and* ***swims****.*

An example of the latter would be the word "to" in:

F: *The woman persuaded **to** answer the door got up from her chair.*[4]

Relative to the ERPs elicited by grammatically preferred control words, the ERPs to the anomalous or dispreferred word exhibits a positivity that is often largest at medial centro-parietal electrodes (e.g., Figures IV.37, IV.45: D) and generally peaks between 500 and 1000 ms after word onset[5].

The P600 has been elicited by many types of ungrammatical stimuli such as violations of subject-verb number agreement (e.g., Sentence E--Coulson, King, & Kutas, 1998b; Hagoort & Brown, 2000; Kemmer, Coulson, De Ochoa, & Kutas, 2004), pronoun case (Coulson, King, & Kutas, 1998b), and phrase structure (Friederici, Hahne, & Mecklinger, 1996; Hagoort & Brown, 2000; Hahne & Friederici, 1999; Neville et al., 1991; Patel et al., 1998). However, the topographies and latency of the P600 can vary somewhat from study to study (for an early review see Coulson, et al., 1998b).

Unlike the N400, it is not clear that the P600 is elicited by nonanomalous words. Although, grammatically dispreferred sentences like F are technically grammatical, there is evidence that they are frequently perceived by experimental participants as ungrammatical or incoherent (Osterhout, 1997; Patel et al., 1998)[6].

---

[4] Readers are strongly biased to interpret *persuaded* as the main verb of the sentence and expect a direct object to follow. However, *persuaded* is not the main verb of the sentence but the verb of a reduced relative clause. This is an example of what is known as a "garden path sentence" in psycholinguistics.

[5] The P600 was originally found to peak approximately 600 ms after word onset and was thus named "P600."

[6] Specifically, in Osterhout's second experiment approximately 92% of the grammatically dispreferred sentences were judged "unacceptable" (i.e., "semantically unusual or ungrammatical"). This is roughly equal to the percentage of semantically anomalous sentences that were judged unacceptable by the same participants in the same experiment. In Patel et al.'s study, 49% of grammatically dispreferred

Coulson and colleagues (1998b) found that grammatical words elicit P600-like effects when ungrammatical sentences are highly frequent within the testing session. However, as Coulson et al. mention, it is possible that this effect is distinct from the P600. More recently, Kaan and colleagues (2000) have shown that some grammatically preferred but relatively difficult constructions elicit P600-like effects. Although this P600-like effect was not significantly different from the canonical P600, their statistical analyses did not show compelling evidence that they were related either[7].

While it is not clear if the P600 is elicited by grammatical words, it is definitely more than simply an index of grammaticality. Several studies have found that the P600 is reduced when the eliciting ungrammatical word is also semantically anomalous (Friederici, et al., 1999; Gunter, et al., 1997; Hagoort, 2003b; Osterhout & Nicol, 1999) or presented in a deviant upper case font (Osterhout et al., 1996)[8]. The P600 is enhanced when ungrammatical words occur relatively infrequently in an experiment (Coulson, et al., 1998b; Gunter, Stowe, & Mulder, 1997; Hahne &

---

sentences were judged unacceptable (i.e., "semantically bizarre or grammatically incorrect"). This is much smaller than the percentage of truly ungrammatical sentences that were judged unacceptable (96%), but is still quite high. Unfortunately, Patel et al. did not report an analysis of the ERPs elicited by grammatically dispreferred sentences that were rated as acceptable. Thus the P600 to the grammatically dispreferred sentences in their data could be entirely due to the sentences perceived as ungrammatical.

[7] To determine if their "difficulty P600" was related to the grammaticality P600, Kaan and colleagues looked for evidence that the effects of difficulty and grammaticality interacted. No interactions were apparent when a conventional ANOVA was used to simultaneously analyze the potentials at rows of electrodes. ANOVAs performed on each individual channel did find an interaction at two electrodes. As Kaan et al. note, given the large number of comparisons (29) in this latter analysis, it is quite likely these interactions are false positives.

[8] The modulatory effect of semantic anomaly was not significant in Friederici et al.'s and Hagoort's studies. In the latter case, it appears that Hagoort would have found a significant effect had he used the most conventional measure of the P600 (i.e., raw ERP amplitude rather than a peak-to-peak measure). The modulatory effect of letter case in Osterhout and colleague's (1996) study also failed to reach significance.

Friederici, 1999), when grammaticality is directly relevant to the experimental task (Gunter & Friederici, 1999; Osterhout & Mobley, 1995; Osterhout et al., 1996), and when the grammatical anomaly is more salient (Coulson, et al., 1998b). Moreover, P600-like effects are elicited by musical anomalies (Patel et al., 1998), rare changes in letter case (Osterhout et al., 1996), misspelled words (Münte et al., 1998), and rarely by incongruous words (Curran, Tucker, Kutas, & Posner, 1993; Münte et al., 1998)[9], though some of these P600-like effects appear to have different topographies from the centro-parietal P600 (Section II.G.2).

In sum, these findings have led to a general consensus that the processes that generate the P600 are not specific to language (Coulson, King, & Kutas, 1998a; Friederici, 2002; Osterhout & Hagoort, 1999). However, there is less agreement as to what functions these processes index when engaged in the service of language comprehension. For example, Friederici and colleagues (2001) proposed that the P600 reflects a process of syntactic "diagnosis" and sometimes a process of "reanalysis" too. The diagnosis process functions to find an alternative parse once an utterance deviates from a preferred parse. Subsequently, the mental representation is modified to reflect the new parse via the reanalysis process. If the preferred parse is particularly difficult to revise, this reanalysis process also contributes to the P600.

[9] Other studies have found an insignificant trend for an enhanced positivity following the N400 to incongruous words (Ganis et al., 1996; Holcomb, Kounios, Anderson, & West, 1999; Kutas & Hillyard, 1980b). However, several studies have found no such trend (Friederici et al., 1993; B. W. Johnson & Hamm, 2000; Kutas & Hillyard, 1980a; Kutas & Hillyard, 1984; Kutas, 1987; Osterhout, 1997; Osterhout & Nicol, 1999; Van Petten, 1993). Thus the few significant instances of this effect may be spurious or the effect must be more sensitive to task demands and/or stimulus properties than the N400.

Alternatively, Kaan and colleagues (2000) have argued that the P600 partially reflects a more general process of "syntactic integration." Specifically they hypothesize that the brain makes predictions about the syntactic categories to come and that the P600 reflects the difficulty of combining novel input with these predictions. They proposed that this process is distinct from diagnosis and revision processes, though they suggest that these too may contribute to the P600 when the eliciting word is ungrammatical.

In contrast to these proposals, which would involve the manipulation of grammatical representations, Coulson and colleagues (1998b) suggest that the P600 indexes the update of a participant's expectations for ungrammatical words. This hypothesis is largely based on the similarities between the P600 and the P3b ERP component reviewed next, and is discussed in detail in Section II.G.2.

II.D Features and Functional Correlates of the P3b

The P3b (sometimes called the "P300" or "P3") was perhaps the first ERP signature of the brain's sensitivity to the probability of the occurrence of a stimulus to be discovered (Sutton, Braren, Zubin, & John, 1965). And it is most likely the single most studied ERP phenomenon (for review see Donchin & Coles, 1988; Verleger, 1997; Kok, 2001; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Polich, 2004).

It is canonically elicited by rare "target" stimuli in a Bernoulli sequence of common "standard" stimuli (e.g., infrequent 2000 Hz tones among a series of 1000

Hz tones--Gonsalvez & Polich, 2002), which experimental participants are instructed to silently count or indicate with a motor response. This simple target detection task is known as the "oddball paradigm" and the ERPs elicited by rare targets (i.e., "oddballs") exhibit a large positivity, largest at centro-parietal electrodes, that peaks between 300 and 900 ms [10] (e.g., Figures IV.12, IV.45). The P3b is elicited by multiple modalities of oddball stimuli (e.g., auditory--Gonsalvez & Polich, 2002; visual--Kutas, McCarthy, & Donchin, 1977; and somatosensory--Yamaguchi & Knight, 2001), and by oddball distinctions that range from the concrete to the abstract (e.g., font size, words that rhyme with a target word, female names vs. male names, synonyms of the word "prod"--Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Kutas et al., 1977). Moreover, similar positivities are elicited by the omission of a probable stimulus (Sutton et al., 1965), task-irrelevant unexpected changes in letter case in a sentence reading task (Kutas & Hillyard, 1980a; Kutas & Hillyard, 1980b; Osterhout et al., 1996), learning nonsense syllable pairings (Horst, Johnson, & Donchin, 1980), and learning artificial categories of geometric shapes (De Swart, Kok, & Das-Smaal, 1981). Although it seems unlikely that all of these late positivities are completely identical (Section II.E.4), the functional and superficial similarities between these ERP effects suggest a common, general-purpose neural process.

Whatever this process is, its average magnitude is a monotonically decreasing function of the probability of the eliciting stimulus. For example, P3b amplitude

[10] The P3b has been frequently studied using a tone oddball paradigm. In this paradigm the P3b peaks around 300 ms, hence the "3" in "P3b" and the name "P300."

strongly negatively correlates with the probability of targets in an oddball paradigm (Duncan-Johnson & Donchin, 1977). In fact, when targets are the norm, standards can elicit larger P3bs (ibid.). Importantly, P3b amplitude does not simply reflect the "objective probability" of a stimulus (i.e., its relative frequency), rather it appears to index a participant's belief that the stimulus will occur (i.e., its "subjective probability"). This is suggested by the fact that the P3b to a stimulus is smaller when participants predict that it will occur (Sutton et al., 1965) and by the way P3b amplitude is affected by a participant's confidence in her prediction of the eliciting stimulus. Specifically, stimuli that confirm a confident prediction elicit smaller P3bs than stimuli that disconfirm that prediction, but the opposite is true when the participant lacks confidence (De Swart et al., 1981; Horst et al., 1980). Finally, in oddball tasks, P3b amplitude positively correlates with the number of standards preceding the eliciting target (Gonsalvez & Polich, 2002). This effect of target sequence (also known as "target-to-target interval") has been interpreted as an effect of subjective probability (Donchin & Coles, 1988).

In addition to subjective probability, P3b amplitude is also strongly affected by attention. Stimuli that would elicit a large P3b in an oddball task fail to do so when participants are instructed to ignore them while performing some other task (e.g., reading--Spencer, Dien, & Donchin, 2001). Instead, such stimuli elicit a more frontal "P3a," discussed in more detail below (Section II.E.4.a). Similarly, when participants perform an oddball task in addition to another task, P3b amplitude to oddballs is reduced (Wickens, Kramer, Vanasse, & Donchin, 1983). In fact, there is

some evidence that P3b amplitude across two simultaneous tasks is "conserved." For example, when stimuli in two simultaneous tasks both elicit a P3b and are presented in distinct sensory modalities (e.g., auditory and visual), increasing the difficulty of the primary task increases the amplitude of the P3b to events in the primary task, and decreases the P3b amplitude to secondary task events (Sirevaag, Kramer, Coles, & Donchin, 1989).

Along these lines, P3b amplitude is also enhanced by the "motivational significance" of a stimulus. For example, when targets and standards are equally likely, targets elicit a larger P3b (Duncan-Johnson & Donchin, 1977) and they elicit a P3b in a greater fraction of individuals (Fabiani, Gratton, Demetrios, & Donchin, 1987). Moreover, emotionally charged stimuli (positive or negative) elicit larger P3bs than emotionally neutral stimuli (Johnston, Miller, & Burleson, 1986; Keil et al., 2002).

A final critical feature of the P3b is that its latency positively correlates with the time necessary to categorize the eliciting stimulus (Bentin et al., 1999; Kutas et al., 1977). For example, the P3b elicited by a target class of letter case (upper or lower) occurs earlier than the P3b elicited by a target class of word meaning (e.g., "animals" versus "colors"), even though the stimuli used to elicit both P3bs are identical (Section IV.E.3.a). Similarly, the addition of obscuring "noise" stimuli slows P3b latency (McCarthy & Donchin, 1981), and, when speed is not emphasized over accuracy, P3b latency positively co-varies with reaction time (Kutas et al., 1977; Makeig et al., 2004). In contrast, P3b latency only weakly correlates with reaction

time when speed is emphasized over accuracy (Kutas et al., 1977) and reaction time but not P3b latency is slowed when the difficulty of a task's motor response is increased (Ragot & Renault, 1981--but see Falkenstein, Hohnsbein, & Hoormann, 1994). These findings suggest that the P3b is primarily an index of perceptual processing or its completion (as opposed to an index of motor processes like response selection).

Based largely on these key properties of the P3b, several thoughtful proposals on the function indexed by the P3b have been developed over the years. Perhaps the most popular is the theory that the P3b reflects an update of a model of the current environment in limited-capacity working memory and that the amplitude of the P3b correlates with the amount of change to the model (Donchin & Coles, 1988). This theory is consistent with the strong inverse relationship between P3b amplitude and subjective stimulus probability, the correlation between P3b latency and speed of stimulus evaluation, and evidence that total P3b amplitude is sometimes conserved across two different perceptual tasks when participants perform them simultaneously. While this interpretation is often agnostic as to the specific computational and neural mechanisms that implement the model or its update, Karl Friston (2005) has recently proposed a general framework of cerebral cortical function that could provide such details. Specifically, Friston argues that the P3b is generated by a mismatch between incoming sensory information and the information that was predicted by a probabilistic cortical generative model. This mismatch is then used by the generative model to improve future predictions. While Friston's interpretation of the P3b is still

highly speculative, the general premise of the theory is well-motivated computationally (Hinton, 1999; Hinton, 2000), has some detailed neurological support, and is consistent with a large body of behavioral evidence that humans perform probabilistic inference when perceiving (Knill & Pouget, 2004).

An alternative interpretation of the P3b is that it reflects a match between a perceived stimulus and an internal representation of an awaited stimulus (Kok, 2001; Verleger, Jaskowski, & Wauschkuhn, 1994). Like the "update" hypothesis, this explanation is consistent with the correlation between P3b latency and speed of stimulus evaluation. Moreover, it can explain why target stimuli elicit larger P3b's than equiprobable standards in oddball tasks. However, it is blatantly contradicted by the inverse relationship between P3b amplitude and the subjective probability of an eliciting stimulus.

Finally, Nieuwenhuis and colleagues (2005) have suggested that the P3b reflects an increase in neural activity enabled by the release of norepinephrine (NE) onto cortical pyramidal cells by the locus coeruleus. This enhanced excitability of cortex "acts to optimize information processing by modulating post decision responses processes" (ibid., pg 523). This theory differs from the aforementioned hypotheses in that the P3b does not reflect a perceptual process. Rather, Nieuwenhuis et al. believe that stimulus categorization occurs much earlier than the P3b and triggers a response by the locus coeruleus that facilitates a response by cortex to the already categorized stimulus event. While this theory has the most detailed neurological support, it appears to be inconsistent with the fact that the P3b can co-

occur with or occur after a motor response to the eliciting stimulus (Kutas et al., 1977; Makeig et al., 2004). Nieuwenhuis and colleagues suggest that the onset of NE facilitation precedes the peak of the P3b and even its onset, thus NE facilitation may still reliably precede participant responses in P3b paradigms.

II.E Evidence for N400, P600, and P3b Subcomponents

While the preceding reviews of the N400, P600, and P3b discussed these phenomena as if they were more or less each a unitary neural response, it is generally believed that they are composed of spatiotemporally overlapping subcomponents generated by multiple, functionally distinct parts of cortex (Coulson, King, & Kutas, 1998b; Kutas & Federmeier, 2000; Halgren, Marinkovic, & Chauvel, 1998; Moores, Clark, Hadfield, Brown, Taylor, Fitzbibbon, Lewis, Weber, & Greenblatt, 2003; Osterhout & Hagoort, 1999; Van Petten & Luka, 2006). The following sections review evidence that this is the case and discuss what some subcomponents of these ERP effects might be.

II.E.1 Types of Evidence

Determining the number of somewhat independent neural sources that contribute to an ERP component is complicated by the fundamentally ambiguous relationship between neural generators and scalp potentials (Section III.D) and by ERP estimation error. To ameliorate these complications, it is valuable to consider

evidence from multiple methods for studying brain function. Unless ERP estimation error is remarkably low, none of these methods alone can prove that an ERP component is complex. However, converging evidence from multiple methods can make a very compelling case for complexity and potentially identify some of these sources.

Unambiguous evidence for multiple subcomponents of an ERP component would be changes in the topography of the component across time or in response to experimental variables (Section III.B.1). Changes in ERP topographies are conventionally tested for using ANOVAs after normalizing the magnitude of the topographies (Urbach & Kutas, 2002). Unfortunately, this method is problematic as it too frequently confuses topography differences due to estimation error with true topography differences when the two sets of scalp potentials compared differ in overall magnitude (i.e., the test has an inflated Type I error rate, ibid.). However, if topography differences prove to be reliable (i.e., they replicate across multiple experiments), then they are most likely real.

Principal components analysis (PCA) of individual participant ERPs is an alternative approach for demonstrating changes in scalp topography (Section IV.E.3.d). However, it too may conflate changes in topographies due to ERP estimation error with true topographic differences.

ERP source localization is another method for potentially identifying ERP subcomponents. If an ERP component cannot be reasonably modeled with a single, focal source or a distributed collection of highly correlated sources, it suggests that

the component is complex. However, source localization results may be inaccurate due to erroneous source assumptions and/or model optimization (Baillet, Mosher, & Leahy, 2001) and they also may confuse ERP estimation error with true sources of the component. With regard to source localization, it is important to note that the localization of "grand average" ERPs (i.e., ERPs that have been derived from multiple participants) is likely to underestimate the complexity of an ERP component, as averaging across participants tends to act as a spatial low-pass filter that blurs ERP scalp distributions (Kayser & Tenke, 2006).

Independent component analysis (ICA--Chapter III) can be thought of as a statistically based analog of source localization. If ICA decomposes an ERP component into multiple subcomponents, this suggests that the component is complex. However, like source localization, ICA may confuse ERP estimation error with true sources of the component and its decomposition may be spurious due to invalid assumptions, inadequate training data, or local minima during training.

Intracranial EEG (iEEG) recordings in human patients and animals can also provide evidence that an ERP component is complex. iEEG can identify intracranial analogs of scalp ERP components (iERPs) localized to specific regions of cortex (Halgren, Marinkovic, & Chauvel, 1998). Multiple, functionally distinct iERP analogs of an ERP component would suggest multiple ERP subcomponents. However, it is difficult to know how much a given iERP component contributes to a scalp ERP component as its amplitude may be too weak to be measured at the scalp and its contribution may be obscured by other EEG sources. Furthermore, iEEG

recordings in patients and animals may not be representative of the activity of healthy human brains.

Functional magnetic resonance imaging (fMRI) can provide more indirect support for the presence of ERP subcomponents by identifying hemodynamic analogs of an ERP component. If the same type of experimental manipulations that elicit an ERP component reliably elicit disparate hemodynamic responses in multiple parts of cortex, then it is possible that those regions are also generating multiple ERP subcomponents. However, the neural activity that generates a hemodynamic response may not significantly contribute to an ERP component for several reasons (e.g., Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Gandhi, Heeger, & Boynton, 1999). For example, given that the temporal resolution of fMRI is on the order of seconds (Dale et al., 2000) the activity may occur before or after the ERP component. In addition, the activity may be too far from electrodes to contribute significantly[11] and/or its contribution to the ERP may be effectively cancelled out by sources of opposite polarity.

Finally, lesions in human patients and animals offer a very different type of evidence for multiple ERP subcomponents. If lesions to multiple parts of the brain affect an ERP component, that suggests that the component is complex. In contrast, if a lesion that is well-circumscribed to a particular region of cortex eliminates an ERP component, that suggests that the component is not complex. However, lesion research in general is difficult to interpret, as lesions may disrupt the generation of the

[11] The strength of the electric potential produced by an EEG generator is an inverse function of the squared distance between the generator and the electrode (Kutas & Dale, 1997).

component indirectly, rather than indicate the source of that component. For example, removal of the eyes will abolish the N400, P600, and P3b elicited by visual stimuli, but obviously the eyes do not generate these phenomena. Like iEEG, lesion research is also complicated by the possibility that human patient and animal brains are not representative of the function of healthy human brains.

II.E.2 N400 Subcomponents

II.E.2.a Evidence for N400 Subcomponents

Twenty-seven years after the discovery of the N400, there is good evidence that it is composed of multiple subcomponents, but it is far from clear what those subcomponents might be (for review see Van Petten & Luka, 2006). The strongest evidence for multiple N400 subcomponents is the fact that concrete words, in the absence of supportive context, reliably elicit a more frontally distributed N400 than abstract words (Kounious & Holcomb, 1994; Holcomb, Kounious, Anderson, & West, 1999; Swaab, Baynes, Knight, 2002; West & Holcomb, 2000; Zhang, Guo, Ding, & Wang, 2005). Given that the effect has replicated across multiple studies, the topographic differences are most likely not an artifact of the aforementioned methodological problems for assessing such differences and indicate that a single source cannot account for the N400 to both concrete and abstract words. As the N400 is frequently measured without regard to the concreteness of the eliciting stimuli, conventional measures of the N400 surely include multiple components that are

dissociated somewhat by concreteness. The N400 to pictures and line drawings are also reliably more frontally distributed than that to words (Ganis et al., 1996; McPherson & Holcomb, 1999). If icon and text N400s share generators, then this would also mean that the N400 is complex, however, it is possible that the two effects are anatomically distinct. The same can be said of the topographic differences between text N400s and speech N400s (the N400 to spoken words is generally bilaterally symmetric--Van Petten & Luka, 2006).

This topographic evidence is corroborated by what is probably the best[12] attempt to localize the magnetoencephalographic[13] analog of the canonical N400 (the N400m) in individual participants (Halgren et al., 2002). This study found that a single focal source or bilateral pair of sources (in the superior temporal sulcus/sulci) could not adequately account for the N400m measured at 122 MEG sensors (i.e., a single equivalent current dipole--ECD--or pair of ECDs only accounted for 56% of the N400m variance on average). When the same data were modeled using thousands of sources, the N400m appeared to be generated by multiple somewhat independent

[12] Other attempts to localize the N400m have been less compelling. Kwon et al. (2005) and Simos et al. (1997) both attempted to localize the N400m using less than 50 sensors limited to the left hemisphere. Furthermore, Kwon et al. localized the grand average N400m, which is problematic (see Section II.E.1). Helenius and colleagues (1998, 2002) have reported two attempts to localize the N400m using full head MEG sensor coverage. However, they do not report how much of the N400m variance was accounted for by their models so their results do not inform us as to the minimal number of sources of the N400m. Moreover, their results are less consistent across participants than those of Halgren et al. and they used less sensors in the 1998 report than used by Halgren and colleagues for localization (Helenius and colleagues do not report the size of the subset of sensors used for localization in their 2002 report).

[13] The EEG is closely related to the magnetoencephalogram (MEG). Both are primarily generated by the synaptic activity of the pyramidal cells of cerebral cortex (Baillet et al., 2001; Kutas & Dale, 1997). MEG is much less sensitive to deep sources and radial sources (sources perpendicular to the scalp) than EEG and EEG is less sensitive to tangential sources than MEG (ibid.), but the generators of analogous EEG and MEG phenomenon may overlap enough that the localization of one can inform our understanding of the sources of the other (Liu et al., 2002).

generators, mostly in the left frontal and temporal lobes. However, one of the two published attempts to localize the N400 appears to be somewhat at odds with the results of Halgren and colleagues. Specifically, Koelsch et al. (2004) found that a bilateral pair of ECDs (in the posterior middle temporal gyri) could account for 90% of the variance of the individual semantic priming N400 effects (on average). Although this suggests that the N400 could be simply generated by a pair of focal sources, Koelsch et al. could have underestimated the number of sources needed to model the N400 as they measured the N400 effect at only 45 electrodes and only modeled a single time point of the effect (the time point that was best explained by the pair of sources in a pre-localization parameter initialization). The other published attempt to localize the N400 using approximately 30,000 sources based on individually derived head models did not find consistent results across participants (Haan, Streb, Bien, & Rosler, 2000).

Less direct evidence for the complexity of the N400 comes from fMRI and lesion studies. Event-related fMRI (erfMRI) studies of N400 paradigms have typically found that incongruous words generally elicit increased hemodynamic activity relative to congruous control words in multiple cortical regions, although there is a good deal of variability in the identify of these regions from study to study (Van Petten & Luka, 2006). In a recent review of 12 such studies, Van Petten and Luka (ibid.) found that the cortical region most frequently reported to show an increase in hemodynamic activity to incongruous words was the left inferior frontal gyrus, and this effect was found only 75% of the time. The second most frequently

reported region, the left superior temporal gryus/planum temporale, was found in only 58% of the twelve studies. Converging evidence of a large temporal lobe and/or temporo-parietal junction contribution to the N400 come from lesion studies (ibid.). However, there is some evidence that left frontal lobe lesions reduce the N400 to semantic anomalies when patients are asked to detect anomalous sentences (Friederici, von Cramon, & Kotz, 1999)[14].

To date, the most convincing iERP analog of the canonical N400 was found bilaterally in the anterior medial temporal lobe (McCarthy, Nobre, Bentin, & Spencer, 1995; Nobre, Allison, & McCarthy, 1994). In the later report, the authors state that although they sampled "virtually the entire cortical mantle," no other regions reliably produced N400 analogs (pg. 1089). While this may seem like evidence against multiple subcomponents of the N400, other studies using lists of words have found N400 analogs in other regions of cortex (Elger et al., 1997; Halgren et al., 1994). Moreover, anterograde amnesics with damage to the anterior medial temporal lobes show normal N400 semantic context and repetition effects (Olichney et al., 2000). Thus the anterior medial temporal lobes must contribute a subcomponent of the N400 if they contribute at all.

[14] Specifically, Friederici and colleagues found that a group of three patients with lesions including the left frontolateral cortex and parts of the basal ganglia did not show a significant N400 effect, although the control group of six participants did show some signs of a significant N400 effect (a Condition X Electrode interaction). However, all three patients did show a trend for an N400 effect (their Figure 5), so the lack of significance may simply be due to the smaller group size.

II.E.2.b Possible N400 Subcomponents

Despite the abundant evidence suggesting that multiple subcomponents comprise the N400, no widely accepted subcomponent has yet been identified. Based on the frontal topographies of the N400 to concrete words and pictures, some investigators have proposed an "imagistic" subcomponent that reflects the use of mental imagery in language comprehension (Holcomb et al., 1999). Other researchers (Chung et al., 1996) have proposed that the N400 partially reflects a delay in the P3b, a possibility that is discussed in more detail below. Van Petten and Luka (2006) have also suggested that the N400 overlaps with a late posteriorly distributed positivity but they did not equate this potential with the P3b nor did they attribute a function to this positivity. They simply proposed that it is generated by the frontal lobes. Besson and Kutas[15] (1993) have suggested that the early phase of the N400 (200-400 ms) reflects the degree to which the eliciting word is expected, while the later phase of the N400 (400-675 ms) reflects expectancy violation or the integration of the word into its preceding context. Given that the N400 elicited by incongruous words is often modulated by the grammaticality of that word, the N400 may share subcomponents with the P600.

Based largely on hemodynamic evidence, several researchers have proposed anatomically distinct processes underlying semantic aspects of language. For example, Peter Hagoort (2005) has proposed that the left temporal lobes serve to access information from semantic memory and that the frontal lobes function to

[15] Kutas (personal communication

"unify" this information with preceding context. Along similar lines, Gabrieli and colleagues (1998) proposed that left prefrontal cortex supports semantic working memory. In contrast, Kuperberg et al. (2003) suggested that medial parietal cortex is *deactivated* as a result of focused attention during semantic processing. However, links between these functions and possible subcomponents of the N400 remain speculative.

II.E.3 P600 Subcomponents

II.E.3.a Evidence for P600 Subcomponents

The evidence for multiple subcomponents of the P600 is weaker than that for the N400 as less is known about its neural substrates (Friederici & Kotz, 2003). There are many examples in the literature of changes in P600 topography across studies (for an early review see Coulson, et al., 1998b) or even across time (Hagoort & Brown, 2000). However, unless specific factors can be identified that systematically change the topography of the P600 it is likely that these topographic differences simply reflect ERP estimation error or between-participant variation (e.g., Section IV.F.3.f, Figures IV.19-20). To the best of my knowledge such a factor has not been well established although some proposals have been made. For example, Hagoort and Brown (Experiment 1, ibid.) found that the P600s elicited by two types of grammatical violations were rather uniformly distributed across the scalp during the 500-750 ms post-stimulus window and posteriorly distributed during the 750-1000 ms

window. They tentatively suggested this difference revealed two different subcomponents of the P600, an earlier (500-750 ms) syntactic integration subcomponent and a later (750-1000 ms) parse failure/reanalysis subcomponent (these functions are discussed in the next section). However, this topography pattern was not found to a third type of grammatical violation in their experiment (subcategorization violations) and it was only found to one of the three violation types (subject-verb number agreement violations) in a previous study that used the exact same stimuli with a somewhat different experimental procedure (Hagoort, Brown, & Groothusen, 1993)[16]. Moreover, three experiments by another research group (Coulson et al., 1998; Kemmer et al., 2005) found that subject-verb number agreement violations did not produce such a topography pattern (these P600 effects were largest at posterior electrodes during the time range that Hagoort and colleagues reported a uniformly distributed P600). It is possible that methodological differences are the root of the discrepant results across studies (e.g., Hagoort and Brown used a very rapid stimulus onset asynchrony in Experiment 1 of the 2000 report). But until the discrepancies can be explained and Hagoort and Brown's proposed subcomponents shown to covary with factors that differentially affect syntactic integration and reanalysis, it is difficult to know if the topography differences reflect anything more than noise.

To the best of my knowledge, there are no published attempts to localize the P600 and only one attempt to localize the magnetoencephalographic analog of the

[16] Hagoort et al. did not report statistical tests of the P600 topographies. My statements are based on the grand average ERPs from that report.

P600, the mP600 (Kwon et al., 2005). Kwon and colleagues fit the grand average mP600 (i.e., the mP600 averaged across all seven participants) with a single source in the left middle temporal gyrus with a residual variance of 4.3%. This appears to indicate a single P600 generator. However, as already mentioned averaging over participants tends to blur and simplify the topographies of the brain's electric and (presumably) magnetic fields. Moreover, Kwon et al. measured the magnetoencephalogram (MEG) at only 40 locations limited to the left side of the head. Thus the localization results are most likely underestimating the number of P600 sources.

Friederici and colleagues (2001) decomposed the P600 using spatial and temporal PCA (Sections III.F & III.H) and identified two spatial principal components, "SF1" and "SF3," of the P600. SF3 projected to occipital electrodes and was interpreted as a subcomponent specific to the "diagnosis" of parsing problems. SF1 projected most strongly to centro-parietal electrodes and was interpreted as a domain general P3b subcomponent whose activity reflects parse revision (these are the proposed diagnosis and reanalysis functions discussed in Section II.C). It is not clear if this analysis provides any evidence for the complexity of the P600. As the analysis was performed on ERPs from multiple participants it could be conflating between-participant variance and true P600 variance[17]. Moreover, the authors provided no evidence that the derived principal components were reliable and it is not

[17] For example, if some participants have a centro-parietally distributed P600 and others an occipital P600, this could lead spatial PCA to find separate centro-parietal and occipital components of the P600. If this hypothetical between participant difference is due to functionally irrelevant anatomical differences, then these two principal components would not be evidence of subcomponents of the P600.

clear if the components are any more meaningful than a random rotation of the principal components (Section III.E.4). The topographies of SF1 and SF3 appear physiologically plausible. However, the topographies were interpolated from only 26 electrodes and the topography of SF1 appears quite different from those of "P3b principal components" derived by some of the authors from other data sets (Dien, Spencer, & Donchin, 2003; Spencer, Dien, & Donchin, 2001). Moreover, it is odd that the activity of the diagnosis subcomponent has the exact same time course as the revision subcomponent[18], since the activity of the former is supposed to trigger that of the latter. Clearly, this analysis should be interpreted only tentatively until validated with additional evidence (Sections III.E.4.c-d).

Event-related fMRI studies of the brain's response to ungrammatical words are the strongest evidence for multiple subcomponents of the P600. Like semantically incongruous words, ungrammatical words modulate hemodynamic activity in multiple regions of cortex. Some studies have found increased parietal (Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Kuperberg et al., 2003), superior frontal (Newman, Pancheva, Ozawa, Neville, & Ullman, 2001), superior temporal (Friederici et al., 2003; Ni et al., 2000), and inferior frontal (Friederici et al., 2003; Ni et al., 2000) activity for ungrammatical words. Such work has led some investigators to hypothesize that syntactic processing is primarily mediated by the left inferior frontal

[18] After applying spatial PCA to their data, Friederici and colleagues performed temporal PCA on the activations of some of their spatial components. For their subject-object ambiguous complement clauses, the activation of SF1 and SF3 to grammatically preferred and dispreferred sentences differed in the same temporal factor, TF7. Friederici et al. proposed that the SF1-TF7 difference indexed revision and that the SF3-TF7 difference indexed diagnosis. Since the waveform of the two effects are exactly the same (i.e., TF7) these differences occur with the exact same time course.

and superior temporal gyri (Friederici & Kotz, 2003; Hagoort, 2005). However, there is also erfMRI evidence that ungrammatical words can actually lead to a decrease in activity (relative to nonanomalous control words) in the left inferior frontal gyrus and left superior temporal sulcus (Kuperberg et al., 2003). Thus while the hemodynamic response to ungrammatical words is clearly distributed, variability in this response across studies makes it difficult to tie to possible subcomponents of the P600.

II.E.3.b Possible P600 Subcomponents

Despite the dearth of convincing evidence, there appears to be a general consensus that the P600 is composed of multiple subcomponents (Coulson, King, & Kutas, 1998a; Kaan et al., 2000; Osterhout & Hagoort, 1999). As discussed in detail below, there is some evidence that the P3b contributes to the P600, though it is not clear if this contribution is a subcomponent of the P600 (Osterhout & Hagoort, 1999) or if it is the entirety of the P600. As just mentioned, Friederici and et al. (2001) have proposed separate diagnosis and reanalysis subcomponents. Hagoort and Brown's (2000) previously mentioned parsing failure/revision subcomponent appears to be functionally similar to the subcomponents proposed by Friederici and colleagues. Moreover, the posterior distribution of Hagoort and Brown's subcomponent might be consistent with the distribution of Friederici et al.'s diagnosis subcomponent or the combination of their diagnosis and reanalysis subcomponents. As discussed in Section II.C, Kaan et al. (2000) have suggested a syntactic integration subcomponent, which is posteriorly distributed. Hagoort and Brown (2000) have also put forward

such a subcomponent (see previous section), though they hypothesized that it is rather uniformly distributed across the scalp. Finally, the fact that the P600 is modulated by semantic congruity suggests that the P600 may share subcomponents with the N400 as well (Hagoort, 2003b; Hagoort & Brown, 2000).

Based largely on erfMRI findings, some investigators have proposed anatomically distinct processes that might contribute to the P600. Peter Hagoort (2005) has suggested mechanisms for syntactic processing that are similar to his previously mentioned proposal for semantic processing. Specifically, he believes that the left superior temporal gyrus accesses syntactic information from memory while the left inferior frontal gyrus "unifies" this information into larger syntactic structures. Kuperberg and colleagues (2003) have speculated that grammatical anomalies lead to a decrease in focused attention (relative to non-anomalous sentence comprehension), reflected in an increase in parietal activation. Links between these functions and possible subcomponents of the P600 remain unproven.

II.E.4 P3b Subcomponents

II.E.4.a Evidence for and Identity of Possible P3b Subcomponents

Discussing subcomponents of the P3b is terminologically complicated as it is commonly believed that the P3b often co-occurs with other potentially overlapping ERP components. For example, some researchers have proposed overlap with a "P3a" that reflects an orienting response (Debener, Makeig, Delorme, & Engel, 2005;

Halgren et al., 1998; Knight & Scabini, 1998; Spencer et al., 2001), a late frontal negative "slow wave" (Squires, Squires, & Hillyard, 1975), a "frontal P3" that indexes speeded responding (Makeig et al., 2004), a slow negativity reflecting "controlled search processes" (Wijers, Mulder, Okita, & Mulder, 1989), and a "positivity-choice response" component that is correlated with response selection or evaluation (Falkenstein et al., 1994)[19]. While these phenomena are often referred to as separate ERP components, research on the P3b often makes no attempt to separate them from the P3b. Thus, by the definitions provided at the start of this chapter, these other phenomena could be considered subcomponents of the P3b if indeed they exist.

There is a large body of converging evidence that the P3b is complex. For example, it is likely that there are subcomponents of the P3b specific to particular sensory modalities. Direct evidence for this is slight topographic differences between auditory and visual P3b's (Ji, Porjesz, Begleiter, & Chorlian, 1999; R. Johnson Jr, 1993) and between P3b's elicited by stimulation of the left or right hand (Bruyant, García-Larrea, & Mauguière, 1993). There is also evidence of iERPs elicited by auditory but not visual oddballs (Halgren et al., 1998) and lesions of the superior temporal and inferior parietal cortex abolish the canonical auditory and tactile P3b, but only diminish the visual P3b (Knight & Scabini, 1998).

[19] Specifically, Falkenstein and colleagues proposed that the P3b is composed of two subcomponents when there are multiple possible motor responses in an experimental paradigm: the "P-choice response" and the "P-stimulus response." Here I am equating the P-stimulus response with the P3b since both are elicited when only one motor response is required of participants, the topographies of both appear to depend on the modality of the eliciting stimulus, and the latencies of both are relatively invariant to the difficulty of the motor component of an experimental paradigm.

There is also converging evidence for the aforementioned P3a (sometimes called the "novelty P3")[20] subcomponent. The P3a is canonically elicited by task irrelevant deviant stimuli (e.g., an ignored stream of oddball stimuli--Spencer, Dien, & Donchin, 2001). It peaks slightly earlier than the P3b and has a somewhat more frontal distribution (ibid.). In a review of a large number of iEEG recordings from P3a and P3b paradigms, Halgren, Markinkovic, and Chauvel (1998) suggested that the P3a is being generated by the supramarginal, dorsolateral prefrontal, and cingulate gyri. This roughly converges with evidence that lesions of prefrontal and temporo-parietal cortex in humans diminish the P3a (Knight & Scabini, 1998). ICA and PCA have also found separate P3a-like and P3b-like components (Debener et al., 2005; Spencer et al., 2001)[21] and both contributed to the canonical P3b.

The other aforementioned subcomponents of the P3b are less well established, but iEEG and fMRI studies of P3b paradigms suggest that disparate regions of cortex contribute to the generation of the P3b. Thus there may well be subcomponents of the P3b besides the P3a. In the previously mentioned review of iERP P3b analogs, Halgren and colleagues (1998) suggested that the P3b is generated primarily by ventrolateral prefrontal cortex, posterior superior parietal cortex, the superior temporal sulcus, and the anterior medial temporal lobes. Both intracranial recordings from and

---

[20] The P3a is canonically elicited by ignored oddball stimuli (e.g., rare tones in a series of standard tones), while the "novelty P3" is canonically elicited by complex task irrelevant stimuli (e.g., a dog bark—Spencer, Dien, & Donchin, 2001). While it is possible that the two components are somehow distinct, the two are sometimes conflated (Friedman, Cycowicz, & Gaeta, 2001) and there is some evidence that the two are the same (Debener et al., 2005; Spencer, et al., 2001).

[21] It is worth noting that neither study demonstrated that their "P3a-like" subcomponents habituated to novel stimuli over time, a defining trait of the P3a. The P3a-like independent components of Debener et al. showed a tendency to habituate that failed to reach significance. Spencer and colleagues do not report an analysis of habituation.

lesions of the locus coeruleus in monkeys suggest that the P3b may reflect the output of the locus coeruleus, which projects widely to inferior parietal, superior temporal, and prefrontal cortex (Nieuwenhuis, Aston-Jones, & Cohen, 2005). In an auditory oddball fMRI study involving 100 participants, Stevens and colleagues (2005) found that the activity of 23 cortical regions correlated with target-to-target interval. The regions were mostly in the frontal and parietal lobes and included the supramarginal gyri, medial frontal gyri, and insula—all of which are regions whose activity Horovitz et al. (2002) had also found to be correlated with target-to-target interval in a much smaller, comparable study. In this smaller auditory oddball study, Horovitz and colleagues also found that the activity in the right supramarginal gyrus and left insula correlated with P3b amplitude from a separate experimental session. In contrast, Eichele et al. (2005) found that the activity of the left supramarginal gyrus and right insula, as well as a few other frontal, parietal, and temporal sites, predicted simultaneously recorded P3b amplitude on single-trials of an auditory oddball paradigm.

Despite considerable evidence for the widely distributed generation of the P3b, it is possible that most of these regions are not significant contributors. Human patient evidence indicates that the P3b at parietal electrodes is not significantly affected by prefrontal lesions (for simple oddball tasks) or damage to the anterior medial temporal lobes (Knight & Scabini, 1998). Moreover, models of individual participant P3bs based on data from 20 to 43 electrodes have found that only two to four focal neural sources (equivalent current dipoles) are typically necessary to

account for more than 92.5% of the P3b variance (Goto, Brigell, & Parmeggiani, 1996; Tarkka & Stokic, 1998). These studies could have underestimated the number of sources due to their meager electrode coverage. However, an attempt to localize a text-elicited P3b using 124 electrodes and individually derived models of the entire cortical mantle (Moores et al., 2003) also suggested a rather concentrated generation of the P3b. Specifically, Moores et al. found that the P3b primarily localized to the superior parietal lobes with small contributions from the left occipital lobes and left fusiform gyrus. Their relatively focal results are remarkable given that the source model assumed approximately 38,000 sources. This could be explained by the fact that their data was pre-processed by a Laplacian spatial filter that attenuates deep and/or diffuse sources.

II.F Subcomponentry and Problems with Conventional ERP Analysis

The high likelihood that the N400, P600, and P3b are composed of multiple subcomponents complicates the interpretation of these phenomena and limits their ability to inform cognitive neuroscience theories. For example, Peter Hagoort (2003b) attempted to use the N400 and P600 to determine if semantic and syntactic information are processed independently or if one type of information is processed first and can then influence the processing of the other. This issue has been a matter of substantial debate in psycholinguistics (Friederici, 2002) and is well suited to the

impeccable temporal resolution of ERPs. To test the question, Hagoort had experimental participants read sentences containing ungrammatical words, semantically anomalous words, and words that were both ungrammatical and semantically anomalous. If semantic and syntactic anomalies have independent effects, then the effect of the double anomaly would equal the sum of the effects of the single anomalies[22]. Hagoort found that this was rather unlikely. The N400 to the doubly anomalous words in sentence internal positions tended to be larger (p=.07) than was predicted by the single anomalies. The P600 to the same doubly anomalous words tended to be smaller than would have been predicted by the single anomalies (Hagoort did not report a statistical analysis of this difference). Thus, according to the most conventional definition of the N400 and P600, it appears that semantic and syntactic processes are not independent and there is no evidence for the precedence of one over the other.

However, Hagoort believed (apparently post-hoc) that the P600 in his data also contained a contribution from the N400. To correct for this, he measured P600 amplitude relative to the peak of the N400. According to this less conventional measure, the P600 to doubly anomalous words was consistent with independent effects of grammaticality and congruity. Hagoort took this as evidence that syntactic processing was independent of semantic processing but that semantic processing was influenced by syntactic information. While this less conventional analysis may be more accurate, it is currently impossible to know. Hagoort could have just as validly

[22] This stems from the fact that the electric potentials produced by different neural sources combine linearly (Section III.B).

applied the conventional analysis of the N400 and P600, or he could have assumed that the P600 was present in the N400 time window[23] and measured the N400 relative to the peak of the P600. Had Hagoort done the latter, it appears that he would have not only have failed to find evidence of semantic and syntactic processing independence, but his results also would suggest that syntactic anomalies can trigger more extensive semantic processing than semantic anomalies (i.e., the N400 to the ungrammatical and doubly anomalous words would have been much bigger than that to simply incongruous words). Without a more principled way for untangling potentially overlapping components from one another, it is impossible to interpret his data as evidence for anything more than some type of an interaction between syntactic and semantic processing.

Another example of the potential limitations of the conventional analysis of the N400 regards the relationship between it and another ERP component that is sometimes elicited by grammatical anomalies, an anterior negativity that occurs during the N400 time window (about 300-500 ms post-stimulus). This negativity is often labeled LAN for "left anterior negativity" as it is sometimes left lateralized. However, there are instances of comparable midline symmetric negativities as well (Friederici et al., 1993; Münte et al., 1998; Osterhout, 1997). Some researchers (Friederici, 2002) have interpreted the LAN as an index of morphosyntactic processing (e.g., computing that the "s" at the end of the verb "runs" indicates a singular third-person subject). However, it is possible that these anterior negativities

[23] Indeed, this assumption seems more reasonable given Hagoort's evidence that his grammatical anomalies could be detected faster than his semantic anomalies.

are really part of the N400 (Osterhout, 1997) and reflect general working memory operations during sentence comprehension (e.g., Kluender & Kutas, 1993). If this is true, then interpretations of these anterior negativities as indices of syntax specific processing (Friederici, 2002) and interpretations of the N400 as simply an index of semantic processing (Hagoort, Wassenaar, & Brown, 2003; Hagoort, 2003b; Kutas & Federmeier, 2000) are not entirely accurate.

As discussed in detail below, similar problems arise in trying to interpret the relationship between the P3b, N400, and P600. While there is clearly a healthy degree of debate as to the functional correlates of the P3b, all of the previously discussed proposed functions could be involved in sentence comprehension. However, it is not clear what the contribution of the P3b is to the ERPs to words in sentences. In particular, it is not clear if the P3b contributes at all to the ERPs to semantically incongruous words. If it does not contribute, then the neural activity that elicits the P3b is not as general-purpose as is frequently assumed. If it does, then our understanding of language comprehension would be greatly enriched by what we already know and will continue to learn about the P3b.

II.G A Contribution of the P3b to the N400 and P600

II.G.1 Evidence for a Contribution of the P3b to the N400

Some investigators (Curran et al., 1993) have suggested that the N400 is produced by a delay of the P3b to semantically unexpected words. In other words,

this hypothesis holds that the relatively negative ERP to semantically unexpected words is produced by the absence of the P3b that is present to semantically more expected words (Figure II.1). This hypothesis was motivated by the fact that the authors of this report found that the ERPs to sentence final incongruous words were small with a "featureless" topography in the N400 time window, which they took to indicate a lack of synchronous neural activity (i.e., activity that would contribute to the ERP--Section III.B). In contrast, the ERPs to congruous endings exhibited a pronounced LPC, which the authors equated with the P3b. Since overlapping synchronous activity of opposite polarity (e.g., the overlap of separate N400 and LPC components) can produce small "featureless" ERPs as well as a lack of synchronous activity, this is not strong evidence for equating the N400 and P3b. A relationship between the two components is better motivated by the fact that P3b latency generally correlates with the time it takes to evaluate a stimulus and N400 amplitude is generally increased by factors that increase the duration of stimulus evaluation. For example, it takes longer to comprehend semantic anomalies than nonanomalous words (Baumgaertner, Weiller, & Buchel, 2002; Cardillo, Aydelott, Matthews, & Devlin, 2004; Kiehl, Laurens, & Liddle, 2002; Kuperberg et al., 2003; McElree & Griffith, 1995; Polich, 1985a; Polich, 1985b), longer to detect semantic anomalies than grammatical ones (Fodor, Ni, Crain, & Shankweiler, 1996; Hagoort, 2003b; Kuperberg et al., 2003; McElree & Griffith, 1995), and longer to detect semantic anomalies than rare changes in letter case (Polich, 1985a). Furthermore, highly probably words in a sentence are generally processed more quickly than less likely

words (Baumgaertner et al., 2002; Cardillo et al., 2004; Fischler & Bloom, 1979). A contribution of the P3b to the N400 is also suggested by the similarity of their scalp distributions, by evidence that the N400 effect is enhanced when detection of semantic anomalies is task relevant (i.e., presumably more attended--Polich, 1985b)[24], and by evidence of common neural substrates. For example as discussed previously, Halgren and colleagues (2002) localized the N400m effect[25] primarily to widespread regions of left frontal and superior temporal regions that possibly contribute to the canonical P3b (Eichele et al., 2005; Halgren et al., 1998; Stevens, Calhoun, & Kiehl, 2005). Moreover, lesions of the left temporo-parietal junction greatly reduce the N400 and the P3b (Knight & Scabini, 1998; Van Petten & Luka, 2006).

While a contribution of the P3b to the N400 seems likely, it cannot entirely account for the N400. As illustrated in Figure II.1, an explanation based purely on a delay of the P3b would mean that the N400 would be followed by an enhanced positivity, and this does not reliably occur (Section II.C). There are also instances of larger N400 effects to stimuli that require less time to comprehend (Hagoort, 2003b; West & Holcomb, 2000) and evidence that LPC latency (i.e., the latency of the positivity that follows the N400) cannot alone account for the N400 (Chung et al.,

[24] Polich measured the N400 to semantically anomalous and nonanomalous sentence endings in two experiments. In the first experiment, participants were instructed simply to read the sentences in order to answer questions later. In the second experiment, participants were instructed to classify the endings of the sentences as anomalous or not. The maximum N400 effects at Cz and Pz appear to about 6 and 10 microvolts in the first and second experiments, respectively. However, Polich did not report a statistical analysis of this apparent difference.

[25] The N400m effect is the difference between two conditions. In this case, it was the difference between the event-related magnetic fields elicited by semantically congruous and incongruous words. Such a measure confounds differences caused by an increase in neural activity with differences caused by a decrease in activity. Thus Halgren and colleagues could well have been localizing the relative absence of a P3b to incongruous words.

1996). Furthermore, there is abundant evidence of increased neural activity in response to stimuli that elicit larger N400s. Event-related fMRI studies of N400 paradigms have generally reported overall greater hemodynamic activity in response to stimuli that elicit larger N400s (Van Petten & Luka, 2006). Several attempts to localize the N400m have found sources that show greater activity in response to incongruous words (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Helenius et al., 2002; Simos, Basile, & Papanicolaou, 1997). Moreover, as previously mentioned, there is convincing iEEG evidence of increased neural activity in response to incongruous words in the anterior medial temporal lobe (McCarthy et al., 1995; Nobre et al., 1994).

Furthermore, there is some direct evidence against a contribution of the P3b to the N400. For example, the effects of semantic congruity (which elicits an N400) and rare changes in letter case (which elicits a P3b-like effect--Section II.D) appear to have additive effects in the N400 time window (Kutas & Hillyard, 1980a). As mentioned in Section II.F, additive effects of two independent variables are consistent with the two effects being generated by independent neural processes. However, this evidence is just as consistent with common generators of the N400 and P3b. It would simply mean that congruity and word case changes have additive effects on those shared generators (a realistic possibility assuming that ERP generators have sigmoidal activation functions--Eichele, et al., 2005). In addition, it is important to note that evidence of additivity is based upon a failure to reject the null hypothesis of additivity. Failure to reject the null hypothesis could be due to noisy data or

insufficient electrode coverage, rather than true effect additivity. In their report, Kutas and Hillyard presented a "synthesized" double anomaly ERPs (i.e., the ERPs to semantically incongruent, surprisingly large words) based upon the ERPs to single anomalies and the additivity assumption. The synthesized ERPs are a good fit to the observed ERPs to double anomalies in the N400 window, thus the failure to reject the additivity hypothesis was not due to noisy data. However, in this study the EEG was only measured at five electrodes. Thus, it is quite possible that with more extensive scalp coverage, an interaction between the two types of anomalies would be found. Finally, it is worth noting that after the N400 window the fit between the synthesized and observed ERPs deteriorated as the synthesized ERPs went more positive than the observed ERPs. Although Kutas and Hillyard did not report if this apparent interaction was statistically significant, it suggests that the late positivity that follows the N400 is related to the P3b[26].

Another piece of evidence against a contribution of the P3b to the N400 effect is the fact that the N400 effect is apparently unaffected by the attentional blink, while the P3b is abolished by it (Vogel, Luck, & Shapiro, 1998). However, it is possible that the N400 effect to words in the attentional blink paradigm, which requires rapid serial visual presentation (e.g., a new visual stimulus presented every 83 ms), is not the same as the canonical N400 effect to words in sentences (presented about every 500 ms).

[26] There is some evidence that the post-N400 positivity is enhanced when the eliciting semantic anomaly is task relevant, which also suggests that this positivity is related to the P3b (Polich, 1985b).

II.G.2 Evidence for a Contribution of the P3b to the P600

The relationship between the P3b and P600 has been actively researched by several groups (Coulson, King, & Kutas, 1998a; Frisch, Kotz, von Cramon, & Friederici, 2003; Gunter et al., 1997; Osterhout & Hagoort, 1999). A contribution of the P3b to the P600 is suggested by the fact that they have similar scalp distributions and are both elicited by unexpected and anomalous stimuli. Moreover, their amplitudes tend to be an inverse function of the relative frequency of such stimuli in the experimental session, are enhanced by stimulus saliency, and stimulus task relevance. Direct evidence for some kind of contribution is the fact that the P3b and P600 effects do not appear to be additive. This was found in a study by Osterhout and colleagues (1996) that was analogous to the aforementioned Kutas and Hillyard (1980a) study. Osterhout et al. compared the ERPs to ungrammatical words (a P600 effect), rare upper case words (a P3b-like effect), and ungrammatical upper case words (a combined P600 and P3b effect). They found that the late positivity to doubly anomalous words was smaller than would have been predicted by additivity, although the difference failed to reach significance.

Despite this evidence, some researchers have argued that the P600 is distinct from the P3b "to an interesting degree" (Osterhout & Hagoort, 1999). Evidence for a distinction comes from the fact that the P3b to rare uppercase words and misspelled words tends to be right lateralized while P600 effects from the same participants are more left lateralized (to the extent that they are lateralized at all--Münte et al., 1998;

Osterhout et al., 1996). As mentioned previously (Section II.E.1), conventional comparisons of the topographies of scalp potentials that differ in magnitude are prone to an inflated Type I error rate. The P3b is larger than the P600 in these studies, but the fact that the trend for this topography difference replicated across four experiments (the report by Osterhout and colleagues included data from three experiments) suggests that the difference is real. However, given that the P600 and P3b are most likely composed of multiple subcomponents, these differences in topography may be caused by a difference in the relative strength and/or timing of these subcomponents (Section III.B.1; Urbach & Kutas, 2002). Thus, they do not prove the qualitative difference between the neural sources of these components that Osterhout and Hagoort propose.

In addition to these topographic differences, Osterhout and Hagoort (1999) have argued that the P600 is distinct from the P3b because it is generally less sensitive to factors of task relevance and anomaly frequency[27]. This argument is not convincing as the weaker relationship is quite possibly due to the fact that the P600 is smaller and more variable than the P3b. Finally, there is some evidence that lesions that include the basal ganglia eliminate the P600 but not the P3b (Frisch et al., 2003). However, this result may simply reflect the fact that these patients had language comprehension difficulties (6 of the 7 patients had comprehension difficulties and the seventh was not tested) rather than a distinction between the neural generators of the

[27] Similarly, there is some evidence that the latency of the P600 is not slowed by age, even though the latency of the P3b is (Kemmer, et al., 2004). The lack of an effect of age on P600 latency may simply reflect the greater difficulty of estimating the latency of the P600 or the fact that age was a between-participants factor (the researchers apparently did not control for differences in linguistic ability across younger and older participants).

P3b and P600[28]. Moreover, in an earlier study by the same research group (Friederici, von Cramon, & Kotz, 1999), the investigators found that two groups of patients with damage to the left basal ganglia[29] each generated P600s to grammatical anomalies, which demonstrates that the basal ganglia are not necessary to generate the P600 and cannot explain the dissociation between the P600 and P3b in their later study[30].

II.G.3 On the Significance of Extracting a P3b Contribution to the N400 or P600

To summarize, decades of research have identified three ERP components reliably elicited by linguistic stimuli whose amplitudes are quite sensitive to the probability of those stimuli. The N400 is sensitive to the probability of a word but is affected by semantic factors as well. The P600 is elicited by grammatically anomalous/dispreferred words and is sensitive to the probability of such words in an experimental session. Finally, the P3b is elicited by a wide range of surprising stimuli, and, like the P600, it is sensitive to the probability of stimuli in an experimental session. Moreover, the P3b is also sensitive to a participant's confidence that the eliciting stimulus will occur. It is likely that all three components are composed of distinct subcomponents and that some of these subcomponents are

[28] For example, a blindfolded person can generate an auditory P3b but not a visual P3b. This does not prove that the two ERP phenomena are neurally and functionally distinct.
[29] A "cortical" group of patients suffered left anterior lesions including parts of frontolateral cortex, the caudate nucleus, and putamen. A "basal ganglia" group of patients suffered lesions restricted to the left basal ganglia (caudate nucleus, putamen, and globus pallidus).
[30] Oddly, Frisch and colleagues do not cite the 1999 report in their later 2003 report, let alone discuss the discrepancy between the two studies.

shared. There is currently no direct evidence that the P3b contributes to the N400 (indeed there is some evidence that they are independent effects) and the P3b alone cannot explain the N400. With regards to the P600, there is a bit more evidence that the P3b contributes somewhat to the effect but the extent of its contribution is still poorly understood. A main goal of this thesis is to try to determine how much (if at all) the P3b contributes to these effects and more importantly, to extract the P3b from these effects using independent component analysis (discussed in Chapter III).

What would the significance of a successful realization of this goal be? As we do not yet have a solid understanding of the P3b at either functional or neural levels, the immediate impact would be primarily "ERPological." In other words we would understand in better detail what these ERP phenomena consist of, what variables they are sensitive to, and we would have a more nuanced way of measuring them. This knowledge would then improve the ability of the N400, P600, and P3b to inform our theories of language comprehension and of how the brain copes with perceptual uncertainty.

For example, if the P3b contributes to the N400, it might help to explain why the N400 can be modulated by grammaticality and why N400 effects are smaller in individuals with high schizotypy[31] (Kiang & Kutas, 2005). If the P3b does not contribute to the N400 at all, then it would mean that the P3b is not as general a

[31] "Schizotypy" is a quantitative measure of the degree to which an individual exhibits schizotypal personality traits (Kiang & Kutas, 2005). Recently, Gassab and colleagues (2006) found that high schizotypy individuals have smaller amplitude and somewhat delayed tone oddball P3b's than comparable low schizotypy individuals. Importantly, earlier auditory ERP components (the N100, P200, and N200) were not found to differ across groups, thus the P3b difference cannot be simply attributed to general differences between participants (e.g., skull thickness, alertness).

response as is widely assumed, and it would flag a need to rethink its functional significance.

With regard to the P600, if it is effectively reducible to the P3b, it could be concluded that the P600 indexes a very general neural function that can be elicited by stimuli that do not have anything like the hierarchical temporal structure of language (Hauser, Chomsky, & Fitch, 2002). Any interpretation of the P600 as an index of the manipulation of grammatical representations (e.g., syntactic integration or parse revision) would need to account for this. However, if there is a component of the P600 that is not shared by the P3b, it would suggest that there is indeed a neural process specialized for dealing with hierarchical temporal structure. The existence of such a process would be significant, even if its function is similar to that of the P3b (Coulson, King, & Kutas, 1998a).

While these immediate ERPological questions are important (the ability for ERP phenomena to inform theories is only as great as our functional and neural understanding of the phenomena), the most important reason for understanding the relationship between the N400 and P600 is that future advances in our understanding of the P3b would inform our understanding of the N400, P600, and the mechanisms of language comprehension. This is significant because, as mentioned in the Chapter I, there are animal models of the P3b but not of language comprehension.

For example, consider Friston's hypothesis that the P3b reflects the function of a probabilistic generative model at a rather abstract level of perception (Friston, 2005). Generative models "perceive" sensory input via "analysis by synthesis." In

other words, generative models try to determine what objects and events caused a particular sensory signal by using a model of how such objects and events generate sensory signals. Generative models are a theoretically appealing language comprehension mechanism as they could explain how the brain exploits the predictability of language to more quickly and more accurately comprehend language (Chapter I; Chater & Manning, 2006). Moreover, they might help to explain how language is learned (Chater & Manning, ibid.) as generative models naturally lend themselves to "unsupervised learning"--that is, learning through mere exposure to sensory signals (Hinton, 1999). Specifically, by learning to reduce the difference between the sensory information it would synthesize and the sensory information it actually observes, a generative model can uncover latent sensory structure. Popular examples of this in the language modeling literature are simple recurrent connectionist networks that are trained to predict the next symbol in a series (Elman, 2004). While not explicitly probabilistic, these algorithms can be thought of as generative models and can learn to detect word boundaries (Elman, 1990), to distinguish syntactic categories (Elman, 2004), and to parse impressively complicated syntactic patterns (Lewis & Elman, 2001).

Given the speed at which infants learn language and the dearth of explicit syntactic instruction (Lightfoot, 1991), it is very likely that the brain learns a great deal about linguistic structure in an unsupervised manner (Kuhl, 2004). For example, infants can learn phonetic categories through two minutes of exposure to language-like babble (Maye, Werker, & Gerken, 2002). At a more abstract level, infants,

humans and cotton-top tamarinds can learn to extract word-like units from a stream of continuous babble based upon the frequency of phoneme sequences (Hauser, Newport, & Aslin, 2001; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). At an even more abstract level, adults appear to be able to learn the meanings of novel words through exposure to only a few examples of those words in manner that is generally consistent with a probabilistic generative model (Tenenbaum & Xu, 2000).

Thus it is likely that the brain implements generative models in the service of language comprehension. However, showing that this is indeed the case and understanding the details of such mechanisms are hampered by the limitations of noninvasive measures and manipulations of brain activity. For example, validating Friston's particular proposal as to how cortex implements generative models requires the ability to independently measure and manipulate the input from feedforward and feedback projections to individual neurons in cortical areas[32]. Such precision is beyond the capacity of the most popular non-invasive measures of brain activity, EEG/MEG and fMRI (Dale et al., 2000), and the non-invasive disruption of neural activity provided by transcranial magnetic stimulation (Daskalakis, Christensen, Fitzgerald, & Chen, 2002). It is much more likely that convincing evidence of neural generative models will come from animal research, but generalizing the results of such research to language function will require the identification of non-invasive human analogues of animal neural function. Although it is not currently clear that the

[32] Friston hypothesizes that there are two distinct populations of neurons within a cortical area: "representational units" that reflect bottom-up inference driven by feedforward projections and "error units" that reflect top-down predictions driven by feedback projections.

P3b is produced by a generative model or that it is involved in naturalistic language comprehension and language learning[33], the appealing possibility that the P3b is produced by a generative model in the service of language comprehension clearly illustrates the importance of being able to better relate brain activity in linguistic experimental paradigms to that in non-linguistic paradigms.

II.G.4 On the Likelihood of Extracting a P3b Contribution to the N400 or P600

Assuming the P3b does contribute in some way to the N400 and P600 effects, its subcomponents should be the easiest to extract and identify of any of the proposed subcomponents of the N400 and P600. This stems from the fact that the P3b is reliable, large, and has the best studied ostensible subcomponents. Evidence of between-participant P3b reliability comes from a report by Fabiani and colleagues (1987), who found that all 54 participants in their study demonstrated at least a trend for a visual P3b at electrode Pz[34]. Moreover, individual participant P3b latency and amplitude from different experimental sessions or different blocks of the same experimental session are generally highly correlated (Pearson's r>=.7, for review see Segalowitz & Barnes, 1993). With regard to P3b amplitude, its size makes it one of the few "cognitive" ERP components that can be reliably detected in the raw EEG

[33] For example, although infant ERPs to familiar and novel stimuli differ, the ERPs to novel stimuli do not resemble the P3b (Thomas, 2003). It is currently not clear to what extent this indicates differences between adult and infant neural processing or anatomical differences (e.g., generator orientation, skull thickness) that can make the same neural electrical activity appear quite different at the scalp (ibid.).
[34] Fabiani and colleagues did not report the percentage of these individual participant trends that were statistically significant.

(Kutas et al., 1977) or in a very limited number of averaged trials (Serby, Yom-Tov, & Inbar, 2005). All of these traits will be particularly important for the subsequent subcomponentry analysis (Chapter VI), which attempts to identify the same subcomponents in individual participants.

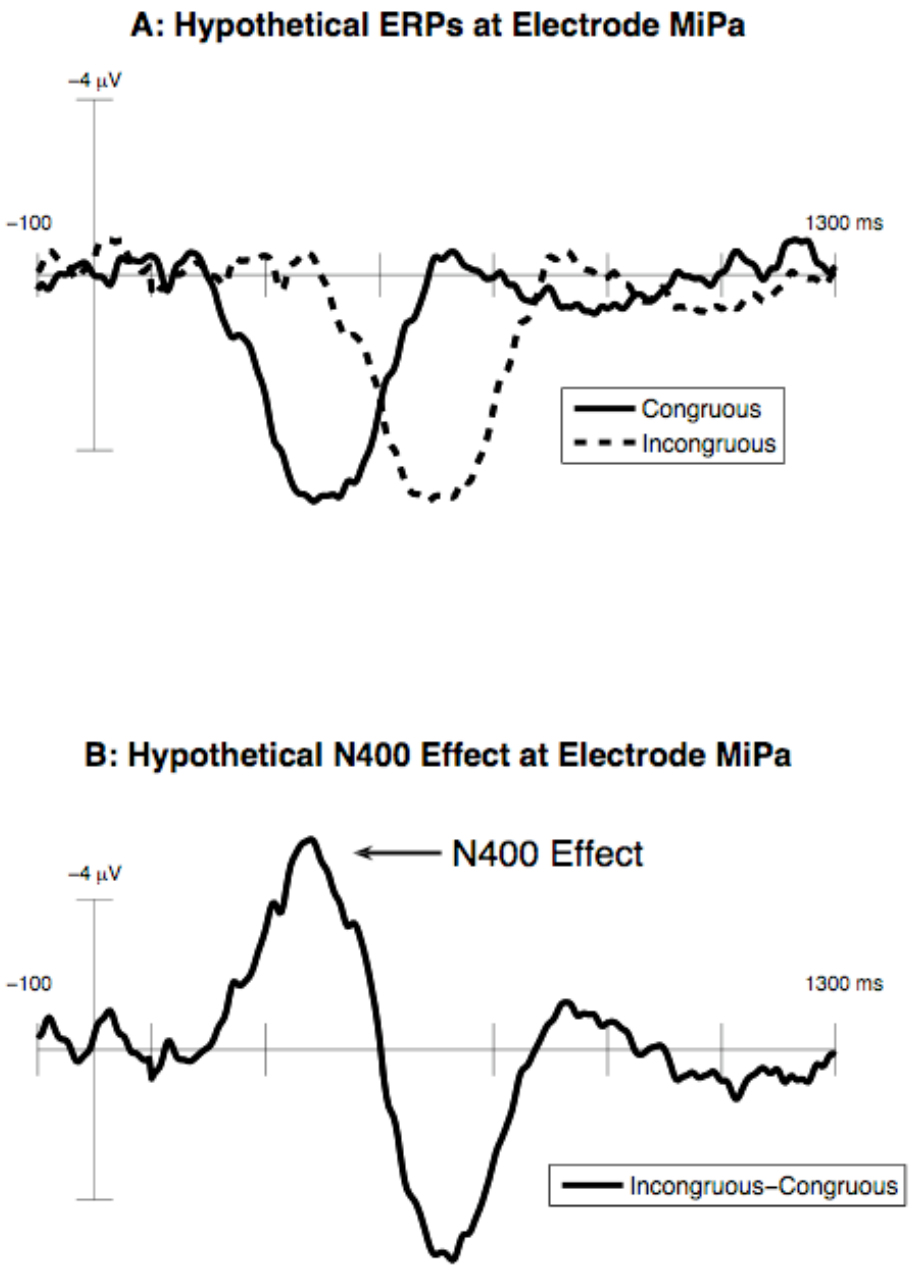

Figure II.1 – Hypothetical ERPs that illustrate how the N400 effect could be caused by a delayed P3b. (**A**) The P3b to incongruous words is delayed relative to the P3b to congruous words. (**B**) The difference between the ERPs to incongruous and congruous words shows an N400 effect that is followed by a greater positivity for incongruous words.

II.H References

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14-30.

Baumgaertner, A., Weiller, C., & Buchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *NeuroImage, 16*(3 Pt 1), 736-745.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235-260.

Besson, M., & Kutas, M. (1993). The many facets of repetition: A cued-recall and event-related potential analysis of repeating words in same versus different sentence contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1115-1133.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology, 24*(1), 14-25.

Bruyant, P., Garcia-Larrea, L., & Mauguiere, F. (1993). Target side and scalp topography of the somatosensory P300. *Electroencephalography and Clinical Neurophysiology, 88*(6), 468-477.

Cardillo, E. R., Aydelott, J., Matthews, P. M., & Devlin, J. T. (2004). Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *Journal of Cognitive Neuroscience, 16*(9), 1552-1561.

Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences, 10*(7), 335-344.

Chung, G., Tucker, D. M., West, P., Potts, G. F., Liotti, M., Luu, P., et al. (1996). Emotional expectancy: Brain electrical activity associated with an emotional bias in interpreting life events. *Psychophysiology, 33*(3), 218-233.

Coulson, S., King, J. W., & Kutas, M. (1998a). ERPs and domain specificity: Beating a straw horse. *Language and Cognitive Processes, 13*(62), 653-672.

Coulson, S., King, J. W., & Kutas, M. (1998b). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology, 88*(3), 188-209.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084*(1), 89-103.

Daskalakis, Z. J., Christensen, B. K., Fitzgerald, P. B., & Chen, R. (2002). Transcranial magnetic stimulation: A new investigational and treatment tool in psychiatry. *Journal of Neuropsychiatry and Clinical Neurosciences, 14*(4), 406-415.

De Swart, J. H., Kok, A., & Das-Smaal, E. A. (1981). P300 and uncertainty reduction in a concept-identification task. *Psychophysiology, 18*(6), 619-629.

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology, 41*(1), 60-74.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*(2), 95-111.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research, 17*(3), 637-650.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 112*(3), 357-374.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology, 14*(5), 456-467.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M. L., Quiroga, R. Q., Nordby, H., et al. (2005). Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 102*(49), 17798-17803.

Elger, C. E., Grunwald, T., Lehnertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., et al. (1997). Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia, 35*(5), 657-667.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences, 8*(7), 301-306.

Fabiani, M., Gratton, G., Demetrios, K., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings & M. G. Coles (Eds.), *Advances in Psychophysiology* (pp. 1-78). London: JAI Press.

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology, 92*(2), 148-160.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469-495.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 1-20.

Fodor, J. D., Ni, W., Crain, S., & Shankweiler, D. (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research, 25*(1), 25-57.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1219-1248.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20 Suppl 1*, S8-17.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(2), 305-323.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170-177.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition, 27*(3), 438-453.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain, 122*(6), 1033-1047.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews, 25*(4), 355-373.

Frisch, S., Kotz, S. A., von Cramon, D. Y., & Friederici, A. D. (2003). Why the P600 is not just a P300: The role of the basal ganglia. *Clinical Neurophysiology, 114*(2), 336-340.

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society of London.Series B, Biological sciences, 360*(1456), 815-836.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 906-913.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 96*(6), 3314-3319.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89-106.

Gassab, L., Mechri, A., Dogui, M., Gaha, L., d'Amato, T., Dalery, J., et al. (2006). Abnormalities of auditory event-related potentials in students with high scores on

the schizotypal personality questionnaire. *Psychiatry Research, 144*(2-3), 117-122.

Gonsalvez, C. L., & Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology, 39*(3), 388-396.

Goto, Y., Brigell, M. G., & Parmeggiani, L. (1996). Dipole-modeling of the visual evoked P300. *Journal of Psychosomatic Research, 41*(1), 71-79.

Gunter, T. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology, 36*(1), 126-137.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*(6), 660-676.

Haan, H., Streb, J., Bien, S., & Rosler, F. (2000). Individual cortical current density reconstructions of the semantic N400 effect: Using a generalized minimum norm model with different constraints (L1 and L2 norm). *Human Brain Mapping, 11*(3), 178-192.

Hagoort, P. (2003a). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage, 20 Suppl 1*, S18-29.

Hagoort, P. (2003b). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883-899.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*(9), 416-423.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*(11), 1531-1549.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439-483.

Hagoort, P., Hald, L., Bastiaansen, M. C., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.

Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Real-time semantic compensation in patients with agrammatic comprehension: Electrophysiological evidence for multiple-route plasticity. *Proceedings of the National Academy of Sciences of the United States of America, 100*(7), 4340-4345.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194-205.

Halgren, E., Baudena, P., Heit, G., Clarke, J. M., Marinkovic, K., Chauvel, P., et al. (1994). Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and rolandic cortices. *Journal of Physiology, Paris, 88*(1), 51-80.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17*(3), 1101-1116.

Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology, 106*(2), 156-164.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78*(3), B53-64.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain, 121*(6), 1133-1142.

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *Journal of Neuroscience, 22*(7), 2936-2944.

Hinton, G. E. (1999). Introduction. In G. E. Hinton, & T. J. Sejnowski (Eds.), *Unsupervised Learning : Foundations of Neural Computation* (pp. vii-xvi). Cambridge, Mass.: MIT Press.

Hinton, G. E. (2000). Computation by neural networks. *Nature Neuroscience, 3 Suppl*, 1170.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721-742.

Horovitz, S. G., Skudlarski, P., & Gore, J. C. (2002). Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: A parametric approach to combining fMRI and ERP. *Magnetic Resonance Imaging, 20*(4), 319-325.

Horst, R. L., Johnson, R., Jr, & Donchin, E. (1980). Event-related brain potentials and subjective probability in a learning task. *Memory & Cognition, 8*(5), 476-488.

Ji, J., Porjesz, B., Begleiter, H., & Chorlian, D. (1999). P300: The similarities and differences in the scalp distribution of visual and auditory modality. *Brain Topography, 11*(4), 315-327.

Johnson, B. W., & Hamm, J. P. (2000). High-density mapping in an N400 paradigm: Evidence for bilateral temporal lobe generators. *Clinical Neurophysiology, 111*(3), 532-545.

Johnson, R.,Jr. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology, 30*(1), 90-97.

Johnston, V. S., Miller, D. R., & Burleson, M. H. (1986). Multiple P3s to emotional stimuli and their theoretical significance. *Psychophysiology, 23*(6), 684-694.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Kayser, J., & Tenke, C. E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns II: Adequacy of low-density estimates. *Clinical Neurophysiology, 117*(2), 369-380.

Keil, A., Bradley, M. M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology, 39*(5), 641-649.

Kemmer, L., Coulson, S., De Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology, 41*(3), 372-384.

Kiang, M., & Kutas, M. (2005). Association of schizotypy with semantic processing differences: An event-related brain potential study. *Schizophrenia Research, 77*(2-3), 329-342.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: An event-related fMRI study of semantic processing. *NeuroImage, 17*(2), 842-850.

Kluender, R., & Kutas, M. (1993). Bridging the gap - evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology, 15*(1), 3-13.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences, 27*(12), 712-719.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302-307.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577.

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1-36.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 804-823.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. In R. Johnson, J. W. Rohrbaugh & R. Parasuraman (Eds.), *Current Trends in Event-Related Potential Research (EEG suppl. 40)* (pp. 406-411) Elsevier Science Publishers B.V. (Biomedical Division).

Kutas, M., & Dale, A. M. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg (Ed.), *Cognitive Neuroscience* (first ed., pp. 197-242). Hove East Sussex, UK: Psychology Press.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99-116.

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Kutas, M., Neville, H. J., & Holcomb, P. J. (1987). A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. *Electroencephalography and Clinical Neurophysiology Supplement, 39*, 325-330.

Kwon, H., Kuriki, S., Kim, J. M., Lee, Y. H., Kim, K., & Nam, K. (2005). MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. *Neuroscience Research, 51*(4), 349-357.

Lewis, J. D., & Elman, J. L. (2001). Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. *Proceedings of the 26th Annual Boston University Conference on Language Development.*

Lightfoot, D. (1991). The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences, 12*, 321-375.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*(1), 47-62.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-11.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*(4477), 77-80.

McCarthy, G., Nobre, A. C., Bentin, S., & Spencer, D. D. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *Journal of Neuroscience, 15*(2), 1080-1089.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 134-157.

McKinnon, R., & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognitive Processes, 11*, 495-523.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53-65.

Moores, K. A., Clark, C. R., Hadfield, J. L., Brown, G. C., Taylor, D. J., Fitzgibbon, S. P., et al. (2003). Investigating the generators of the scalp recorded visuo-verbal P300 using cortically constrained source localization. *Human Brain Mapping, 18*(1), 53-77.

Münte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research, 30*(3), 339-364.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12*(1), 120-133.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510-532.

Nobre, A. C., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature, 372*(6503), 260-263.

Olichney, J. M., Van Petten, C., Paller, K. A., Salmon, D. P., Iragui, V. J., & Kutas, M. (2000). Word repetition in amnesia. electrophysiological measures of impaired and spared memory. *Brain, 123*(9), 1948-1963.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*(1), 1-14.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8*, 413-437.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8*, 507-526.

Osterhout, L., & Mobley, L. A. (1995). Event-related potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739-773.

Osterhout, L., & Nicol, J. L. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283-317.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Polich, J. (1985a). N400s from sentences, semantic categories, number and letter strings? *Bulletin of the Psychonomic Society, 23*(4), 361-364.

Polich, J. (1985b). Semantic categorization and event-related potentials. *Brain and Language, 26*(2), 304-321.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical overview. In N. C. Moore, & K. Arikan (Eds.), *Brainwaves and Mind: Recent Developments* (pp. 15-29). Wheaton, IL: Kjellberg Inc.

Ragot, R., & Renault, B. (1981). P300, as a function of S-R compatibility and motor programming. *Biological Psychology, 13*, 289-294.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*(4), 606-621.

Segalowitz, S. J., & Barnes, K. L. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology, 30*(5), 451-459.

Serby, H., Yom-Tov, E., & Inbar, G. F. (2005). An improved P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(1), 89-98.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656.

Simos, P. G., Basile, L. F., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research, 762*(1-2), 29-39.

Sirevaag, E. J., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica, 70*(1), 77-97.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343-358.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*(4), 387-401.

Stevens, M. C., Calhoun, V. D., & Kiehl, K. A. (2005). fMRI in an oddball task: Effects of target-to-target interval. *Psychophysiology, 42*(6), 636-642.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(700), 1187-1188.

Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research, 15*(1), 99-103.

Tarkka, I. M., & Stokic, D. S. (1998). Source localization of P300 from oddball, single stimulus, and omitted-stimulus paradigms. *Brain Topography, 11*(2), 141-151.

Tenenbaum, J. B., & Xu, F. (2000). Word learning as Bayesian inference. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society,*

Thomas, K. M. (2003). Assessing brain development using neurophysiologic and behavioral measures. *Journal of Pediatrics, 143*(4 Suppl), S46-53.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology, 39*(6), 791-808.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes, 8*(4), 485-531.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394-417.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19*(1), 95-112.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279-293.

Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology, 34*(2), 131-156.

Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: On the relationship between expectancies and P3. *Psychophysiology, 31*(4), 359-369.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1656-1674.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Wickens, C., Kramer, A. F., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science, 221*(4615), 1080-1082.

Wijers, A. A., Mulder, G., Okita, T., & Mulder, L. J. (1989). Event-related potentials during memory search and selective attention to letter size and conjunctions of letter size and color. *Psychophysiology, 26*(5), 529-547.

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654-667.

Yamaguchi, S., & Knight, R. T. (1991). P300 generation by novel somatosensory stimuli. *Electroencephalography and Clinical Neurophysiology, 78*(1), 50-55.

Zhang, Q., Guo, C., Ding, J., & Wang, Z. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language, 96*, 59-68.

Chapter III: Independent Component Analysis for ERP Subcomponentry

III.A Introduction

A method for dealing with the problem of spatiotemporally overlapping ERP subcomponents is independent component analysis (ICA). As typically applied to EEG/ERP datasets, ICA decomposes an ERP into a set of "independent components" (ICs) via a basis of linear spatial filters (LSFs). To the extent that ICs correspond to latent ERP subcomponents, ICA provides a cleaner measure of these subcomponents and allows their identification across different experimental paradigms (e.g., sentence reading and oddball tasks). The main purpose of this chapter is to illustrate how ICA could in principle extract spatiotemporally overlapping ERP subcomponents, briefly explain how ICA learns a basis of LSFs for a particular dataset, and review evidence that ICA is a well motivated methodology that can extract latent ERP subcomponents in practice. I close the chapter by noting potential problems with ICA and by comparing it to three alternative methods for decomposing ERPs into subcomponents (spatial principle component analysis, source localization, and temporal principle component analysis) to ICA.

III.B Sources of Scalp Potentials

To understand why LSFs are useful for ERP analysis, one must first understand the relationship between electric potentials recorded at the scalp and their neural and non-neural generators.

III.B.1 Neural Sources

The electrical activity of the brain that is measurable at the scalp is called the electroencephalogram (EEG). Event-related potentials (ERPs) are the portion of the EEG whose phase is related to the onset of a class of events and is typically derived by averaging multiple samples of EEG time-locked to the occurrence of such events (e.g., the onset of a visual stimulus or a button press). Much of the brain's electrical activity is practically invisible at the scalp (due to the small magnitude of the activity or interactions between sources of opposite polarity—Kutas & Dale, 1997). Indeed, it is generally believed that the great bulk of the EEG is generated by the synchronous activation of post-synaptic potentials of adjacent cortical pyramidal cells (Baillet, Mosher, & Leahy, 2001; Kutas & Dale, 1997). This activity can be accurately approximated as a finite set of equivalent current dipoles (ECD) or multipoles (ibid.). In this framework, the EEG is the linear combination of the electrical activity of a discrete set of neural electrical sources (ibid.).

To illustrate, consider a situation in which there is only a single neural source[35], y_1, and two scalp electrodes, x_1 and x_2. The voltage recorded at each scalp electrode is simply a scaled version of the source's voltage (Figure III.1). The polarity and magnitude of the scaling is determined by the location and orientation of the source relative to the electrode and the conductive properties of the medium between the source and the electrode. This relationship can be represented with a linear function where the source voltage is multiplied by a "gain vector," **g**, to obtain the scalp voltages:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} .04 \\ -.15 \end{bmatrix} y_1(t) = \mathbf{g} y_1(t) \qquad (1)$$

As long as the position of the source relative to the electrodes and the conductive properties of the head do not change, this linear relationship will remain constant for each time point, *t*. Note that the gain vector defines the scalp topography of a source. With only one source, the topography of scalp voltages will remain constant (Figure III.2: B & C).

To make our example slightly more complicated, let us add two more sources. Equation 1 becomes:

[35] By "neural source" I mean extremely correlated patches of cerebral cortex. This could be a single localized patch of cortex that could be accurately modeled as a single ECD or it could be a distributed set of patches that would require multiple ECDs to be accurately modeled.

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} .04 & -.08 & -.05 \\ -.15 & -.04 & .01 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \mathbf{G}\mathbf{y}(t) \quad (2)$$

where the three gain vectors form a "gain matrix," **G**. Now each scalp electrode measures the sum of the scaled voltages of the three sources (Figure III.3). As before, the scalp topography of each source is static and defined by its gain vector. However, as there is more than one source and the sources are at least somewhat independent, the topographies of the individual sources combine such that the topography of the scalp potentials is not constant but changes over time (Figure III.2: D).

In reality, the neural sources of a segment of EEG are likely to be far more numerous than three. Assuming that a patch of cortex on the order of 25 mm^2 is the smallest neural unit whose electrical activity is measurable from the scalp (Baillet et al., 2001) and that the surface area of a hemisphere of human cerebral cortex is approximately 80,000 mm^2 (Van Essen, Drury, Joshi, & Miller, 1998), the number of EEG sources could be as high as 6,400. However, the number of sources contributing to an ERP is probably much smaller than this upper bound. The activity of multiple (especially adjacent--Destexhe, Contreras, & Steriade, 1999) patches of cortex may be correlated enough that they can be approximated as a single source. Moreover, the activity of some sources may not be sufficiently phase-dependent on the onset of the event used to derive the ERP to contribute to the ERP or may be cancelled out by sources of opposite polarity. In addition, some sources may be too weak and too far from scalp electrodes to contribute significantly (the strength of a source potential falls

off as an inverse function of distance squared--Kutas & Dale, 1997). In fact, many ERP phenomena can be modeled with only a handful of sources (e.g., Sections IV.E.3.c and IV.F.3.e). Unfortunately, it is impossible to know with certainty how many sources generated a particular segment of an ERP. For example, multiple EEG sources with similar scalp distributions may be inaccurately modeled by fewer sources. On the other hand, a few EEG sources can be inaccurately modeled by too many sources (Liu, Dale, & Belliveau, 2002).

III.B.2 EEG Artifacts

Attempts to measure the EEG are complicated by the fact that the brain is not the only source of the electric potentials recorded at the scalp and by technical disturbances (Talsma & Woldorff, 2005). Phenomena such as blinks, eye movements, and head muscle activity generate electrical potentials at amplitudes that can be orders of magnitude bigger than ERP amplitudes and can seriously confound ERP analysis. Sensor noise, caused by poor connections between the scalp and electrodes, and 60 Hz line noise, also commonly pollute EEG recordings.

The effect of many of these phenomena on recorded scalp potentials can be well approximated by the same framework used to model the EEG: each artifact is modeled as a set of discrete electrical sources with fixed scalp distributions whose activity combines linearly at the scalp with other electrical sources. In terms of Equation 2, each artifact can be approximated by one or more elements in the source vector, **y**, and a corresponding number of columns in the gain matrix, **G**.

For example, blink potentials (Ille, Berg, & Scherg, 2002; Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000), scalp muscle activity (Jung, Makeig, Humphries et al., 2000), and 60 Hz line noise (Jung, Makeig, Humphries et al., 2000; Tang, Sutherland, & McKinney, 2005) each can be well approximated as a single source. Eye movements can be successfully modeled by two sources, one each for movement in the horizontal and vertical planes (Ille et al., 2002).

III.C Linear Spatial Filters

Given that the electric potentials of different EEG and EEG artifact sources linearly combine and can have distinct topographies, linear spatial filters are a potentially powerful tool for teasing apart these different sources despite their spatiotemporal overlap. A linear spatial filter (LSF) is a weighted sum of the potentials at each electrode. For example, the vector, **w**, below would be an LSF for two electrode data:

$$u(t) = \begin{bmatrix} .06 & -.14 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{w}^T \mathbf{x}(t) \tag{3}$$

The weights to each electrode define the LSF and determine the output, or "activation," of the filter, *u*. While each electrode measures a weighted sum of the electrical activity of neural and EEG artifact sources, the LSF provides a re-weighted sum of those sources and functions like a "virtual electrode" (Spencer, Dien, &

Donchin, 2001). Figure III.4 (A) plots the activation of the LSF from Equation 3 when input the data from Figure III.3. The activation is very similar to the activity of Source 1 (r_{cos}=0.89)[36] but smaller in amplitude. In other words, the filter sums the activity of Source 2 and Source 3 with small weights and Source 1 with a relatively large weight.

We can compute the weight that an LSF, **w**, assigns to a source with gain vector, **g**, from the following equation:

$$\cos(\theta)|\mathbf{w}||\mathbf{g}| = \mathbf{w}^T\mathbf{g} \tag{4}$$

where θ is the angle between the filter and source gain vectors, |**w**| is the length of the filter's gain vector, and |**g**| is the length of the source's gain vector. The first term in this equation, cos(θ), takes a value between -1 and 1. If the filter and source gain vectors are perpendicular (i.e., θ=90 or 270 degrees), then cos(θ) is zero and the filter will not pick up the source's activity at all. As the filter and source gain vectors come closer together, the angle between them shrinks and cos(θ) approaches one. As the angle between the two gain vectors approaches 180 degrees, cos(θ) approaches negative one. The first term interacts with the final term, |**g**|, to make the filter more sensitive to some sources than others. The remaining term in the equation, |**w**|, simply

[36] r_{cos} is the "cosine similarity" between two vectors (i.e., the cosine of the angle between the two vectors) and is analogous to Pearson's linear correlation coefficient, *r*. An r_{cos} of 1 indicates that two vectors are identical save for a possible difference in magnitude (i.e., they point in the same direction). An r_{cos} of –1 indicates that two vectors point in exactly opposite directions and an r_{cos} of 0 indicates that vectors are orthogonal. More specifically, the cosine similarity between two vectors **x** and **y** is $\mathbf{x}^T\mathbf{y}/(|\mathbf{x}||\mathbf{y}|)$.

scales the output of the filter and will affect the weights between the filter and all sources equally.

In light of Equation 4, we can now understand why our example filter in Figure III.4 works as it does by visualizing the filter and source gain vectors in "electrode space" (Figure III.4: B). The filter is nearly perpendicular to the gain vectors of Source 2 and Source 3, and nearly parallel to the gain vector of Source 1. Hence the filter gives greater weight to Source 1 than to the other sources. The angle between the filter and Source 1's gain vector is acute, hence the filter picks up Source 2's activity with the same polarity. Finally, the filter is short, which contributes to the amplitude of the filter output being much smaller than the amplitude of Source 1.

III.C.1 LSF Limitations and Difficulties

Note that in this example, the filter cannot perfectly capture one of the sources and completely filter out the other two. That would require the filter to be perpendicular to two of the sources and non-perpendicular to the third. When there are more than n sources (where n is the number of electrodes) with linearly independent gain vectors[37], it is impossible to perfectly filter out the activity of one source from that of all the other sources[38]. Given that the measurements of an

[37] A set of linearly independent gain vectors means that none of the gain vectors equals a weighted sum of the other gain vectors. For example, if there are two sources and two electrodes, the gain vectors are linearly independent if they are not parallel.

[38] For an LSF, $\mathbf{w}$, to perfectly filter out one source from n other sources whose gain vectors ($\mathbf{g}_i$, $1 \leq i \leq n$) are linearly independent, the following would have to be true: $\mathbf{w}^T\mathbf{g}_i=0$ for all $1 \leq i \leq n$. However, if this were true, the inner product of $\mathbf{w}$ and any possible gain vector would be zero (i.e., $\mathbf{w}$ would be the zero

electrode will be polluted by a degree of sensor noise that will be somewhat independent of the sensor noise at other electrodes (e.g., next section), there will always be a set of *n* linearly independent sources in addition to any neural or other artifact sources. Thus, it will be impossible to perfectly isolate any single source. While steps can be taken to minimize this sensor noise (e.g., averaging multiple samples of the EEG responses to the same class of events), there may still be more neural and EEG artifact sources than electrodes and these may force a degree of filter error, sometimes called "misallocation of variance" (Wood & McCarthy, 1984). Also notice that if the gain vectors of two of the sources were parallel (i.e., the angle between the vectors was 0 or 180 degrees), that no spatial filter would be able to differentiate between the two of them. Spatial filters depend on topography differences to selectively re-weight sources.

While it is important to keep these possible LSF limitations in mind, in practice, LSFs may be able to quite accurately extract sources with large activations or gain vectors especially when numerous electrodes are used to measure the EEG. When a source's activity or gain vector is large, its contribution to the LSF output can effectively overpower that of other sources that the LSF is not able to ignore (Ghahremani, Makeig, Jung, Bell, & Sejnowski, 1996; Kobayashi, James, Nakahori, Akiyama, & Gotman, 1999; Makeig, Jung, Ghahremani, & Sejnowski, 2000). Moreover, with each additional electrode, an LSF gains a free parameter that can improve (or at worst, leave the same) the LSF's ability to extract a particular source.

---

vector) since the *n* gain vectors of the unwanted sources span the space of possible vectors. Thus **w** would not be able to extract any sources.

Additional electrodes also reduce the likelihood that multiple sources have indistinguishable topographies.

Indeed, the greatest difficulty with using LSFs to extract EEG sources is probably not these fundamental limitations. Rather, it is knowing how to derive an LSF that accurately extracts large EEG sources of interest. This difficulty stems from the fact that it is impossible to know, with certainty, what the sources of a particular set of scalp potentials are. As there are possibly many more scalp potential sources than electrodes, there are an infinite number of source configurations that could have generated a given segment of EEG data (this is known as the "EEG inverse problem"—Kutas & Dale, 1997). Because the exact source configuration is uncertain, it is not clear what the latent sources are to be extracted let alone what LSF, if any, could extract them.

To deal with this problem, it is necessary to make assumptions about the sources of an EEG dataset in order to derive LSFs that might extract some of these sources. The sections that follow review three methods for deriving LSFs from EEG datasets. The first two of these, independent component analysis (ICA) and principal components analysis (PCA), derive LSFs primarily (or in ICA's case, entirely) from statistical assumptions. In contrast, the lattermost class of techniques, source localization, relies primarily on anatomical models to derive LSFs. The emphasis of the review will be on ICA, the method used in this dissertation.

III.D Independent Component Analysis for Deriving LSFs

ICA is a blanket term for a family of algorithms that learns a set of linear filters from the statistics of a dataset by trying to make the output of the filters temporally independent (i.e., the output of one filter at time *t* provides no information about the output of any of the other filters at time *t*) or relatively so (Hyvärinen et al., 2001). The logic of ICA, as typically applied to EEG data, is clear when formalized as an attempt to fit a linear generative model to a time series of observations (Roweis & Ghahramani, 1999):

$$\mathbf{x}(t) = \mathbf{A}\mathbf{u}(t) \tag{5}$$

where **u** is an *n*-dimensional vector of independent sources whose activity linearly, instantaneously combines via a full rank "mixing matrix," **A**, to create the *n*-dimensional observations, **x**[39], which in our case is an EEG dataset. The problem for ICA is then to learn **A** or equivalently, the "unmixing matrix," **W**, which is the inverse of **A** and consists of a basis of *n* LSFs that extract the latent source activity.

$$\mathbf{u}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{A}^{-1}\mathbf{x}(t) \tag{6}$$

The intuitive reason why it should be possible to learn such "unmixing" LSFs is that the dimensions of a mixture of independent sources will generally be somewhat

[39] For some ICA algorithms, the dimensionality of the sources and data do not have to be the same, but when ICA is applied to EEG/ERP data it is typically assumed that $n=m$ and that **A** is full rank.

dependent, since the activity of a single source is apparent in multiple dimensions. Thus, by trying to unmix the data in a way that minimize these temporal dependencies, ICA may be able to recover the latent sources (e.g., Figure III.5). The specifics of how this unmixing is done differs from ICA algorithm to algorithm. Some ICA algorithms find **W** analytically by estimating statistics of **x** (e.g., second-order blind identification--Belouchrani, Abed Meraim, Cardoso, & Moulines, 1997) while others iteratively improve the fit between an estimate of **u** and the assumed generative model (e.g., extended INFOMAX ICA--Lee, Girolami, & Sejnowski, 1999).

These different algorithms correspond to different assumptions about the statistical properties of the latent sources. To learn **W**, each algorithm has to assume that the sources are somehow non-Gaussian or are somehow non-white (i.e., the activity at each time point is not independent of all other time points), because if the sources are white Gaussian, then there are an infinite number of unmixing matrices that will make the unmixed data independent and the problem is under-constrained[40]. Perhaps the most popular ICA algorithm applied to EEG is extended INFOMAX ICA, which assumes independent supergaussian or subgaussian sources but makes no additional assumptions about temporal structure. Supergaussian random variables (e.g., random samples of the potentials generated by occasional bursts of muscle activity) have a probability density function (pdf) with a much stronger peak at the

[40] Gaussian random variables are independent if they are uncorrelated. For any particular mixture of random variables, one can find an unmixing matrix, **A**, that "spheres" the mixture such that each dimension is uncorrelated and of equal variance. Post-multiplying **A** with an orthonormal matrix **J** will simply rotate the output of **A**. Since a rotated sphere of probability density is indistinguishable from the original sphere, the output of **JA** is also uncorrelated and of equal variance. Since there are an infinite number of orthonormal matrices **J**, there are an infinite number of sphering matrices, **JA**.

mean and heavier tails than a Gaussian distribution with the same variance (Figure III.6). A subgaussian random variable (e.g., random samples of a sine wave) has a much flatter pdf around the mean than a Gaussian distribution with the same variance (Figure III.6). Alternatively, another ICA algorithm, second-order blind identification (SOBI), assumes that the sources are uncorrelated at multiple time lags (e.g., the activity of a source at time *t* is uncorrelated with the activity of another source at time *t*+100 ms) and have unique normalized spectra, but assumes nothing about the Gaussian nature of the sources.

III.D.1 An Example of ICA

To illustrate how ICA might be useful for extracting spatiotemporally overlapping ERP components, consider a hypothetical 12 electrode EEG dataset that was generated by 8 sources and 12 mild uncorrelated white Gaussian sensor noise sources (SD=1.25 μV; in comparison the SD of each neural source at the electrode to which it maximally projected was at least 3.42 μV) in two experimental conditions. The average activations of the neural sources (i.e., "source ERPs") in the two conditions are plotted in Figure III.7 and the linear correlations between each possible pair of sources are presented in Table III.1 (A). The average responses of many sources appear to differ significantly across the two conditions, are generally mildly correlated with one another (median $|r|$=.25) and can be quite correlated. For example, the ERPs of Sources 5 and 8 are effectively identical and the ERPs of Source 2 are rather correlated to those of Sources 1 and 3. However, the strong source ERP

correlations are not necessarily reflective of the degree of correlation in the single trial activity from which the source ERPs are derived. These correlations are presented in Table III.1 (B) and are generally near zero (median $|r|$=.02); Figures III.8-9 illustrate how this is possible. The one pair of sources whose single trial activations are correlated, Sources 5 and 8, represents two sources that are cross-hemisphere homologues of one another.

These sources, plus the 12 sensor noise sources, combine at the scalp via a 20 x 12 gain matrix. Each gain vector is visualized topographically in Figure III.8 and the similarity of each possible pair of gain vectors is presented in Table III.2. Notice that the gain vectors of the sources generally overlap (median $|r_{cos}|$=.27) and can be quite similar. For example, the gain vector of Source 1 is quite similar to those of 2, 6, and 7.

The resulting scalp data are plotted in Figure III.11. A difference between conditions is most apparent at fronto-central electrodes. Figure III.12 visualizes the difference between conditions more clearly by plotting the "difference wave" between the two conditions, the data from Condition Blue minus the data from Condition Red. Based on the statistics of the EEG from both conditions (not shown), extended INFOMAX ICA unmixed the EEG into 12 components. The average activations of these components in the two conditions are plotted in Figure III.13. The average activations of the first eight ICs correspond quite closely to those of the eight neural sources (Table III.3: Column 4) although all of the ICs capture a bit of sensor noise. The single trial activations of these ICs are also generally similar to the eight neural

sources though much less so for ICs 1, 6, and 8 (Table III.3: A--Column 5). For the first seven ICs, the less accurate single trial activations stems from the zero-mean sensor noise, whose contribution to IC activation is mitigated by averaging (Table III.3: B--Column 4). However, the poor match between the single trial activation of IC 8 and Source 8 reflects the failure of ICA to extract Source 8 because it is highly correlated with its right hemisphere homologue, Source 5 (Table III.2: B). The last four ICs correspond to residual sensor noise.

Figure III.14 presents the topographies of the 12 ICs. The topographies of the first eight ICs correspond quite closely to those of the eight neural sources (Table III.3: Column 4). In fact the topographies of ICs and sources are generally more similar than their activations. This illustrates that while it is generally impossible to perfectly extract the activation of a source using an LSF when there are more sources than electrodes, IC topographies can still perfectly match those of EEG sources. The two ICs that most poorly match the topographies of their corresponding sources are ICs 5 and 8. This reflects the fact that the two sources are highly correlated. Indeed, the topography of IC 5 appears to be a blend of the topographies of Sources 5 and 8. Note the splotchy, non-dipolar topographies of the last four ICs that correspond to residual sensor noise.

To get a sense of what each IC contributes to the scalp data, it is useful to compute the "scalp-projection" (or "back-projection") of a component's activity:

$$\hat{\mathbf{x}}(t) = \mathbf{a}_i u_i(t) \tag{7}$$

where $u_i(t)$ is the activation of the ith component at time t and $\mathbf{a}_i$ is the mixing vector for the ith component. For example, IC 2's contribution to the difference wave is plotted in Figure III.15 on top of the difference between the ERPs from the two conditions. IC 2 contributes greatly to the difference at fronto-central midline electrodes, but contributes very little at the posterior and leftmost electrodes.

One can quantify the magnitude of this contribution by measuring the "percentage variance accounted for," ν, by the scalp-projected activity:

$$\nu = 1 - \frac{\sum_j \sum_t (x_j(t) - \hat{x}_j(t))^2}{\sum_j \sum_t x_j(t)^2} \tag{8}$$

where $x_j(t)$ is the ERP voltage at time t and electrode j, $\hat{x}_j(t)$ is the component's scalp-projection at time t and electrode j. ν is one minus the "residual variance," error measure. Residual variance is a unitless measure that is simply the sum squared difference between the scalp and scalp-projected data normalized by the power of the scalp data. It is 0 if the scalp-projected activity perfectly matches the scalp data and accounts for all of the variance of the data. Residual variance approaches infinity (and ν negative infinity) as the difference between the two grows increasingly large without bound. The variance of the difference wave accounted for by Component 2 is 39%.

Figure III.12 plots Component 5's contribution to the difference wave, which is negligible at all but the lateral posterior electrodes (Electrodes 9 and 12). This contribution accounts for 6% of the difference wave's variance. The scalp-projection of Components 2 and 5 combined accounts for 44% of the difference wave's variance (not shown). Note that the variance accounted for by both components together does not equal the sum of the variance accounted for by each component alone. This will generally be the case, unless the basis of LSFs is orthogonal (e.g., as in principal components analysis, Section III.F), and is a quirk of the percentage variance accounted for measure.

III.D.2 ICA Ambiguities

III.D.2.a ICA Scaling Ambiguity

In this simple example, readers may have noticed two ambiguities in ICA decompositions. The first being that the polarity and magnitude of the LSFs are arbitrary. To illustrate, consider two unmixing matrices **W** and **W'** that are identical except that the first row of **W'** is –2 times the first row of **W**:

$$\mathbf{w}_1 = -2\mathbf{w}_1' \tag{9}$$

The output of the first component of the **W'** will be twice as big as that of **W** and of the opposite polarity, but the outputs of the two matrices will be just as independent. Thus the two matrices are equally valid by ICA criteria.

The corresponding mixing matrices, **A** and **A'**, will have the inverse relationship. The first column of **A'** will be -0.5 times the first column of **A**:

$$\mathbf{a}_1 = -\frac{\mathbf{a}_1'}{2} \tag{10}$$

Thus the scalp-projections of the first component of the two decompositions will be exactly the same:

$$\mathbf{a}_1'\mathbf{w}_1'\mathbf{x}(t) = -\frac{\mathbf{a}_1'}{2}(-2\mathbf{w}_1')\mathbf{x}(t) = \mathbf{a}_1\mathbf{w}_1\mathbf{x}(t) \tag{11}$$

The scaling ambiguity thus only applies to component activations and scalp topographies, not to their scalp-projections. This means that when comparing the activations or scalp topographies of multiple components, some type of normalization must be done (e.g., normalizing activations to unit variance, normalizing scalp maps to unit length), which may be inaccurate. Normalization, however, can be avoided by only comparing component scalp-projections.

III.D.2.b ICA Component Order Ambiguity

In addition to IC scaling, the order of ICs is also arbitrary. Take for example two unmixing matrices found by ICA, **W** and **W'** that are identical except that the first row of **W** equals the second row of **W'** and vice versa. The outputs of the two matrices will be just as independent and equally valid by ICA criteria.

This ambiguity makes it less than straightforward to compare ICs from different ICA decompositions as it is not necessarily obvious which ICs in two or more ICA bases are equivalent. The order ambiguity is commonly resolved by ranking ICs in order of the magnitude of their scalp-projected activity (e.g., the magnitude of their scalp-projected variance). However, such rankings may not be very stable as one component's scalp-projection may be about as large as several others and the relative amplitudes of latent subcomponents in two datasets may vary due to differences between participants, recording conditions, or task demands. Thus identifying homologous ICs from different decompositions is typically done based on component features (e.g., scalp topographies, activation power spectra).

III.D.3 Physiological Motivation for ICA

While ICA was able to successfully extract almost all of the latent ERP sources in the previous example, this was because the example was constructed to largely meet the assumptions of extended INFOMAX ICA (i.e., all the neural sources were supergaussian, relatively large and few in number, and, save for two sources,

relatively independent of one another). The success of any technique for deriving LSFs to extract EEG sources depends upon how well the technique's assumptions are met by the true data generators and there is good physiological motivation for some of ICA's assumptions. Specifically there are good a priori reasons and some empirical evidence to think that some EEG sources might be independent enough for ICA to extract. A priori, the activity of functionally distinct ERP subcomponents should be somewhat independent if the independent variables that they are sensitive to are manipulated independently. For example, if one source is sensitive to word predictability and another is sensitive to word concreteness, those sources should be at least somewhat independent if concreteness and expectancy vary independently in the stimuli. Empirically, there is a great deal of latency variability in ERP components in the raw EEG (Kutas, McCarthy, & Donchin, 1977; Makeig et al., 2004), which should also produce a degree of independence (e.g., Figures III.8-9). In addition, there is evidence that the correlation between the activities of nearby cortical areas (as measured by 20 to 240 second segments of unaveraged local field potentials) generally falls off on the order of millimeters in the awake brain (Destexhe, Contreras, & Steriade, 1999).

In practice, ICA's general assumption of non-Gaussian or temporally structured sources is valid to some extent. If all EEG sources were Gaussian, then the EEG ICs (i.e., ICs that appear to correspond to EEG sources rather than EEG artifacts) derived by ICA algorithms such as extended INFOMAX ICA (Lee et al., 1999) and Fast ICA (Hyvärinen, 1999) would not be reliable and there are examples of reliable

EEG ICs found by these algorithms (Chapter V; Himberg, Hyvärinen, & Esposito, 2004). Likewise, if all EEG sources had identical normalized power spectra, then the EEG ICs derived by SOBI would not be reliable (Belouchrani et al., 1997) and there are examples of reliable EEG ICs from SOBI (Groppe, unpublished; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002). Thus these assumptions are generally met to some extent though there may be some EEG sources that violate the assumptions of particular ICA algorithms and are inextricable by those algorithms. The more specific ICA assumptions about the statistical nature of the sources (i.e., specific source probability density functions or uncorrelated autocorrelation matrices at specific time lags) are more difficult to justify, but they may not be so problematic as many sources may fit the assumptions of multiple algorithms. Also, different ICA algorithms may be used to cross-check assumptions. For example, if an algorithm that can find Gaussian sources (e.g., SOBI) does not find any such sources in a dataset, one is more justified in using an algorithm that assumes non-Gaussian sources (e.g., Extended INFOMAX ICA).

The final potentially problematic assumption of ICA is the assumption of a specific number of sources (typically as many as there are electrodes). As mentioned earlier, it impossible to know how many neural and non-neural electrical sources significantly contribute to scalp potentials. Research on simulated EEG data (Ghahremani et al., 1996; Kobayashi et al., 1999; Makeig et al., 2000) has shown that ICA is capable of accurately extracting strong EEG sources when there are more sources than electrodes.

III.D.4 Past ICA Findings

The physiological motivation for using ICA to decompose the EEG is corroborated by several successful applications. ICA was first applied to the EEG by Scott Makeig, Tzyy-Ping Jung, and colleagues in 1995 (for review see Onton, Westerfield, Townsend, & Makeig, 2006). Since then ICA has proved useful for removing EEG artifacts such as potentials caused by blinks and 60 Hz line noise (Joyce, Gorodnitsky, & Kutas, 2004; Jung, Makeig, Humphries et al., 2000; Jung, Makeig, Westerfield et al., 2000; Tang et al., 2005), improving the discrimination of ERPs in the raw EEG/MEG[41] (Serby, Yom-Tov, & Inbar, 2005; Tang, Pearlmutter, Malaszenko, & Phung, 2002; Tang et al., 2005; Tang, Sutherland, & Wang, 2006)—but see Section III.D.5.b), and accurately isolating individual early sensory (Tang et al., 2005) and epileptic spike sources (Kobayashi, Merlet, & Gotman, 2001).

It also appears that ICA is useful for extracting ERP subcomponents as multiple studies have found that ICA decomposes the EEG/MEG and ERPs into physiologically plausible, functionally distinct subcomponents (Debener et al., 2005; Makeig et al., 1999; Makeig et al., 2004; Onton, Delorme, & Makeig, 2005; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002). Since the true subcomponents of ERPs are generally unknown it is difficult to evaluate the accuracy of these

[41] MEG is primarily the linear combination of the magnetic fields generated by the synaptic activity of cerebral cortical pyramidal cells (Baillet et al., 2001; Kutas & Dale, 1997). Thus LSFs are just as potentially useful for parsing the MEG into subcomponents as EEG and the two modalities may share subcomponents. Some applications of ICA to MEG data are included in this chapter as evidence of ICA's ability to extract latent subcomponents and as evidence of possible subcomponents of the N400.

decompositions but their physiological plausibility is often quite compelling (e.g., Figure V.10). Probably the most convincing extraction of a subcomponent of a "cognitive ERP" to date is a subcomponent of the "error-related negativity" (ERN) ERP component elicited by erroneous experimental participant responses (Debener et al., 2005). Debener and colleagues found a class of central medial independent components (ICs) that contributed to the ERN and whose single trial activity predicted future behavior. Moreover, the average topography of the ICs and single trial IC activity was consistent with simultaneously recorded hemodynamic activity in the rostral cingulate zone.

Some researchers have used ICA to decompose the P3b into subcomponents. Makeig and colleagues (2004) applied ICA to the EEG recorded from 15 participants performing a visual target detection task. The ERPs to targets showed a clear P3b, which was decomposed into nine types of ICs, five of which are worth mentioning here. The "P3f" ICs had prefrontal medial distributions, tended to peak before participants' responses, and were taken to index decision making. "P3b" ICs had broad posterior medial distributions, tended to peak after participants' responses, and seemed to correspond to the conventionally-defined P3b ERP component. Left and right "mu" ICs had focal posterior-central lateral distributions and single trial power spectra characterized by peaks at 10 and 22 Hz. They contributed most strongly to the slow-wave component of the P3b, and were believed to reflect hand somatosensory and/or motor processing (e.g., tactile feedback from button pressing). Finally, the "central-midline" ICs had broad central medial topographies reminiscent of the P3a

and contributed somewhat to the P3b. These results corroborated and extended an earlier ICA decomposition of grand average ERPs[42] from the same data (Makeig et al., 1999).

Debener et al. (2005) found similar IC types in data from 33 participants in an auditory novelty oddball paradigm. Rare target and task irrelevant sounds in this experiment elicited P3b and P3a[43] ERP components. Debener and colleagues found a class of "posterior P3" ICs that contributed to both the P3b and P3a ERP components and were somewhat similar to the P3b ICs of Makeig and colleagues. The other major contributor to the ERP components was a class of "frontal P3" ICs that resembled the "central-midline" class of ICs from Makeig, et al. (2004). These ICs had broad central medial topographies, their contribution to the P3a tended to habituate over the course of the experiment, and they were interpreted as reflecting a frontal automatic brain process related to the detection of stimulus novelty and salience.

There have been two published attempts to apply ICA to EEG/MEG data from N400-eliciting paradigms. Fogelson and colleagues (2004) performed ICA separately on two EEG datasets from the same participants: one recorded during the reading of sentences ending in semantically congruous or incongruous words and the other during the viewing of number sequences that ended in expected or unexpected numbers. Although the authors did not report an N400 effect, they identified ICs that were more active in an "N400 time window" (250-550 ms) than outside of it. They

[42] Grand average ERPs are averaged over participants as well as trials.
[43] Strictly speaking, Debener and colleagues recorded a "novelty P3" not a "P3a." As discussed in Section II.E.4.a, these two ERP components are often equated and there is some evidence to support this belief.

claimed that they found similar pairs of such ICs across the two datasets although they did not report any measures of component similarity nor did they attempt to summarize the characteristics of the ICs (e.g., mean topography, ERPs, etc.). Salustri and Kronberg (2004), likewise, applied ICA to MEG recorded from participants performing a visual lexical decision task[44]. ICA was applied separately to word and pseudoword trials. There was a "left-frontal" IC in all participants when words were used as stimuli, but no such ICs in the MEG responses to pseudoword stimuli. As the MEG homologue of the N400 (N400m) was not measured, it is not known what this class of ICs contributed to it. Moreover, it is difficult to evaluate the credibility of their classes of ICs as their classes were formed by hand, measures of within class IC similarity were not reported, and the characteristics of only one or two ICs per class were reported.

III.D.5 ICA Caveats

III.D.5.a Component Reliability

While ICA is a potentially powerful tool for ERP analysis, there are a couple caveats that should be kept in mind when interpreting ICA results. The first of these is that any given IC may not be reliable (Chapter V). Since ICs are learned from the statistics of a dataset, noisy estimates of those statistics or local minima during training (for iterative ICA algorithms) may lead to spurious ICs. Moreover, there are

[44] In a lexical decision task, participants are presented with words (e.g., "dog") and pseudowords (e.g., "deg") and asked to discriminate between the two classes.

some sources that are impossible for ICA to reliably decompose (e.g., two white Gaussian sources or two out of phase sine waves--Figure V.1) as multiple decompositions provide equally independent results. This problem is far from insurmountable though as the reliability of ICs can be assessed by replicating an analysis on comparable data sets. For example, the reliability of EEG ICs can been established by finding similar ICs across multiple participants (Debener et al., 2005; Makeig et al., 2004; Onton et al., 2005) or subsets of a dataset (Chapter V; Makeig et al., 1999).

III.D.5.b Component Accuracy

The second caveat to the interpretation of LSFs derived by ICA of any other technique (see below) is that because we rarely know what the latent subcomponents of an ERP are, it is difficult to know how accurately ICs capture those subcomponents. In the absence of this knowledge, the accuracy of ICs must be assessed indirectly.

One method that has been used to assess IC accuracy is the degree to which the topography of an IC is consistent with a single ECD or a bilaterally symmetric pair of ECDs, which is quantified by modeling the source of the IC (Section III.F). The motivation for this criterion is the theory that "an independent component should account for synchronous activity within a connected cortical domain, and accordingly its scalp projection should match a single equivalent current dipole (or sometimes two bilaterally symmetric dipoles)" (Debener, Makeig, Delorme, & Engel, 2005, pg. 312). In practice, the ICs that account for the most EEG and ERP variance (and are thus

most likely to correspond to ERP subcomponents) often meet this criterion (e.g., Figures III.17 & VI.3-5). The number of such "dipolar" ICs is quite striking, especially when compared to the results of another statistically based method for decomposing the EEG, principal components analysis (e.g., Figures III.18-19; Makeig & Delorme, 2004), which is discussed in the next section.

Nevertheless, dipolarity is probably not a necessary condition for IC accuracy. For example, if the P3b is partially mediated by the widespread release of norepinephrine in cortex (Section II.D), the activity of multiple, distributed patches of cortex may be highly correlated and the topography of their contribution to the P3b might be non-dipolar. On the other hand, dipolarity is probably not a sufficient criterion for IC accuracy either, as dipolar ICs may simply reflect the fact that the potentials recorded by nearby electrodes tend to be correlated. Consider, for example, the components of the principal square root of the estimated covariance matrix (i.e., "psPCs") that the EEGLAB toolbox (Delorme & Makeig, 2004) implementation of ICA uses as its first learning step. The set of psPCs decomposes a dataset into temporally uncorrelated components and can be thought of as a version of principal components analysis (next section). The topographies of the psPCs of a participant's EEG data from the two experiments described in the next chapter are displayed in Figure III.20. Many of the topographies appear to be dipolar. Indeed, in a comparison of multiple statistically-based algorithms for decomposing the EEG, Makeig and Delorme (2004) found that the topographies of the psPCs produced more dipolar components on average than any of the ICA or other PCA algorithms in their analysis.

However, while ICA appears to have successfully extracted some readily identifiable EEG artifacts as single components, the psPCs fail to do so. For example, IC 8 (Figures III.17 & IV.6) corresponds to horizontal eye movement potentials, but the psPC decomposition splits the artifact into two components, psPC 21 and 36 (Figure III.20). Similarly, eye blink potentials are not well captured by a single independent component (IC 1--Figure III.17), but are split among several psPCs (1, 2, 3, 5, 7, 8, 9--Figure III.20). This indicates that the psPCs are not accurately extracting large EEG artifacts, which suggests that they are not accurately extracting smaller EEG sources either.

More compelling evidence for IC accuracy can come from corroborating findings from other indices of brain function. Although the relationship between other measures of brain function (e.g., fMRI, iEEG) and the EEG is ambiguous (Section II.E.1), compelling converging evidence is possible. For example the previously mentioned ICA subcomponent of the ERN (Debener et al., 2005) was consistent with simultaneously recorded hemodynamic activity in the rostral cingulate. Similarly, ICs corresponding to the SI sensory evoked potential elicited by median nerve stimulation have been found to be consistent (within 2 cm on average) with estimates of SI generator locations from previous EEG, MEG, and fMRI studies (Tang et al., 2005).

Alternatively, evidence that ICs exhibit a tighter relationship to experimental variables and events than scalp recordings boosts IC credibility. For example, Tang and colleagues have shown that the activations of ICs corresponding to sensory evoked potentials showed less trial-to-trial variability than at comparable sensors

(Tang et al., 2005). Moreover, the presence of different types of sensory evoked components were easier to detect in the activations of such ICs than in the scalp data (Serby et al., 2005; Tang, Pearlmutter, Malaszenko, & Phung, 2002; Tang et al., 2006). Such evidence may suggest that the ICs are a more accurate measure of source activity than the scalp electrodes. However, the comparisons by Tang and Serby et al. were all made using scalp data that had not been screened for EEG artifacts. Thus it is not clear if these results indicate anything more than ICA's ability to isolate EEG artifacts.

Finally, the most convincing evidence that a class of ICs is accurate is evidence that confirms predictions about the behavior of such ICs. If a class of ICs truly reflects a functionally independent EEG process, then it should be possible to design experiments that selectively manipulate that particular function. For example, in Makeig and colleague's initial application of ICA to the P3b (1999), they found a "P3f" subcomponent that they suggested reflected the activity of a fronto-parietal system for orienting to spatial stimuli and "brain systems associated with speeded manual responding." If these interpretations are correct, then participants that show a P3f in spatial target detection tasks[45] should not show a P3f in non-spatial target detection tasks. Moreover, instructing participants to emphasize response accuracy over speed should diminish the P3f. If predictions like these are confirmed by further experiments, it would increase the credibility of this class of ICs.

[45] Makeig et al. found the P3f in data obtained from participants who performed a visual target detection task in which target identity was partially determined by the spatial location of a stimulus.

III.E PCA for Deriving LSFs

An alternative method for finding a basis of LSFs to decompose ERPs into subcomponents is principal components analysis, (PCA—see Dien & Frishkoff, 2005). In the ERP literature, PCA is often called "spatial PCA" when used to find spatial filters. Like ICA, PCA learns a basis from the statistics of the data such that each component is somewhat independent of other components and is subject to the same ambiguities and caveats as ICA. PCA differs from ICA in that it is only sensitive to second order statistics (i.e., correlation) and ignores any additional temporal structure (e.g., autocorrelation). In other words, PCA only assumes that the sources of a dataset are temporally uncorrelated, which is a weaker statistical assumption than that of ICA algorithms[46]. Because of the weaker assumption, the results of PCA are fundamentally ambiguous, as for any dataset there are an infinite number of bases of LSFs that will unmix the data into uncorrelated components (Footnote III.6). The canonical way to resolve this ambiguity is to define the first principal component (PC) as the LSF that accounts for the most variance of any possible LSF. The second PC captures the most variance of any possible LSF that is orthogonal to and uncorrelated with the first PC. The third PC captures the most variance of any possible LSF that is orthogonal to and uncorrelated with both the first and the second PC, and so on. Thus, the canonical PCA solution produces a basis of

[46] While two independent random variables are uncorrelated, two uncorrelated random variables may not be independent. The lack of dependence between two dependent but uncorrelated variables will be detectable in "higher order statistics" that ICA is sensitive to (Hyvärinen, Karhunen, & Oja, 2001).

orthogonal LSFs whose outputs are uncorrelated and the top q components account for the maximum amount data variance.

Often, the canonical PCA solution is modified by ignoring the bottom PCs and rotating the top few PCs to produce a new basis. For example, the varimax rotation produces orthogonal components that tend to have extreme mixing/unmixing vector weights[47] (e.g., strong weights for a few electrodes and weak weights to the rest). When applied to ERP data, this equivalent to assuming that ERPs are composed of only a few uncorrelated subcomponents, with maximally different (i.e., orthogonal) topographies with a highly variable distribution (i.e., extreme gain vector weights).

PCA can be understood in the same generative model framework as ICA. In this framework, PCA is the optimal unmixing of a dataset when the sources, **u**, are white Gaussian random variables, possibly less than the number of electrodes, and embedded in spherical white Gaussian sensor noise, **z** (Roweis & Ghahramani, 1999).

$$\mathbf{x}(t) = \mathbf{A}\mathbf{u}(t) + \mathbf{z}(t) \tag{12}$$

If this were an accurate model of the EEG/ERPs, than PCA would be better than ICA for deriving LSFS. However, this is clearly not the case. The fact that ICA reliably finds the same components in a dataset shows that EEG and ERP datasets are not strictly white or Gaussian (Section III.D.3). Sensor noise is also not uniform across all electrodes. Peri-ocular, temporal, and occipital electrodes tend to have more

[47] PCA unmixing and mixing vectors are qualitatively the same. In other words, they are parallel and the magnitude of one is the inverse of the magnitude of the other.

noise than other electrodes in ERP datasets due to blinks/eye movements, muscle activity, and EEG alpha activity (respectively). Finally, it is unlikely that ERP subcomponents have orthogonal (maximally different) topographies and they may not fit the desideratum of rotations. For example, ERP sources that are deep within the head can have very broad scalp distributions with rather uniform weights at many electrodes. A varimax rotation would not be likely to return such a component.

Despite the implausibility of PCA's assumptions, PCA might be a useful decomposition method when there are a priori reasons to expect that there are few subcomponents with topographies that are nearly orthogonal and consistent with a particular rotation. For example, as mentioned in the previous chapter, Spencer and colleagues (Dien, Spencer, & Donchin, 2003; Goldstein, Spencer, & Donchin, 2002; Spencer et al., 2001) used PCA to decompose ERPs from variants of the auditory oddball paradigm designed to elicit P3a and P3b ERP components. In two separate datasets, they found a "P3b" PC, with a medial posterior distribution, and a "P3a" PC, with a frontal-central medial distribution. Both PCs contributed to P3b and P3a ERP components to different degrees. These results are consistent with the previously mentioned ICA decompositions of the P3b and P3a (Debener, Makeig, Delorme, & Engel, 2005; Makeig et al., 2004) at a superficial level. As mentioned previously (Section II.E.3.a), Spencer and colleagues (Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001) also applied PCA to a P600 dataset and found a "P3b" PC and a "parsing problem diagnosis" PC that both contributed to the P600 ERP component. The P3b PC had a notably more anterior distribution than the P3b PCs

found by Spencer and colleagues in the oddball datasets and the “parsing problem diagnosis” PC had a laterally symmetric occipital distribution. At present it is difficult to know how reliable or meaningful this decomposition of the P600 is as they have yet to be corroborated by other studies.

It is currently impossible to say if ICA or PCA generally provides more accurate ERP subcomponents, since we rarely know what the subcomponents of an ERP are. To my knowledge, the most careful comparison of ICA and PCA derived ERP subcomponents (Makeig et al., 1999), found that the 3 main ICs of the P3b correlated more strongly with reaction time and were more reliable than their PCA equivalents. Some researchers have compared ICA and PCA's ability to remove artifacts from EEG recordings. In contrast to ERP subcomponents, artifact components reflecting blinks, eye movements, muscle activity, and line noise are readily identifiable and ICA appears to be much better at capturing these sources as distinct components (Jung, Makeig, Humphries et al., 2000; Tang et al., 2005). I have chosen to use ICA in this dissertation rather than PCA primarily because the assumptions of PCA are less physiologically plausible and because of these empirical comparisons, although clearly more comparisons of the two methods need to be done.

III.F Source Localization for Deriving LSFs

A very different alternative for decomposing ERPs into subcomponents using LSFs is anatomically-based source localization (Baillet et al., 2001). In contrast to

ICA, which learns filters using only the statistics of scalp potentials, source localization builds a model of the EEG gain matrix that can be based on a participant's anatomy, information about probable source locations, the statistics of the scalp potentials, and the biophysics of EEG generation. Once this gain matrix, or "forward model," has been derived, an "inverse model," a set of LSFs, can be computed which estimates the activity of the putative sources. The forward and inverse models are analogous to the ICA mixing and unmixing matrices.

One of the most popular methods for source localization is to approximate the brain, skull, and scalp with a series of concentric, homogeneous spheres and to assume a handful of dipolar generators (Baillet et al., 2001). The locations and orientations of the equivalent current dipoles (ECDs) are then iteratively fit to the data until a sufficiently small degree of error is reached. The dipoles provide a forward model, $\hat{\mathbf{G}}$, which is an *n* x *m* matrix that maps the activity of *m* sources, $\hat{\mathbf{y}}$, to the potentials at *n* electrodes, **x**, which are embedded in noise, **z**.

$$\mathbf{x}(t) = \hat{\mathbf{G}}\hat{\mathbf{y}}(t) + \mathbf{z}(t) \tag{13}$$

To estimate the source activities from the data, $\hat{\mathbf{y}}$, typically the pseudoinverse of $\hat{\mathbf{G}}$ is used.

$$\hat{\mathbf{y}}(t) = (\hat{\mathbf{G}}^T\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^T\mathbf{x}(t) \tag{14}$$

This minimizes the sum squared error between the model and the scalp data.

While such an iterative method assumes fewer sources than electrodes, a different approach is to assume more sources than electrodes, which removes the need for iteration. An example of such a method is dynamic statistical parametric mapping (DSPM—Dale, et al., 2000; Liu, et al., 2002). DSPM approximates the cortical surface, inner skull, outer skull, and skin with a boundary element model and a dipolar EEG source is placed about every 1 cm on the cortical surface, oriented perpendicular to it. This approach provides a forward model, $\hat{\mathbf{G}}$, with around three thousand sources per hemisphere. Given the forward model, assumptions about the covariance of the sources, and an estimate of the noise, an inverse model can be computed:

$$\hat{\mathbf{y}}(t) = \mathbf{R}\hat{\mathbf{G}}^T(\hat{\mathbf{G}}\mathbf{R}\hat{\mathbf{G}}^T + \mathbf{C})^{-1}\mathbf{x}(t) \quad (15)$$

where **R** and **C** are the covariance matrices of the neural and noise sources respectively. This estimation of source activity is optimal when the sources and noise are Gaussian[48]. To correct for the additive noise, an additional step is taken: the estimated activity of each source is divided by the standard deviation of the noise contributing to that activity.

Relative to ICA, the advantages of source localization for ERP subcomponentry are that one can make use of knowledge about the location of ERP

[48] This is a reasonable assumption for localizing ERPs as they are derived from the mean of a large number of epochs of EEG (usually between 30 and 100). Thus, by the central limit theorem, the sampling distribution of the average source and noise activity at each time point should be approximately Gaussian (Wilcox, 2002). Gaussian assumptions may not be accurate for localizing single epochs of EEG or ERPs constructed from a limited number of epochs.

sources and the statistics of the noise when it is available. Furthermore, one does not have to make strong independence assumptions, localization does not suffer from scaling ambiguities (only polarity ambiguity), and it is easier to compare results across participants, studies, and other neuroimaging methods (e.g., functional magnetic resonance imaging, intracranial recordings).

The disadvantage of source localization is that inaccuracies in the forward model and statistical assumptions may lead to erroneous results. For example, while the boundary element model used by DSPM is a more accurate approximation of the head than the concentric spheres used in iterative dipole fitting, both ignore the fact that some parts of the brain are anisotropic and that different regions of skull differ in their conductive properties (Baillet et al., 2001). More problematic for source localization are assumptions about the number and locations of sources, which are rarely known with much certainty. Iterative dipole fitting may produce accurate results when fitting a very small number of dipoles (e.g., Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002). However, as the number of dipoles increases, the stability of the solutions rapidly degrades and different initial estimates of source locations and orientation can produce very different results (Baillet et al., 2001). While DSPM does not the have the stability problems of iterative solutions, by potentially greatly over-estimating the number of sources, DSPM will sub-optimally localize activity (Liu et al., 2002).

Source localization and ICA are complementary as they exploit different sources of information and can make compatible assumptions; source localization

depends on strong anatomical assumptions and ICA depends on strong statistical assumptions. Given this, localization and ICA can be used to corroborate one another's results. Moreover, ICA can function as a useful preprocessing step in iterative source localization. Iterative source localization works best when modeling data with few dipoles. As mentioned previously, the topographies of many ICA components are consistent with a single dipole or dual-symmetric dipolar source (e.g., 28% of non-EEG artifact ICs were consistent with a single dipole in one 64 channel study—Debener, Makeig, Delorme, & Engel, 2005). Thus while an ERP component may be too spatially complex for iterative localization to find a reliable solution, localizing ICs that contribute to that ERP may simplify the problem enough to make it tractable and thereby improve accuracy. In fact, research on simulated and real data have shown that ICA can improve localization accuracy relative to conventional iterative methods (Kobayashi, Akiyama, Nakahori, Yoshinaga, & Gotman, 2002; Tang, Pearlmutter, Malaszenko, Phung, & Reeb, 2002), although not always (Kobayashi et al., 2001).

Conversely, source localization may assist ICA (Onton et al., 2005). Localizing ICs will remove ICA's magnitude ambiguity, facilitate comparison with results from other imaging methods, and ease the comparison of ICs across participants. The last of these benefits results from the fact that anatomically equivalent sources in two different participants can have significantly different scalp distributions due to potentially functionally irrelevant differences in cortical folding and head shape. Thus ICs that represent the activity of such a pair of sources will

have dissimilar topographies. Localizing can correct for these irrelevant differences and can reduce the dimensionality of an IC's topography (e.g., from 64 electrodes to 6 Cartesian coordinates), which facilitates analysis (e.g., clustering) and visualization. Iterative source localization will be used in conjunction with ICA in this thesis for the former purpose.

III.G Temporal PCA for Defining ERP Subcomponents

In contrast to the previously mentioned methods for defining ERP subcomponents with LSFs, some researchers have used linear temporal filters (LTFs) derived by PCA to decompose ERP subcomponents. This method, called "temporal PCA," is simply PCA (see section III.G) where the individual time samples (e.g., 100, 104, 108 ms post-stimulus) are treated as random variables (Dien & Frishkoff, 2005). As typically applied, the generative model for the data is:

$$\mathbf{x}(e_{ac}) = \mathbf{A}\mathbf{u}(e_{ac}) + \mathbf{z}(e_{ac}) \tag{16}$$

where each element of the vectors **x** (scalp data), **u** (source), and **z** (noise) represents a different time sample and e_{ac} indexes an electrode, *e*, on participant *a* in experimental condition *c*.

Applied this way, PCA decomposes ERPs into subcomponents whose time courses are fixed, mutually orthogonal, and whose topographies are uncorrelated and

can change across experimental conditions and participants. As with spatial PCA, the top few PCs account for the great bulk of data variance and may be rotated to be "simpler," while the lower PCs are ignored.

In contrast to LSFs, temporal PCA is not well motivated physiologically. According to the biophysics of ERP generation, ERPs are a linear combination of temporally varying source activity with fixed topographies; sets of LSFs can selectively filter out some of those sources. Nothing about the physiology of ERP generation suggests that it makes sense to think of ERPs as linear combinations of temporally fixed waveforms with fluid topographies. In fact such subcomponents violate conventional definitions of ERP components as having fixed scalp distributions but varying latencies. Because of this lack of physiological justification, I believe LSFs are more suitable for defining ERP subcomponents.

III.H Summary

If an ERP component is composed of multiple, spatiotemporally overlapping subcomponents that are generated by distinct sets of cortical areas that have at least somewhat distinct topographies, linear spatial filters (LSFs) can potentially extract those subcomponents. However, because of the ambiguous relationship between scalp electric potentials and their neural and non-neural sources, it is difficult to know how to derive LSFs that successfully do this. By making assumptions about the sources of

scalp electric potentials it is possible to derive successful LSFs if the assumptions are valid.

One technique for deriving LSFs from EEG data is independent component analysis (ICA), which learns LSFs by making assumptions about the statistics of the underlying generators. More specifically, ICA assumes that the underlying generators are temporally independent and somehow non-Gaussian or temporally structured. In addition, ICA, as typically applied to EEG data, assumes the number of sources equals the number of electrodes. There is both a priori and empirical evidence that ICA assumptions about the statistics of the sources are somewhat valid. Moreover, in practice, ICA appears to be able to successfully extract EEG artifacts (e.g., blinks, muscle potentials) and ICA often derives EEG components that are compellingly physiologically plausible. Thus it appears that ICA may be generally able to extract ERP subcomponents and it is used in this dissertation for this purpose (Chapter VI). However, the results of ICA, or any other technique for deriving LSFs, should be qualified by the fact they may be erroneous due to inaccurate assumptions and should be validated by converging evidence from other methods and tested predictions. Moreover, since ICA LSFs are learned from the statistics of a data set, it is important to assess the reliability of the derived LSFs.

ICA differs from two other methods for deriving LSFs to extract subcomponents of the EEG, principal components analysis (PCA) and source localization. Like ICA, PCA derives LSFs by assuming that the sources of an EEG dataset are somewhat temporally independent. Specifically, PCA assumes that the

sources are simply uncorrelated, which is a weaker form of independence than is assumed by ICA. Because of the weaker assumption, PCA results are ambiguous and users are forced to make additional assumptions (e.g., orthogonal LSFs and highly variable filter weights) to resolve the ambiguity. These additional assumptions are harder to justify than the stronger statistical assumptions of ICA and there is some evidence that ICA is better at extracting latent non-neural and neural sources of EEG datasets. Moreover, principal components are just as subject to the concerns of component validity and reliability as independent components. Thus ICA seems to be a more promising approach for extracting ERP subcomponents though clearly more comparisons of the two methods need to be done.

In contrast to ICA and PCA, source localization relies upon an anatomical model of EEG generation to derive LSFs (in addition to assumptions about source and/or noise statistics). The anatomical model provides some advantages over ICA and PCA in that one does not have to make strong independence assumptions, localization does not suffer from scaling ambiguities (only polarity ambiguity), and it is easier to compare results across participants, studies, and other neuroimaging methods (e.g., functional magnetic resonance imaging, intracranial recordings). However, inaccuracies in the anatomical model may lead to erroneous results, and since the locations of the generators of ERP components are rarely known, this can be a serious problem for source localization.

Finally, an alternative method for extracting subcomponents of an ERP is "temporal principal components analysis" (tPCA), which extracts ostensible

subcomponents with linear temporal filters (LTFs). As typically applied, PCA decomposes ERPs into subcomponents whose time courses are fixed, mutually orthogonal, and whose topographies are uncorrelated and can change across experimental conditions and participants. In contrast to LSFs, temporal PCA is not well motivated physiologically as nothing about the biophysics of ERP generation suggests that it makes sense to think of ERPs as linear combinations of temporally fixed waveforms with fluid topographies. In fact such subcomponents violate conventional definitions of ERP components as having fixed scalp distributions but varying latencies. Because of this lack of physiological justification, I believe LSFs are more suitable for defining ERP subcomponents.

Table III.1 – (**A**) Correlation between the hypothetical source ERPs visualized in Figure III.7. (**B**) Correlation between the hypothetical source unaveraged activity from which the source ERPs in Figure III.8 are derived.

A: Source ERP Correlation (Pearson's *r*)

y_1	1.00							
y_2	.61	1.00						
y_3	-.56	-.71	1.00					
y_4	.23	.07	.11	1.00				
y_5	.14	.25	-.33	-.31	1.00			
y_6	.28	.20	-.04	.39	-.42	1.00		
y_7	.16	.32	-.14	0	.02	.11	1.00	
y_8	.15	.26	-.34	-.32	.99	-.41	.02	1.00
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

B: Source EEG Correlation (Pearson's *r*)

y_1	1.00							
y_2	-.01	1.00						
y_3	-.01	.03	1.00					
y_4	.04	-.03	-.01	1.00				
y_5	-.02	.01	.03	-.03	1.00			
y_6	.01	.03	.02	.04	0	1.00		
y_7	.01	.05	0	.03	.01	0	1.00	
y_8	-.01	.02	.02	-.02	.83	.01	.01	1.00
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

Table III.2 – Gain vector similarity of the hypothetical neural ERP sources visualized in Figure III.10. Similarity is measured as the cosine of the angle between two gain vectors, r_{cos}. This metric is analogous to Pearson's linear correlation coefficient, *r* (i.e., 1=exactly the same, 0=unrelated, -1=exactly the same but polarity reversed).

Source Gain Vector Similarity (r_{cos})

y_1	1.00							
y_2	.96	1.00						
y_3	.37	.56	1.00					
y_4	.18	.37	.67	1.00				
y_5	.03	.02	-.18	.30	1.00			
y_6	.88	.90	.65	.21	-.19	1.00		
y_7	-.89	-.84	-.19	-.14	-.29	-.66	1.00	
y_8	.02	-.13	-.24	-.41	.15	.04	.04	1.00
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8

Table III.3 – Topography and activation similarity of each IC and its analogous source (from the hypothetical data and ICA decomposition presented in Figures III.7-16). ICs were paired with sources according their topographies. The IC and source with the most similar topographies were paired first. Then, of the remaining ICs and sources, the IC and source with the most similar topographies were paired and so on until each IC was paired with exactly one source. Similarity is defined as the cosine of the angle between feature vectors, r_{cos}. The similarities in the second table (**B**) were computed after removing the 12 sensor noise sources.

A: Similarities Between ICs and Sources

IC	Best Source Match	Topography r_{cos}	ERP r_{cos}	EEG r_{cos}
1	1	.99	.93	.74
2	2	1.00	.92	.88
3	3	.99	.99	.95
4	4	1.00	.98	.98
5	5	.90	.99	.97
6	6	.97	.92	.83
7	7	.97	.91	.86
8	8	.84	.94	.52
9	12	.71	.61	.60
10	16	.69	.77	.74
11	18	.86	.80	.81
12	19	.76	.86	.85

B: Similarities Between ICs and Sources (without Sensor Noise)

IC	Best Source Match	ERP r_{cos}	EEG r_{cos}
1	1	.98	.96
2	2	.94	.95
3	3	.99	.98
4	4	.98	.99
5	5	.99	.98
6	6	.98	.96
7	7	.93	.94
8	8	.98	.65

Table III.4 – Activation similarity of each IC and its analogous source (from the hypothetical data and ICA decomposition presented in Figures III.7-16). These IC-source pairings are identical to those used in Table III.3. The difference between the tables is that the IC activations used here were computed from noiseless EEG data. The last four ICs are not included in the table as they correspond primarily to sensor noise. Similarity is defined as the cosine of the angle between feature vectors, r_{cos}.

Similarities Between ICs and Sources (without Sensor Noise)

IC	Best Source Match	ERP r_{cos}	EEG r_{cos}
1	1	.98	.96
2	2	.94	.95
3	3	.99	.98
4	4	.98	.99
5	5	.99	.98
6	6	.98	.96
7	7	.93	.94
8	8	.98	.65

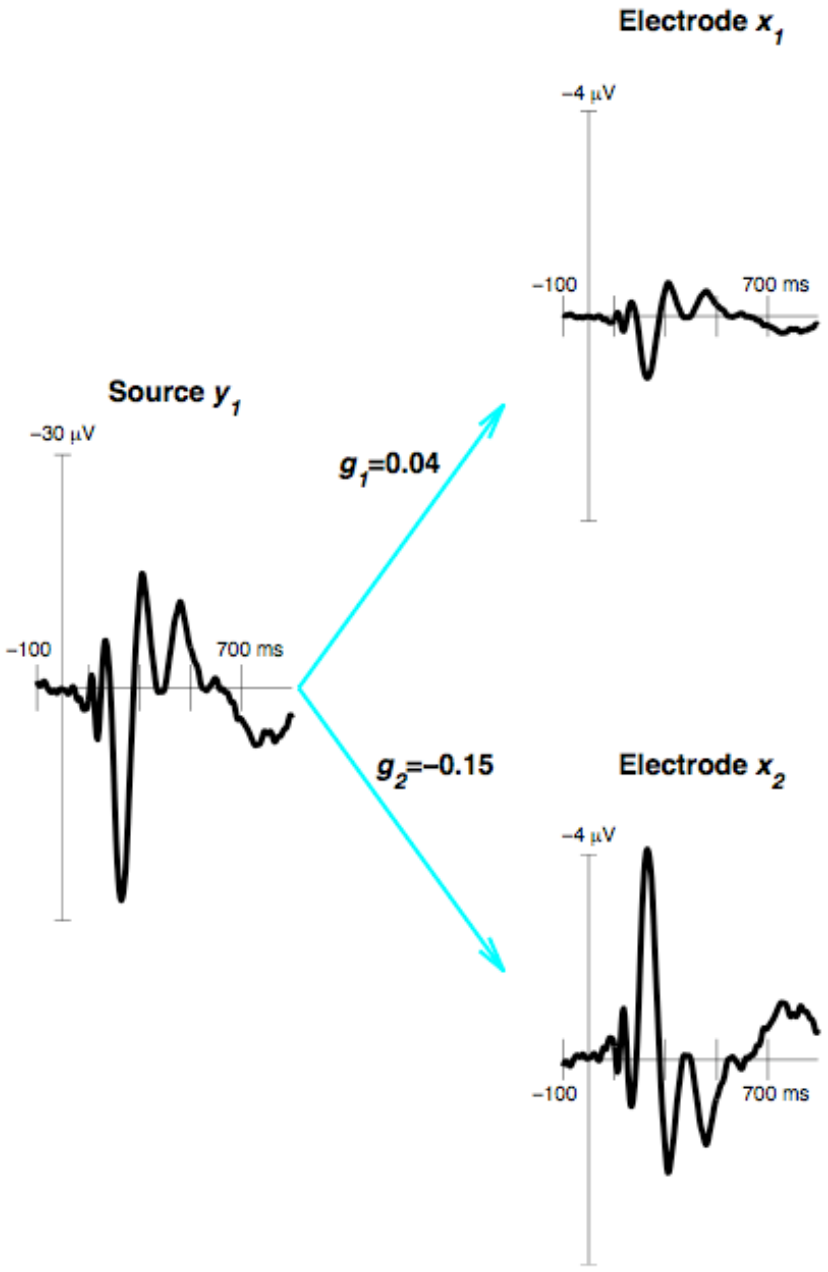

Figure III.1 – A single EEG source, y_1, and its projection to two electrodes, x_1 and x_2. The gain values between the source and each electrode are g_1 and g_2. Note the difference in voltage scale between the source and the electrodes.

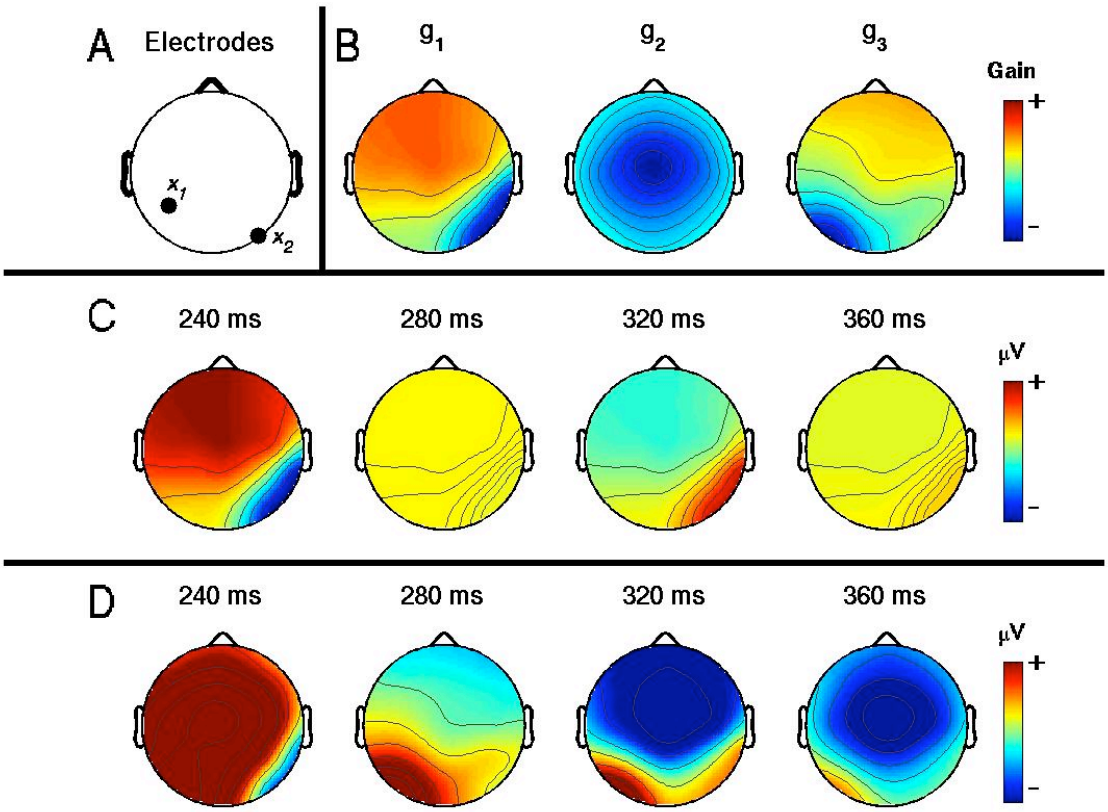

Figure III.2 – (**A**) Black dots indicate scalp positions of electrodes in Figures III.1 and III.3. (**B**) The unitless gain vectors (topographies) of the three EEG sources in Figures III.1 and III.3. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1. (**C**) Topographies of the scalp data from Figure III.1 at four points in time. Because there is only one source, the topography only changes in magnitude and polarity. (**D**) Topographies of the scalp data from Figure III.3 at four points in time. Because there is more than one somewhat independent source, the topographies qualitatively change across time.

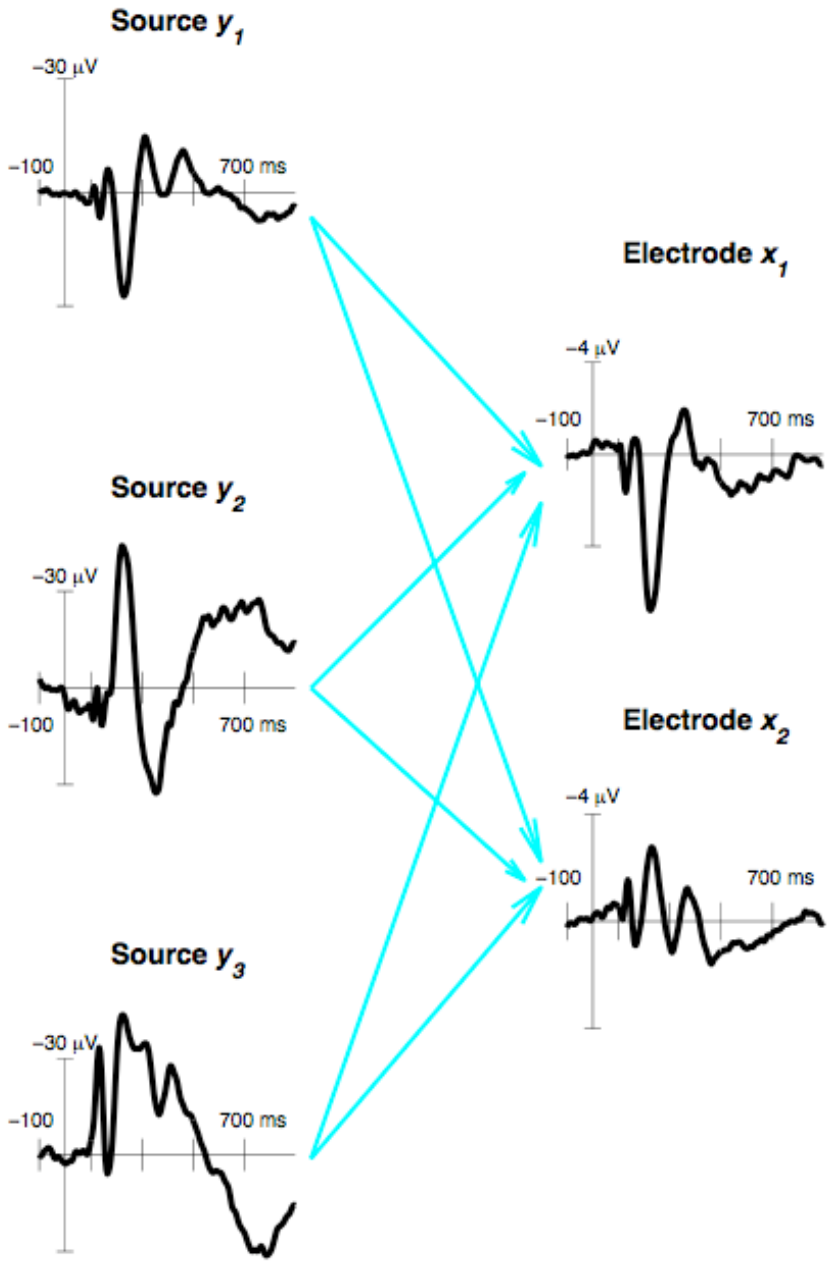

Figure III.3 – Three EEG sources (y_1 ,y_2 , and y_3), and their combined projections to two electrodes, x_1 and x_2. Note the difference in voltage scale between the sources and the electrodes.

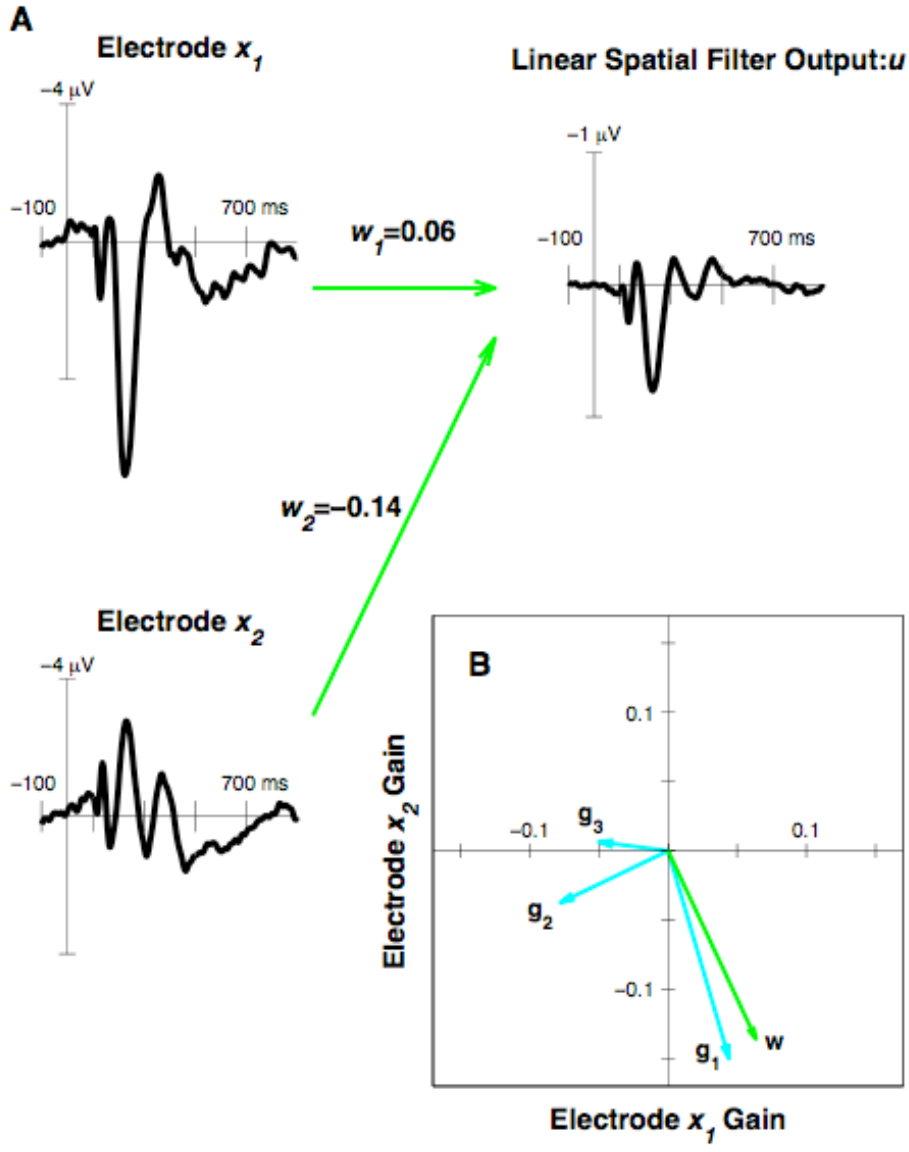

Figure III.4 – (**A**) The output, *u*, of a linear spatial filter, **w**, applied to the potentials at two electrodes from Figure III.3. Note the difference in voltage scale between the electrodes and the linear spatial filter (LSF). (**B**) The gain vectors ($\mathbf{g_1}$, $\mathbf{g_2}$, $\mathbf{g_3}$) of the three sources in Figure III.3 and the gain vector of the LSF, **w**. Since **w** is nearly perpendicular to $\mathbf{g_2}$ and $\mathbf{g_3}$, and nearly parallel to $\mathbf{g_1}$, the filter output closely correlates with that of Source $\mathbf{y_1}$ (shown in Figure III.3).

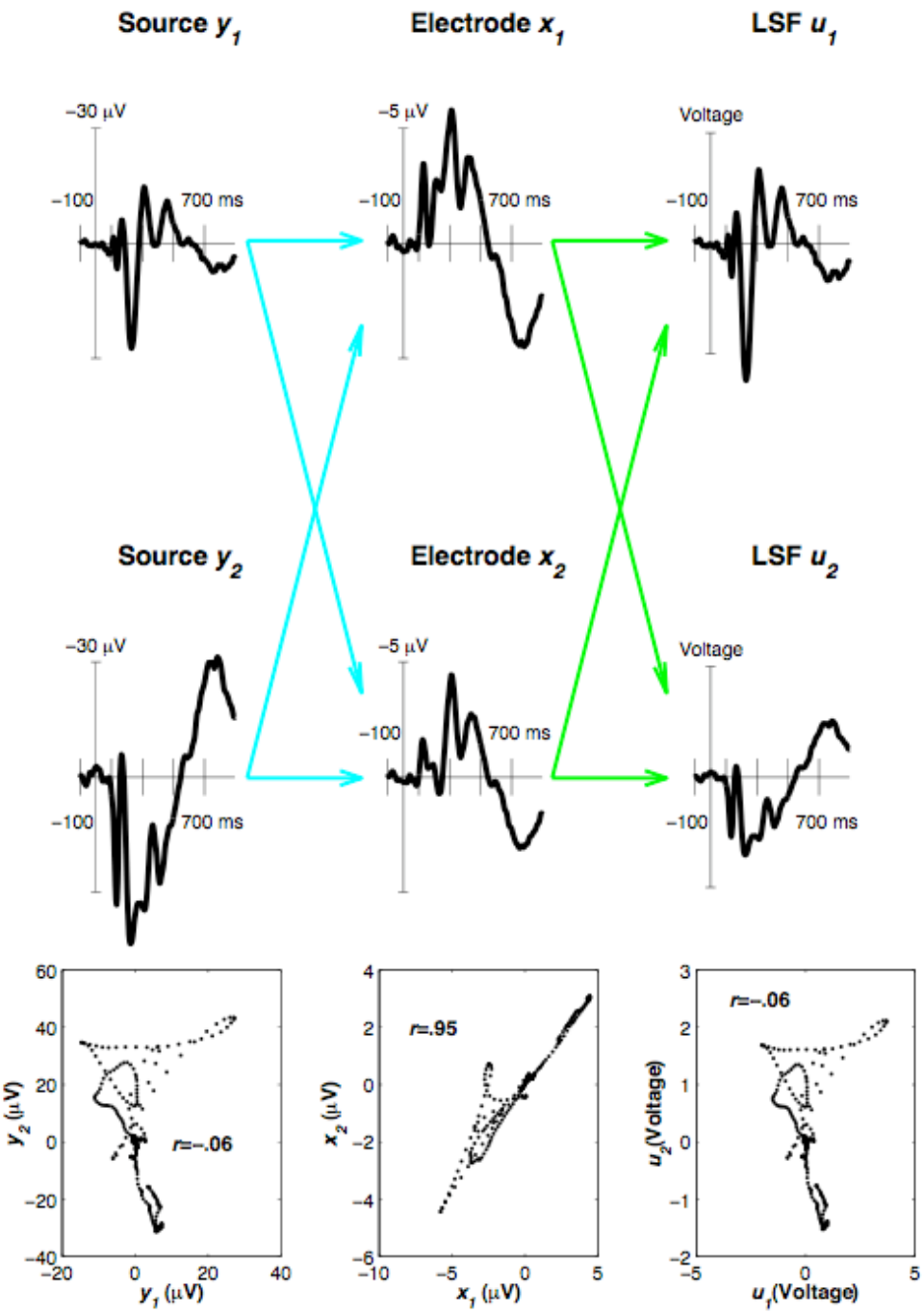

Figure III.5 – (**Left Column**) Two hypothetical EEG sources. A scatter plot of the source activity (**Left Column-Bottom**) illustrates a degree of independence between the sources. (**Middle Column**) The combined projection of the two sources at two scalp electrodes. The potentials at the two electrodes are highly correlated. (**Right Column**) The output of two linear spatial filters perfectly match the original sources and are thus somewhat independent. Blue arrows represent gain weights. Green arrows represent filter weights. Scale of LSF output is arbitrary (Section III.D.2.a).

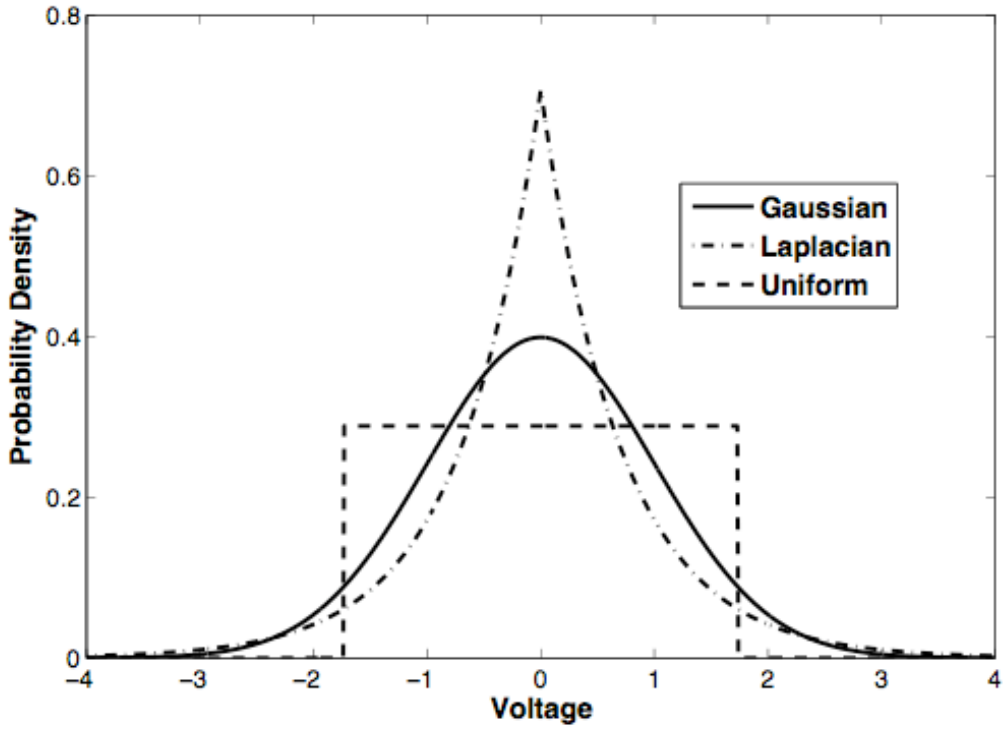

Figure III.6 – Examples of Gaussian, supergaussian, and subgaussian distributions. All distributions are zero mean and unit variance. Supergaussian random variables (in this case one with a Laplacian pdf) are more likely to take values close to the mean or extreme values than Gaussian variables. Random samples of the electric potentials generated by occasional bursts of muscle activity have supergaussian distributions because most of the time their activation is near zero, but occasionally they produce very strong potentials. Subgaussian distributions (in this case one with a uniform pdf) are more uniformly distributed around the mean than Gaussian variables. Random samples from a sine wave have a subgaussian distribution.

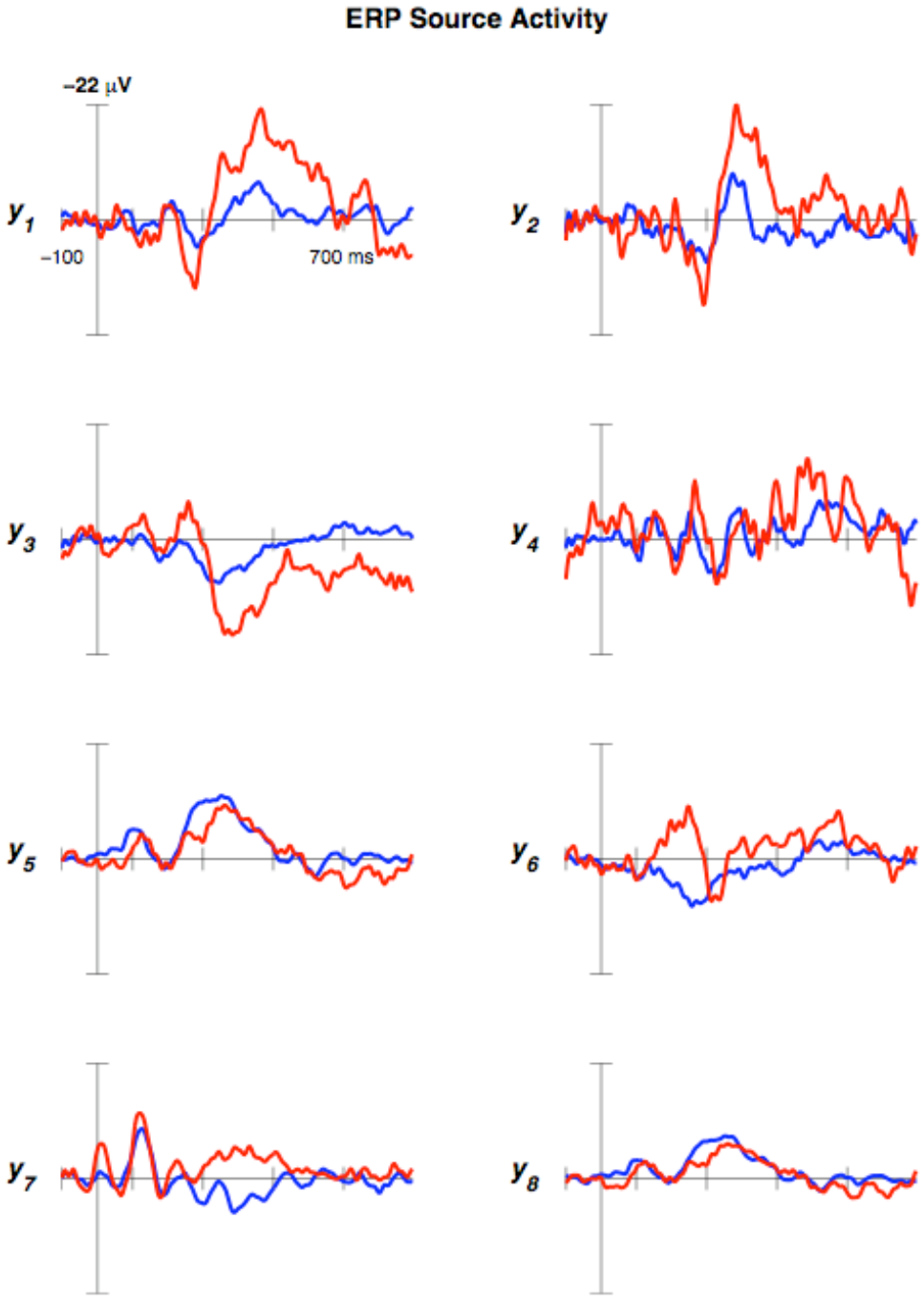

Figure III.7 – Average activity of eight hypothetical sources in two experimental conditions (Red and Blue). The activity of 12 additional noise sources are not shown.

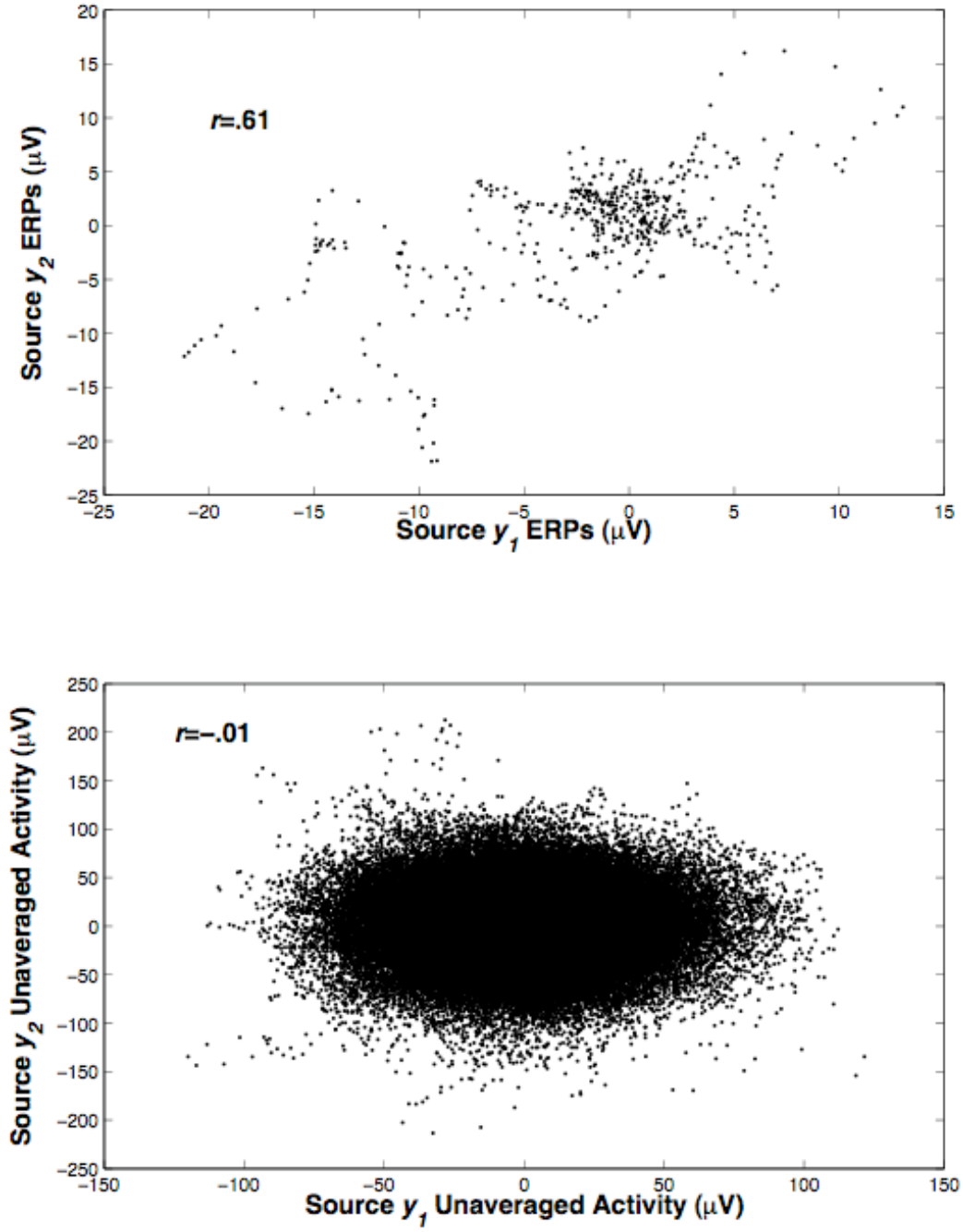

Figure III.8 – (**Top**) A scatter plot of the mean activity of Sources 1 and 2 in the two experimental conditions shows that the ERPs of the two sources are correlated. (**Bottom**) A scatter plot of the unaveraged activity of the same sources reveals that the each source's contribution to the raw EEG is somewhat independent of the other. *r* is Pearson's linear correlation coefficient.

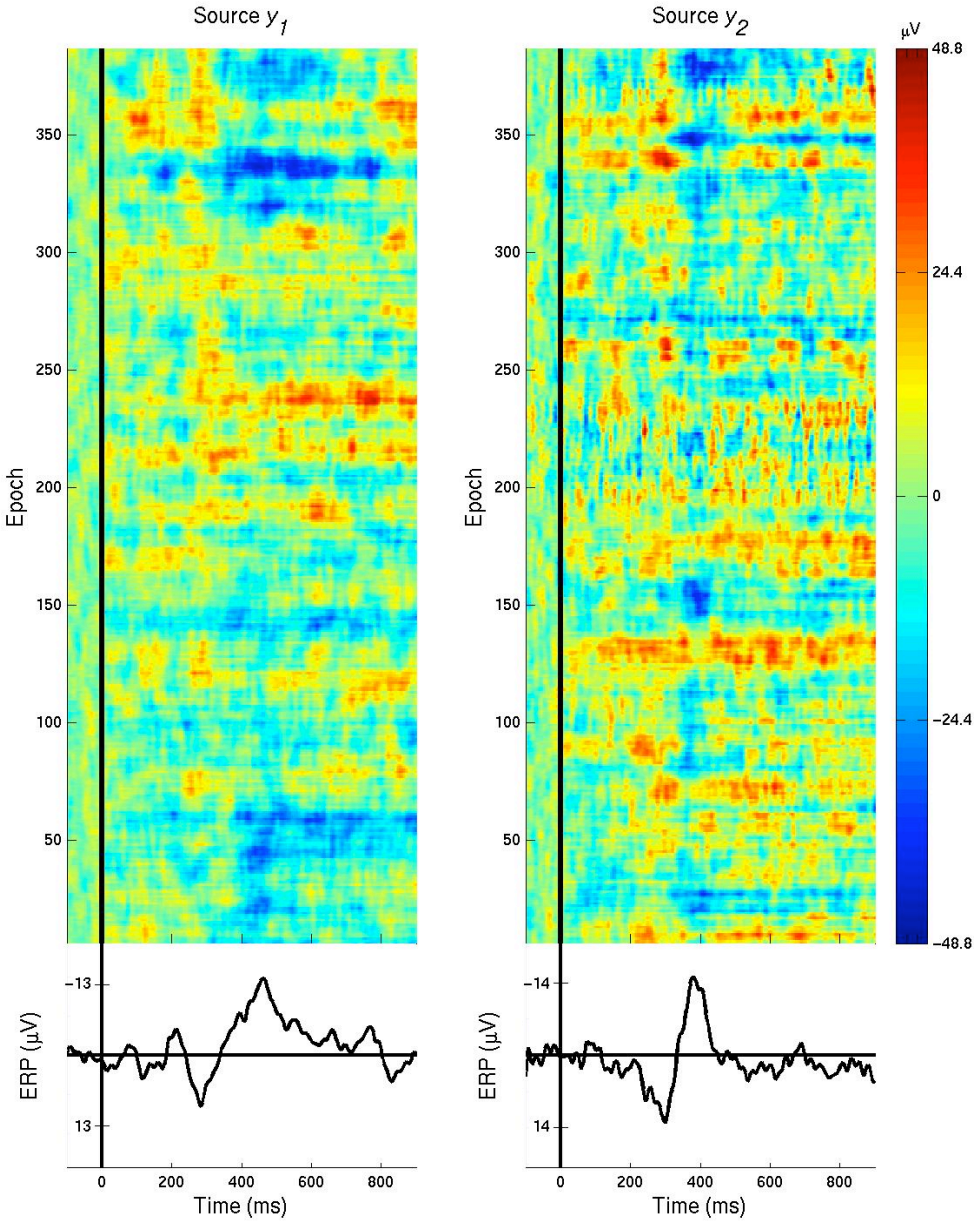

Figure III.9 – **(Top)** Single trial activity of Source 1 (**Left**) and Source 2 (**Right**) in both experimental conditions. Each horizontal line represents source activity from a single experimental trial or "epoch." Voltage is represented with color. Single trials are smoothed using a 10 trial vertical moving average. **(Bottom)** Mean single trial activity (ERP) of Source 1 (**Left**) and Source 2 (**Right)**. Although the average activations of the two sources are somewhat similar, the single trial activations are rather independent due to the large amount of single trial variation.

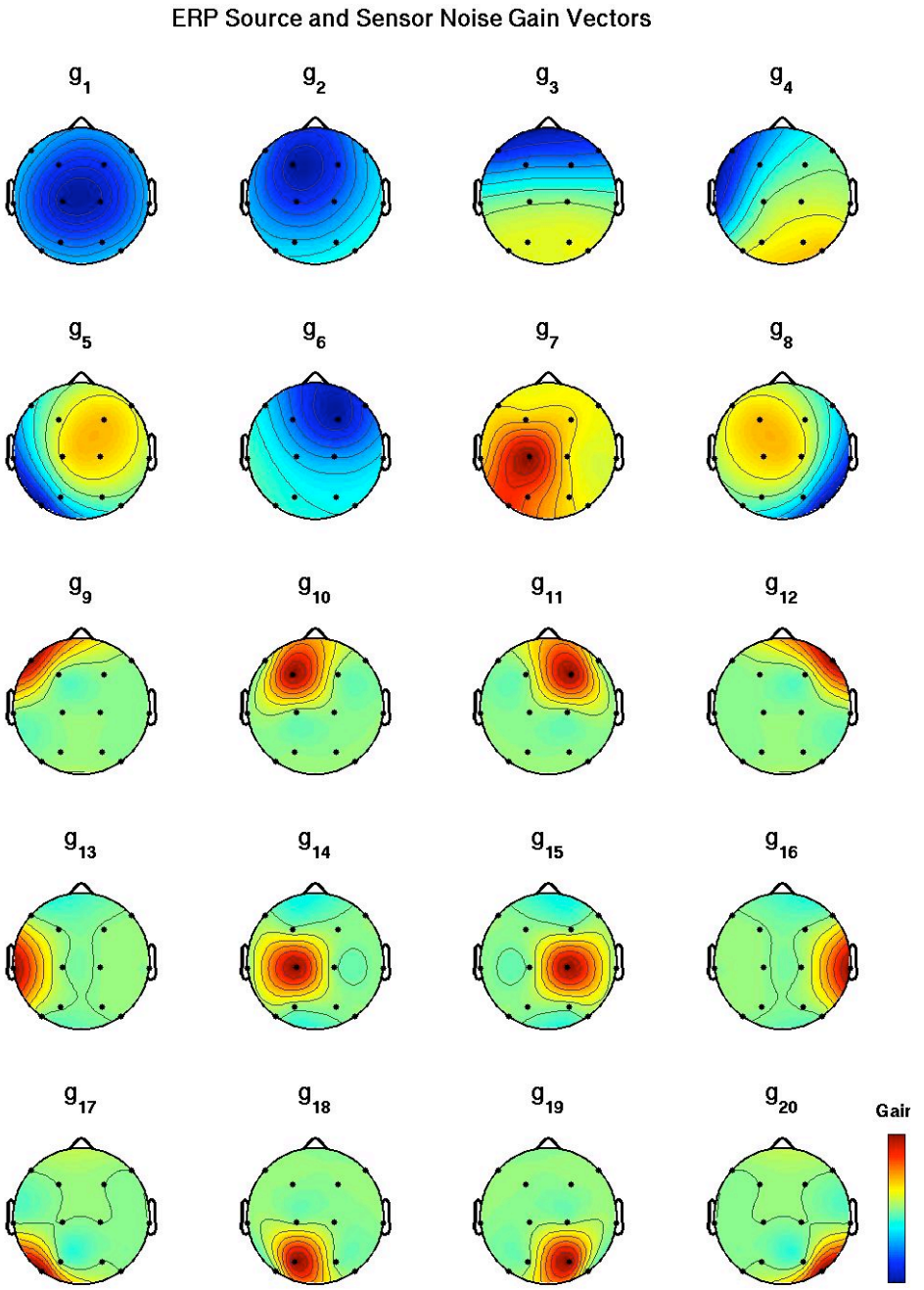

Figure III.10 – The first eight topographies correspond to the ERP sources in Figure III.7 and the last 12 correspond to sensor noise. Black dots indicate the locations of the 12 electrodes. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1. Note that although the sensor noise is confined to single electrodes, their gain weights appear to extend beyond each electrode due to the interpolation used for visualization.

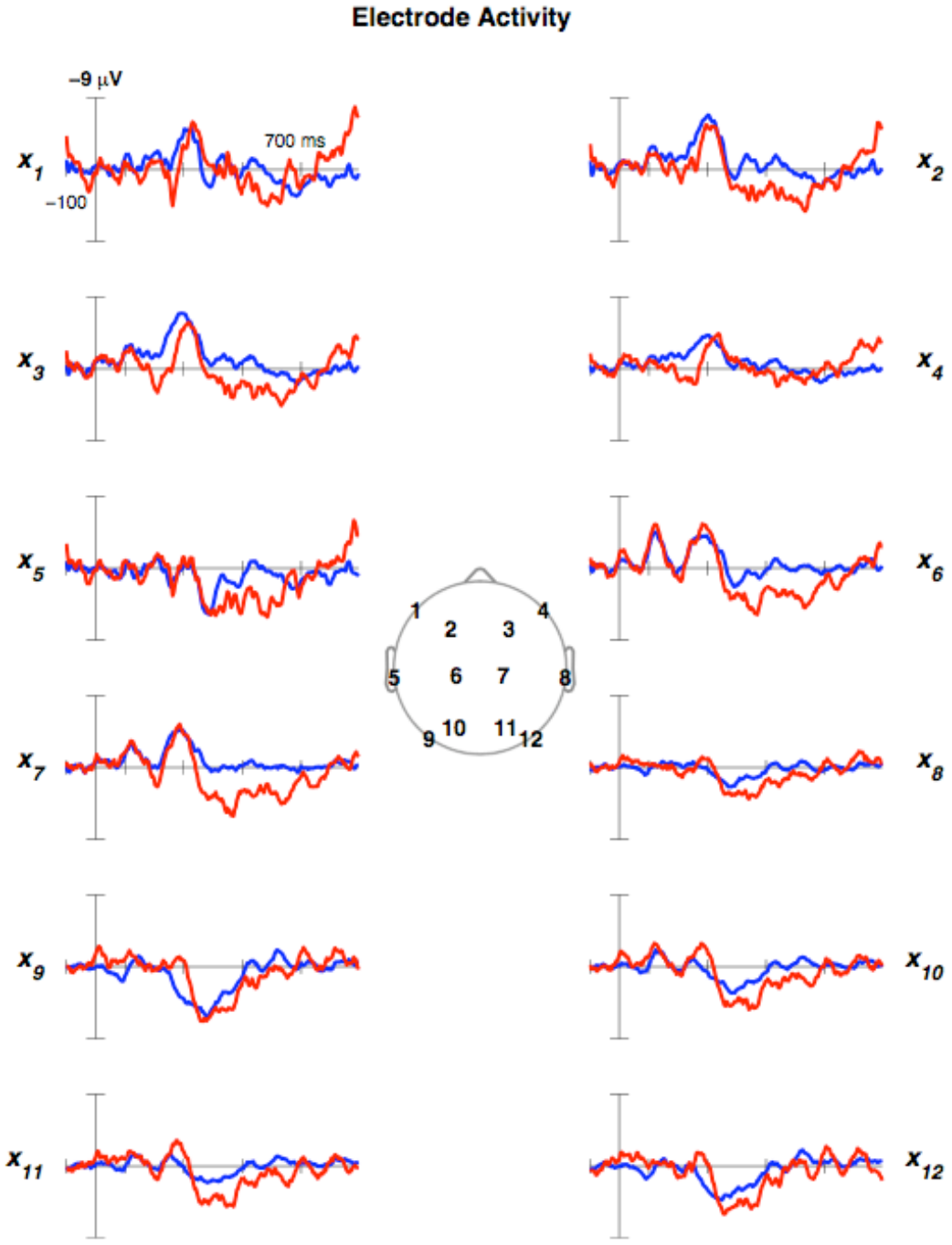

Figure III.11 – Mean electrode activity in two experimental conditions (Red and Blue) generated by the sources whose mean activations and topographies are illustrated in Figures III.7 and III.10, respectively. Cartoon head shows electrode locations.

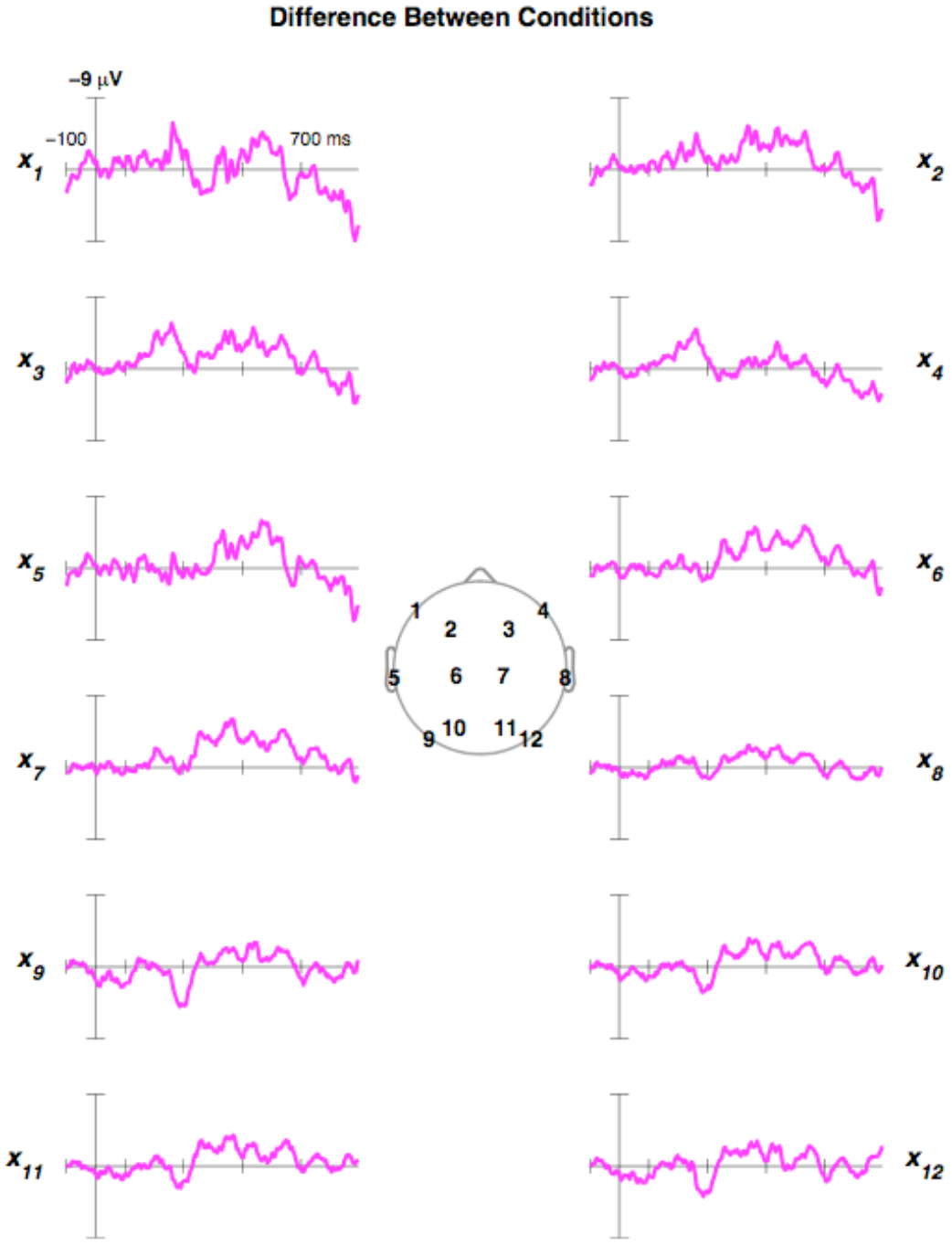

Figure III.12 – "Difference waves" obtained by subtracting the ERPs from Condition Red from those from Condition Blue (the two sets of ERPs shown in Figure III.11). Cartoon head shows electrode locations.

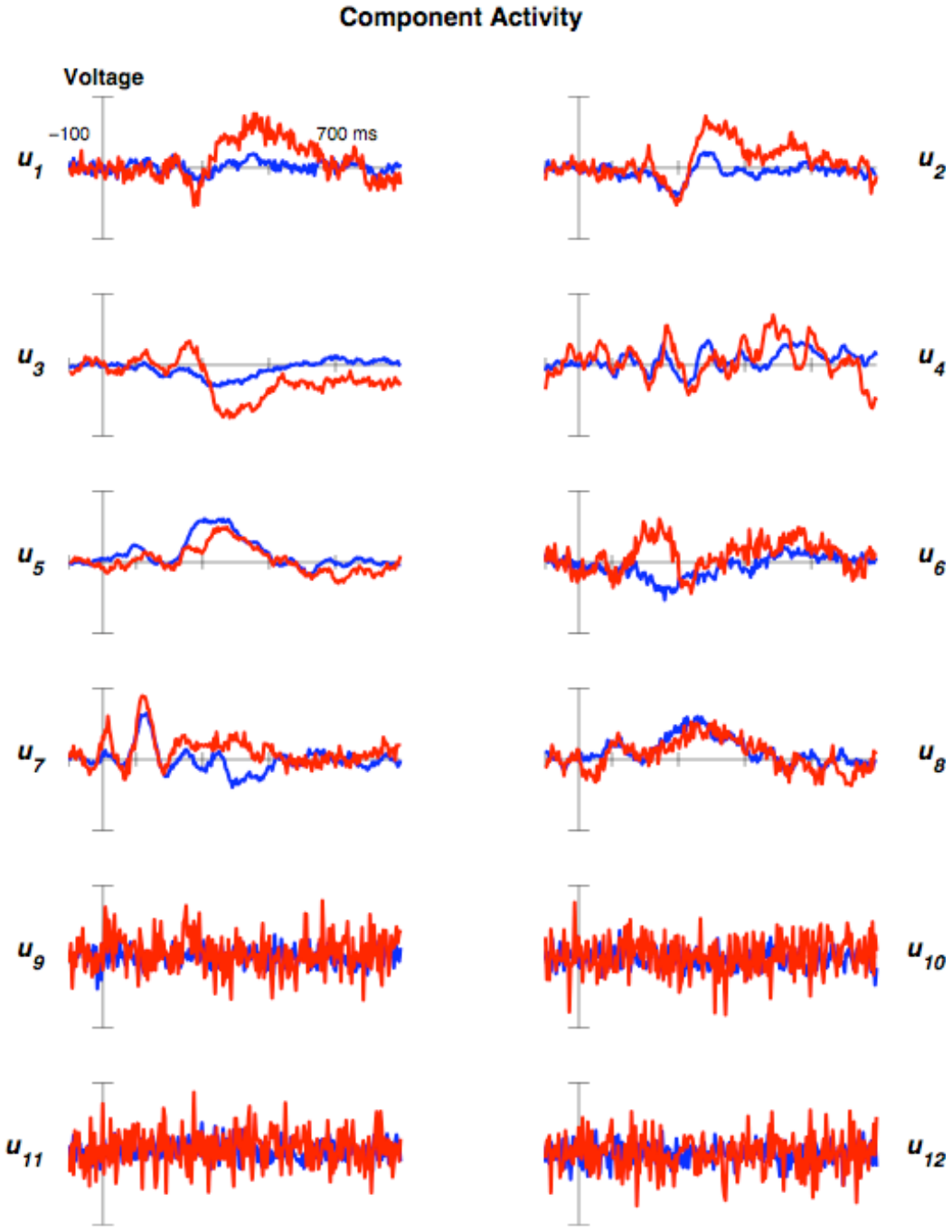

Figure III.13 – Independent component activity derived from the simulated EEG from which the ERPs in Figure III.11 were derived. For visualization, the average activation of each component has been normalized to unit variance.

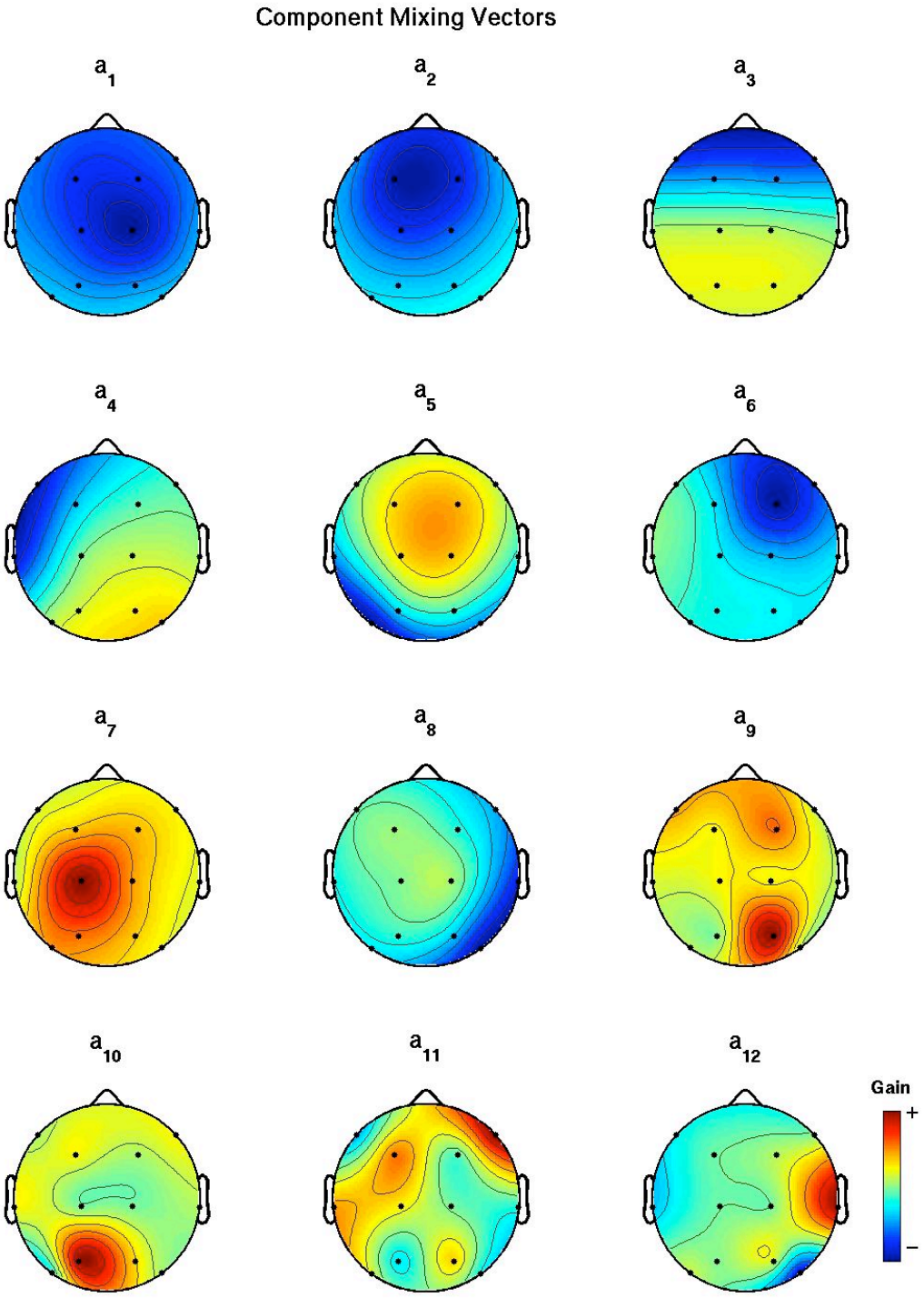

Figure III.14 – Gain vectors (topographies) of the 12 independent components whose activations are shown in Figure III.13. Black dots indicate the locations of the 12 electrodes. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1.

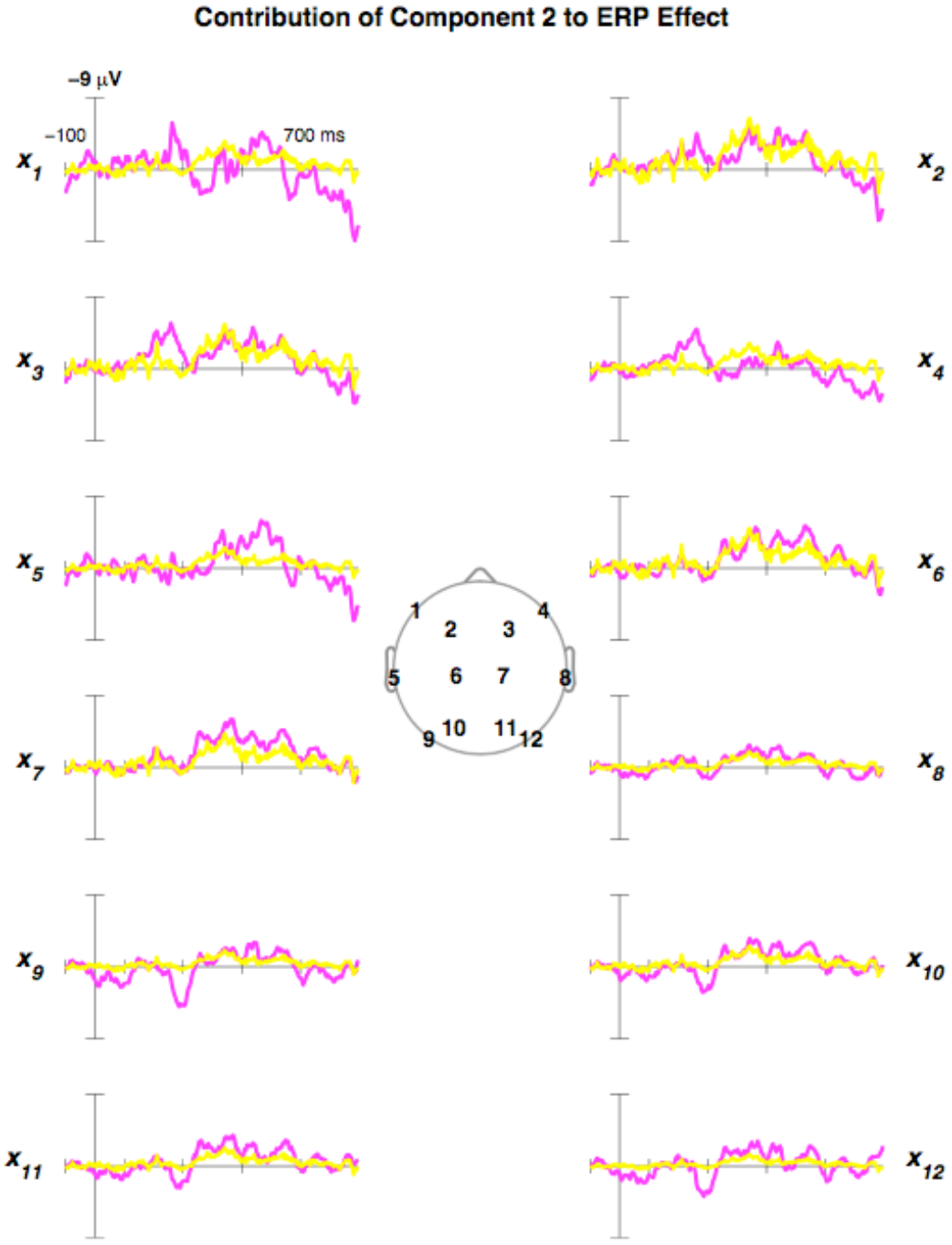

Figure III.15 – The contribution of Component 2 (yellow) to the difference waves from Figure III.12 (pink). Cartoon head shows electrode locations.

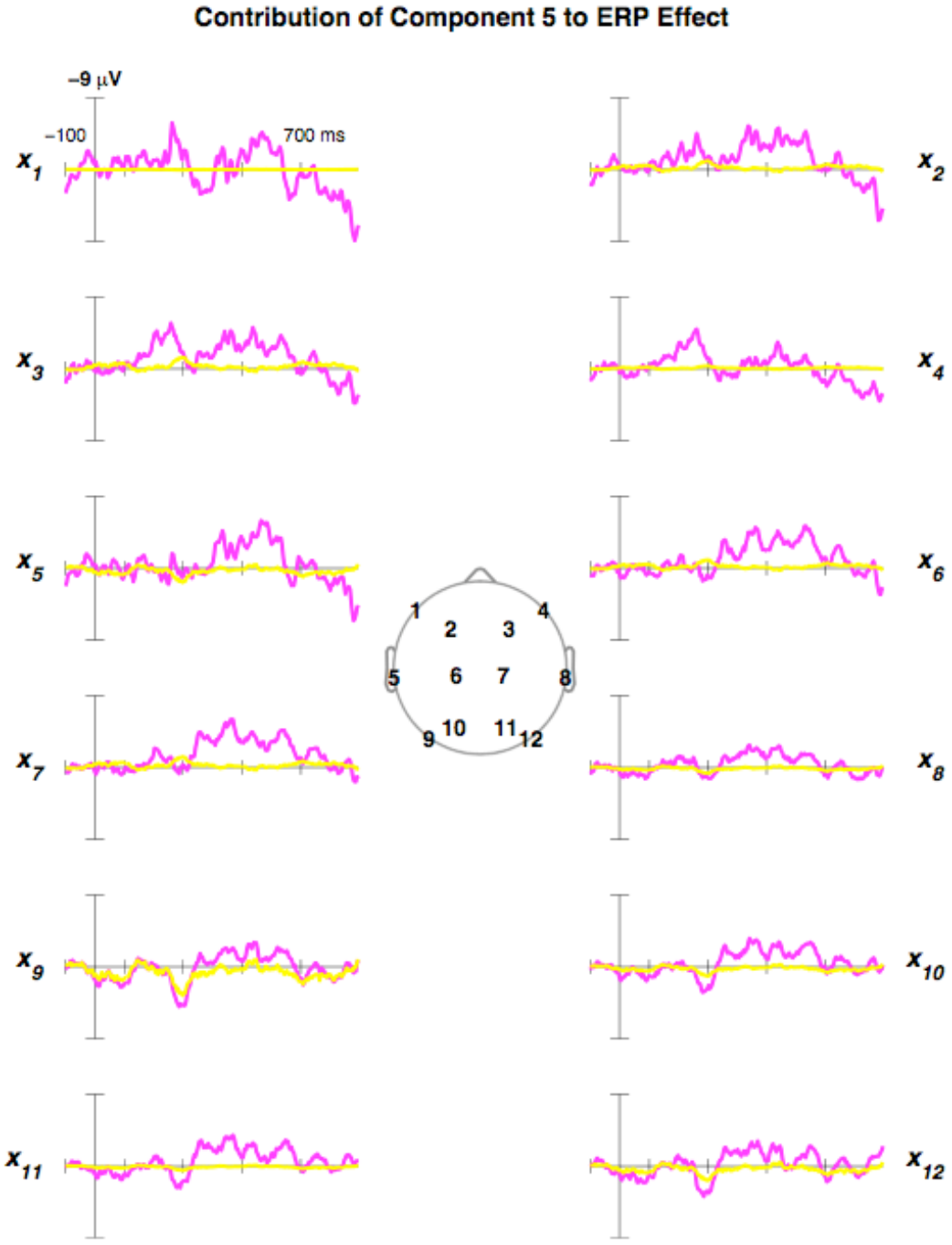

Figure III.16 – The contribution of Component 5 (yellow) to the difference waves from Figure III.12 (pink). Cartoon head shows electrode locations.

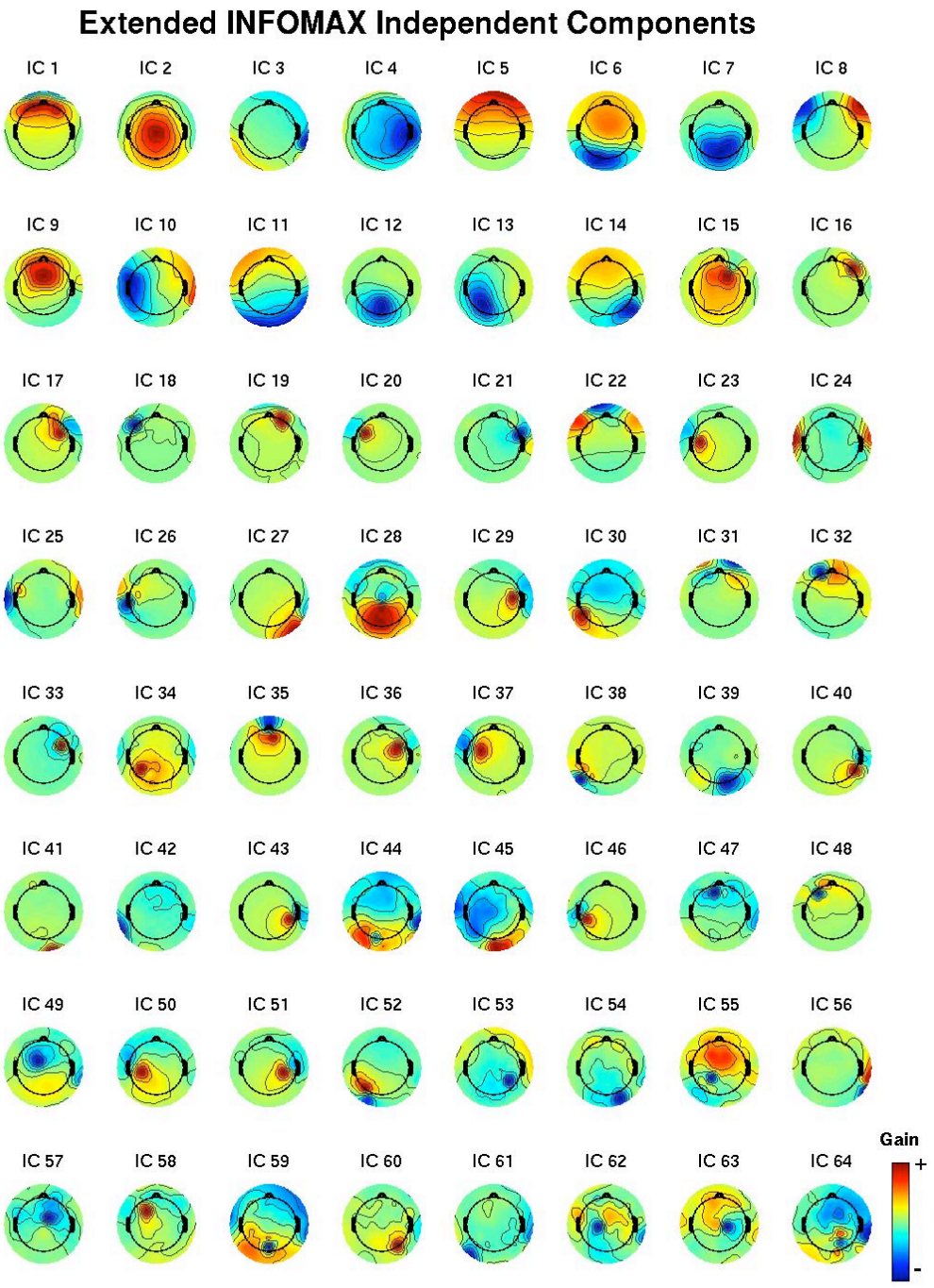

Figure III.17 – Mixing vectors (topographies) of the 64 independent components derived from Participant 15's data from the two experiments described in Chapter IV (for ICA details see Section IV.D). For visualization, each gain vector has been normalized so that its absolute maximal weight is 1 and the 64 electrodes are not shown. ICs are ranked in decreasing order of scalp-projected variance (summed across all electrodes).

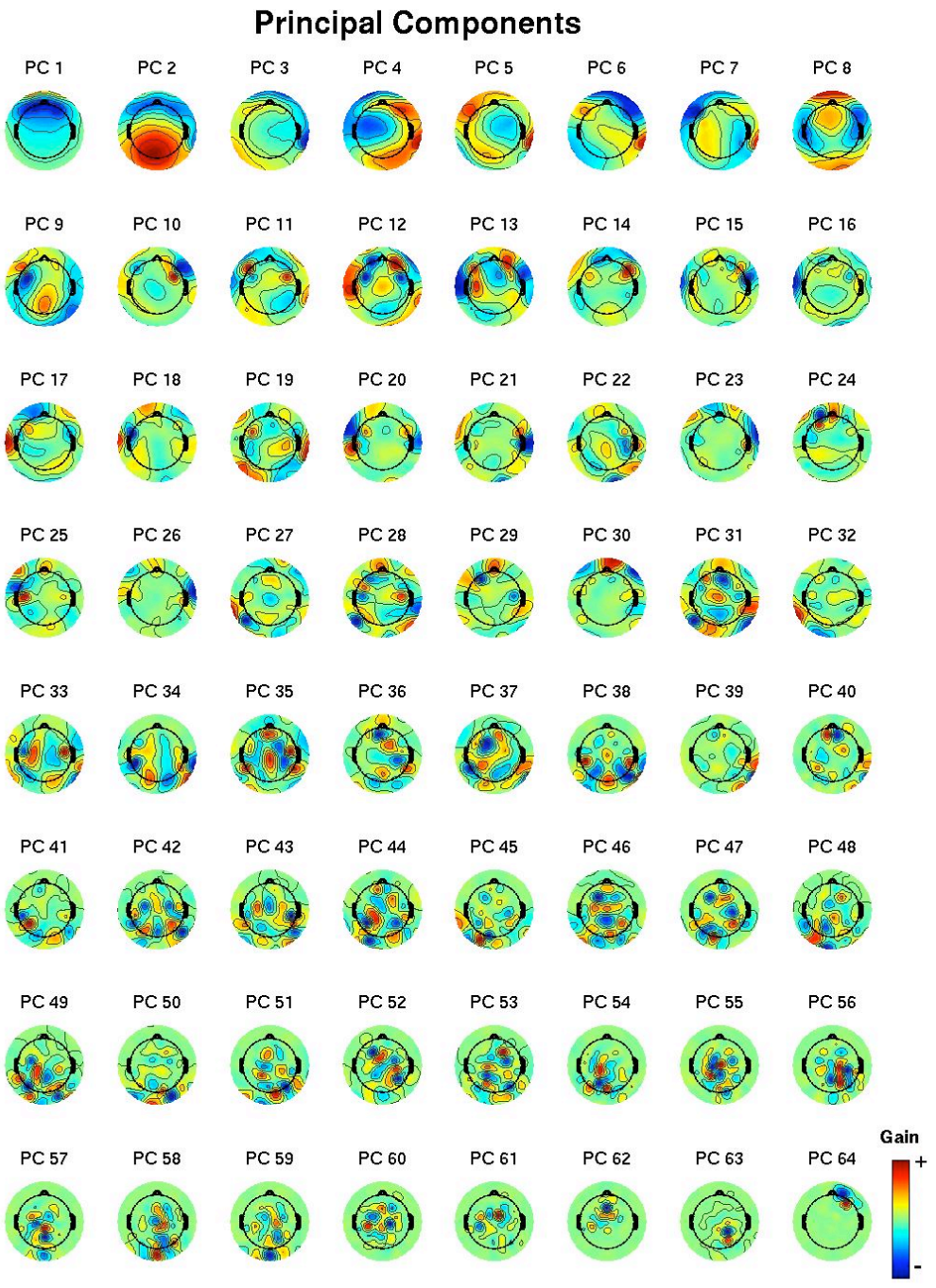

Figure III.18 – Mixing vectors (topographies) of the 64 principal components derived from the same data from which the ICs in Figure III.17 were derived. These PCs are the "canonical" PCs (Section III.E). For visualization, each gain vector has been normalized so that its absolute maximal weight is 1 and the 64 electrodes are not shown.

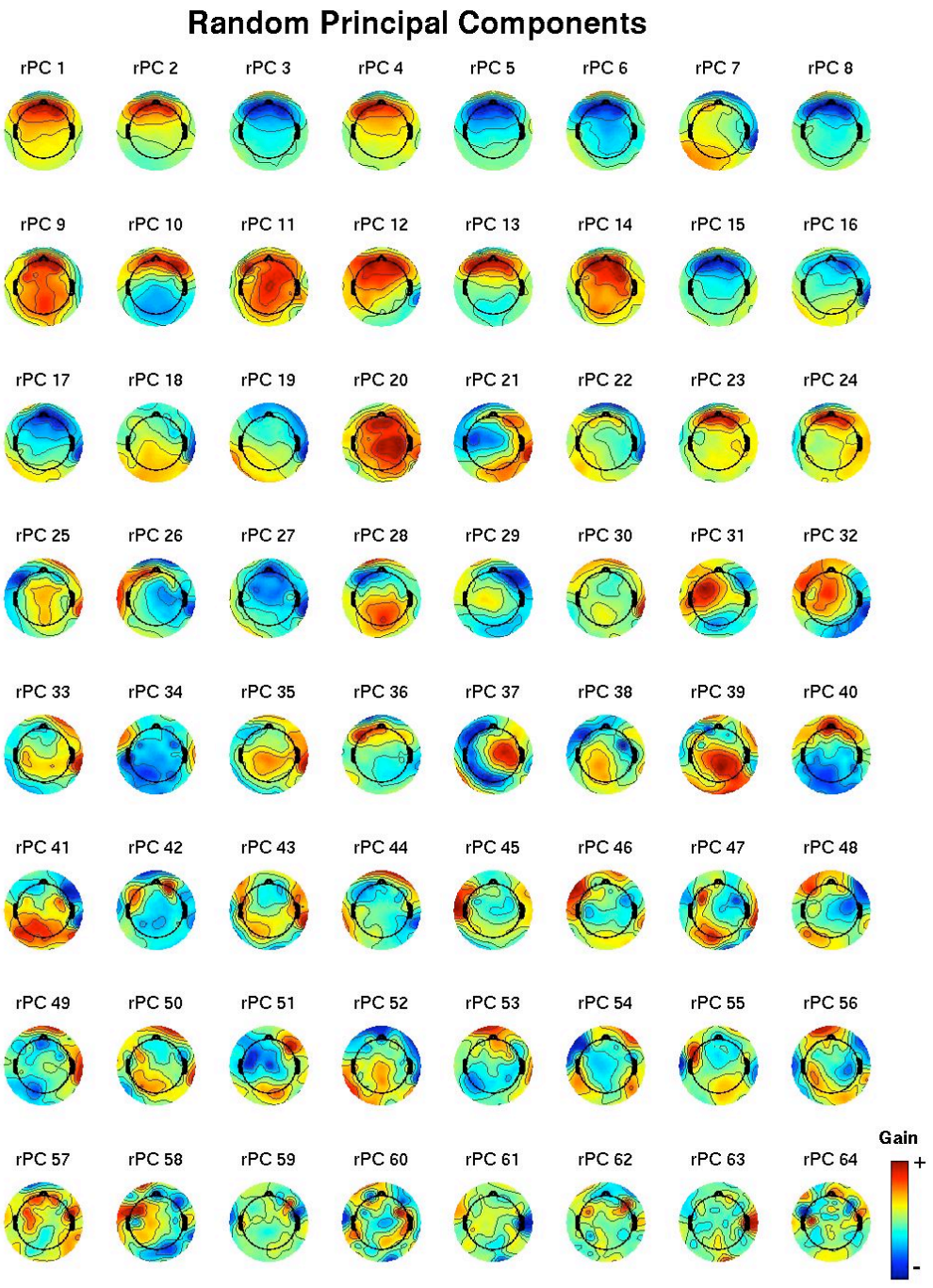

Figure III.19 – Mixing vectors (topographies) of a random rotation (Devroye, 1986, pg. 607) of the principal components displayed in Figure III.18. Like the original PCs, this decomposition unmixes the data from which they were derived into uncorrelated components. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1 and the 64 electrodes are not shown. Random PCs are ranked in decreasing order of scalp-projected variance (summed across all electrodes).

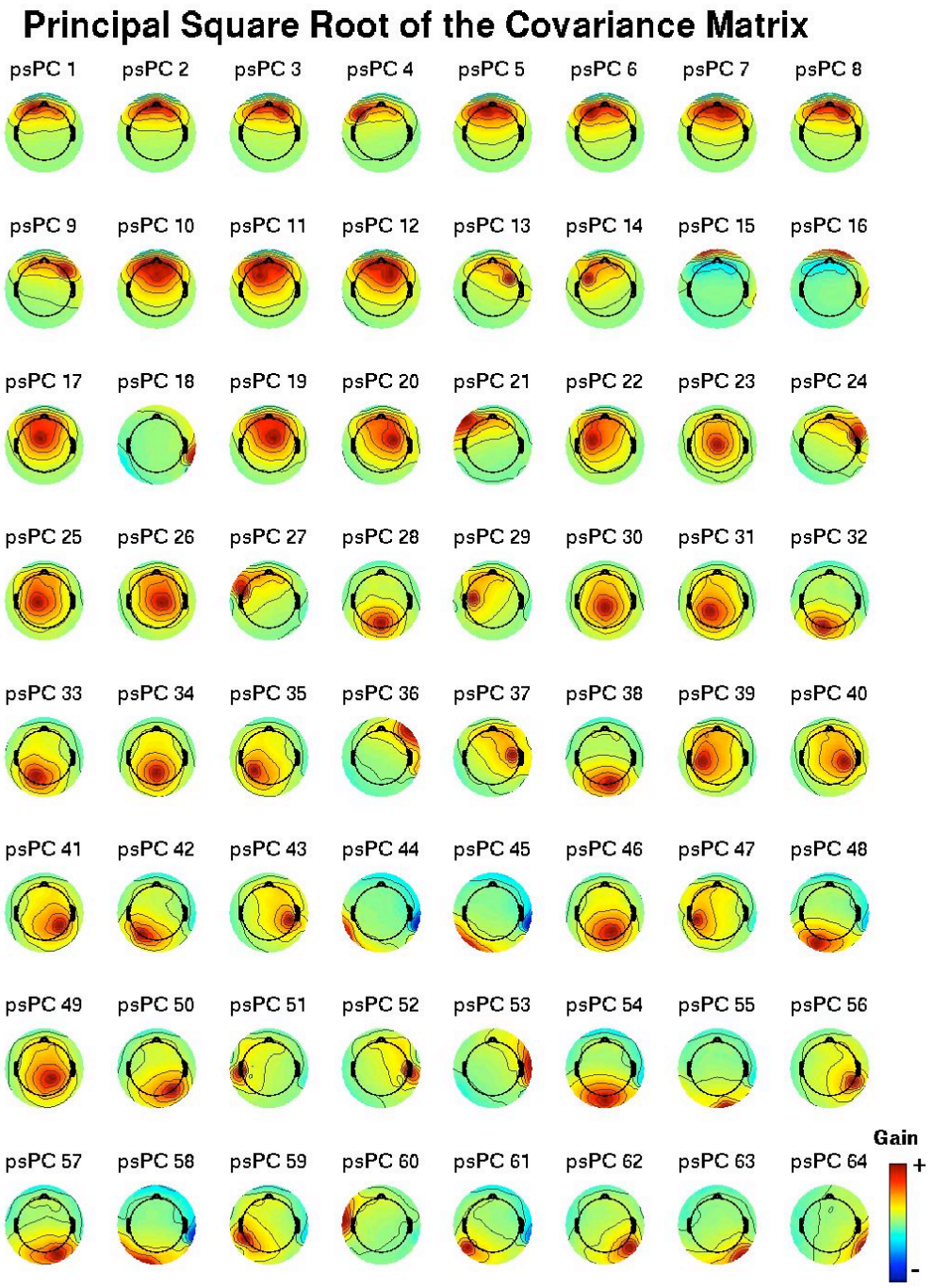

Figure III.20 – Mixing vectors (topographies) of the 64 components of the principal square root of the covariance matrix of the same data used to derive the decompositions in Figures III.17-19. Like PCs, this decomposition unmixes the data from which they were derived into uncorrelated components. For visualization, each gain vector has been normalized so that its absolute maximal weight is 1 and the 64 electrodes are not shown. Components are ranked in decreasing order of scalp-projected variance (summed across all electrodes).

III.I References

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14-30.

Belouchrani, A., Abed Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing, 45*, 434-444.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Destexhe, A., Contreras, D., & Steriade, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *Journal of Neuroscience, 19*(11), 4595-4608.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*(2), 95-111.

Dien, J., & Frishkoff, G. (2005). Principal components analysis of ERP data. In T. C. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (1st ed., pp. 189-207). Cambridge, Mass.: MIT Press.

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research, 17*(3), 637-650.

Fogelson, N., Loukas, C., Brown, J., & Brown, P. (2004). A common N400 EEG component reflecting contextual integration irrespective of symbolic form. *Clinical Neurophysiology, 115*(6), 1349-1358.

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research, 11*(2), 305-323.

Ghahremani, D., Makeig, S., Jung, T.-P., Bell, A. J., & Sejnowski, T. J. (1996). *Independent component analysis of simulated EEG using a three-shell spherical head model* (Technical Report No. 96-01). La Jolla, California, USA: Institute for Neural Computation, University of California San Diego.

Goldstein, A., Spencer, K. M., & Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and novelty P3. *Psychophysiology, 39*(6), 781-790.

Himberg, J., Hyvärinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage, 22*(3), 1214-1222.

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626-634.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. New York: J. Wiley.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *Journal of Clinical Neurophysiology, 19*(2), 113-124.

Joyce, C. A., Gorodnitsky, I. F., & Kutas, M. (2004). Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology, 41*(2), 313-325.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V. J., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2000). Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clinical Neurophysiology, 111*(10), 1745-1758.

Kobayashi, K., Akiyama, T., Nakahori, T., Yoshinaga, H., & Gotman, J. (2002). Systematic source estimation of spikes by a combination of independent component analysis and RAP-MUSIC. I: Principles and simulation study. *Clinical Neurophysiology, 113*(5), 713-724.

Kobayashi, K., James, C. J., Nakahori, T., Akiyama, T., & Gotman, J. (1999). Isolation of epileptiform discharges from unaveraged EEG by independent component analysis. *Clinical Neurophysiology, 110*(10), 1755-1763.

Kobayashi, K., Merlet, I., & Gotman, J. (2001). Separation of spikes from background by independent component analysis with dipole modeling and comparison to intracranial recording. *Clinical Neurophysiology, 112*(3), 405-413.

Kutas, M., & Dale, A. M. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg (Ed.), *Cognitive Neuroscience* (first ed., pp. 197-242). Hove East Sussex, UK: Psychology Press.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation, 11*(2), 417-441.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*(1), 47-62.

Makeig, S., & Delorme, A. (2004). Physiological plausibility and stability of independent component analysis. *Cognitive Neuroscience Society Annual Meeting Program,* San Francisco. 216.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Makeig, S., Jung, T.-P., Ghahremani, D., & Sejnowski, T. J. (2000). Independent component analysis of simulated ERP data. In T. Nakada (Ed.), *Integrated Human Brain Science* (pp. 123-146). New York: Elsevier.

Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience, 19*(7), 2665-2680.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage, 27*(2), 341-356.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews, 30*(6), 808-822.

Roweis, S., & Ghahramani, Z. (1999). A unifying review of linear Gaussian models. *Neural Computation, 11*(2), 305-345.

Salustri, C., & Kronberg, E. (2004). Language-related brain activity revealed by independent component analysis. *Clinical Neurophysiology, 115*(2), 385-395.

Serby, H., Yom-Tov, E., & Inbar, G. F. (2005). An improved P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(1), 89-98.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343-358.

Talsma, D., & Woldorff, M. G. (2005). Methods for the estimation and removal of artifacts and overlap in ERP waveforms. In T. C. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (pp. 115-148). Cambridge, Mass.: MIT Press.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., & Phung, D. B. (2002). Independent components of magnetoencephalography: Single-trial response onset times. *NeuroImage, 17*(4), 1773-1789.

Tang, A. C., Pearlmutter, B. A., Malaszenko, N. A., Phung, D. B., & Reeb, B. C. (2002). Independent components of magnetoencephalography: Localization. *Neural Computation, 14*(8), 1827-1858.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage, 25*(2), 539-553.

Tang, A. C., Sutherland, M. T., & Wang, Y. (2006). Contrasting single-trial ERPs between experimental manipulations: Improving differentiability by blind source separation. *NeuroImage, 29*(1), 335-346.

Van Essen, D. C., Drury, H. A., Joshi, S., & Miller, M. I. (1998). Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 788-795.

Wilcox, R. R. (2002). *Applying Contemporary Statistical Techniques*. London: Academic Press.

Wood, C. C., & McCarthy, G. (1984). Principal components analysis of event-related potentials: Simulation studies demonstrate misallocation of variance across components. *Electroencephalography and Clinical Neurophysiology, 59*, 249-260.

Chapter IV: Electrode Level Analysis of the P3b, N400, and P600

IV.A Introduction

To best compare the P3b, N400, P600 and their ostensible subcomponents two interwoven experiments were carried out to elicit these phenomena within the same experimental participants and within the same experimental testing session. This design minimized potential confounds of between-participant variation (which can be significant, see Sections IV.H.2) and between-session differences in electrode placement and cognitive state. This chapter presents the details of the experiments and an "electrode level" analysis of these ERP phenomena using their conventional definitions.

The results of the analysis indicate that the experiments were successful at eliciting the canonical P3b, N400, and P600. Averaging across participants, all three ERP components were statistically reliable, and exhibited standard temporal morphologies and scalp distributions. However, analysis of individual participant data revealed significant individual variation in the topographies these effects and not all participants reliably exhibited N400 and P600 effects. Such analyses are generally not reported in ERP studies, but they will help to interpret the individual participant

variation in the subsequent independent component analysis of these data (Chapter VI).

In addition to demonstrating the replication of the three ERP components, analyses were performed to look for evidence of subcomponents of these effects and correlations between the features of each effect and the post-sentence late positive component (LPC), which some investigators have suggested is the P3b (Chung et al., 1996; Curran, Tucker, Kutas, & Posner, 1993). There was some evidence that the N400 and P600 are necessarily composed of subcomponents. On average, it was not possible to approximate the two effects with a single source to the degree of accuracy conventionally seen in ERP source modeling. However, the apparent source complexity could be caused by ERP estimation error. What is clear is that no more than a handful of sources are necessary to model any of these three ERP effects in these data.

Although the P3b, N400, and P600 effects and the LPC all exhibited somewhat similar topographies, the amplitudes and peak latencies of the effects were not correlated, save for the amplitude of the N400 and the LPC. In light of the striking similarity of the grand average N400 and LPC topographies, the amplitude correlation and a supplementary single trial analysis of the LPC suggest that the N400 effect was at least partially caused by a delayed onset of the LPC or that the N400 and LPC are part of the same biphasic source activity. The lack of correlations between the P600 and the other phenomena was quite possibly due to estimation error. There is also

evidence that P600 peak latency, like that of the P3b, tended to correlate with the time necessary to evaluate a stimulus, but the tendency was not significant.

In addition to these results that inform the central questions of this dissertation, the analyses reported in this chapter revealed other notable ERP effects. Specifically, the target sequence effect in the oddball tasks (the P3b experiment) consisted of a frontal negativity in addition to the posterior P3b-like positivity. The frontal negativity was not apparent in some simpler oddball tasks (Gonsalvez & Polich, 2002) and may reflect the linguistic or relatively complex nature of the oddball tasks used here. In the sentence acceptability task (the N400 & P600 experiment), ungrammatical pronouns elicited a focal anterior negativity but ungrammatical verbs elicited an N400-like effect, which tended to be left lateralized. These results challenge some existing interpretations of such effects and their differences may have been caused by the fact that participants are faster to detect ungrammatical pronouns. Incongruous words tended to elicit a small enhanced positivity that may reflect expectancy violation and appeared somewhat dissimilar to the P600. Finally, the comparison between nonanomalous noun and verb endings suggests that differences in the comprehension of these two parts of speech (Federmeier, Segal, Lombrozo, & Kutas, 2000) disappear with supportive context.

IV.B Methods: Oddball Tasks (Experiment 1)

IV.B.1 Overview and Logic of Experimental Design: Oddball Tasks

To elicit the P3b two text-based oddball paradigms were used that have successfully elicited the P3b in previous studies. The first of these was an oddball task based on differences in letter case (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Moores et al., 2003) and the other, an oddball task based on the semantic category of a word (Kotchoubey & Lang, 2001). For both tasks, participants were asked to silently count oddballs rather than to make an overt response. This was to make the oddball task ERPs as comparable as possible to those in the sentence acceptability task (Section IV.C) in which overt responses were delayed until 1.4 seconds after the onset of critical stimuli.

These two specific oddball tasks were primarily chosen because of the subsequent subcomponentry analysis. The contrast of the two tasks should allow the identification of subcomponents that reflect the reading of a word, not just the identification of its physical characteristics. Furthermore, the P3b should occur earlier in the letter case task (Bentin et al., 1999), as it presumably takes less time to discriminate uppercase and lowercase words. This difference in P3b latency should facilitate the characterization of subcomponents and the ability of ICA to tease apart P3b subcomponents from one another. Finally, having two different P3b's provides a sense of the robustness of the P3b in these data. If the P3b does not clearly replicate

across oddball tasks within participants, then it is unlikely that its contribution, if any, to the N400 and P600 will be discernable within participants.

IV.B.2 Participants: Oddball Tasks

Nineteen UCSD undergraduate volunteers participated in the experiment for academic credit and/or financial compensation. All were right-handed, native English speakers with no history of reading difficulties or neurological/psychiatric disorders. One participant's data was excluded from analysis as he had had very little sleep the night before. Another participant did not finish the experiment as she complained of a headache early into the recording session and a third participant's data was rejected due to excessively artifact-ridden EEG recordings. This left a total of sixteen participants for the subsequent analyses (10 women, 18 to 22 years of age, mean age=19.7).

IV.B.3 Stimuli: Oddball Tasks

Stimulus materials consisted of common names for 20 different types of animal (e.g. "cat," "zebra") and for 20 different colors (e.g. "blue," "magenta"). Words ranged in length between 3 and 8 letters (mean=5). Word frequency of use ranged between 0 and 365 per million with a median frequency of 13 (Francis & Kucera, 1982).

From this set of 40 unique words a set of 200 words was constructed for each of four oddball tasks: Uppercase Target, Lowercase Target, Animal Target, and Color Target. In the Uppercase Target set, each word occurred four times in all lowercase letters (e.g., “cat”) and once in all uppercase letters (e.g., “CAT”). The Lowercase Target set contained the same words but the proportion of uppercase and lowercase occurrences was reversed. In the Animal Target set, each animal word occurred twice and each color word appeared eight times. Animal and color words occurred in opposite proportions in the Color Target set. In the Animal Target and Color Target sets, each word occurred an equal number of times in all lowercase and all uppercase letters.

IV.B.4 Procedure: Oddball Tasks

Volunteers were tested in a single experimental session conducted in a soundproof, electrically shielded chamber. They were seated in a comfortable chair approximately 1 meter in front of a CRT monitor and performed four visual oddball tasks (Lowercase Target, Uppercase Target, Color Target, and Animal Target) in four blocks that alternated with four blocks of a sentence acceptability task (Experiment 2). Half of the participants began the experiment with an oddball block.

Each oddball block was composed of four sub-blocks, one for each oddball task. The stimuli for sub-blocks were generated for each task by dividing the task's set of 200 stimulus words into 4 lists of 50 words. Words were randomly assigned to each list and randomly ordered with the constraint that a word could not immediately

follow another occurrence of itself, unless the second occurrence of the word was in a different case. Each list made a sub-block. The order of the sub-blocks in the first block was random with the constraint that a Lowercase Target or Uppercase Target block had to follow an Animal Target or Color Target block. The order of the sub-blocks rotated from block to block such that the first sub-block in the first block became the last sub-block in the second block. To illustrate, Figure IV.1 presents an example series of tasks that a participant might perform in her/his experimental session.

For each visual oddball sub-block, participants were instructed to count silently the number of occurrences of the rare "target" class of word. For example, in Lowercase Target sub-blocks they were to count occurrences of lowercase words and in Animal Target sub-blocks they were to count words that referred to animals. Across all the sub-blocks of each oddball task, the target class of word occurred 20% of the time. The percentage of target words in each sub-block randomly varied.

At the beginning of each oddball sub-block, participants were given the target class for that sub-block. Then a fixation cross appeared in the center of the CRT for 700 ms where the participants were instructed to foveate for the duration of the sub-block. After 300 ms, the stimuli for that block began to appear in the center of the CRT for a duration of 100 ms. SOA in the oddball tasks was 1000 ms plus a random delay generated from a uniform distribution ranging between 0 and 500 ms. At the end of the sub-block, participants were asked to verbally report how many targets they had counted.

Participants were allowed to blink whenever necessary during the oddball blocks, but were instructed to remain still and to not move their eyes. Before performing the oddball tasks, participants were given a list of all 40 unique words used in the tasks and asked to circle the animal words, in order to familiarize the participants with the oddball stimuli. After circling the animal words, participants were given a brief practice trial of the Lowercase Target oddball task to familiarize them with the task.

IV.B.5 EEG Recording Parameters: Oddball Tasks

The electroencephalogram (EEG) was recorded from 59 electrodes embedded in an Electro-cap. To obtain relatively even coverage of the scalp, electrode locations were determined by approximating the head with an icosahedron with roughly 4 cm edges. Each of the faces of the icosahedron was divided into 9 equilateral triangles. An electrode was placed at the vertex of each triangle on all the top faces of the icosahedron and on some of the vertices on the side faces.

In addition to the Electro-cap, five additional electrodes were placed on the outer canthus and infraorbital ridge of each eye (EOG electrodes), and on the right mastoid. The sites of all electrodes are plotted in Figure IV.2 and their approximate locations in spherical coordinates are given in Appendix 1 (Section IV.I). All electrodes were tin and referenced to the left mastoid. Electrode impedances were kept below 5 KΩ. EEG was processed through Grass amplifiers set at a bandpass of

0.01-100 Hz. EEG was continuously digitized at 250 Hz and stored on hard disk for later analysis.

IV.C Methods: Sentence Acceptability (Experiment 2)

IV.C.1 Overview and Logic of Experimental Design: Sentence Acceptability

Relatively predictable sentence ending words and semantically or grammatically anomalous endings of the same sentences were used to measure the N400 and P600 respectively. These stimuli were chosen to maximize the magnitude of the N400 and P600 effects[49] (Coulson et al., 1998; Hagoort, 2003; Kutas & Hillyard, 1984; Patel et al., 1998), which should facilitate the ability of independent component analysis to extract subcomponents of these effects. Two types of sentence final words were used for each ERP effect. Nouns and verbs were used to measure the N400 and pronouns and verbs were used to measure the P600. Two types of sentence ending anomalies for each effect were used to keep the occurrence of anomalies from being too predictable. The use of semantically and grammatically anomalous verbs was chosen to elicit comparable N400 and P600 effects.

[49] Using predictable endings should minimize the N400 to nonanomalous endings (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984) and using semantically anomalous endings should produce a larger N400 than simply unlikely endings (Kutas & Hillyard, 1980b). There is evidence that ungrammatical words elicit larger P600s than grammatically unlikely words (Patel, Gibson, Ratner, Besson, & Holcomb, 1998) and that ungrammatical pronouns elicit particularly large P600s (Coulson, King, & Kutas, 1998). There is also evidence that sentence final words tend to elicit larger N400 and P600 effects than sentence internal words (Hagoort, 2003).

Experimental participants were asked to read the sentences and to determine if they contained an anomaly or not. Participant responses were delayed until 1.4 seconds after the onset of the sentence final word. This delayed-response "sentence acceptability" task is a common N400 and P600 paradigm (Kuperberg et al., 2003; Osterhout, 1997). Participant response is delayed to avoid contamination of the N400 and P600 with response related activity.

IV.C.2 Participants: Sentence Acceptability

The same 16 participants who served in Experiment 1 also served in Experiment 2. Four participants were each randomly assigned to one of the four sets of sentence stimuli (AX, BX, AY, BY) described below.

IV.C.3 Stimuli: Sentence Acceptability

Stimulus materials in the Sentence Acceptability task consisted of 78 noun final sentence stems, 78 pronoun final sentence stems, 156 verb final sentence stems and multiple stem completions. Each noun final stem had two possible completions: a nonanomalous ending and a semantically anomalous (i.e., incongruous) ending. Each pronoun final stem had two possible completions: a nonanomalous ending and a grammatically anomalous (i.e., ungrammatical) ending. Each verb final stem had three possible completions: a nonanomalous ending, an incongruous ending, and an ungrammatical ending. All pronoun final stems required an accusative case pronoun

completion. Ungrammatical pronoun completions were the nominative case of the nonanomalous completions. All sentence final verbs were in the present tense. Ungrammatical verb completions violated number agreement between the verb and its subject. Examples of each type of stem and completion are given below:

NOUN FINAL STEM AND COMPLETIONS

(Nonanomalous/Incongruous Completion):

The tweeting in the shed sounded like baby birds/tasks.

PRONOUN FINAL STEM AND COMPLETIONS

(Nonanomalous/Ungrammatical Completion):

Once a month, Carol goes to the theater with me/I.

VERB FINAL STEM AND COMPLETIONS

(Nonanomalous/Incongruous/Ungrammatical Completion):

If Julia misses the bus, she usually just walks/builds/walk.

To generate experimental stimuli from the set of stems and completions, the stems were divided in half and assigned to one of two categories, A and B. The verb final stems were further divided in half and assigned to one of two categories, X and Y. These categories were used to generate 4 unique sets of sentences: AX, AY, BX, BY. In the AX and AY sets of sentences, A stems were completed with anomalous

completions and the B stems completed with nonanomalous completions. In the BX and BY sets the opposite was true. In the AX and BX sets the anomalous X verb ending stems ended with semantic anomalies and the anomalous Y verb ending stems ended with grammatical anomalies. In the AY and BY sets the opposite was true. Thus each set of stimuli contained 156 sentences with nonanomalous completions (39 noun final, 39 pronoun final, 78 verb final) and 156 sentences with anomalous completions (39 noun final semantic anomalies, 39 pronoun final grammatical anomalies, 39 verb final semantic anomalies, 39 verb final grammatical anomalies).

IV.C.3.a Cloze Probability

To quantify the predictability of the sentence endings, cloze probabilities were estimated for the possible completions of the 312 sentence stems by a standard procedure (e.g., Federmeier & Kutas, 1999b). To earn academic credit, native English speaker undergraduate volunteers (none of whom participated in either EEG experiment) were asked to complete each stem with "the first completion that came to mind." Completions could be one or multiple words and should form a sensible, grammatically correct sentence. The stems were presented in random order on a personal computer and each participant was given 50 minutes to type completions for as many of the sentence stems as possible. Between 65 and 74 undergraduates wrote completions for each stem (mean=70). The cloze probability of a stem completion was calculated as the percentage of individuals who used that word as the first word of their completion. A word's cloze probability is commonly used as an index of how

much a reader expects that word in the sentence and is highly correlated with the amplitude of the N400 elicited by the word (DeLong et al., 2005; Kutas & Hillyard, 1984).

Table IV.1 (A) lists the mean cloze probability for each type of stem completion. The cloze probabilities of nonanomalous completions were much higher than their corresponding anomalous completions. 100%, 97%, and 98% of anomalous noun, pronoun, and verb completions (respectively) had cloze probabilities less than .05. Furthermore, 87%, 58%, and 52% of the nonanomalous noun, pronoun, and verb completions (respectively) were the most frequently supplied next word.

To determine if cloze probabilities were balanced across part of speech and list, the cloze probabilities of nonanomalous and incongruous endings were submitted to a three-way ANOVA with variables of Congruity, Part of Speech (noun, verb), and List (AX, AY, BX, BY). Nonanomalous nouns had higher cloze probabilities than nonanomalous verbs (Part of Speech x Congruity interaction, $F(1,764)=17.47$, $p\approx 0$). But there were no effects of List ($p>.82$).

A similar ANOVA was done for nonanomalous and ungrammatical verb and pronoun endings. This found that nonanomalous verbs were more expected than nonanomalous pronouns (Part of Speech x Grammaticality interaction, $F(1,764)=16.21$, $p<.0001$). Again, there were no effects of List ($p>.27$).

A final comparison of interest was across anomaly type. The cloze probability of ungrammatical verb endings was greater than that of incongruous verb endings, $t(310)=-2.19$, $p<.03$, but the difference was negligible (.01 vs. 0 cloze probability). A

two-way ANOVA with variables Anomaly (nonanomalous, incongruous, ungrammatical) and List (AX, AY, BX, BY), found that cloze probabilities of verb endings did not differ across lists (p>.95).

IV.C.3.b Word Length & Frequency

Word length and frequency of use have been found to affect the ERPs to written words, especially in the N400 time window (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; King & Kutas, 1995; Van Petten & Kutas, 1991). Table IV.1 (A) lists the mean lengths for each type of stem completion. While pronouns were clearly shorter than nouns and verbs, the differences in length across nonanomalous, incongruous, and ungrammatical endings were negligible. Three-way ANOVAs, like those done for cloze probability (above), found no effect of stimulus list on the length of endings (p>.36). A two-way ANOVA with variables Anomaly (nonanomalous, incongruous, ungrammatical) and List (AX, AY, BX, BY), found that the lengths of verb endings did not differ across lists (p>.82).

Each completion's frequency of use was estimated using the Francis and Kucera corpus (1982). Median completion frequencies are presented in Table IV.1 (A). Again, three-way ANOVAs were used to detect differences in completion frequency. For the nonanomalous and incongruous endings, completion frequency was not affected by Part of Speech (noun, verb), Congruity, or List (AX, AY, BX, BY) in any way (p>.17). Pronouns were obviously more frequent than verbs and ungrammatical pronouns were clearly more frequent than nonanomalous pronouns. A

two-way ANOVA showed that the frequencies of verb completions were not affected by Anomaly or List ($p>.10$).

IV.C.3.c Lexical Association

Two words are said to be "lexically associated" if perceiving one of the words brings the other to mind (e.g., "cat" and "dog"). Lexical association is typically not controlled for in N400/P600 paradigms but it has been found to decrease N400 magnitude (Van Petten, 1993) and it is likely to confound effects of word predictability as high cloze probability endings tend to be lexically associated with preceding words. To test for this confound, the degree of lexical association between the words in a stem and a completion for that stem was quantified by summing the association strengths between all stem words and completions. Association strengths were taken from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998). 40% of stem completions and 49% of stem words were not normed (i.e., used as stimuli in a discrete word association task). When possible, the uninflected version of an un-normed word (e.g., "break" is the uninflected form of "breaks") was used to approximate the associations of that word. Subsequently, it was possible to estimate the associations of all but three stem completions ("them," "they," and "us") and 37% of the stem words. Given these remaining un-normed words, the degree of lexical association between stems and completions were most likely greater than the current estimates.

Mean sum association strengths are presented in Table IV.1 (A). Noun and verb completions were generally not associated with any stem words. Pronoun completions were slightly more associated with words in their stems. This was largely due to the fact that pronouns are associated with other closed class words (e.g., prepositions and other pronouns). To determine if association strengths were balanced across lists, two three-way ANOVAs and a two-way ANOVA, like those previously used to analyze cloze probabilities, were applied to the sum association strengths and found no significant effects of List (p>.23).

IV.C.3.d Concreteness

The "concreteness" of words has also been shown to affect ERPs to words in sentences (Holcomb et al., 1999; West & Holcomb, 2000). In particular, highly concrete words (e.g., "dog," "chair") often have larger and more frontally distributed N400s than abstract words (e.g., "democracy," "justice"). To check for any unintentional concreteness differences across the various types of stem completions, the concreteness of each completion was estimated according to a standard procedure (e.g., Altarriba, Bauer, & Benvenuto, 1999). Native English speaker undergraduate volunteers (none of whom participated in either EEG experiment) were asked to rate the concreteness of each completion on a scale from 1 to 7 ("1" being most abstract, "7" being most concrete). The completions were presented in random order on a computer and volunteers were allowed to proceed at their own pace. 47 volunteers rated each completion to earn academic credit.

Mean stem completion concreteness ratings are shown in Table IV.1 (A). While nouns tended to be more concrete than verbs, which in turn were more concrete than pronouns, these differences were negligible[50]. To determine if concreteness was balanced across lists, two three-way ANOVAs and a two-way ANOVA, like those previously used to analyze cloze probabilities, were applied to the sum association strengths and found no significant effects of List (p>.30).

IV.C.3.e Repetition

The final potentially confounding lexical factor considered was the number of times stem completions appeared in the entire set of experimental stimuli. The N400 to a written word tends to decrease with repeated presentation (Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). While no noun or verb completion was used more than once to complete a stem, pronoun completions were necessarily used multiple times. Noun, verb, and pronoun stem completions occurred in stems as well, and two nonanomalous noun completions, "bear" and "elephant," appeared in the visual oddball tasks. These repeated words were generally high frequency (median frequency=135.5 occurrences/million).

Table IV.1 (A) presents the mean number of repeated occurrences for each type of completion (e.g., the number of repeated occurrences for a completion that was only used as a completion would be 0). Nonanomalous and anomalous noun and verb

[50] For comparison, studies of word concreteness have found ERP differences around 2-4 microvolts for mean concreteness differences of about 3 (Holcomb, Kounios, Anderson, & West, 1999; West & Holcomb, 2000).

completions were repeated about equally as often. Ungrammatical pronoun completions were repeated more often than nonanomalous pronoun completions, due to the greater frequency of the nominative case of pronouns (e.g., "I", "we") in English. To determine if repetition was balanced across lists, two three-way ANOVAs and a two-way ANOVA, like those previously used to analyze cloze probabilities, were applied to the sum association strengths. A significant List x Congruity interaction ($F(3,764)=4.1$, $p<.007$) was found among the nonanomalous and incongruous nouns and verbs. This was due to the difference in repetitions between nonanomalous and incongruous completions being in the opposite direction for lists AX, AY, and BX than for BY. These differences were small though (maximum difference of 8). This effect of list was smaller when only verb completions were analyzed (List x Anomaly interaction, $F(6,612)=2.1$, $p>.05$). No other effects of list were significant ($p>.11$).

IV.C.3.f Sentence Stem Constraint/Entropy

To quantify the degree to which a stem makes it possible to predict the next word in its completion two measures were derived from the cloze probability estimates. The first measure, "constraint," is simply the maximum cloze probability of all possible next words. A stem with high sentence constraint strongly predicts the next word.

The second measure is the entropy of the probability distribution of possible next words. Entropy is the average information provided by witnessing the value of a

random variable (i.e., $\log_2$(probability of that outcome). Entropy is minimal (0 bits) when the next word is perfectly predictable from the stem. Entropy is maximal ($\log_2$(N) bits, where N is the total number of words in the language) when the next word is maximally unpredictable (i.e., each word in the language is equally likely to complete the stem). Because one must estimate the probability distribution of possible next words from a finite number of participants in a cloze probability norming study, one's estimates of entropy are bounded by $\log_2$(n), where n is the total number of participants who wrote completions for that stem (for a good primer on entropy in the context of studying natural language see Manning & Schütze, 1999).

In the sentence processing ERP literature, constraint is the most commonly used measure of the degree to which a stem predicts subsequent words. However, it reflects the predictability of only one of its possible completions. In contrast, entropy reflects the uncertainty across all possible completions. This is illustrated by two hypothetical cloze probability distributions for two stems plotted in Figure IV.3. Half of the participants completed Stem A with one word ("cat") and the other half completed the stem with another word ("dog"). Half of the participants also completed Stem B with a single word ("cat"), but four other words were each used by 12.5% of the participants to complete the stem. These two stems have identical constraints but clearly the word following Stem A is more predictable than the word following Stem B. This difference in predictability is reflected in the entropy measure. Note that for very high and for very low constraint stems, entropy and constraint will be almost perfectly correlated because the cloze probability of the most

frequently supplied word captures all the variation in the distribution (i.e., the cloze probabilities of the other words will be nearly 0 or equal to that of the most frequently used word).

The constraint and entropy for all stems are plotted in Figure IV.4. While the two measures are highly correlated, the interdependence of the two measures varies with constraint. The constraint and entropy of high constraint stems (.8 to 1) are nearly perfectly correlated. The correlation between the remaining stems is much weaker.

The mean constraint and entropies of each type of sentence stem are presented in Table IV.1 (B). According to both measures, stems ending in nouns had the most predictable completions (independent sample *t*-tests, $p<.001$ for all comparison). Stems ending in pronouns tended to have the least predictable completions than verbs, but this difference was not significant ($|t(232)|<1.23$, $p>.22$ for both measures) Because all stimuli lists used the same sentence stems, constraint and entropy were perfectly balanced across lists.

IV.C.3.g Summary: Sentence Stimuli Properties

The cloze probabilities of nonanomalous completions were much higher than their anomalous counterparts, but were generally balanced on the lexical factors of frequency, length, lexical association, concreteness, and repetition that are known to affect ERPs. The one exception to this were ungrammatical pronoun completions, which were much more frequent and repeated than nonanomalous pronoun

completions. This difference was expected and allowed in order to have a non-verb sentence final grammatical anomaly.

While all anomalous endings appeared equally unexpected, nonanomalous nouns were more predictable than nonanomalous verbs, which were in turn more predictable than nonanomalous pronouns. Likewise, noun ending sentence stems were more constraining than verb ending stems, which were more constraining than pronoun ending stems. Noun and verb completions were balanced on the other examined lexical factors, but pronouns differed from their verb counterparts in most of these. Pronoun completions were shorter, more frequent, more associated with stem words, and more repeated than their verb counterparts. As with the nonanomalous vs. ungrammatical pronoun differences just noted, these differences were expected and allowed.

Finally, the completions of the four different lists of stimuli were effectively balanced on all of the factors examined.

IV.C.4 Procedure & EEG Recording Parameters: Sentence Acceptability

Testing conditions and EEG recording parameters were the same as described for Experiment 1. The stimuli for each of the four sentence acceptability blocks consisted of 78 sentences. These were obtained from the participant's full set of 312 sentences by randomly splitting the full set into four lists. Participants were shown the

sentence stimuli one word at a time in the center of the monitor. Each sentence was preceded by a fixation cross for a duration of 550 ms plus a random delay generated from a uniform distribution ranging between 0 and 950 ms. The fixation cross directed the participant's gaze to the center of the monitor. Words were presented for 100 ms with an SOA of 300 ms. After 1300 ms, the final word in each sentence was followed by the words "acceptable" and "unacceptable" presented simultaneously on the left and right sides of the screen. Upon seeing this cue, participants were to press a button to indicate the acceptability of the sentence. "Acceptable" sentences were defined as grammatically well-formed, nonanomalous sentences. The location of the cue words corresponded to the meaning of the buttons (e.g., if "acceptable" appeared on the left side of the monitor, pressing the left button indicated that the sentence is acceptable) and the meaning of the buttons was rotated halfway through each block. Half of the participants began the experiment using the left button to indicate acceptable sentences. After a button press, the participant would be shown the next sentence in the block. They were told to take as much time as they needed to make a response.

Participants were instructed not to blink during the presentation of a sentence (from the presentation of the fixation cross to the presentation of the response cue). Participants were offered breaks after every 19 sentences to blink and rest their eyes. Before the first sentence acceptability block, participants performed a brief practice trial to familiarize themselves with the task.

IV.D EEG Data Preprocessing: Both Experiments

EEG data was re-referenced off-line to the algebraic sum of the left and right mastoids and divided into 1.4 second epochs extending from 100 ms before to 1300 ms after stimulus onset. 1.4 second epochs were chosen to avoid activity related to making a button press in the sentence acceptability task. Data from the oddball tasks were also divided into 1.4 second epochs so that data preprocessing across the two types of task would be comparable. Relevant stimuli were target and standard words in the oddball tasks and all sentence final and some sentence internal words in the sentence acceptability task. Only a subset of the sentence internal words were chosen to avoid extensive overlap across epochs (i.e., the same time points appearing in multiple epochs). Each epoch was 50 Hz low-pass filtered and the mean of each epoch was removed. After filtering, individual epochs were rejected via a combination of visual inspection and objective tests designed to detect blocking, drift, and extreme values. These tests are included in the EEGLAB toolbox (Delorme & Makeig, 2004). 20% of the epochs were rejected on average leaving an average of 1487 epochs per participant (minimum=1279). On average, 95.4% (minimum=95.0%) of a participant's time points were unique (i.e., appeared in only one epoch) and each participant had 122 (minimum=105) unique time points per number of electrodes squared.

To further clean the data, extended INFOMAX independent component analysis (Lee, Girolami, & Sejnowski, 1999) using the binica function from the EEGLAB toolbox was individually performed on each participant's data. The

extended INFOMAX ICA algorithm was set to estimate the number of subgaussian sources after every training block (every 6000 data points). Besides this, the default ICA training parameters were used. Specifically ICA training stopped when the cosine of the angle between the unmixing matrix change of the current of the previous training step was less than 1e-7 or after 512 training steps (whichever came first). The initial ICA learning rate was .0001. The ICA output unit bias was updated online and the initial state of the ICA unmixing matrix was a sphering matrix[51] (two times the inverse of the principal square root of the data covariance matrix).

Independent components (ICs) corresponding to blinks (e.g., Figure IV.5), eye movements (e.g., Figure IV.6), heart beats (e.g., Figure IV.7), drift (e.g., Figure IV.8), and 60 Hz line noise (e.g., Figure IV.9) were manually identified by their characteristic scalp maps and patterns of activity. To automatically identify ICs corresponding to muscle activity, the power spectra for each IC was computed and converted into t-scores (this was done separately for each participant and frequency). ICs whose mean power in the 30-45 Hz range was greater than the mean power in the .5-25 Hz range by .5 *t*-scores or more were classified as muscle noise (e.g., Figure IV.10). In addition to these automatically identified muscle ICs, five ICs that failed the automatic rejection test were labeled as muscle ICs by hand. ICs with topographies dominated by a single electrode (i.e. the weight to a single electrode accounts for 60% or more of the topography vector's length) were classified as mechanical electrode artifact (e.g., Figure IV.11). On average, 31 ICs (out of 64 total)

[51] A sphering matrix makes each dimension of the data uncorrelated and of equal variance. Thus a sphering matrix gives ICA a "head start" by initially decomposing the data into somewhat independent components.

were removed per participant, which is slightly higher than the number of EEG-artifact ICs found by previous researchers in a comparable study (Debener, Makeig, Delorme, & Engel, 2005)[52]. Table IV.2 displays the number of artifact ICs of each class found per participant and the percentage of data variance accounted for by each class.

Averages of artifact-free ERP trials that preceded accurate participant responses were calculated for each type of "critical word" (standards and targets in the oddball tasks, sentence final words in the sentence acceptability task) after subtraction of the 100 ms prestimulus baseline.

IV.E Results: Oddball Tasks (Experiment 1)

IV.E.1 Behavioral Accuracy: Oddball Tasks

Participants' target counts in the oddball tasks were generally quite accurate. Data from three sub-blocks were thrown out because participant counts were off by 7 or more. In the remaining sub-blocks, participant counts were within 3 of the true number of targets (on average, 10 out of the 50 words in each sub-block were targets). Mean absolute count error per sub-block in both the letter case task and semantic category tasks was approximately 0.3 and did not significantly differ across tasks, $t(15)=.1471$, $p>.88$.

[52] Debener and colleagues applied extended INFOMAX ICA to approximately 64 electrode data from 33 participants. On average, they rejected 41% (26 out of 64) of ICs as EEG artifacts.

IV.E.2 Event-Related Potentials: Oddball Tasks

Grand average ERPs (i.e., the mean of the ERPs of all participants) to standards and targets at all electrodes are presented in Figures IV.11 and IV.13 for the letter case task and semantic category task respectively. Early components to standards and targets in both oddball tasks included, at lateral posterior electrodes, a positivity that peaked around 100 ms (P1), a negativity that peaked around 170 ms (N170 or posterior N1), and a positivity that began around 240 ms (P2). The N170 was noticeably bigger on the left side of the head. Early components at frontal electrodes include a negativity that peaked around 100 ms (N1) and a positivity that peaked around 180 ms (P2) These components are more clearly visible in Figures IV.12 and IV.14, which present the oddball task ERPs at 15 representative electrodes.

Standards

For standards, the early components of the ERP were followed at lateral posterior electrodes by a negativity that peaked around 700 ms. At frontal and central electrodes the early components were followed by a negativity that peaked around 300 ms and a positivity that peaked around 650 ms. This frontal negativity lasted longer in the semantic category task.

Targets

For targets, the early components of the ERP were followed by a large, broad positivity (P3b or P300) at central and posterior electrodes that peaked around 450 ms in the letter case task (Figure IV.12) and approximately 100 ms later in the semantic category task (Figure IV.14). The P3b was slightly larger in amplitude and more peaked in the letter case than in the semantic category task. At medial central and lateral posterior electrodes, the P3b was followed by late negativities that begin at about 750 ms and 800 ms in the letter case and semantic category tasks, respectively. At frontal electrodes, early ERP components in both tasks were followed by a negativity that peaked around 450 ms in the letter case task and 470 ms in the semantic category tasks.

Targets vs. Standards

Figures IV.16-17 show the grand average difference waves (target ERPs minus standard ERPs) for both oddball tasks. In the letter case task, the ERPs to targets began to diverge from those to standards around 200 ms at frontal electrodes where the target ERPs were positive relative to the negative going standard ERPs. At more posterior electrodes, the difference between conditions was even more striking with the ERPs to targets being much more positive (the P3b effect). This effect began around 250 ms and peaked about 460 ms, and was largest at medial posterior electrodes (Figure IV.20: A). On the tail end of the P3b effect was an enhanced negativity for targets at medial frontal electrodes that began around 550 ms, peaked about 660 ms, and continued to the end of the epoch as its focus became more central.

The difference between targets and standards in the semantic category task looks very much like that in the letter case task, but the difference wave features were generally delayed and slightly smaller in amplitude. The frontal positivity began at about 260 ms and peaked around 320 ms. The P3b effect began about the same time as it did in the letter case task but it peaked about 60 ms later and was smaller in amplitude and more diffuse in time. This P3b effect had a slightly more posterior distribution than that of the letter case task (Figure IV.21: B). The post-P3b medial frontal negativity for targets was about the same as in the letter case task, but delayed by 50 ms.

Target Sequence Effects

As mentioned previously (Section II.D) the magnitude of the P3b to targets is diminished if the eliciting target is immediately preceded by another target. To determine if this occurred in the present experiment, the ERPs to targets preceded by targets ("targets->targets") and targets preceded by standards ("standard->targets") were separately computed. Because of the limited number of targets preceded by targets, these ERPs were computed by averaging trials from both oddball tasks. As expected, the P3b to targets preceded by standards was larger (Figures IV.18-19). In addition a frontal negativity was also enhanced. Together, these two effects make the topography of this sequence effect (Figure IV.20: C) appear quite different than the canonical P3b effect (Figure IV.21: A, B).

IV.E.3 ERP Analyses: Oddball Tasks

IV.E.3.a P3b Effect Peak Latency

The peak latency of the P3b effect was measured by first computing the target–standard difference wave for each individual participant and task. These 32 difference waves were low-pass filtered at 7 Hz and the latency of the most positive value between 150 and 900 ms at each electrode was measured. The median peak latency across all electrodes was taken as the P3b effect peak latency for that participant and task.

Mean P3b effect peak latencies were 443 (SD=58) ms and 518 (SD=108) ms in the letter case and semantic category tasks, respectively. This difference was significant, $t(15)=-4.10$, $p<.001$. P3b effect peak latency across tasks was strongly linearly correlated, $r(14)=.78$, $p<.0004$.

IV.E.3.b Mean Amplitude Analysis: Oddball Effects

Based on the latency and durations of the P3b's in the grand average ERPs, mean voltage measures were taken in 250 ms windows for the letter case (325 to 575 ms) and semantic category (425 to 675 ms) oddball tasks. These measures were analyzed using an ANOVA with three repeated measures: Task (letter case, semantic category), Stimulus (target, standard), and Electrode (59 electrodes--data from the right mastoid and the 4 EOG electrodes were not included in this nor all later mean

amplitude ANOVAs[53]). All *p*-values for this and all later ANOVAs are reported after Epsilon correction (Greenhouse-Geiser) for departures from sphericity.

Targets were significantly more positive than standards, $F(1,15)=35.39$, $p<.0001$, and this difference, the P3b effect, was greatest at posterior electrodes (especially medial parietal) causing a Stimulus x Electrode interaction, $F(58,870)=22.09$, $p<.0001$. There was a tendency for the P3b effect in the letter case task to be slightly larger (Task x Stimulus interaction, $F(1,15)=3.80$, $p>.07$). This tendency was greatest at central and parietal electrodes (Figure IV.20: A, B), causing a Task x Stimulus x Electrode three way interaction, $F(58,870)=5.59$, $p<.0007$, and was quite consistent within-participants. Individual participant P3b amplitudes[54] across the two tasks were highly linearly correlated, $r(14)=.87$, $p<.00002$.

While there was no significant main effect of Task, $F(1,15)=.28$, $p>.60$, there was an interaction between Task and Electrode, $F(58,870)=7.10$, $p<.004$, as the potentials at lateral posterior electrodes (especially on the right) were more positive in the letter case task. Finally, there was a main effect of Electrode, $F(58, 870)=19.66$, $p<.0001$, primarily because the potentials at posterior electrodes were much more positive on average than those at anterior electrodes.

To get a sense of how reliable the P3b effect was in individual participants, each participant's data was analyzed with a one-tailed t_{max} multivariate permutation

---

[53] EOG electrodes and the right mastoid are conventionally omitted from ERP mean amplitude analyses.

[54] To determine the amplitude of the P3b effect for each participant, the mean voltage of individual participant target-standard difference waves in a 150 ms window centered around the peak was measured. The mean amplitude of the electrode with the maximal P3b effect was used as the amplitude measure.

test (Hemmelmann et al., 2004). Specifically, the mean amplitude of each participant's single trial EEG responses to targets and standards was computed in the P3b window for each oddball task (i.e., there was one sample for each single trial). The t_{max} test then determined if the trials to targets were significantly more positive ($p<.05$) than those to standards at at least one electrode. This was done by performing *t*-tests at each electrode and comparing the maximum *t* value to the maximum *t* values obtained from 2000 random permutations of the single trials. The one-tailed t_{max} test was chosen because it is sensitive to the direction of a difference between conditions (i.e., more positive or more negative) and because Hemmelmann and colleagues empirically found that the t_{max} test has relatively good power for multivariate data, like EEG, with highly correlated dimensions. According to the test, all 16 participants had a significantly enhanced positivity to targets at at least one electrode.

IV.E.3.c Mean Amplitude Analysis: Sequence Effects

To determine if the effects of target sequence were reliable, mean voltage measures were taken in a 250 ms window (350 to 600 ms) and analyzed using an ANOVA with two repeated measures: Sequence (targets->targets, standards->targets), and Electrode (59 electrodes). Because the sequence effect was negative at some electrodes and positive at others, the main effect of Sequence was not significant ($F(1,15)=0.10$, $p>.75$). However, the sequence effect was indeed significant as revealed by the Sequence x Electrode interaction ($F(58,870)=9.21$, $p<.0001$). Unsurprisingly, the main effect of Electrode was also significant ($F(58,870)=19.58$,

$p<.0001$) as the potentials at posterior electrodes were generally much more positive than those at frontal electrodes.

IV.E.3.d Anatomical Complexity

As explained in Section III.C.1, it is impossible to determine the neural sources of scalp electrical activity with certainty. However, given a particular error metric, it may be possible to determine a lower bound on the number of sources necessary to generate a segment of EEG/ERP. An error metric commonly used for ERP source localization is "percentage variance accounted for," ν (see Section III.D.1):

$$\nu = 1 - \frac{\sum_j \sum_t (x_j(t) - \hat{x}_j(t))^2}{\sum_j \sum_t x_j(t)^2} \quad (1)$$

where $x_j(t)$ is the ERP voltage at time t and electrode j, $\hat{x}_j(t)$ is the model's estimate of the ERP voltage at time t and electrode j. The maximum ν possible is 100% (a perfect fit to the data) and ν decreases towards negative infinity as a model's fit becomes increasingly worse.

When the model sources are restricted to be spatially fixed and m in number, $\hat{\mathrm{x}}(t)$ is given by:

$$\hat{\mathbf{x}}(t) = \sum_{k=1}^{m} \hat{\mathbf{g}}_k \hat{y}_k(t) \quad (2)$$

where $\hat{\mathbf{g}}_k$ is the gain vector of the model's kth source and $\hat{y}_k(t)$ is the voltage of that source at time t. An optimal model for the ν metric (Duda, Hart, & Stork, 2001, pg.

114; Kutas & Dale, 1997) is to define the *m* model gain vectors as the first *m* principal components (PCs) of the sum of products matrix, **L**, whose *ij*th entry is:

$$\mathbf{L}_{ij} = \sum_{t} x_j(t) x_i(t) \tag{3}$$

Thus, for an arbitrary degree of error (e.g., v=95%), the minimal number of PCs of **L** needed to achieve at least that degree of error is the minimal number of sources that could have possibly generated that data. It is important to note that the only anatomical constraint of the hypothetical sources corresponding to the PCs is that they are spatially fixed relative to the electrodes. So while it may be possible for *m* sources to model an ERP component under a certain degree of error, these sources may have to be distributed across cortex (i.e., not dipolar).

Figure IV.22 (Top) shows the mean cumulative percentage variance accounted for by the largest PCs of the P3b windows of the target-standard difference waves for both oddball tasks. The P3b windows were identical to those used for the mean amplitude analysis. In both tasks, on average, just a single PC provided a borderline reasonable fit to the P3b effect, accounting for a bit more than 90% of the variance. Adding a second PC markedly improved the percentage of variance accounted for and provided a quite good fit to both P3bs. Additional PCs slightly improved results such that by the fifth PC a nigh perfect fit was achieved.

There is no clear minimum v that a source model needs to meet in order to be acceptable. Some researchers have been satisfied by models that account for ~90% of the variance (Debener et al., 2005; Del Gratta et al., 2002; Potts, 2004) though much

better fits (greater than 95%) are common (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Tang, Sutherland, & McKinney, 2005). Thus it is unclear if more than one source is necessary to account for these ERP effects. What is evident though, is that the minimal number of sources needed to model these effects is quite low (on the order of a handful).

IV.E.3.e P3b Effect Topographies

The grand average P3bs from both oddball tasks exhibited canonical visual P3b scalp topographies (Figure IV.21: A, B), but the Task x Stimulus x Electrode interaction found in IV.E.3.b suggests that the two sets of topographies significantly differed. This interaction may have been produced by genuine topography differences or it may be due to the tendency of for the P3b effect to be larger in the letter case task (McCarthy & Wood, 1985). Conventionally, to correct for confounding magnitude differences, the potentials across the conditions would be scaled to eliminate magnitude differences and the ANOVA would be repeated. If the interaction remained, then the null hypothesis that the two topographies are the same would be rejected (McCarthy & Wood, 1985). Urbach and Kutas (2002) showed that this method has an inflated Type I error rate and thus will mistakenly indicate that topographies are different more often than the α level of the ANOVA. Other methods for comparing topographies that are invariant to magnitude differences have been proposed (Galan, Biscay, Rodriguez, Perez-Abalo, & Rodriguez, 1997; Greenblatt & Pflieger, 2004), but applications of these tests to simulated data show that they can

also have inflated Type I error rates when condition magnitudes differ (Groppe, unpublished). In light of these methodological limitations, I will refrain from testing hypothesis about the identity of ERP topographies. Instead, I will simply describe topography similarity using the cosine similarity metric, r_{cos}, which is analogous to Pearson's linear correlation coefficient, r (i.e., 1=exactly the same, 0=unrelated, -1=exactly the same but polarity reversed--see Footnote 36).

The grand average P3b topographies were nearly identical (r_{cos}=.98). The similarity of the topographies across tasks was evident even in the individual participant data. Figures IV.23-24 (leftmost two columns) present each participant's P3b topographies. These were derived from each individual's mean target-standard difference waves in the P3b time windows used in the mean amplitude analysis (Section IV.E.3.b). To maximize detail, the color scale used to represent voltage is unique to each map. Maximal red equals the maximum absolute voltage of all 63 non-mastoid electrodes while maximal blue equals the negative of maximal red. The median r_{cos}, of a participant's P3b topographies was .92 (SIQR=.11). However, across participants there was more variation. While most participants have medial central/parietal P3b topographies, some (e.g., participants 1, 9, & 15) have markedly frontal P3bs. To quantify this variation across participants, the r_{cos}, of all 120 possible pairs of topographies (within task, across participants) was computed. Median r_{cos}, was .80 (SIQR=.12) and .80 (SIQR=.16) in the letter case and semantic category tasks respectively.

IV.E.4 Single Trial Analysis: Oddball Tasks

Primarily to determine an effective way to analyze the onset of the late positive component (LPC) in Experiment 2 (Section IV.F.3.g), the onset of the P3b in each single trial EEG response to targets in both oddball tasks was determined using a Woody filter (Woody & Nahvi, 1973), which has previously been successful at determining P3b latency (Kutas, McCarthy, & Donchin, 1977). A template of the P3b for each participant in each oddball task was obtained from a 400 ms window of each participant's ERP to targets in each task. The 400 ms sample included all 64 electrodes and extended 250 ms before and 150 ms after the peak of the P3b (see Section IV.E.3.a). The cosine similarity, r_{cos}, between the P3b template and each single trial at each time point (starting at 64 ms before the onset of the stimulus) was computed. The latency of the P3b was simply defined as the latency of maximal r_{cos}. Thus applying the Woody filter assumes that the P3b has a fixed, 400 ms long wave shape and a fixed scalp distribution and polarity for each time point, but variable latency and amplitude.

Because the P3b may not be present in every trial or may be obscured by noise, a minimal cross-correlation cutoff, λ, was determined for each participant and oddball task by similarly computing the maximal r_{cos}, between the P3b template and the participant's single trial EEG responses to standards in that oddball task. λ was selected by maximizing the percentage of target trials with a r_{cos}, above it while minimizing the percentage of standard trials with a r_{cos}, above it. The average λ across participants and oddball tasks was .58 (min=.22; max=.85). On average, the filter

detected the P3b in only 60% (SD=17%) of the letter case target trials and in 55% (SD=14%) of the semantic category target trials. This modest detection rate indicates a considerable amount of single trial variation in the wave shape or topography of the P3b in both tasks and/or P3b-like responses to standards.

Figures IV.27-28 present ERPimages (Jung et al., 2001) of the single trial EEG response to targets in the letter case and semantic category oddball tasks at a midline parietal electrode. ERPimages visualize single trial EEG data by using color to represent voltage, the horizontal axis to represent time, and the vertical axis to index each single trial. Typically, the single trials are sorted by some variable (in this case, latency of the P3b) and smoothed by replacing each single trial with the average of it and the trials immediately above and below it on the vertical axis (in this case, 25 trials above and 25 trials below).

The mean of each participant's median P3b peak latency in the single trials was 443 (SD=58) ms and 516 (SD=93) ms in the letter case and semantic category oddball tasks respectively. On average, an individual's median P3b latency was faster for the letter case task, $t(15)=5.04$, $p<.001$, and median latency across tasks were correlated, $r(14)=.81$, $p<.001$.

IV.E.5 Results Summary: Oddball Tasks

Participant performance in the oddball tasks was quite accurate and their ERPs to targets and standards displayed the expected P3b effect–-namely, a greater late positivity to targets relative to standards with a broad, posterior distribution, largest at

medial parietal electrodes. P3b effect latency was later and P3b effect magnitude tended to be smaller in the more difficult semantic category oddball task. Within-participants, the two P3b effects were highly similar. Topographies of the effects were nearly identical, and effect amplitude and latency were highly correlated. Given these strong similarities, it is reasonable to interpret these P3b effects as an index of a roughly equivalent physiological process. Furthermore, the anatomical complexity analysis shows that this process may be implemented by only a handful of neural sources, perhaps as few as one.

As expected, P3b amplitude was diminished when the eliciting target stimulus was preceded by another target. In addition, targets preceded by targets were less negative at frontal electrodes than targets preceded by standards. Thus the target sequence effect does not appear to be a simple change in the magnitude of the P3b.

Finally, while all participants exhibited a significantly greater positivity to targets than to standards in each oddball task, there was a fair amount of reliable variation in P3b topography across participants and great deal of variation in the single trial EEG response to targets.

IV.F Results: Sentence Acceptability (Experiment 2)

IV.F.1 Analyses of Behavior: Sentence Acceptability

IV.F.1.a Behavioral Accuracy

Participant performance on the sentence acceptability task was near ceiling with participants correctly classifying 95% of the sentences on average (SD=6%). To test for effects of congruity, participant accuracy for noun, nonanomalous verb, and incongruous verb sentence completions (Figure IV.29: Top) was subjected to an ANOVA with one between-participants variable, List (AX, AY, BX, BY), and two repeated measures: Part of Speech (noun, verb) and Congruity (congruous, incongruous). The analysis found a main effect of List, $F(3,12)=9.67$, $p<.0017$, and a List x Part of Speech x Congruity interaction, $F(3,12)=4.31$, $p<.0281$. Post hoc *t*-tests found that the main effect was due to accuracy on list BX being lower than all other lists ($p<.003$ for each of the three comparisons). The interaction reflects the fact that this effect of list was greater for nonanomalous nouns and incongruous verbs than the other two types of endings. The analysis also found a main effect of Part of Speech, $F(1,12)=15.63$, $p<.0019$, as participants were 3% more accurate when evaluating noun endings. No other effects or interactions were significant ($p>.14$).

To test for effects of grammaticality, participant accuracy for pronoun, nonanomalous verb, and ungrammatical verb sentence completions (Figure IV.29: Middle) were subjected to an identical analysis, but with Grammaticality (grammatical, ungrammatical) replacing the Congruity factor. This ANOVA found a

main effect of Part of Speech, $F(1,12)=17.33$, $p<.0014$, and a Part of Speech x Grammaticality interaction ($F(1,12)=7.76$, $p<.0166$). This indicates that participants were more accurate for pronouns than verbs (by 4%) and that this difference was larger for ungrammatical endings. There was a trend for this Part of Speech effect to differ by list (List x Part of Speech, $F(3,12)=2.57$, $p<.1029$). No other effects or interactions were significant ($p>.49$).

Finally, an ANOVA was performed on participant accuracy for all verb endings to compare effects of congruity and grammaticality (Figure IV.29: Bottom). This ANOVA had one between-participants variable, List (AX, AY, BX, BY), and one repeated measure, Ending Type (nonanomalous, incongruous, ungrammatical). There was a main effect of List, $F(1,12)=4.75$, $p<.0209$, as accuracy on list BX was worse than the other lists (post hoc t-tests, $p<.005$). No other effects or interactions reached significance ($p>.38$).

In summary, while participant accuracy was generally near ceiling, participants were slightly less accurate when evaluating verbs than nouns and pronouns (especially for ungrammatical pronouns and verbs). However, these differences were small and are unlikely to have had much of an effect on the ERPs. In addition, accuracy on one of the stimulus lists, BX, was generally worse than the other lists. Given that these participants were still quite accurate (mean accuracy for list BX nouns and verbs=89%, SD=7%) and that this effect only minimally interacted with the other factors, it should not confound the ERP effects of congruity and grammaticality that are the main focus of this study.

IV.F.1.b Reaction Time

Although participant responses were delayed beyond the amount of time it should have taken participants to identify anomalous sentences (about 1 second--Hagoort, 2003; Kiehl, Laurens, & Liddle, 2002; Kuperberg et al., 2003), an analysis of reaction time (RT) was performed as RT may still have weakly reflected the duration of stimulus evaluation. Moreover, RT may be important for interpreting the latest segment of the ERPs. Because participants were allowed to take as much time as desired to respond, reaction time was highly positively skewed. To accommodate the skew, each participant's median reaction time was taken to represent their reaction times to each of the major types of sentence endings. Only the reaction times to trials included in the EEG analysis were included to make the measures as comparable as possible to the ERPs. The across-participant mean of these median reaction times are presented in Figure IV.30.

As expected, there was a trend for incongruous endings to take longer to respond to than nonanomalous endings. However, an ANOVA with one between-participants variable, List (AX, AY, BX, BY), and two repeated measures: Part of Speech (noun, verb) and Congruity (congruous, incongruous), found no significant differences in the reaction times to congruous and incongruous noun and verb endings (all $p>.12$).

A similar analysis was performed on the pronoun and verb grammatical and ungrammatical endings and found that pronoun endings were responded to

significantly faster than verbs (main effect of Part of Speech, $F(1,12)=6.40$, $p<.03$). There was also a trend for a List x Grammaticality interaction ($F(3,12)=3.12$, $p>.06$). No other effects or interactions were significant ($p>.10$). Thus while participants were faster to evaluate sentences ending in pronouns (by about 100 ms), there was no clear effect of grammaticality or stimulus list/participant. Given that participants were also more accurate at classifying sentences that ended in pronouns, it is clear the pronoun ending stimuli were easier to evaluate.

Finally, an ANOVA was performed on participant response time for all verb endings to compare effects of congruity and grammaticality. This ANOVA had one between-participants variable, List (AX, AY, BX, BY), and one repeated measure, Ending Type (nonanomalous, incongruous, ungrammatical). The trend for incongruous verb endings to take the longest to detect was in the expected direction, but it is far from significant (main effect of Ending Type, $F(2,24)=1.04$, $p>.32$). No other effects or interactions were significant ($p>.11$).

IV.F.2 Event-Related Potentials: Sentence Acceptability

Grand average ERPs to all sentence final words are presented in Figures IV.31 to IV.38. The early components of these ERPs included, at lateral posterior electrodes, a positivity (P1) that peaked about 110 ms and a negativity (N170 or posterior N1) that peaked about 180 ms and was left-lateralized. At central and frontal electrodes, there was a broadly distributed negativity (N1) that peaked around 110 ms and a positivity (P2) that peaked at about 240 ms (e.g., Figure IV.32).

Congruous/Grammatical Endings

ERPs to nonanomalous sentence final nouns, pronouns, and verbs were largely identical (Figures IV.31-32). After the early components, a brief negative deflection after the P2 gave rise to a broadly distributed late positive component (LPC) that was largest at medial parietal electrodes and peaked about 600 ms.

Semantic Congruity

ERPs to incongruous and comparable nonanomalous endings diverged around 250 ms as ERPs to incongruous words went negative (Figures IV.32-35). This broadly distributed negativity (N400) peaked about 400 ms, was greatest at central and posterior lateral electrodes, and was similar for nouns and verbs (Figures IV.38-39). This greater negativity for incongruous words was the N400 effect. It was broadly distributed across the head, but largest medially and posteriorly, with a tendency to be right lateralized (Figure IV.44: A, B).

Like their nonanomalous controls, incongruous endings elicited a pronounced LPC. However, the post-incongruity LPC tended to be larger and peak later. This difference was more prominent to incongruous nouns (Figures IV.39-40), largest posteriorly, and broadly distributed to nouns and spottily distributed to verbs (Figure IV.45: E & F).

Grammaticality

ERPs to nonanomalous and ungrammatical verbs began to diverge around 250 ms when the ERPs to ungrammatical verbs became relatively more negative (Figures IV.35-36). This negative going deflection peaked around 400 ms and was similarly distributed to the N400 effect of semantic congruity (Figure IV.45: D). The ERPs to nonanomalous and ungrammatical pronouns diverged around 300 ms as the ERPs to ungrammatical pronouns went slightly negative at frontal electrodes (Figures IV.37-38). This anterior negativity (AN) tended to be slightly left lateralized (Figure IV.45: C) and may correspond to what is generally referred to as the left anterior negativity (LAN).

The most pronounced difference between ungrammatical and comparable nonanomalous endings was that the LPC, elicited by both types of endings was larger for ungrammatical words. This difference, the P600 effect, was medial-posteriorly distributed (Figures IV.45: G, H) and began and peaked earlier for pronouns (Figures IV.41-42)

Congruity vs. Grammaticality

To better contrast the effects of congruity and grammaticality, these effects on the ERPs to verbs are plotted together in Figures IV.35-36. Both types of anomalies resulted in greater negativities in the N400 time window, followed by enhanced positivities in the P600 window. The negativity to ungrammatical verbs was smaller than that to incongruous verbs and this difference was greatest at right frontal electrodes. This caused a slightly different distribution of the two effects (Figure

IV.45: D, B), with the congruity effect being left lateralized and the grammaticality effect being right lateralized.

In contrast, grammaticality had a larger effect in the late window. This difference was greatest at medial parietal electrodes where the grammaticality effect was strongest, creating distinct topographies for the effects (Figure IV.45: F, H). While the enhanced positivity to ungrammatical verbs had a single focus over medial parietal electrodes, that to incongruous verbs was spotty with two posterior, bilateral foci.

Late Slow Negativity

The ERPs to all sentence final words ended in slow negativities, greatest at left frontal electrodes. These began around 900 ms and peaked shortly after the cue to respond at 1400 ms (not shown). These negativities grew most quickly for nonanomalous endings and ungrammatical pronouns. They rose more slowly for ungrammatical verbs and most slowly for incongruous endings. As the onset of these negativities roughly follows participant reaction time, they may have been readiness potentials preceding participant responses (for a brief review see Kutas & Van Petten, 1994).

IV.F.3 ERP Analyses: Sentence Acceptability

IV.F.3.a N400 and P600 Effect Peak Latency

The peak latency of the N400 effect was measured by first computing the incongruous–nonanomalous difference wave for each individual participant and part of speech (noun and verb). These 32 difference waves were low-pass filtered at 7 Hz and the latency of the most negative value between 250 and 700 ms at each electrode was measured. The median peak latency across all electrodes was taken as the N400 effect peak latency. P600 effect peak latency was similarly measured, identifying the most positive peak between 400 and 1100 ms in the low-pass filtered ungrammatical-nonanomalous difference waves for each participant and part of speech (pronoun and verb).

Mean N400 effect peak latency was 440 (SD=67) ms and 441 (SD=73) ms for nouns and verbs, respectively. Mean P600 effect latency for pronouns and verbs was 709 (SD=158) ms and 810 (SD=135) ms, respectively. N400 effect latency did not differ across parts of speech, $t(15)=.08$, $p>.93$. The tendency for the pronoun P600 effect to peak earlier than that to verbs was not significant, $t(15)=1.88$, $p>.08$.

IV.F.3.b Mean Amplitude Analysis: Congruity Effects

Mean voltage measures were taken between 275 and 550 ms (N400 window) and 650 and 1100 ms (P600 window) for sentence final nonanomalous and incongruous words. Each window was separately analyzed with an ANOVA with one

between-participants variable, List (AX, AY, BX, BY), and three repeated measures: Congruity (nonanomalous, incongruous), Part of Speech (noun, verb), and Electrode (59 electrodes).

In the N400 window, incongruous endings were significantly more negative than comparable nonanomalous endings, $F(1,12)=62.6$, $p<.0001$. This difference was greatest at medial central and posterior electrodes causing a Congruity x Electrode interaction, $F(58,696)=22.89$, $p<.0001$. Posterior electrodes were generally more positive than others (main effect of Electrode, $F(58,696)=7.18$, $p<.0007$). No other effects or interactions reached significance ($p>.27$).

To get a sense for how reliable this effect was in individual participants, the mean amplitudes (in the N400 window) of a participant's single trial EEG responses to all incongruous and comparable nonanomalous endings were analyzed with a one-tailed t_{max} multivariate permutation test (see Section IV.E.3.b). The analysis included both noun and verb endings to maximize test power and was performed using 2000 permutations per test. According to the t_{max} test, 14 of the 16 participants have at least one electrode that is significantly ($p<.05$) more negative to incongruous words than nonanomalous words in the N400 window.

In the P600 window, incongruous endings were significantly more positive than nonanomalous endings, $F(1,12)=10.7$, $p<.007$. As in the earlier window, there were no significant differences between nouns and verbs, but there was a trend for verbs to be more positive (main effect of Part of Speech, $F(1,12)=2.98$, $p>.10$) and for the congruity effect to be larger for nouns (Part of Speech x Congruity, $F(1,12)=4.19$,

p>.06). Finally, posterior medial electrodes were generally more positive than others (main effect of Electrode, F(58,696)=11.74, p<.0001). All other main effects and interactions were not significant (p>.23).

IV.F.3.c Mean Amplitude Analysis: Grammaticality Effects

Because grammaticality effects tended to qualitatively differ between pronouns and verbs within the N400 window, each part of speech was analyzed separately in this latency range. The mean voltages between 275 and 550 ms following sentence final nonanomalous and ungrammatical verbs were subjected to an ANOVA with one between-participants variable, List (AX, AY, BX, BY), and two repeated measures: Grammaticality (nonanomalous, ungrammatical) and Electrode (59 electrodes).

Ungrammatical verbs were more negative than nonanomalous verbs (main effect of Grammaticality, F(1,12)=49.66, p<.0001) and medial posterior electrodes tended to be more positive than others (main effect of Electrode, F(58,696)=3.20, p<.037). Mean voltages differed somewhat across List and therefore across participants (main effect of List, F(3,12)=3.6, p<.046). None of these factors interacted (p>.15).

An identical analysis was done for mean voltages between 350 and 550 ms following sentence final nonanomalous and ungrammatical pronouns. While there was no main effect of grammaticality (F(1,12)=.01, p>.92), the grammaticality effect did differ across electrodes (Grammaticality x Electrode, F(58, 696)=5.59, p<.0034). Because of this interaction the ANOVA was repeated, post hoc, using just the 14 most

frontal electrodes. Now the main effect of Grammaticality was significant ($F(1,12)=7.75$, $p<.0166$), there was a tendency for this effect to differ across list (List x Grammaticality, $F(3,12)=2.81$, $p>.08$) and the most frontal electrodes were more negative in general (main effect of Electrode, $F(13,156)=12.15$, $p<.0001$). All other effects and interactions were not significant ($p>.18$).

Separate P600 effect windows were used for pronouns and verbs because of the tendency for the P600 effect to peak later for verbs, and because the verb P600 effect was preceded by a broad N400-like effect. Mean voltage measures were taken between 550 and 950 ms for pronouns and 650 and 1100 ms for verbs and subjected to an ANOVA with one between-participants variable, List (AX, AY, BX, BY), and three repeated measures: Grammaticality (congruous, ungrammatical), Part of Speech (pronoun, verb), and Electrode (59 electrodes) [55].

Ungrammatical endings were more positive than comparable nonanomalous endings (main effect of Grammaticality, $F(1,12)=19.47$, $p<.0009$) and this effect was greatest at posterior (especially medial posterior) electrodes (Grammaticality x Electrode interaction, $F(58,696)=9.04$, $p<.0014$). Pronouns were generally more positive than verbs (main effect of Part of Speech, $F(1,12)=17.55$, $p<.0014$) and there was a trend for pronouns to be more positive over frontal electrodes (Part of Speech x Electrode, $F(58,696)=2.19$, $p>.13$). Finally, medial posterior electrodes were more

[55] Because of the variability in individual participant P600 effect peaks, this analysis was also repeated with individually tailored P600 windows for each participant and part of speech (48 ms before and after the P600 effect peak). This analysis qualitatively replicated the results of the more conventional grand window analysis, save that the main effect of Part of Speech failed to reach significance.

positive than others (main effect of Electrode, $F(58,696)=16.79$, $p<.0001$). All other effects and interactions were not significant ($p>.53$).

As with the N400 effect above, a one-tailed t_{max} multivariate permutation test was used to assess how reliable the P600 effect was in individual participants. The mean amplitudes (in the P600 window) of a participant's single trial EEG responses to all ungrammatical and comparable nonanomalous endings were computed using the pronoun and verb P600 windows used for the ANOVA. The analysis included both pronoun and verb endings to maximize test power and was performed using 2000 permutations per test. According to the t_{max} test, 10 of the 16 participants had at least one electrode that is significantly ($p<.05$) more positive to ungrammatical words than nonanomalous words in the P600 window.

IV.F.3.d Mean Amplitude Analysis: Congruity vs. Grammaticality

To compare the effects of congruity and grammaticality, mean voltage measures were taken between 275 and 550 ms (N400 window) and 650 and 1100 ms (P600 window) for sentence final incongruous, and ungrammatical verbs. Each window was separately analyzed with an ANOVA with one between-participants variable, List (AX, AY, BX, BY), and two repeated measures: Anomaly Type (incongruous and ungrammatical) and Electrode (59 electrodes).

In the N400 window, there was a main effect of Anomaly Type ($F(1,12)=12.2$, $p<.005$) and an Ending Type x Electrode interaction ($F(58,696)=3.47$, $p<.03$). This indicates that although incongruous and ungrammatical verbs were both more negative

than nonanomalous verbs, the effect differed in size and/or distribution. Because the topographies of the two effects were nearly identical (see Section IV.F.3.f), the interaction was probably due to the fact that the grammaticality N400 effect was about half as large as the congruity N400 effect. All other main effects and interactions failed to reach significance ($p>.18$).

In the P600 window, there was again a main effect of Ending Type ($F(1,12)=6.53$, $p<.03$) and a significant effect of Electrode ($F(58,696)=15.31$, $p<.0001$). All other main effects and interactions failed to reach significance ($p>.08$). Thus while the incongruous and ungrammatical verbs were both more positive than nonanomalous verbs (see previous two sections), this effect was greater for ungrammatical verbs.

IV.F.3.e Anatomical Complexity

Principal components analysis (PCA) was performed on the N400 and P600 effects to determine a lower bound on the number of possible neural sources (see Section IV.E.3.d). Specifically, PCA was applied to the N400 window of each participant's noun and verb incongruous-nonanomalous difference waves and the P600 window of each participant's pronoun and verb ungrammatical-nonanomalous difference waves. N400 and P600 windows were identical to those used in the mean amplitude analysis.

Figure IV.22 (Bottom) shows the mean cumulative percentage variance accounted for by the largest PCs of the four ERP effects. In contrast to the P3b effects

(Figure IV.22: Top), it appears less likely that these ERP effects can be adequately modeled with a single neural source. The largest PC accounted for approximately 86% of the variance of the N400 effects and only about 77% of the variance of the P600 effects. With the addition of a second PC, the percentage of variance accounted for falls within the range of acceptable source models (see Section IV.E.3.d). By the sixth PC, a nearly perfect fit could be achieved for all effects.

The mediocre fit of the first PC may reflect the presence of multiple generators of the N400 and P600. However, the poor fit could also be caused by ERP estimation error. Without a measure of the fraction of data variance due to estimation error it is impossible to say. What is clear is that the data could have been generated by only a few sources.

IV.F.3.f N400 and P600 Effect Topographies

N400 and P600 effects matched canonical topographies. The grand average noun and verb N400 scalp maps (Figure IV.45: A, B) were approximately identical (r_{cos}=.995), both being broadly distributed across the scalp, slightly bigger on the right, and largest at medial parietal electrodes. Pronoun and verb P600 grand average topographies (Figure IV.45: G, H) were slightly less similar (r_{cos}=.96), but both had broad, medial posterior foci. The slight frontal negativity in pronoun P600 topography stems from the temporal overlap between the pronoun P600 and AN effects.

The similarity of these effects across parts of speech was much less for individual participants (Figures IV.23-24). Median within-participant topography

similarity was .78 (SIQR=.30) for noun and verb N400 effects and .61 (SIQR=.23) for pronoun and verb P600 effects. Between-participant topography similarity was about the same for the N400 effects, but was much lower (median r_{cos}= ~.45) for the P600 effects (Table IV.4: C). It is difficult to know how much the greater variability in the individual participant topographies was caused by estimation error or true individual variation (see Section IV.G.3).

The topography of the grand average effect of grammaticality on verb ERPs in the N400 window (Figure IV.45: D) was nearly identical (r_{cos}=.98) to the grand average congruity N400 effects (Figure IV.45: A, B). The grand average topographies of the P600-like effects of congruity on nouns and verbs (Figure IV.45: E, F) were rather similar (r_{cos}=.88) despite the spottiness of the verb topography. The noun effect (Figure IV.45: E) was similarly distributed to canonical pronoun and verb P600 effects (r_{cos}=.96 and .88, respectively). The verb effect (Figure IV.45: F) was much less similar to the canonical pronoun and verb P600 effects (r_{cos}=.81 and .73, respectively).

IV.F.3.g Effects of Cloze Probability

To examine the effects of cloze probability on the N400 and LPC, only the data from nonanomalous nouns and verbs were analyzed. Anomalous endings were excluded because of the effects of anomaly (see Sections IV.F.b-d, h). Pronouns were excluded because such closed-class words are differently affected by preceding context than open-class words (Van Petten & Kutas, 1991) and their frequencies and

cloze probabilities markedly differed from that of the nouns and verbs in this experiment.

High cloze and low cloze ERPs for nonanomalous nouns and verbs were computed for each participant by separately averaging the trials from the 50% of stimuli with the highest and lowest cloze probabilities (i.e., median splits). Table IV.6 presents summary statistics for these two classes of stimuli. While the high cloze probability endings were roughly twice as likely as the low cloze endings, they were rather balanced across other features. Frequency, concreteness, and length sometimes differed between high and low cloze endings, nouns and verbs, or lists, but these differences were small and were unlikely to affect the ERPs. Specifically, previous research has found that mean frequency differences as large as 357 occurrences per million had no effect on the N400 (or by eye, the LPC) by the fifth word of sentences with high cloze probability endings (Van Petten & Kutas, 1991). For comparison, the largest difference in mean frequency of any of the stimulus groupings presented in Table IV.5 was 209 occurrences per million (between low cloze verbs in List A and low cloze verbs in List B). The same study also found that by the fifth word in each sentence differences in length as large as 1.7 characters had no effect on the N400 (or by eye, the LPC). Regarding concreteness, other researchers have found that the N400 and LPC to relatively predictable final words of well-formed sentences were not affected by differences in concreteness as large as 3.7 (Holcomb et al., 1999).

The mean amplitudes of these ERPs in the N400 window (275-550 ms) and the LPC window (400-900 ms) were separately subjected to ANOVAs with one between-

participants variable, List (AX, AY, BX, BY), and three repeated measures: Part of Speech (noun and verb), Cloze Probability (high and low), and Electrode (59 electrodes). The N400 window was identical to that used to examine effects of congruity because there is no clear N400 in the ERPs to nonanomalous sentence final words (Figures IV.31-32). The LPC window was chosen to include most of the component, even though it overlapped somewhat with the N400 window.

In the N400 window, there was a main effect of Electrode ($F(58, 696)=18.26$, $p<.0001$) and a barely significant Electrode x Cloze Probability interaction ($F(58, 696)=2.85$, $p<.05$). The results were similar in the LPC window where again there was a main of effect of Electrode ($F(58, 696)=13.94$, $p<.0001$) and a significant Electrode x Cloze Probability interaction ($F(58, 696)=3.20$, $p<.04$). Again, no other effects or interactions in either window were significant ($p>.08$). The main effects of Electrode were due to the ERPs at medial posterior electrodes being more positive than at other electrodes. The significant interactions were caused by the fact that the ERPs to low cloze probability words were more positive at frontal (especially left) electrodes (Figures IV.47-48). In the N400 window these words were also slightly more negative at posterior electrodes.

In light of the interactions, these ANOVAs were repeated, post-hoc, on the mean amplitude measures from the 34 post-central electrodes only. In the N400 window, the ERPs at medial electrodes were more positive than others (main effect of Electrode, $F(33, 396)17.34$, $p<.0001$) and Cloze Probability interacted with List ($F(3,12)=3.53$, $p<.05$). No other effects or interactions were significant ($p>.22$). In

the LPC window, ERPs at medial electrodes were again more positive than others (main effect of Electrode, $F(33, 396)=11.42$, $p<.0004$), but no other effects or interactions were significant ($p>.10$). Thus low cloze probability endings tend to have a larger N400 but this difference varies across stimulus list (and therefore participants). The mean amplitude of the LPC was not affected by cloze probability.

As an alternative way to visualize possible effects of cloze probability, ERPimages were constructed for select electrodes (see Section IV.E.4). Figure IV.49 (Top) shows an ERPimage of all the participants' data at electrode MiPa, where the grand average N400 effect was largest. The bottom part of the figure plots the ERPs to the top and bottom 50% of the trials. Though there was considerable trial-to-trial variation, there appears to be a tendency for the LPC to low cloze words to begin and end later. This effect was most evident when comparing the extremely high cloze and relatively low cloze words and suggests that the small effect of cloze probability in the N400 window was at least partially due to an earlier LPC for more predictable words.

To quantify this trend a Woody filter like that previously used to determine P3b latency (Section IV.E.4) was used to estimate the latency of the LPC in these single trials. Specifically, a template of the LPC for each participant was obtained from a 400 ms window of each participant's ERPs to nonanomalous noun and verb endings. The 400 ms sample included all 64 electrodes and extended 250 ms before and 150 ms after the peak of the LPC (see Section IV.G.1 for details as to how the LPC peak latency was determined). The cosine similarity, r_{cos}, between the LPC template and each single trial at each time point (starting at 64 ms before the onset of

the stimulus) was computed. The latency of the LPC was simply defined as the latency of maximal r_{cos}. Thus applying the Woody filter assumed that the LPC had a fixed, 400 ms long wave shape and a fixed scalp distribution and polarity for each time point, but variable latency and amplitude. Rank correlation (Kendall's tau) was used to determine if LPC latency was a monotonically decreasing function of the eliciting word's cloze probability for each individual participant. In general, the correlations were small and negative (median tau: -.10, min/max tau: -.23/.09), with 14 of the 16 participants showing tendencies for a negative correlation. The LPC latency of five participants showed a significantly negative correlation (one-tailed test, tau<-.15, p<.05) with cloze probability, which was significant overall (binomial(16,.05)=5, p<.001).

In addition to the effect of cloze probability on the LPC, there was a larger trend for an effect of cloze probability at frontal sites. Figure IV.50 shows an ERPimage and ERPs from a left frontal electrode where this effect was largest. There appeared to be no clear monotonic relationship between cloze probability and the increased positivity for less predictable words in the N400 and LPC windows. The onset of the late negativity did appear to negatively covary with cloze probability and possibly reflected response selection (i.e., a readiness potential).

IV.F.3.h Effects of Entropy/Constraint

While the previous analysis explored the effects of expectation violation on the N400 and LPC to acceptable sentence endings, this section examines how expectation

violation modulated the N400 and P600 to anomalous endings. The degree to which the final word of each sentence was expected was measured by the entropy of the probability distribution of possible completions. As discussed in Section IV.C.2.f, entropy is conceptually similar to the commonly used measure of constraint, and for these stimuli, the measures were highly correlated. Consequently, splitting these data by entropy was approximately equivalent to splitting the data by constraint.

High entropy and low entropy ERPs for incongruous and ungrammatical verbs were computed for each participant by separately averaging the 50% of trials with the highest and lowest entropies (i.e., median splits). Analyzing the ERPs to only verb stimuli provided a more controlled comparison of the N400 and P600.

Table IV.7 presents summary statistics for the sentence stems and anomalous sentence final verbs in the high and low entropy groupings. The possible endings of high entropy stems were roughly twice as unpredictable as those in the low entropy bins, but they were still relatively predictable (mean constraint of 45%). The anomalous ending features were generally balanced across the high and low entropy splits and anomaly type. Although endings to high entropy stems were less frequent and more concrete than low entropy stem endings, the differences were quiet small and were unlikely to have a significant effect on the ERPs (see previous section). Ungrammatical endings had a higher cloze probability than incongruous endings, but again the difference was minute.

Figures IV.51-52 show the ERPs to incongruous verb endings of high and low entropy stems. The most noticeable effect of entropy was an enhanced frontal, right

lateralized negativity for endings of low entropy stems. This effect occurs during and then again after the N400 window. There was a smaller, opposite effect of entropy on the N400 at occipital electrodes and the LPC appears to be slightly enhanced by low entropy stems. Figures IV.53-54 show the analogous ERPs to ungrammatical endings. As with the incongruous endings, the ERPs following low entropy stems were more positive at posterior electrodes, especially over the right hemisphere, in both the N400 and P600 window. Low entropy stems also tended to produce a larger P2 and a small late enhanced negativity at left frontal electrodes.

The mean amplitudes of these ERPs in the N400 window (275-550 ms) and the P600 window (650-1100 ms) were separately subjected to ANOVAs with one between-participants variable, List (AX, AY, BX, BY), and three repeated measures: Anomaly (incongruous and ungrammatical), Entropy (high and low), and Electrode (59 electrodes). In the N400 window, there was a main effect of Anomaly ($F(1,12)=11.88$, $p<.005$) and an Anomaly x Electrode interaction ($F(58,696)=3.56$, $p<.05$) as the ERPs to incongruous endings were more negative than ungrammatical endings and this difference was smallest at prefrontal and left lateral electrodes. The only effect of Entropy was an Electrode x Entropy interaction ($F(58,696)=3.46$, $p<.03$) as the ERPs to the anomalous endings of the low entropy stems were generally more negative frontally and more positive posteriorly than those of high entropy stems. All other effects and interactions failed to reach significance ($p>.19$).

In the P600 window, there was a main effect of Anomaly ($F(1,12)=6.11$, $p<.03$) and a trend for an Anomaly x Electrode interaction ($F(58,696)=2.81$, $p>.08$) as

ERPs to ungrammatical endings were generally more positive than those to incongruous endings and this difference tended to be greatest at central and posterior electrodes. Again, the only effect of Entropy was an Electrode x Entropy interaction ($F(58,696)=4.14$, $p<.006$) due to the ERPs to anomalous endings of the low entropy stems being more negative frontally and positive posteriorly than those of high entropy stems. Finally, ERPs at posterior electrodes tended to be more positive than at anterior electrodes, (main effect of Electrode, $F(58,696)=15.17$, $p<.0001$). No other effects or interactions were significant ($p>.15$).

To get a richer look at these effects of entropy, the single trials to anomalous verbs were visualized as ERPimages (see Section IV.E.4). Figures IV.54-55 show ERPimages for incongruous and ungrammatical verbs at the right dorsal posterior parietal electrode, where the effect of entropy on ungrammatical verbs was largest. For the incongruous verbs, the N400 was visible and had a tendency to begin later as entropy decreased (i.e., as the next word became more predictable). The N400 was not visible in the ERPimage to ungrammatical verbs. The effect of entropy in the N400 window of the ERPs to ungrammatical words appeared to be due to alpha blocking (i.e., the reduction of 12 Hz EEG activity) following high entropy stems. The LPC to incongruous verbs had a tendency to increase with decreasing entropy, but it was not a monotonic relationship as the smallest LPC occurred around mid-range entropies. The opposite appeared to be true for ungrammatical verbs; for these the LPC was slightly larger and longer lasting around mid-range entropies. The LPC to

both incongruous and ungrammatical verbs had a tendency to begin later for high entropy stem completions.

Figures IV.57-58 show ERPimages for incongruous and ungrammatical verbs at the right dorsal prefrontal electrode, where the effect of entropy on incongruous verbs was largest. The greater negativity for incongruous verbs in the N400 window following low entropy stems appeared to be due to a longer lasting N400. However, the clearest effect of entropy was on the onset of the late negativity. Neither effect was visible in the equivalent ERPimage for ungrammatical verbs. Here the most striking effect of entropy was the lack of a P2 following high entropy stems. Such an effect was not seen in response to nonanomalous words (Figure IV.48).

To summarize, both the N400 and LPC were slightly enhanced by expectation violation, but the topographies of these effects differed from each component's canonical distribution. The N400 enhancement was at right frontal electrodes and that of the LPC at right posterior electrodes. Single trial visualization suggested that these effects were not a monotonic function of entropy. However, these effects were small and should be interpreted only tentatively.

IV.F.4 Results Summary: Sentence Acceptability

The ERP analyses show that this experiment replicated the canonical N400 and P600 effects. The ERPs to incongruous words were most strikingly more negative than nonanomalous control words. This broadly distributed difference peaked around 400 ms, was greatest at central and posterior electrodes, and tended to be larger over

the right hemisphere. The ERPs to ungrammatical words were markedly more positive than nonanomalous control words. This posteriorly distributed difference peaked between 600 and 850 ms. The P600 effect to pronouns tended to peak earlier than that to verbs, but there were no significant effects of part of speech or stimulus list on these canonical effects. Furthermore, as few as two neural sources could have generated these canonical effects.

The grand average N400 and P600 effect topographies were nearly identical across parts of speech but were less similar within-participants. Across participants N400 effect topographies were about as similar as within participants, but P600 effect topographies were more variable.

In addition to the canonical N400 and P600, other, smaller effects were elicited by anomalous sentence final words. An N400-like effect was elicited by ungrammatical verbs. In the same time window, an anterior negativity was elicited by ungrammatical pronouns, which may correspond to what is generally referred to as the lateral anterior negativity (LAN). An enhanced LPC followed incongruous sentence completions. This effect had a P600-like topography for nouns, but for verbs the effect had an idiosyncratic, spotty distribution.

Relatively low cloze probability nonanomalous noun and verb completions tended to have a larger N400 than comparable high cloze probability completions, but this effect was not significant. Single-trial analysis suggested that this tendency was partially due to a delayed LPC for low cloze words. These low cloze probability

completions also tended to elicit more positive potentials at frontal electrodes in both the N400 and P600 time windows.

In general, ERPs to anomalous verb completions to highly constraining (low entropy) stems were more positive over the back of the head and more negative over the front than comparable completions to less constraining stems. However, the topographies and temporal shape of these effects did not resemble the canonical N400 or P600 effects and they tended to differ between incongruous and ungrammatical verbs.

IV.G Results: P3b, N400, P600 and LPC Comparisons

IV.G.1 Peak Latency and Amplitude Correlation

If the features of two ERP components are reliable predictors of one another, then it suggests that the two components function as part of the same physiological system. For example, evidence that the amplitude and latency of a participant's P3b predict the amplitude and latency of a participant's P600 is consistent with the possibility that the P3b and P600 are physiologically similar or have similar antecedent causes. However, feature correlations may also reflect very general differences between participants that have little or nothing to do with the functions indexed by particular ERP components. ERP amplitude correlations may be caused by differences in head conductivity (i.e., all else being equal, participants with thinner skulls will have bigger ERPs), and latency correlations may simply reflect differences

in age (i.e., the latency of some ERP components increase with age—e.g., Sections II.B & II.D).

That being said, a pattern of selective correlation between some ERP components and not others does suggest that the correlated components are physiologically related. For this reason, the rank correlations[56] between the latencies and amplitudes of the P3b, N400, and P600 effects and the LPC were calculated and are presented in Table IV.5. To minimize ERP estimation error, each participant's P3b effect was derived by averaging her/his P3b effects from both oddball tasks. Similarly, the N400 and P600 effects were derived by averaging the N400 and P600 effects to different parts of speech. The LPC was derived by averaging a participant's ERPs to three types of nonanomalous sentence endings (nouns, verbs, and pronouns).

The method of peak latency measurement for the ERP effects was described in Sections IV.E.3.a and IV.F.3.a. To determine the latency of the LPC, each participant's ERP to nonanomalous endings was low-pass filtered at 7 Hz and the latency of the most positive value between 300 and 1100 ms at each electrode was measured. The median latency of these peaks across all electrodes was taken as the LPC peak latency. To determine the amplitude of the ERP effects and the LPC, the mean voltage of each difference wave or ERP in a 150 ms window centered around the peak was measured. The mean amplitude of the electrode with the maximal effect (P3b, N400, & P600) or magnitude (LPC) was used as the amplitude measure. To

[56] Rank rather than linear correlation was used because rank correlation measures the monotonic relationship between two variables; this includes linear and some non-linear relationships and is thus more general. Furthermore, rank correlation is less sensitive to outliers than linear correlation, an important feature given that not all participants exhibit N400 and P600 effects.

estimate a lower bound on the reliability of these measures, they were repeated on the same ERPs from 50% splits of the data (every other trial from each of the 12 types of stimuli) and correlated within the same type of component.

According to these measures, the latencies of the four ERP phenomena were not correlated. The fact that the N400 and P600 did not correlate with other components could be due to estimation error, as the 50% splits of these components do not reliably correlate with themselves. The P3b and LPC do correlate with themselves but clearly do not correlate with one another. ERP amplitudes were reliably measured for each ERP component and there is a tendency for the amplitude of all ERP components to be correlated with one another. However, the only two components whose amplitudes are significantly correlated are the N400 and LPC.

IV.G.2 Least-Squares Source Model of the N400

The amplitudes of the N400 effect and the LPC could be correlated for several reasons as they overlap in time and have similar distributions. It could be that a large N400 effect predicts a large LPC because participants with large N400s to incongruous words have small N400s to nonanomalous words (assuming that the N400 and LPC overlap somewhat in time). Another possibility is that a large LPC could predict a large N400 effect because the N400 effect is primarily the product of a delayed LPC to incongruous words. To explore the likelihood of these two possibilities, each participant's N400 effect was modeled with the single source, $\hat{\mathbf{g}}$, that would account for the maximal amount of N400 effect variance. This source

model is the first PC of the previously described sum of products matrix, **L** (Section IV.E.3.d), which is equivalent to a non-zero scalar, *h*, times the sum of the N400 difference wave in the N400 window:

$$\hat{\mathbf{g}} = h\sum_{t}\mathbf{x}(t) \tag{4}$$

where **x**(t) represents the N400 difference wave at all electrodes at time *t*. As mentioned previously (Section III.F), the least-squares estimate of this source's activity, $\hat{y}$, is simply:

$$\hat{y}(t) = (\hat{\mathbf{g}}^T\hat{\mathbf{g}})^{-1}\hat{\mathbf{g}}^T\mathbf{x}(t) \tag{5}$$

Note that the polarity and scale of $\hat{\mathbf{g}}$ are underdetermined because of the arbitrary non-zero scalar, *h*. Consequently, the polarity and scale of the source activity is also ambiguous. However, the ratio of source activity magnitude between two conditions is unambiguous and it is possible to determine which condition elicits greater source activity.

The average contribution of each participant's single N400 source to her/his ERPs to all incongruous and all nonanomalous noun and verb sentence endings were computed and their absolute magnitudes compared. Figure IV.59 (A) displays the number of participants who had greater absolute average activity to incongruous endings at each time point in the epoch. Curiously, at the beginning of the N400 window there was no significant difference between conditions. By the end of the window, nonanomalous endings elicited more activity but then the trend reversed such that by 850 ms, incongruous endings elicited more activity (on average).

These results are not consistent with either a pure "enhanced negativity" or "delayed positivity" explanation of the N400 effect, as during most of the N400 window there was no clear difference between conditions. Rather, inspection of individual participant source activities shows that incongruous endings generally elicited a predominantly scalp-negative wave and nonanomalous endings a predominantly scalp-positive wave in the N400 window (Figure IV.59: B). Thus, to a first approximation, it appears that the N400 effect was created by both a scalp-negative wave to incongruous words and a scalp-positive wave to nonanomalous words of about equal magnitude. Moreover, it is possible that the negativity and positivity were part of the same biphasic source activity.

The results of this analysis could be quite misleading as this minimal source model is not capable of dissociating ERP sources with similar topographies. All that it definitely shows is that the ERPs to incongruous words were generally characterized by a negativity that was somewhat similar in topography to the N400 effect and that the ERPs to nonanomalous words were generally characterized by a positivity that was somewhat similar in topography to the N400 effect. The negativity and positivity could have been generated by distinct sources, and one or the other may have been primarily responsible for the N400 effect. Nonetheless, the possibility of a biphasic N400 is worth considering.

IV.G.3 Topography Similarity

In the absence of estimation error, if two ERP components are generated by the same neural sources to the same relative degree, they will have identical topographies (Urbach & Kutas, 2002). Thus if two ERP phenomena have highly similar topographies, they may well have identical neural underpinnings. For this reason it is worthwhile to compare the topographies of the P3b, N400, P600, and LPC. Unfortunately, the comparison of ERP topographies is complicated by estimation error. As explained in Section IV.E.3.e, if the two ERP components differ in magnitude, the conventional ANOVA-based technique for determining if the two topographies differ has been shown to have an inflated Type I error rate. Since the magnitudes of the P3b, N400, and P600 effects and the LPC are clearly not equal (Figures IV.21), I will refrain from testing the identity of these effect topographies. Rather, as done in the preceding topography analyses, I will simply describe topography similarity using r_{cos}.

Unlike the previous analyses, this analysis compares an ERP feature (the LPC) with difference ERP features (the P3b, N400, & P600 effects). It is important to note that the similarity of the LPC topography to that of the other components will be somewhat skewed because the LPC to sentence final words reflects activity from preceding words in the prestimulus baseline (Urbach & Kutas, 2002). In other words, the LPC topography reflects the gain vectors of the underlying EEG sources in the LPC window minus that of the prestimulus baseline activity. This is not an issue for

comparing difference waves as the prestimulus contribution should be about equal in the two contrasted conditions and cancel out (ibid.).

For this analysis, the P600 effects were estimated separately for each part of speech as the pronoun and verb P600 topographies tended to differ. The N400 effects were also estimated separately by part of speech and the P3b effects estimated separately for each oddball task to have a comparable signal to noise ratio with that of the P600 estimates. The LPC topography was estimated from the average of the ERPs to nonanomalous noun, verb, and pronoun endings (each part of speech is weighted equally) as these ERPs were all highly similar on average. Based on the grand average of these ERPs, 450 to 800 ms was chosen as the LPC time window.

The similarity of the grand average P3b, N400, and P600 effect and LPC topographies are displayed in Table IV.3. Nearly all of the effects were highly similar ($|r_{cos}| \geq .90$). In fact the P600 effects were about as similar to some of the other effects as they were to themselves. Notably, the LPC topography was most similar to the N400 topographies of all the ERP effects (Figure IV.46). However, the grand average topographies probably overestimate the degree of similarity as averaging across participants tends to act as a low-pass spatial filter (Kayser & Tenke, 2006), which would blur-away potentially fine distinctions between these topographies.

To avoid this problem with grand averaging, the individual participant topographies were compared and were generally found to be much less similar (Figures IV.23-26). Table IV.4 (A) displays the median within-participant topography similarities of these ERP phenomena. While the two P3b effects were highly similar

to one another, the other comparisons revealed more moderate similarities. In general, the noun and verb N400s were about as similar to the P3b and LPC as they were to one another. The P600s to pronouns and verbs tended to be more similar to the P3b and LPC than to one another. The N400 and P600 were the most dissimilar of all.

As with the previous estimates of peak latency and amplitude, a lower bound on topography reliability was estimated by computing each type of topography twice from 50% splits of the data (every other trial from each of the 12 types of stimuli). These estimates are presented in Table IV.4 (B) and show that while the P3b and LPC topographies were highly reliable, the N400 topographies were less reliable and the P600 topographies were generally the least reliable. Thus, to some extent the lack of similarity between the N400 and P600 effects and the other ERP phenomena probably reflected estimation error.

IV.G.4 Results Summary: P3b, N400, P600 and LPC comparisons

In summary, comparisons of the peak latency, amplitude, and topographies of these four ERP phenomena generally revealed only modest or weak similarities. The exception to this was the N400 and LPC, which had significantly correlated amplitudes and relatively similar topographies. The P3b and LPC also had relatively similar topographies, but their amplitudes and peak latencies were clearly not well correlated. The topography of the P600 was somewhat similar to the topographies of the P3b and LPC, but there was no evidence of a relationship between their peak latencies and amplitudes. The lack of similarity between the P600 and the other ERP

phenomena may largely have been due to ERP estimation error. To a lesser extent, this may have been true for the N400 as well.

IV.H Discussion

IV.H.1 Replication of the Canonical P3b, N400, and P600

The primary result of this chapter is that the two experiments succeeded in replicating the canonical P3b, N400, and P600 ERP effects, an obvious prerequisite for the subsequent subcomponentry analysis of these effects. On average, these effects exhibited the ERP morphologies and topographies characteristic of these phenomena. Moreover, consistent with previous research (Bentin et al., 1999; Kutas et al., 1977), the P3b effect in the presumably more difficult semantic category oddball task was slightly smaller and peaked later than in the letter case oddball task. The oddball task data appeared to deviate in one respect from the canonical P3b. Specifically, while the P3b to targets preceded by standards was enhanced relative to the P3b to targets preceded by other targets, the posterior enhanced positivity was accompanied by a frontal enhanced negativity. Previous research using simple auditory (tones) or visual (X's & O's) oddball paradigms found only an enhanced positivity that was largest at posterior electrodes (Gonsalvez & Polich, 2002). The difference could be due to the more difficult or linguistic nature of the oddball paradigms used here. This result suggests that equivocating effects of stimulus sequence and target probability on the P3b (Donchin & Coles, 1988) is mistaken.

However, since effects of stimulus probability could not be analyzed in these data, this conclusion is merely speculative at present.

The P600 effect also deviated somewhat from previous findings. While the amplitude of the P600 effects to pronouns and verbs were approximately equal in this study, Coulson and colleagues (1998) found that pronouns elicited a larger P600 than verbs (by about 2 microvolts at a left parietal electrode) using identical types of grammatical anomalies. The difference may be due to the fact that Coulson and colleagues used much higher (80%) or lower (20%) frequencies of anomalous sentences than used here, which may enhance the P600 to the easier/more salient pronoun anomalies. A later study by Kemmer et al. (2002) used equi-frequent grammatical and ungrammatical sentences and found comparable P600 amplitudes for ungrammatical pronouns and verbs, though the pronouns used in that study were different than those used here and by Coulson and colleagues (i.e., they used reflexive pronouns). The difference between studies may also be due to the fact that Coulson et al. used sentence internal (rather than sentence final) anomalies and probably used less likely nonanomalous control words[57].

Finally, the effect of cloze probability on the N400 to nonanomalous noun and verb endings appeared to be consistent with that of previous research (DeLong et al., 2005; Kutas & Hillyard, 1984). In the N400 window, the ERPs to highly probably nonanomalous endings tended to be more positive at posterior electrodes and more negative at far frontal (especially on the left) than those to less probable endings.

[57] Coulson et al. did not report cloze probability estimates for their nonanomalous control words. In general, sentence internal words should be less probable than sentence final words.

However, this tendency was barely significant, which probably reflects the limited range of nonanomalous ending probabilities (i.e., all were rather probable).

IV.H.2 Between-Participant and Single Trial Variation

While the P3b, N400, and P600 ERP effects were reliable on average, it is important to note the variability in these effects across participants. This variation is most clearly apparent in the P3b effects. All participants reliably showed some sort of enhanced positivity to targets relative to standards, but the topography of the P3b effect could vary greatly between participants. Some participants elicited P3b effects that were much more frontal or posterior than the grand average P3b effect. Furthermore, these differences typically replicated across oddball tasks, so they are not simply the product of estimation error.

The sentence acceptability task effects were somewhat more variable between participants than the P3b. Almost all (88%) of the participants showed some type of reliably greater negativity to incongruous than congruous endings and the topography of the N400 effect across participants was about as variable as that of the P3b. In contrast, the P600 effect was remarkably unreliable. Only 63% of the participants exhibited a significantly greater positivity to ungrammatical words than grammatical controls and the topography of the P600 effect was generally not similar across participants (median r_{cos} of ~.45). Based on the current comparisons, it is difficult to tell how much of this variability reflects estimation error and how much reflects true individual variation.

In addition to a degree of estimation error, the between participant variation in these effects could reflect functionally irrelevant differences in the location and orientation of the effect generators due to individual differences in cortical folding and head shape (Homan, Herman, & Purdy, 1987). However, as discussed further in Chapter VII, it is likely that these superficial differences index real differences in participant neural and cognitive function, which have important implications for the study of the subcomponentry of these effects. At present, it suffices simply to note the degree of inter-participant variation as it will be a useful yardstick by which to interpret individual variation in the subsequent independent component analysis of these data (Chapter VI).

In addition to the inter-participant variability, single trial analysis/visualization revealed striking inter-trial variation of these ERP effects. Consistent with previous studies (Kutas et al., 1977; Makeig et al., 2004), the ERPimages of the P3b (Figures IV.26-27) showed a broad range of P3b peak latencies (about 600 ms in this study). Moreover, the fact that the Woody filter was such a poor match to the P3b in the single trials indicates a great deal of inter-trial variation in the topography and/or temporal morphology of the P3b and/or a P3b-like responses to standards. The ERPimages of the verb N400 and P600 (Figures IV.55-56) showed some variation in the onset and duration of these components as well. While this variation is problematic for conventional ERP analysis (and rarely reported or addressed), it should facilitate the ability of independent component analysis to separate subcomponents of these effects.

IV.H.3 Evidence for Subcomponents of the P3b, N400, and P600

It is clear from the anatomical complexity analyses that the P3b, N400, and P600 effects may be generated by only a handful of coherent neural sources, an order of magnitude less than the maximal number necessary. A single source provided a borderline reasonable model of the P3b effect, but two sources are needed to achieve conventional levels of approximation of the N400 and P600 effects. This latter result suggests that the N400 and P600 are necessarily composed of multiple, somewhat independent subcomponents but the apparent complexity could be due to ERP estimation error. With more trials it is possible that the N400 and P600 could be well approximated by a single principal component.

The fact that so few sources may have generated these effects is somewhat surprising given the previously reviewed evidence that they are composed of multiple subcomponents (Section II.E). However, given the moderate number of electrodes used in this study, the use of spatial filters (ICA) to remove EEG artifacts, and the fact that PCA provides a lower bound on the number of sources, it is highly likely that these results are an underestimate of the true number of sources. At most, *n*-*m* sources are needed to account for *n*-*m* or more time points of data from *n* electrodes after *m* EEG artifact independent components have been removed from the data[58]. Thus, at most, 33 sources on average would have been needed to model these effects and

[58] Removing an independent component (IC) of a data set effectively removes an electrode from a data set. In other words, after *m* ICs have been removed from an *n* electrode data set, given the voltage at *n*-*m* electrodes, one can perfectly interpolate what the voltage is at the remaining *m* electrodes.

multiple sources may be accounted for by a smaller number of principal components if the sources have similar topographies or if their average activity tends to be correlated (despite being somewhat functionally independent).

While it is very likely that PCA is underestimating the number of sources, these results demonstrate that the number of sources may still be rather small and attempts to model them with thousands of ECDs (Dale et al., 2000; Haan, Streb, Bien, & Rosler, 2000; Halgren et al., 2002; Moores et al., 2003) may be quite suboptimal. In particular, they may be overestimating the extent of the underlying neural activity (Liu, Dale, & Belliveau, 2002).

IV.H.4 On the Relationship Between the P3b and the LPC

The topographies of the P3b and LPC were generally quite similar and the fact that the latency of the LPC negatively correlated with the probability of a word suggests that its timing, like that of the P3b, reflects the duration of stimulus evaluation. While these two findings might indicate a relationship between the two effects, the P3b and LPC are clearly dissociable. This is apparent in the fact that the peak latencies and magnitudes of these phenomena did not correlate, and this lack of correlation was not due to estimation error. This dissociation could reflect qualitatively different physiological underpinnings of the two effects, which would contradict their past equivocation by some investigators (Chung et al., 1996; Curran et al., 1993). Of course, the dissociation may also have been caused by some subcomponents of the P3b or LPC that are unique to one of the effects and obscured

the similar subcomponents. Alternatively, the dissociation may simply indicate different "sensory routes" to the generation of the same effect. For example, a blindfolded individual can elicit an auditory P3b but not a visual P3b, yet there may well be EEG processes common to these two effects. In the case of the LPC and P3b, the different sensory routes would be the comprehension of sentences versus the comprehension of single words. The two perceptual problems may involve different antecedent processes that can lead to a dissimilar engagement of the subsequent P3b/LPC.

IV.H.5 On the Relationship Between the N400, P3b, and LPC

The analyses in this chapter indicate a relationship between the N400 effect and the LPC, but there is no evidence that the N400 effect is related to the P3b effect, save for their somewhat similar topographies. The best evidence that the N400 and LPC are related comes from the fact that N400 effect magnitude strongly correlated with LPC amplitude and the grand average topographies of the two phenomena were nearly identical. The topographies of these phenomena were also rather similar in individual participants but less remarkably so. To some extent, this lesser similarity may have been due to the somewhat unreliable estimates of the N400 effect. Likewise, N400 estimation error may have contributed to the lack of correlation between the peak latencies of the two phenomena.

The relationship between the N400 effect and LPC amplitudes is consistent with the theory that the N400 effect at least partially reflects a reduced or delayed LPC

to incongruous sentence endings (Chung et al., 1996). This position is somewhat corroborated by the single trial visualization (ERPimage) of the EEG response to nonanomalous nouns and verbs as a function of cloze probability. According to the ERPimage, the LPC tended to begin and end later for less likely endings. Moreover, this relationship was small but reliable in a significant fraction of participants when a Woody filter was used to quantify the latency of the LPC.

Alternatively, since the N400 effect and the LPC overlap in time, it is possible that a large LPC simply reflects a small N400 to congruous endings and that individuals with small "congruous N400s" tend to have large "incongruous N400s." Furthermore, a larger N400 to less probable endings could make it appear as if the LPC begins later to such words. However, such a "pure N400 negativity" explanation would not explain the similar distributions of the N400 and LPC effect or the delayed offset of the LPC to less probable endings. These results could be explained by a biphasic source of the N400 and LPC, as suggested by the least-squares single source modeling of the N400 effect. In other words, it could be that the N400 and LPC are really two parts of the same biphasic response and that the N400 effect is created by a negativity to incongruous words and a positivity to congruous words whose latency tends to be an inverse function of word probability. Further discussion of these possibilities and their implications will be postponed until Chapter VII in order to incorporate the results of applying independent component analysis to these data (Chapter VI).

IV.H.6 On the Relationship Between the P600, P3b, N400, and LPC

Assessing the relationship between the P600 and the other three ERP components was complicated by the unreliability of the P600. The amplitude and peak latency of the P600 effect did not reliably correlate with any of the other components and the P600 topography was only moderately similar to that of the others. While the lack of amplitude correlation was convincing, the degree of independence between the peak latencies and topographies may simply have been caused by P600 estimation error.

There is some evidence that the P600 and P3b are at least functionally similar if not anatomically similar. Specifically, it appears that P600 peak latency correlates with reaction time. Although participant responses were delayed until 1400 ms after the presentation of the sentence final word, participants were still faster to respond to ungrammatical pronouns than ungrammatical verbs. On average, the P600 to pronouns peaked before that to verbs, though this trend was not significant. Again the lack of significance probably reflects P600 estimation error. Indeed, the trend is corroborated by a study by Coulson and colleagues (1998) who also found that the P600 to ungrammatical pronouns tended to precede that to ungrammatical verbs using the same types of grammatical anomalies. However, Coulson et al. did not report an analysis of this trend.

IV.H.7 Other Effects and Null Effect of Note

In addition to the main ERP effects of interest, there were other effects of expectancy violation in the sentence acceptability task. Ungrammatical verbs elicited an enhanced negativity that was smaller and left-lateralized, but very similar to the N400 effect triggered by incongruous endings. Almost all participants showed a tendency for this N400-like effect and the later P600 effect. These data corroborate previous findings of an N400-like effect after sentence final grammatical anomalies (Osterhout, 1997; Friederici, et al., 1993;--but see Gunter, Stowe, & Mulder, 1997). This negativity may reflect the fact that both ungrammatical verbs and incongruous endings appear to take longer to evaluate than pronoun endings or nonanomalous endings. Other investigators (Osterhout & Holcomb, 1992) have suggested that such negativities reflect "sentence wrap-up" semantic processing as sentence-embedded grammatical anomalies often elicit N400 effects following sentence final words (for review see: Osterhout, 1997) and sentence-embedded ungrammatical verbs do not reliably elicit N400 effects (Kaan, Harris, Gibson, & Holcomb, 2000; Kemmer, Coulson, De Ochoa, & Kutas, 2004; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2005; Kutas & Hillyard, 1983; Osterhout & Nicol, 1999;--but see Coulson et al., 1998). However, the lack of an N400 effect to those sentence-embedded anomalies may have been due to relatively unpredictable nonanomalous control words (none of those studies report cloze probability estimates for control words). Moreover, the theory that grammatical violations lead to semantic difficulty with

sentence wrap-up does not explain the lack of an N400 effect to ungrammatical pronouns, which we turn to next.

Rather than an N400-like effect, ungrammatical pronouns elicited a small anterior negativity (AN) that may correspond to an ERP component commonly called the "left anterior negativity" (LAN). The LAN is sometimes elicited by ungrammatical words or words that tax working memory (Coulson et al., 1998; Kutas & Van Petten, 1994; Osterhout, 1997). Despite its name, the topography of the effect is variable and has sometimes been found to be symmetric across the midline (Friederici, Pfeifer, & Hahne, 1993; Munte, Heinze, Matzke, Wieringa, & Johannes, 1998; Osterhout, 1997). The presence of such an effect to ungrammatical pronouns but not to ungrammatical verbs adds to the substantial body of evidence (Coulson et al., 1998; Gunter et al., 1997, Exp. 1; Kaan et al., 2000, Exp. 2; Kemmer et al., 2004; Kemmer & Kutas, 2006; Kolk, Chwilla, van Herten, & Oor, 2003, Exp. 1; Kuperberg et al., 2003, 2005) against the theory that such focal anterior negativities reflect morphosyntactic processing (Friederici, 2002). However, the superficial differences between pronouns and verbs in this study may be misleading. The AN to pronouns may also have contributed to the N400-like effect to ungrammatical verbs and its presence may simply have been obscured by the earlier onset of the P600 to pronouns.

In addition to the canonical N400 effect, incongruous completions were followed by enhanced late positivities. This positivity to incongruous nouns was broader and weaker over midline electrodes than the P600 effects. The positivity to incongruous verbs was quite small and spottily distributed. Enhanced late positivities

(or at least trends for late positivities) following incongruous words are not uncommon (Curran, Tucker, Kutas, & Posner, 1993; Federmeier & Kutas, 1999; Ganis, Kutas, & Sereno, 1996, mixed stimuli experiment; Holcomb, et al., 1999; Kuperberg, et al., 2004; Kutas & Hillyard, 1980b, Exp. 2; Kutas & Hillyard, 1983; Munte, et al., 1998). However, their topographies are quite variable and often the positivities are absent (Friederici, et al., 1993; Johnson & Hamm, 2000; Kutas, 1987; Kutas & Hillyard, 1980a, 1984; Osterhout, 1997; Osterhout & Nicol, 1999; Van Petten, 1993). The unreliability of these effects makes them difficult to interpret. They may be contingent on individual strategies or anatomy, or subtle differences between task and stimuli. This study suggests that they may at least partially reflect expectation violation, as low entropy verbs tend to elicit larger late positivities than high entropy verbs, and incongruous nouns elicit larger late positivities than incongruous verbs (recall that noun final stems have a lower entropy than verb final stems). Previous research on the effect of constraint on the N400 (Federmeier & Kutas, 1999b; Kutas & Hillyard, 1984) have not unequivocally found that semantically incongruous endings of high constraint (i.e., low entropy) sentences elicit larger late positivities than incongruous endings of low constraint (i.e., high entropy) sentences. These null effects may reflect the fact that the task in both of those studies did not require an immediate response and/or target detection (participants were asked to read the stimuli in order to answer a questionnaire later).

In addition to this posterior-positive effect of entropy, there was an anterior enhanced negativity to incongruous verb endings of low entropy sentence stems.

However, this effect was small and did not replicate in the ERPs to ungrammatical verb endings.

Finally, it is worth noting the lack of an effect of part of speech on the ERPs to nonanomalous noun and verb endings. In less supportive sentence contexts, sentence intermediate nouns elicit generally more negative ERPs than verbs of comparable frequency and length (Federmeier et al., 2000). According to the data presented here, any such differences disappear when words are relatively predictable. This null effect is quite convincing as the grand average ERPs to the two types of nonanomalous endings are practically identical and it is remarkable given that some investigators have proposed that the representation of noun meaning qualitatively differs from that of verbs (for discussion see Federmeier et al, 2000). For example, Pulvermuller (1999) has proposed that comprehending the meaning of verbs commonly involves frontal areas used in motor processing and comprehending concrete nouns (like the majority of noun endings used in this study) involves visual cortical areas.

The noun/verb null effect reported here is similar to the effect of supportive context on processing concrete and abstract words. In the absence of supportive context, ERPs to concrete words are generally more negative than those to comparable abstract words (Holcomb et al., 1999; Kounios & Holcomb, 1994; Swaab, Baynes, & Knight, 2002; West & Holcomb, 2000). However, when a word is relatively predictable from previous context, the concreteness of the word has no effect on the ERPs it elicits (Holcomb et al., 1999). Again, this is striking given proposals that

comprehending abstract words requires more frontal lobe (Barsalou, 1999) or less right hemisphere (Pulvermuller, 1999) involvement.

While there may be some truth to such proposed representational distinctions, the results reported here and by Holcomb et al. suggest that the primary determinant of cerebral cortex's immediate response to a word is the fit between an incoming word and previous context, not the representation of a word's meaning. This is consistent with the theory that the brain predicts incoming sensory information in general (Friston, 2005) and linguistic information in particular (DeLong et al., 2005; Federmeier & Kutas, 1999a; Pickering & Garrod, 2007) in several ways. Firstly, if the brain pre-activates word meaning, differences in meaning representation might not be apparent in ERPs as the activity will be present in the prestimulus baseline and consequently subtracted from the poststimulus EEG. Another possibility is that effects of word concreteness or part of speech are generated by activity that is inhibited by the higher order sentence interpretation. This possibility is analogous to a proposal by Kersten and colleagues (2004) as to how higher-order interpretations influence lower level feature processing in visual object perception. Specifically, Kersten and colleagues propose two possible top-down inhibitory mechanisms. The first is that the higher order interpretations inhibit the activation of lower levels when the lower levels have nothing to add to the higher level interpretation. While this may seem counterintuitive, inhibiting uninformative action potentials could conserve metabolic resources. The other proposed mechanism is that the neural activity that is inconsistent with higher order interpretations is suppressed. In this light, the

differential activation of nouns and verbs or concrete and abstract words in unsupportive contexts reflects differences in their multiple possible interpretations. When a narrow meaning of the word is perceived based on previous context, the elicited neural activity may be too small to significantly influence ERPs.

Table IV.1 – Summary statistics for stem completion **(A)** and stem **(B)** properties. Median (semi-interquartile range) are used to describe stem completion frequency. Mean (standard deviation) are used for all other properties.

A: Stem Completion Properties

	Cloze Probability	Frequency (occurrences/ million)	Length (characters)
Congruous/ Grammatical Nouns	.70(.24)	75(54)	5.7(1.9)
Incongruous Nouns	.00(.00)	75(52)	5.8(1.8)
Grammatical Pronouns	.44(.34)	1173(333)	2.8(0.8)
Ungrammatical Pronouns	.01(.03)	3621(1138)	2.4(1.0)
Congruous/ Grammatical Verbs	.58(.25)	52(66)	5.2(1.3)
Incongruous Verbs	.00(.00)	33(48)	5.3(1.3)
Ungrammatical Verbs	.01(.03)	52(66)	5.1(1.3)

	Sum Stem Association	Concreteness	Repetition
Congruous/ Grammatical Nouns	.00(.00)	5.2(1.3)	0.8(3.3)
Incongruous Nouns	.00(.00)	5.1(1.1)	0.2(0.6)
Grammatical Pronouns	.04(.15)	4.2(0.2)	7.8(9.8)
Ungrammatical Pronouns	.02(.17)	4.4(0.3)	35.5(8.1)
Congruous/Grammatical Verbs	.00(.00)	4.6(0.9)	0.8(1.4)
Incongruous Verbs	.00(.03)	4.4(0.8)	0.5(1.1)
Ungrammatical Verbs	.00(.00)	4.6(0.9)	0.8(1.4)

B: Stem Properties

	Constraint (in units of probability mass)	Entropy (in bits)
Noun Final Stems	.71(.22)	1.5(1.0)
Pronoun Final Stems	.58(.23)	2.0(1.0)
Verb Final Stems	.61(.21)	1.9(0.9)

Table IV.2 – Summary statistics describing how many EEG artifact independent components (ICs) were identified and what percentage of a participant's data was contributed by these components (and subsequently removed). It is important to note that the variance accounted for by two sets of ICs is not necessarily the sum of the variance accounted for by each set of ICs alone. This follows from the fact that ICs scalp projections are not necessarily orthogonal to one another. Nevertheless, variance accounted for does give a sense of the fraction of the data attributed to each artifact class.

	Mean (SD) # of each Type of Artifact ICs per Participant	**Mean (SD) Percentage Data Variance Accounted for by each Type of Artifact ICs per Participant**
Blink	1.4(0.6)	.243(.193)
EOG	1.0(1.4)	.022(.040)
Heart Beat	0.9(1.1)	.023(.032)
Drift	6.0(2.5)	.013(.015)
Muscle	20.6(9.0)	.102(.096)
Bad Electrode	1.0(1.4)	.002(.003)
60 Hz	0.1(0.2)	.000(.000)

Table IV.3 – Similarity of grand average ERP effect and LPC topographies. Topography similarity is measured as the cosine of the angle between two topography vectors. This metric is analogous to Pearson's linear correlation coefficient, *r* (i.e., 1=exactly the same, 0=unrelated, -1=exactly the same but polarity reversed).

Grand Average Topography Similarity

	Letter Case P3b	**Category P3b**	**Noun N400**	**Verb N400**	**Pronoun P600**	**Verb P600**
Category P3b	.98					
Noun N400	-.96	-.92				
Verb N400	-.94	-.90	1.00			
Pronoun P600	.94	.97	-.89	-.86		
Verb P600	.96	.96	-.97	-.95	.96	
LPC	.95	.91	-.97	-.97	.87	.94

Table IV.4 – Similarity of ERP effect and LPC topographies. Topography similarity is measured as the cosine of the angle between two topography vectors. This metric is analogous to Pearson's linear correlation coefficient, *r* (i.e., 1=exactly the same, 0=unrelated, -1=exactly the same but polarity reversed). **(A)** Median (semi-interquartile range) of the similarity of the topographies of ERP effects within participants (16 comparisons for each cell). **(B)** Median (semi-interquartile range) of the similarity of the topographies of ERP effects across participants (120 comparisons for each cell).

A: Within-Participant Topography Similarity

Category P3b	.92(.11)					
Noun N400	-.76(.15)	-.79(.24)				
Verb N400	-.66(.27)	-.78(.18)	.78(.30)			
Pronoun P600	.73(.17)	.70(.25)	-.58(.39)	-.54(36)		
Verb P600	.78(.20)	.59(.15)	-.67(.26)	-.53(.44)	.61(.23)	
LPC	.78(.15)	.87(.30)	-.85(.27)	-.78(.15)	.71(.26)	.66(.15)
	Letter Case P3b	**Category P3b**	**Noun N400**	**Verb N400**	**Pronoun P600**	**Verb P600**

B: Within-Participant Topography Similarity (50% splits)

Topography Similarity	.98(.03)	.97(.02)	.60(.24)	.80(.31)	.43(.41)	.65(.31)	.97(.03)
	Letter Case P3b	**Category P3b**	**Noun N400**	**Verb N400**	**PronounP600**	**Verb P600**	**LPC**

C: Between-Participant Topography Similarity

Topography Similarity	.80(.12)	.80(.16)	.75(.30)	.86(.23)	.42(.36)	.48(.29)	.79(.15)
	Letter Case P3b	**Category P3b**	**Noun N400**	**Verb N400**	**PronounP600**	**Verb P600**	**LPC**

Table IV.5 – Peak latency and amplitude rank correlations (Kendall's tau) for ERP effects and the LPC. Diagonal correlations are derived from correlating ERP effects from 50% splits of single trials. The P3b effect was measured by averaging standards and targets from both oddball tasks. The N400 effect was measured by averaging nonanomalous and incongruous sentence final nouns and verbs (each part of speech weighted equally). Similarly, the P600 effect was measured by averaging nonanomalous and ungrammatical sentence final pronouns and verbs (each part of speech weighted equally). The LPC was measured by averaging nonanomalous sentence final words (each part of speech weighted equally). Reported *p*-values are not corrected for multiple comparisons.

A: Peak Latency Rank Correlation

P3b	.55**			
N400	-.10	.37		
P600	.07	.15	.11	
LPC	-.09	.14	.13	.73***
	P3b	**N400**	**P600**	**LPC**

B: Amplitude Rank Correlation

P3b	.58**			
N400	-.25	.43*		
P600	.25	-.30	.57**	
LPC	.35	-.63***	.37	.82******
	P3b	**N400**	**P600**	**LPC**

*p<.05.
** p <.01.
*** p <.001.
**** p <.0001.
***** p <.00001.
****** p <.000001.

Table IV.6 - Summary statistics for the top and lower 50th cloze probability percentiles of congruous/grammatical nouns and verbs. Normal text indicates mean (standard error). Italicized text indicates median (semi-interquartile range). The following characters indicate which features differ by part of speech or cloze probability. Differences in median frequency were tested using Mann-Whitney U-tests. All other features were analyzed with two-way ANOVAs. Frequency and length were the only factors affected by List.

A: Factors that do not Differ Across Lists

	High Cloze Nouns	Low Cloze Nouns	High Cloze Verbs	Low Cloze Verbs
Cloze Probability*@	.88(.01)	.49(.03)	.78(.01)	.36(.02)
Concreteness*	5.4(.2)	5.0(.2)	4.5(.1)	4.7(.1)

B: Frequency (occurrences per million)#$

List AX, AY

	High Cloze	Low Cloze
Nouns	*72.5(59.5)*	*83.0(55.0)*
Verbs	*61.0(65.5)*	*20.0(25.2)*

List BX, BY

	High Cloze	Low Cloze
Nouns	*80.0(67.5)*	*62.0(35.0*
Verbs	*92.0(91.0)*	*39.5(91.0)*

C: Length (characters)*χ

List AX, AY

	High Cloze	Low Cloze
Nouns	5.7(.4)	6.0(.6)
Verbs	4.7(.2)	5.4(.2)

List BX, BY

	High Cloze	Low Cloze
Nouns	6.0(.4)	5.3(.3)
Verbs	5.3(.2)	5.2(.2)

*=Main effect of Part of Speech (noun, verb), $F(1,227)>3.88$, $p<.05$
@=Main effect of Cloze Probability (high, low), $F(1,227)>3.88$, $p<.05$
#=Effect of Cloze Probability (high, low), $U(114,120)$, $p<.05$
$=Effect of List, $U(117,117)$, $p<.05$
χ=Cloze Probability x List interaction $F(1,227)>3.88$, $p<.05$

Table IV.7 – Summary statistics of the features of sentence stems **(A)** and of incongruous and ungrammatical verb endings **(B)** of the upper and lower 50th entropy percentiles of sentence stems. Normal text indicates mean (standard error). Italicized text indicates median (semi-interquartile range). The following characters indicate which features differ across stimulus type. Differences in median frequency were tested using Mann-Whitney U-tests. Differences in entropy and constraint were analyzed with a one-way ANOVA. All other features were analyzed with three-way ANOVAs with factors Entropy (high, low), Anomaly (incongruous, ungrammatical), and List (AX, AY, BX, BY).

A: Sentence Stems

	High Entropy Stems	**Low Entropy Stems**
Entropy [bits]*	2.58(.07)	1.19(.04)
Constraint [probability mass]*	.45(.02)	.78(.01)

B: Sentence Final Words

	Incongruous Verbs, High Entropy Stems	**Incongruous Verbs, Low Entropy Stems**	**Ungrammatical Verbs, High Entropy Stems**	**Ungrammatical Verbs, Low Entropy Stems**
Frequency [occurrences per million]#	*26(29.6)*	*43(63.8)*	*30(53.3)*	*64(70.5)*
Cloze Probability@	0(0)	0(0)	.01(.01)	.00(.00)
Length [characters]	5.3(.2)	5.3(.1)	5.1(.1)	5.2(.2)
Concreteness$	4.5(.1)	4.3(.1)	4.7(.1)	4.5(.1)

*=Main effect of Entropy (high, low), $F(1,154)>3.90$, $p<.05$
$=Main effect of Entropy (high, low), $F(1,294)>3.87$, $p<.05$
@=Main effect of Anomaly (incongruous, ungrammatical), $F(1,294)>3.87$, $p<.05$
#=Effect of Entropy (high, low), $U(154,156)$, $p<.002$

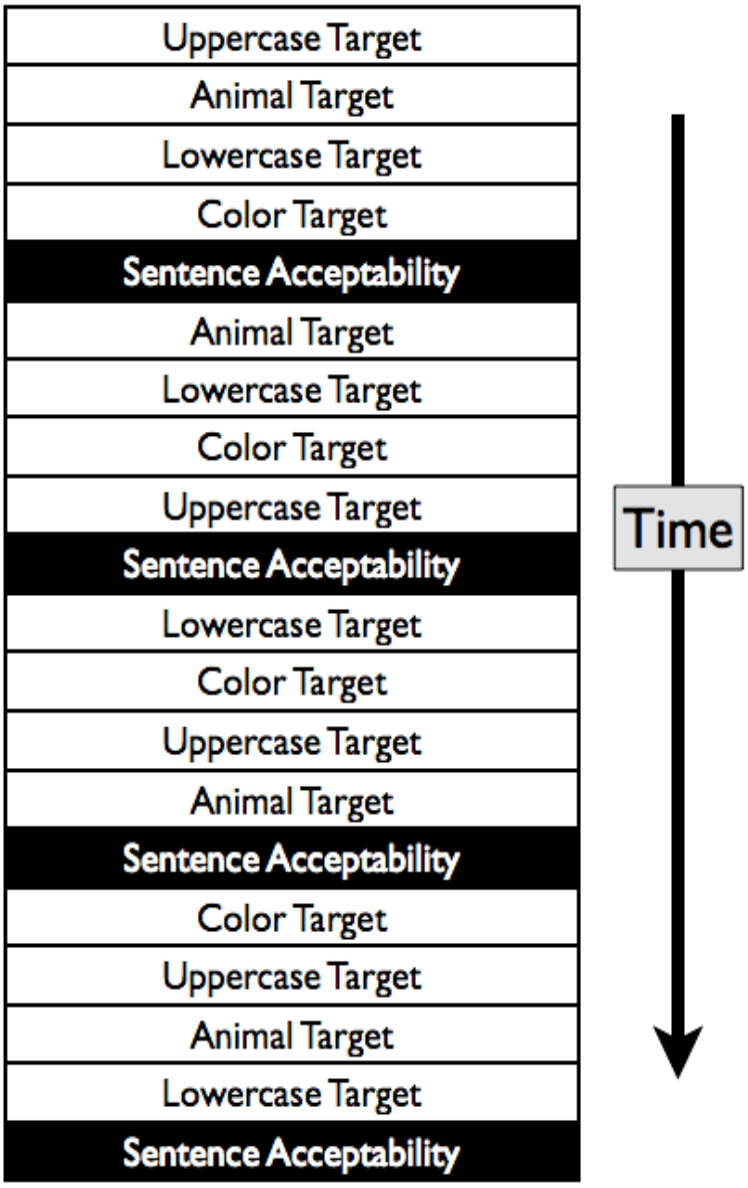

Figure IV.1 – An example series of tasks that a participant would perform in the experimental session (time runs from top to bottom). Each cell represents a task sub-block. White and black backgrounds denote Experiment 1 (Visual Oddball Tasks) and Experiment 2 (Sentence Acceptability Task) sub-blocks, respectively.

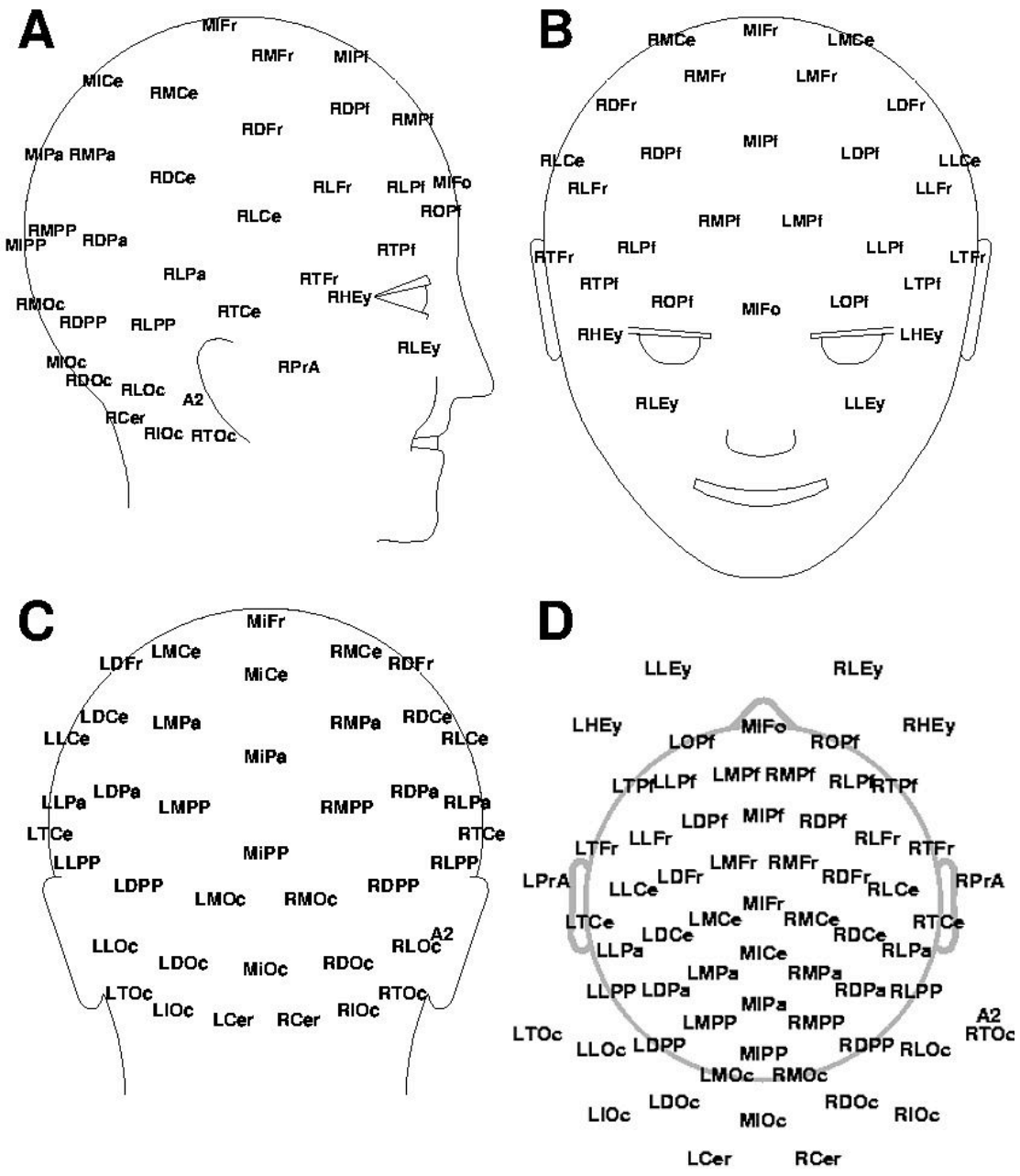

Figure IV.2 – (A-C) The approximate location of each electrode is represented on cartoon heads by each electrode's abbreviated label (e.g., MiCe="Midline Central"). Electrode coordinates and full electrode labels are given in Appendix 1 (Section IV.I). **(D)** Representation of electrode locations used for visualization of scalp distributions. The head is approximated by a sphere and viewed from above (anterior=towards top of page, right=towards right of page). Electrodes below the head's equator are plotted progressively beyond the radius of the head so that all electrodes may be visualized at once.

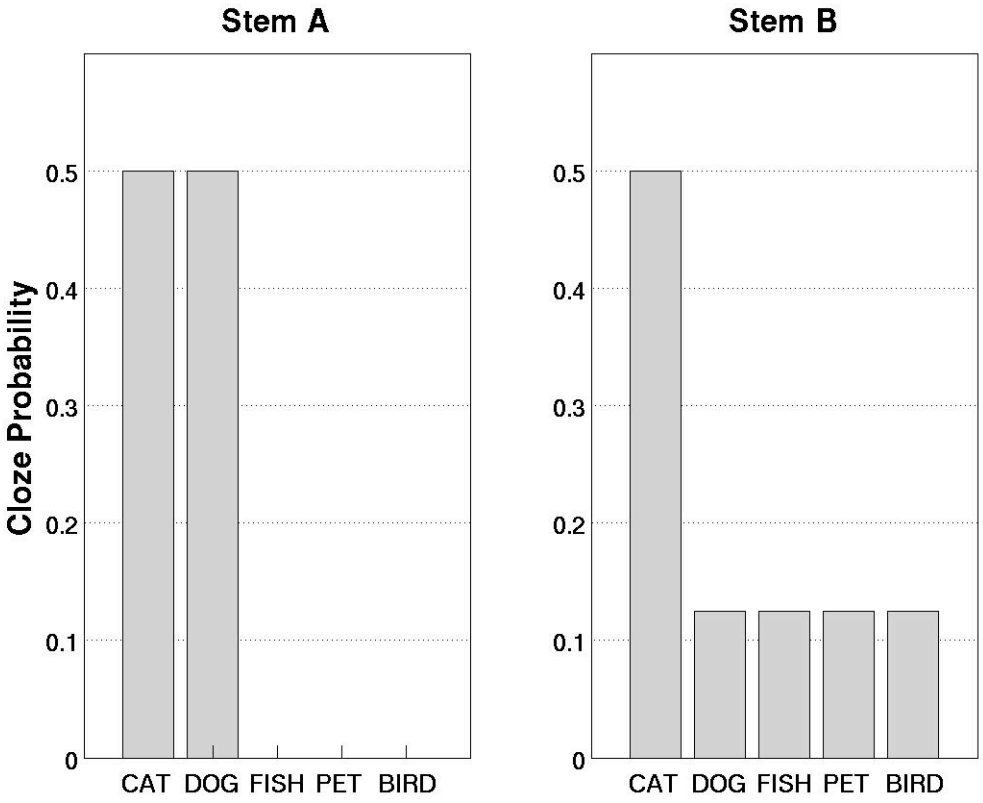

Figure IV.3 – A hypothetical example to illustrate the difference between constraint and entropy. (**Left**) The cloze probabilities of continuations of hypothetical sentence Stem A. This stem has a constraint of .50 and an entropy of 1 bit. (**Right**) The cloze probabilities of continuations of hypothetical sentence Stem B. This stem also has a constraint of .50, but an entropy of 1.39 bits.

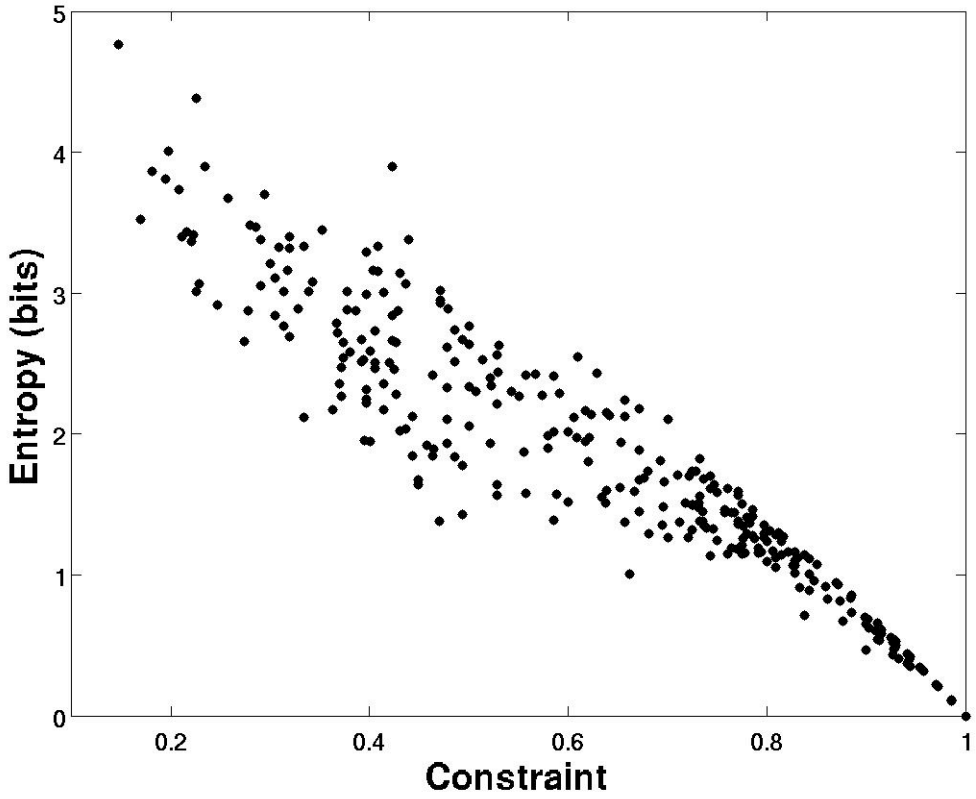

Figure IV.4 – Scatter plot of the 312 sentence stem constraints and entropies. The two features are highly linearly correlated (*r*=-.94).

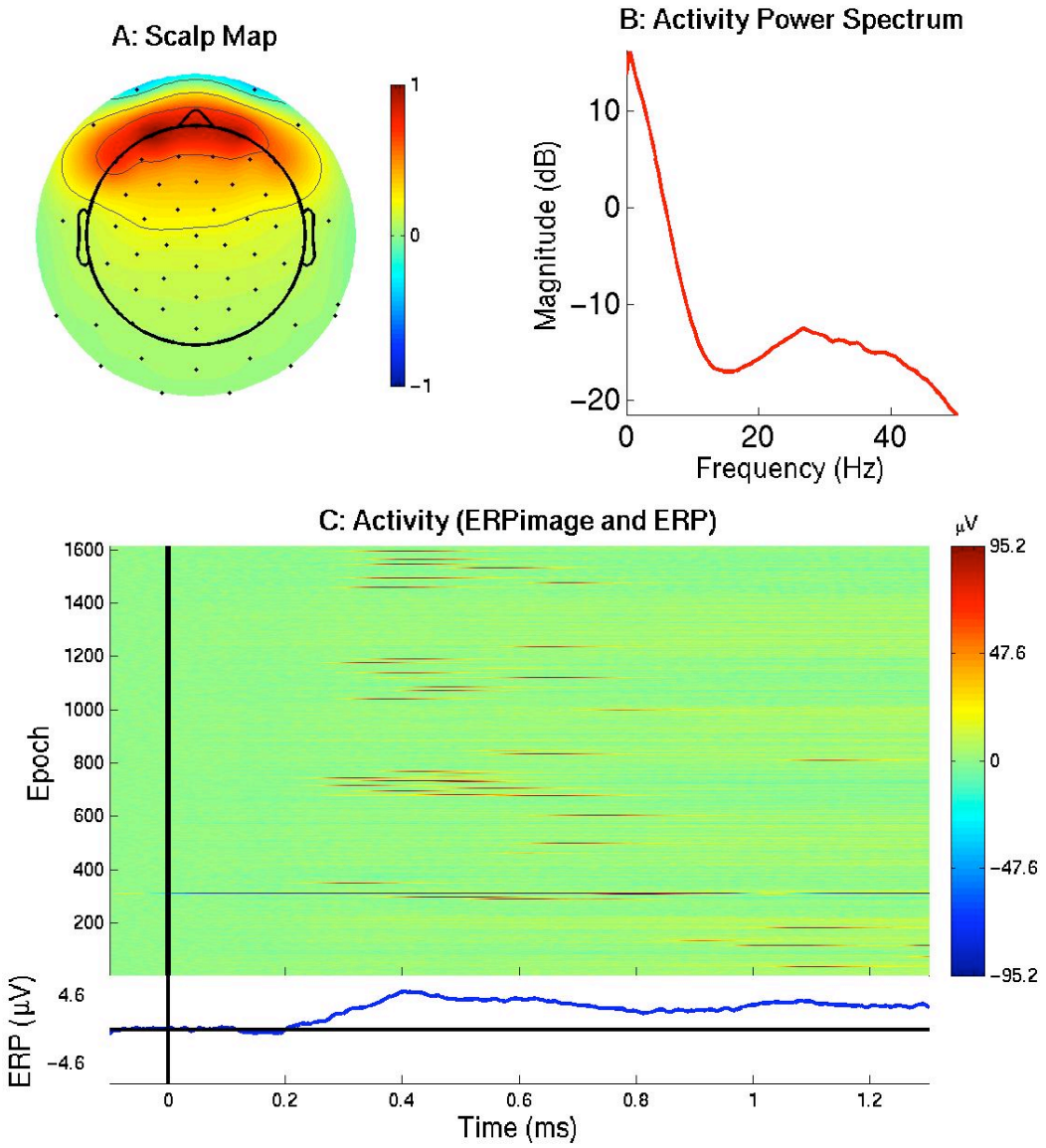

Figure IV.5 – Properties of an independent component (Participant 1, IC 1) that primarily represents EEG blink artifact. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. Note the polarity reversal between the electrodes below (the two topmost electrodes) and the electrodes above the eyes that is characteristic of blink potentials. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

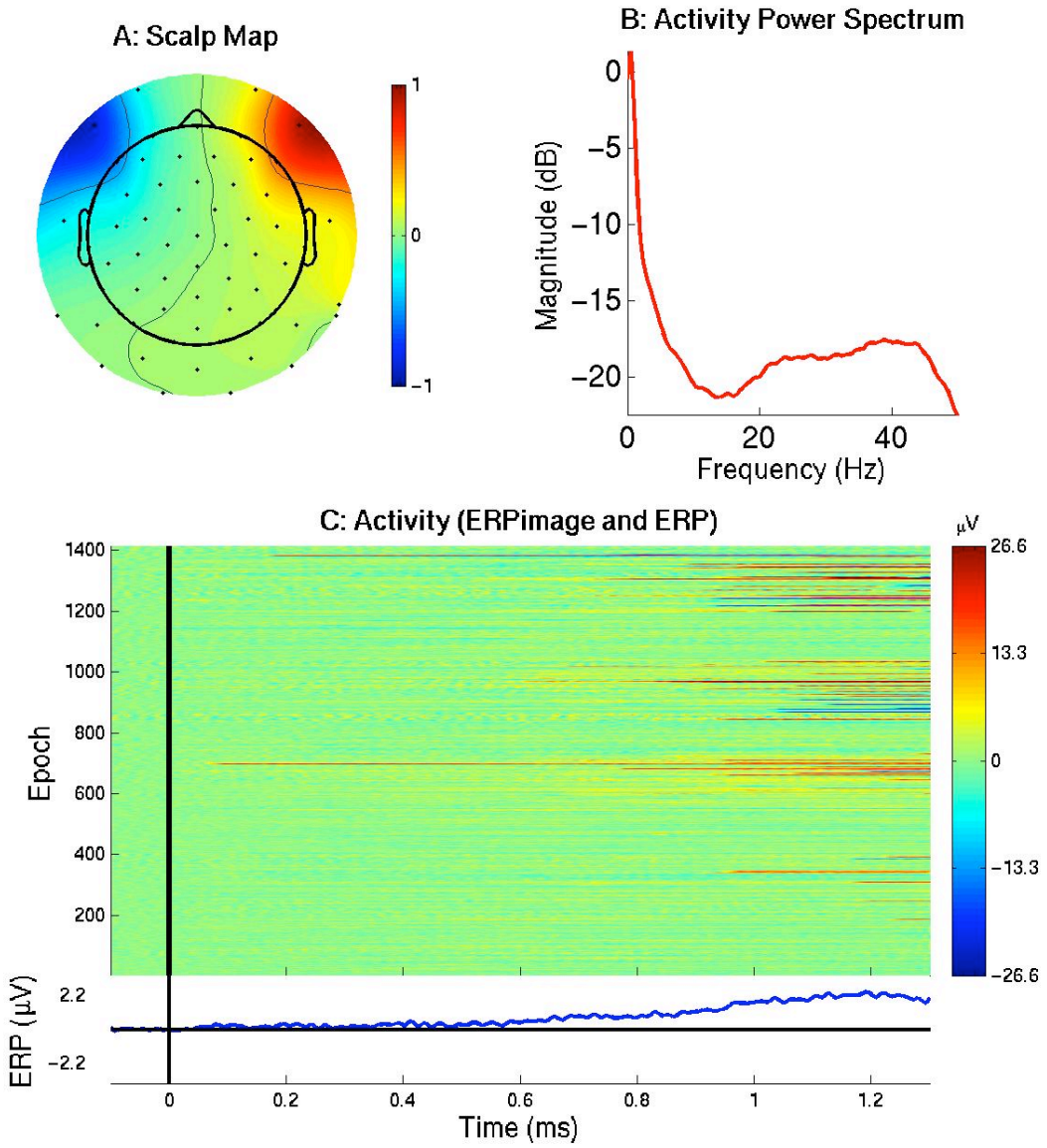

Figure IV.6 – Properties of an independent component (Participant 15, IC 8) that primarily represents EEG eye movement artifact. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. Note the polarity reversal between the electrodes to the right and the left of the eyes that is characteristic of horizontal eye movement potentials. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

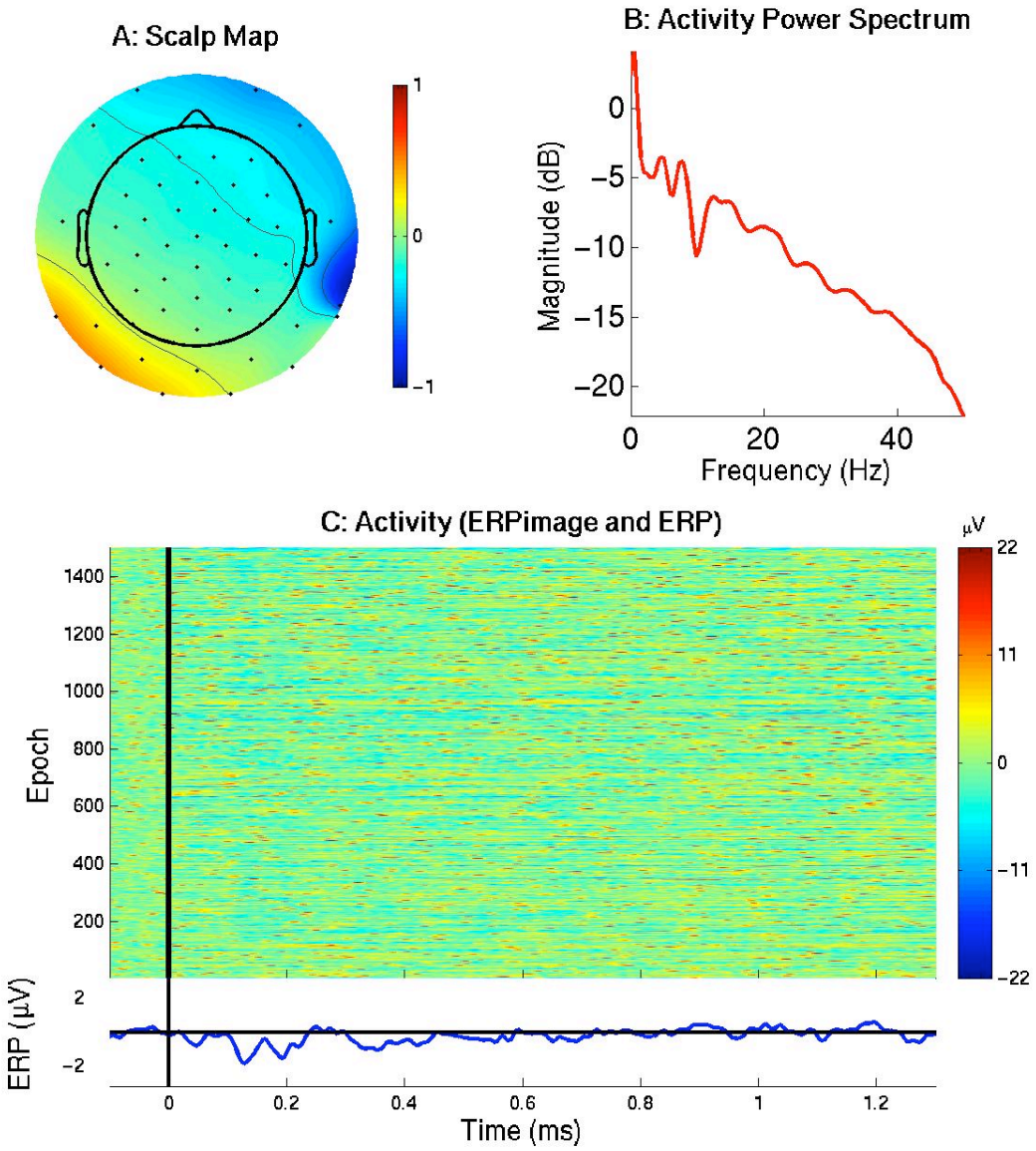

Figure IV.7 – Properties of an independent component (Participant 2, IC 3) that primarily represents EEG heart beat artifact, which occurs when the reference electrode is placed too close to an artery. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

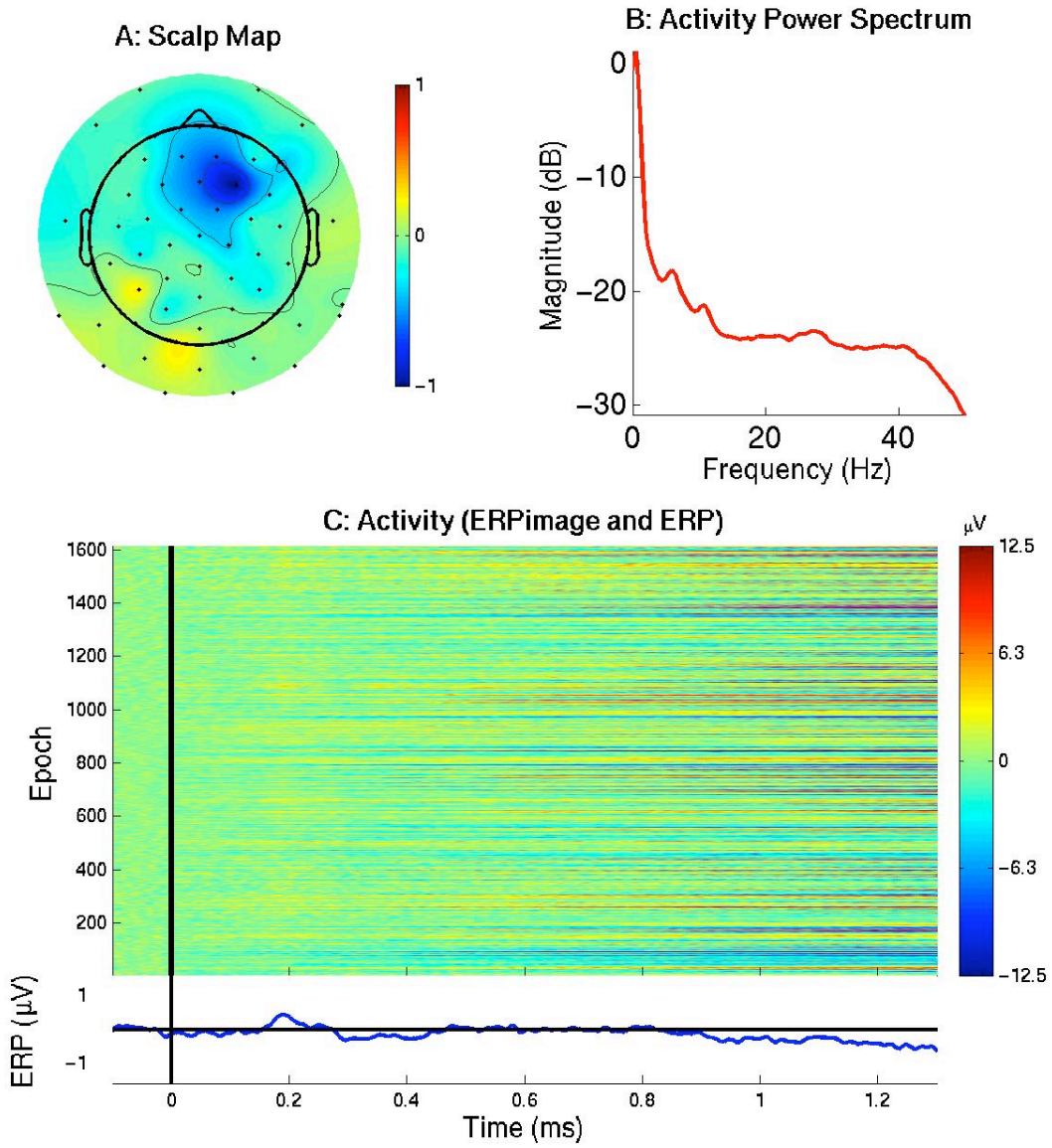

Figure IV.8 – Properties of an independent component (Participant 1, IC 33) that primarily represents EEG drift artifact, (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

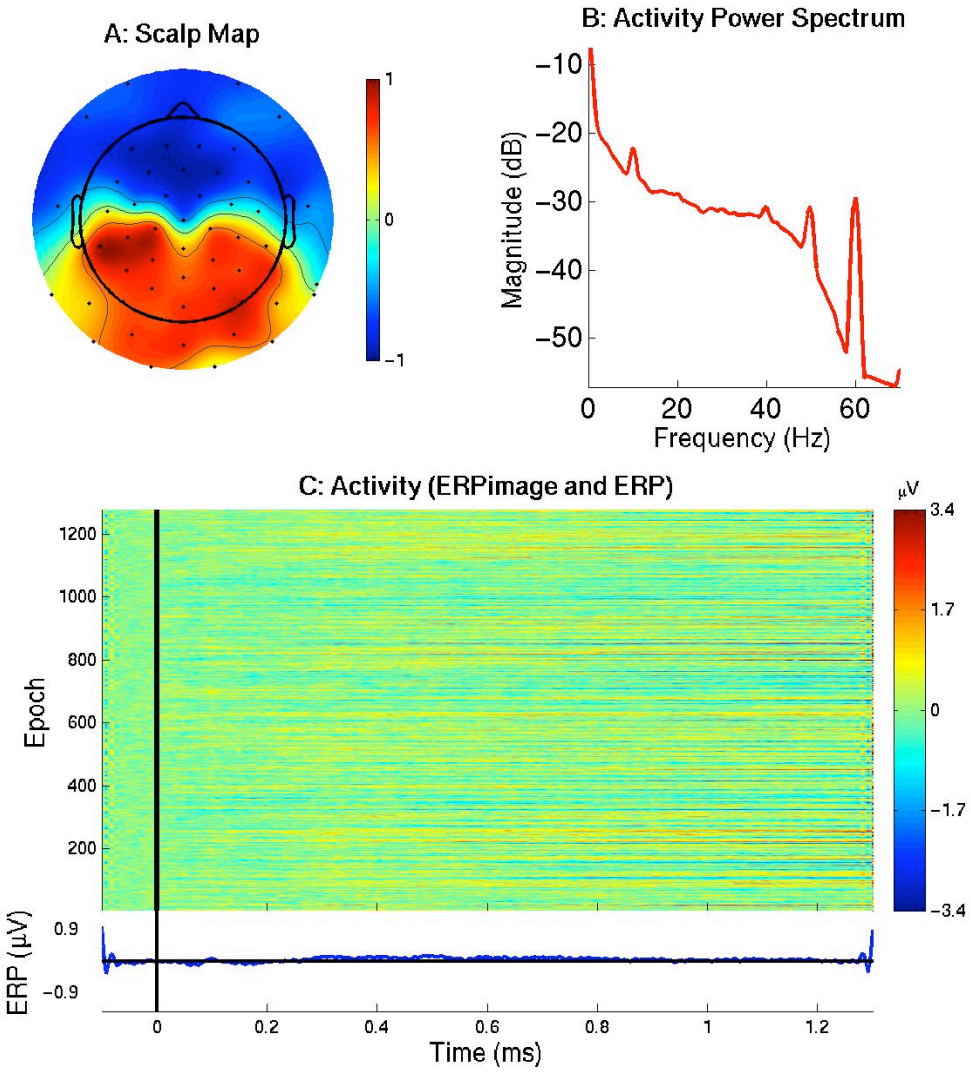

Figure IV.9 – Properties of the one independent component (Participant 11, IC 46) that primarily represents 60 Hz line noise. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. The topography presumably reflects the fact that anterior and posterior electrodes were connected to separate amplifiers. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs). The relatively high amplitude activity at the beginning and end of the recording epoch presumably reflect edge effects of the 50 Hz low pass filter that was used to preprocess the EEG.

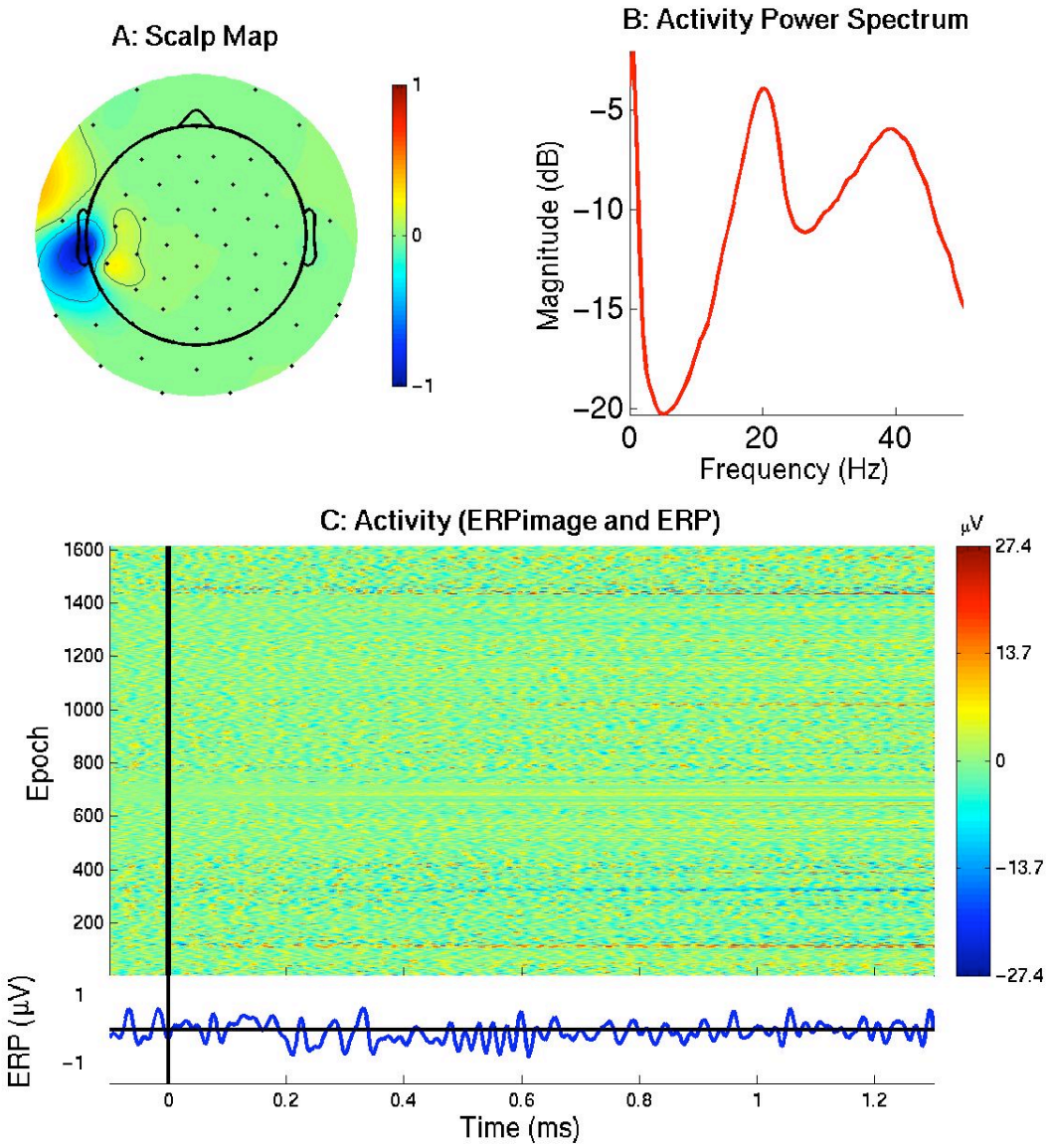

Figure IV.10 – Properties of an independent component (Participant 1, IC 12) that primarily represents muscle activity. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

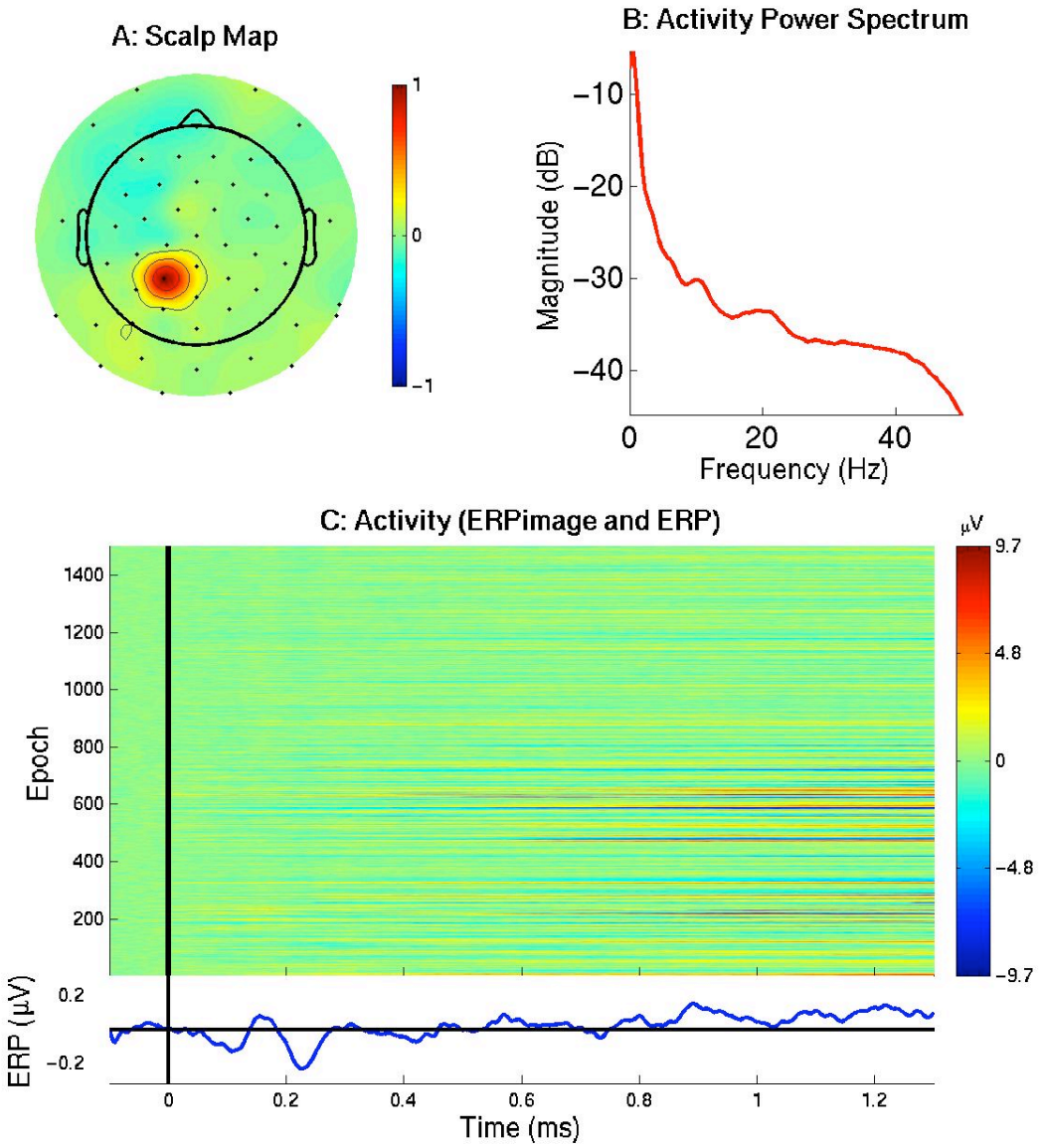

Figure IV.11 – Properties of an independent component (Participant 2, IC 63) that primarily represents a faulty electrode. (**A**) The IC's scalp projection measured in unitless weights, represented with color. The scalp projection is normalized such that the maximum absolute weight is of unit magnitude. (**B**) Power spectrum of component activity. (**C-top**) ERPimage of component activity. X-axis represents time and Y-axis represents epochs (1=the first epoch of the recording session). The magnitude of activity is represented with color (red=most positive, blue=most negative). Each epoch has been smoothed by averaging it with the immediately preceding and subsequent epochs. (**C-bottom**) Component event-related activation (i.e., the mean activity of all epochs).

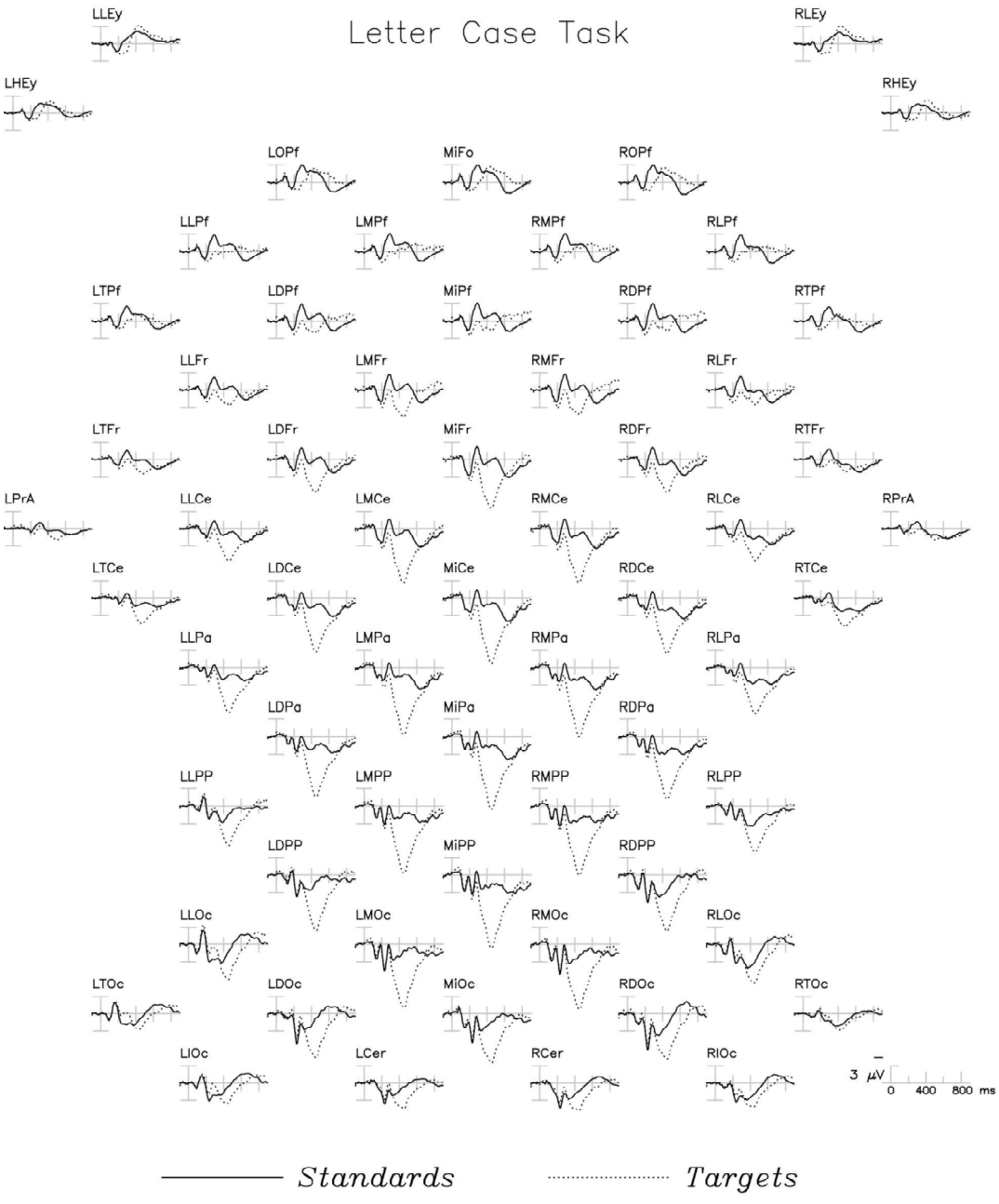

Figure IV.12 – Grand average ERPs from the letter case oddball task. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

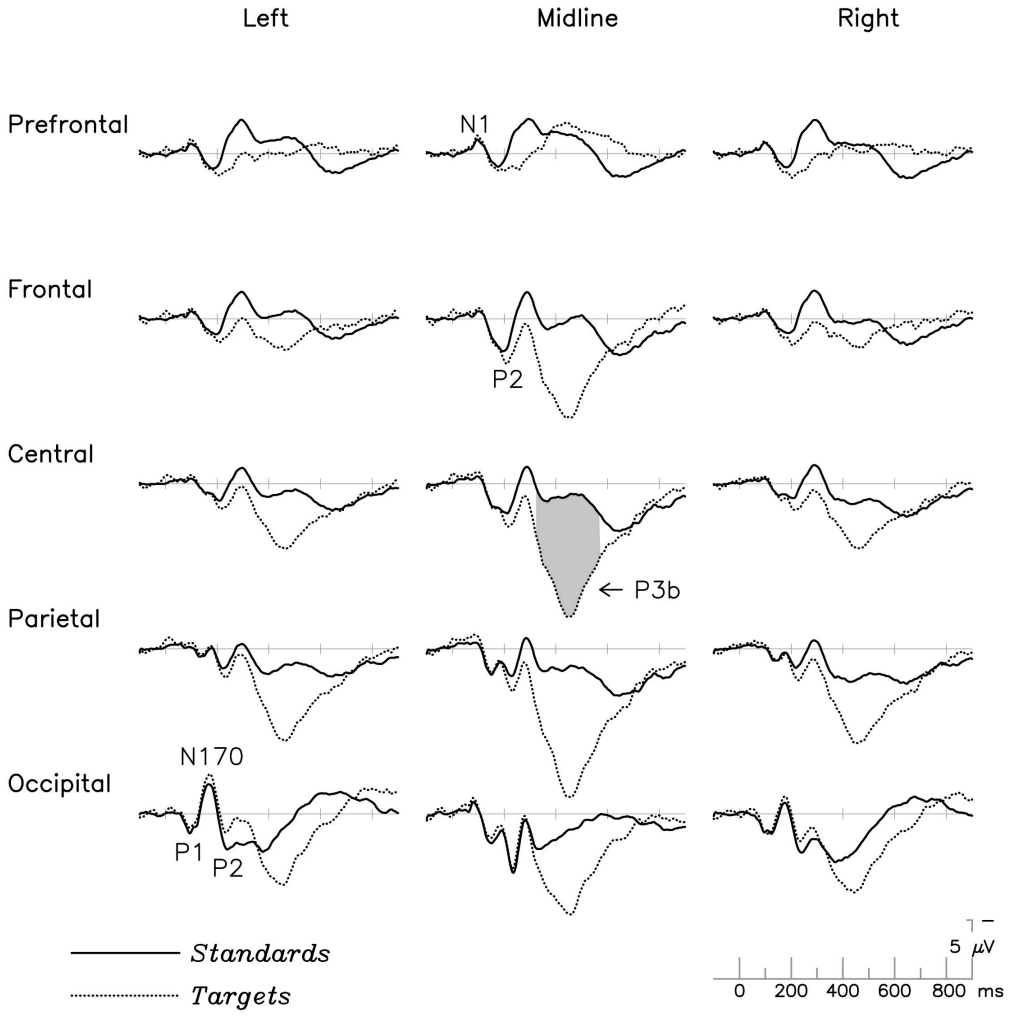

Figure IV.13 – Example grand average ERPs from the letter case oddball task showing anterior N1/P2, posterior P1/N170/P2, and P3b components. Shaded region indicates P3b window (325-575 ms).

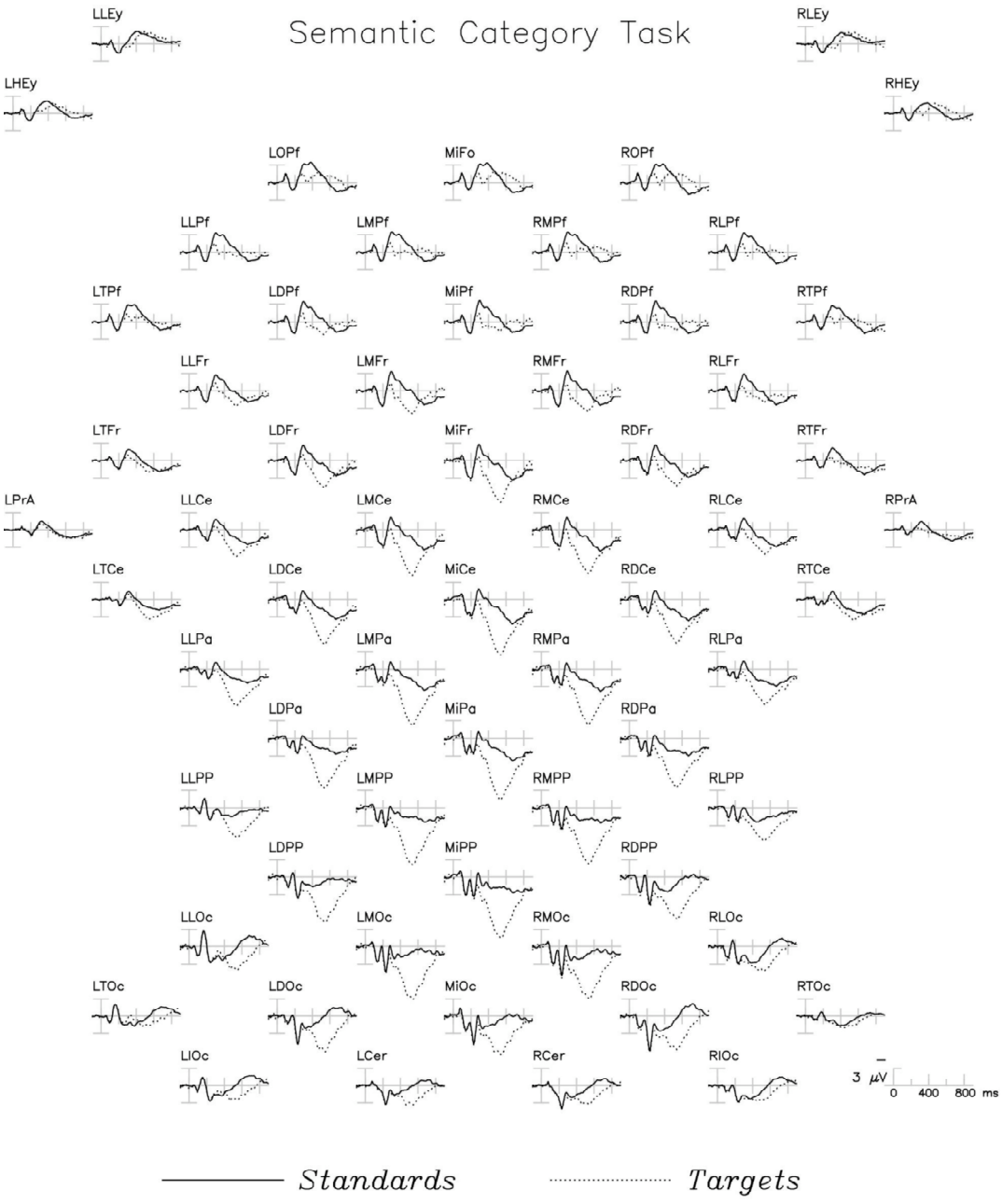

Figure IV.14 – Grand average ERPs from the semantic category oddball task. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

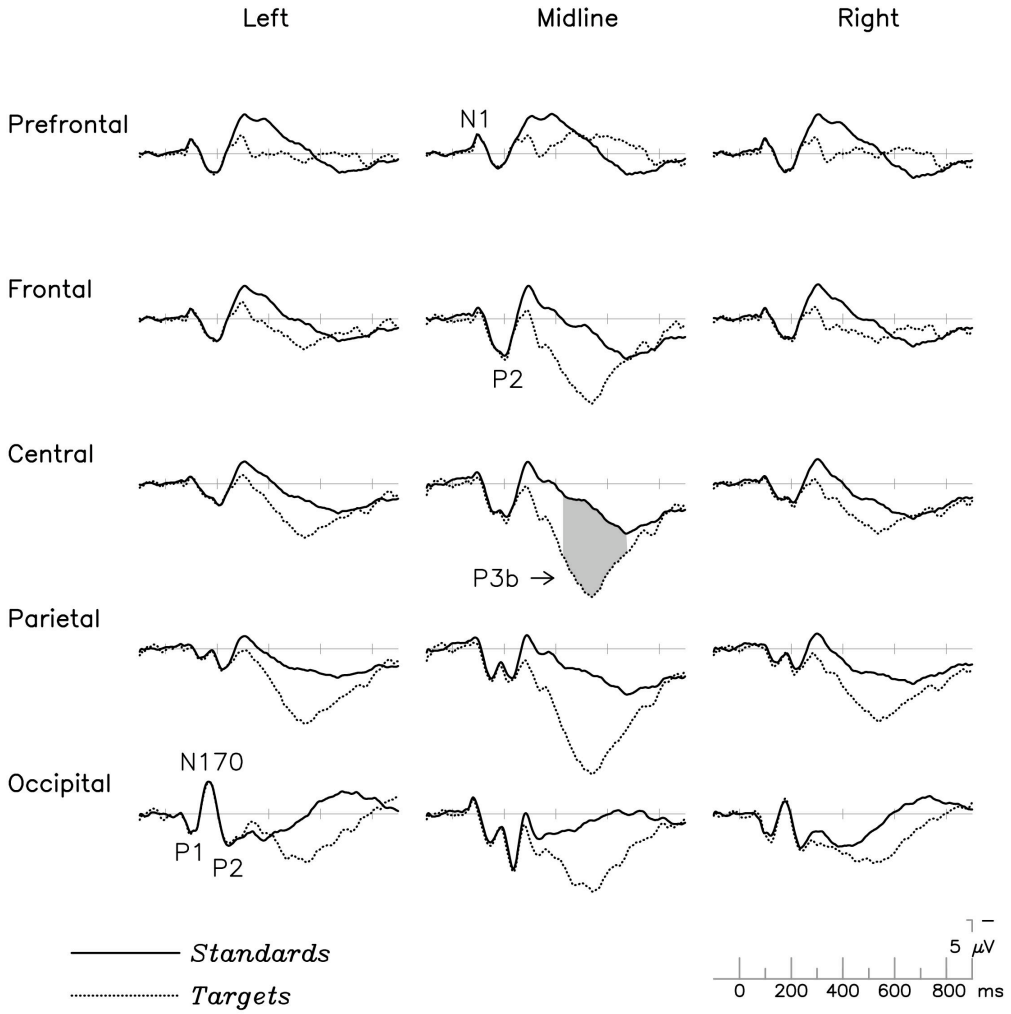

Figure IV.15 – Example grand average ERPs from the semantic category oddball task showing anterior N1/P2, posterior P1/N170/P2, and P3b components. Shaded region indicates P3b window (425-675 ms).

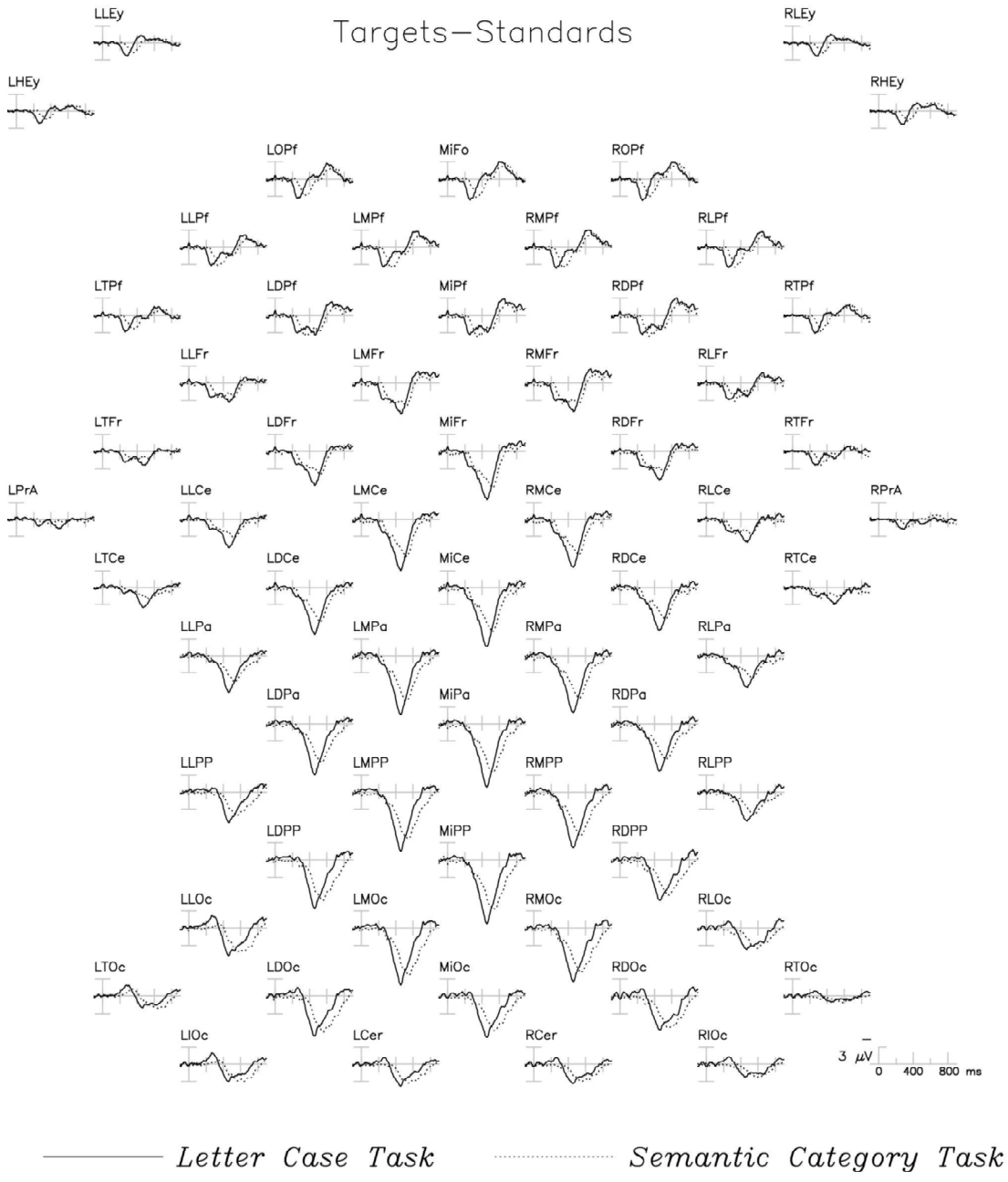

Figure IV.16 – Grand average difference waves from both oddball tasks. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

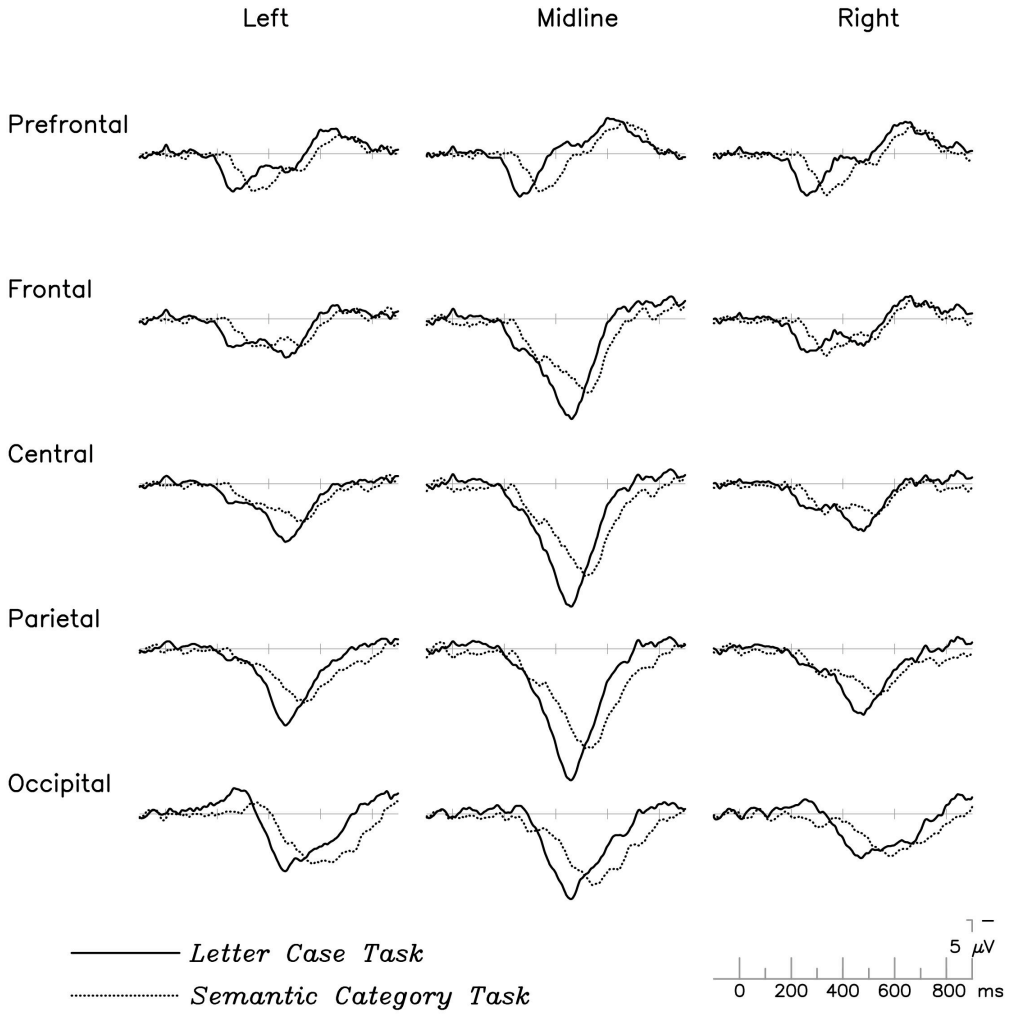

Figure IV.17 – Example grand average ERP difference waves (targets-standards) from both oddball tasks.

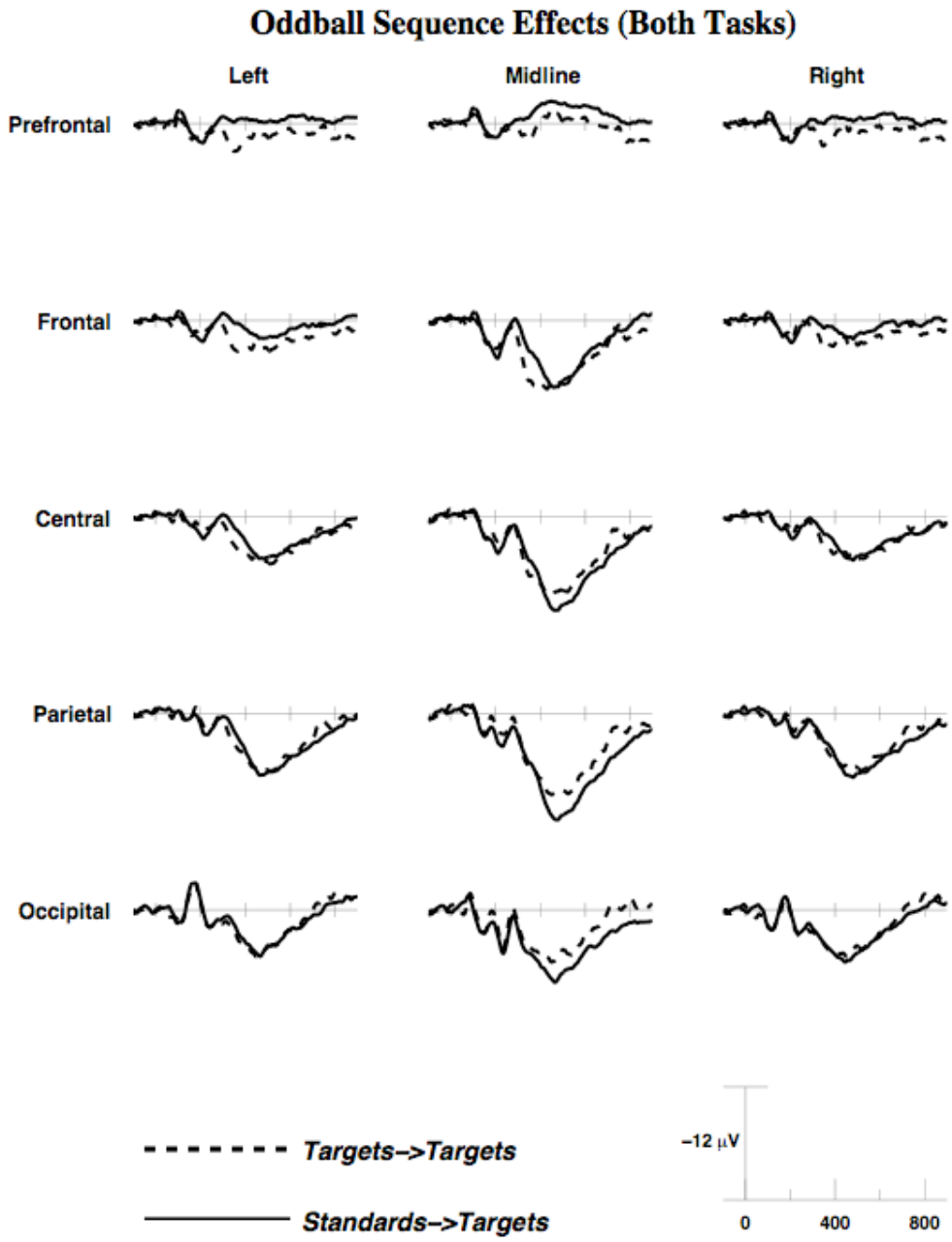

Figure IV.18 – Example grand average ERPs from both oddball tasks showing P3b sequence effect. The P3b is larger to targets preceded by standards (Standards->Targets) than to targets preceded by targets (Targets->Targets).

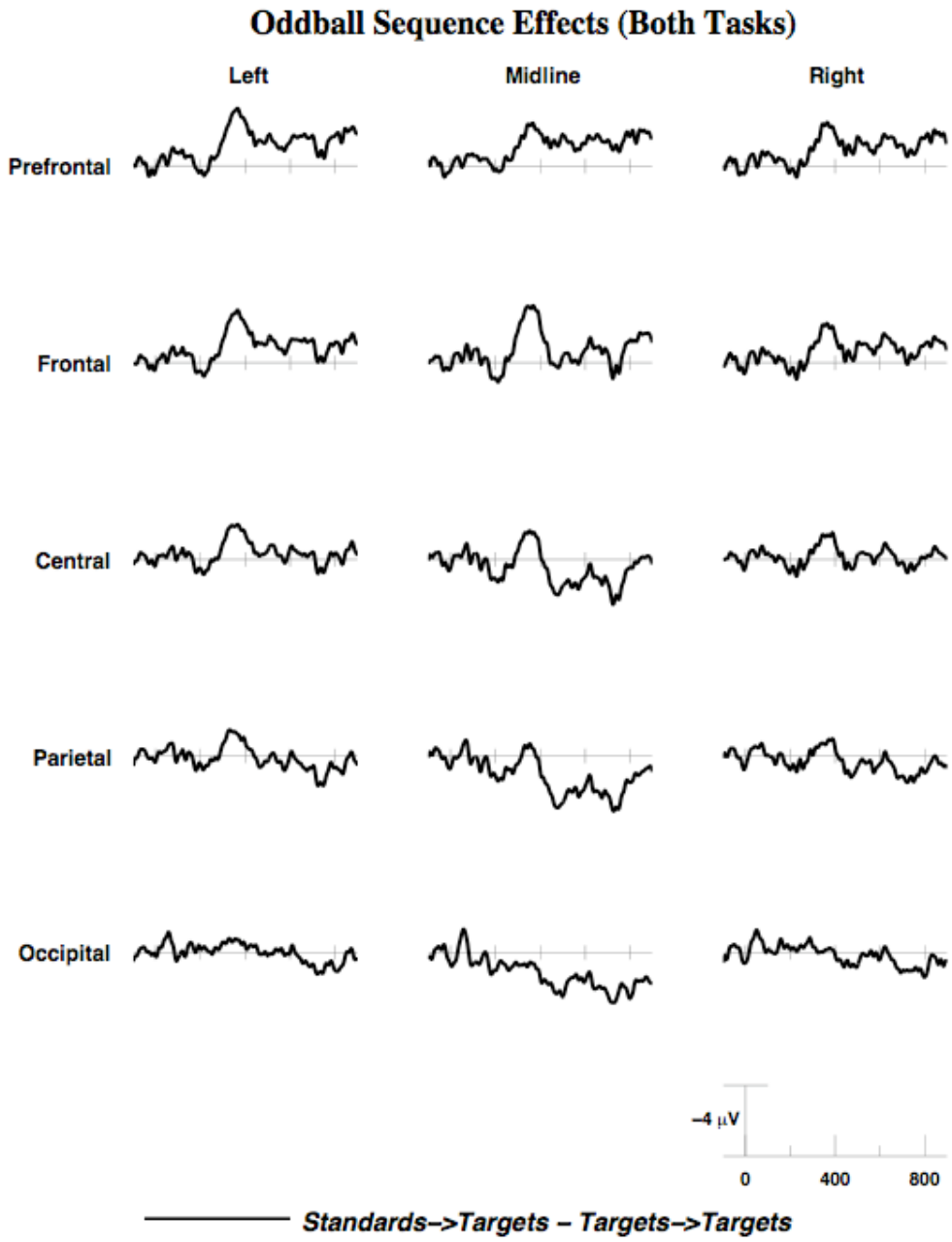

Figure IV.19 – Example grand average ERP difference waves (targets preceded by standards minus targets preceded by targets) from both oddball tasks.

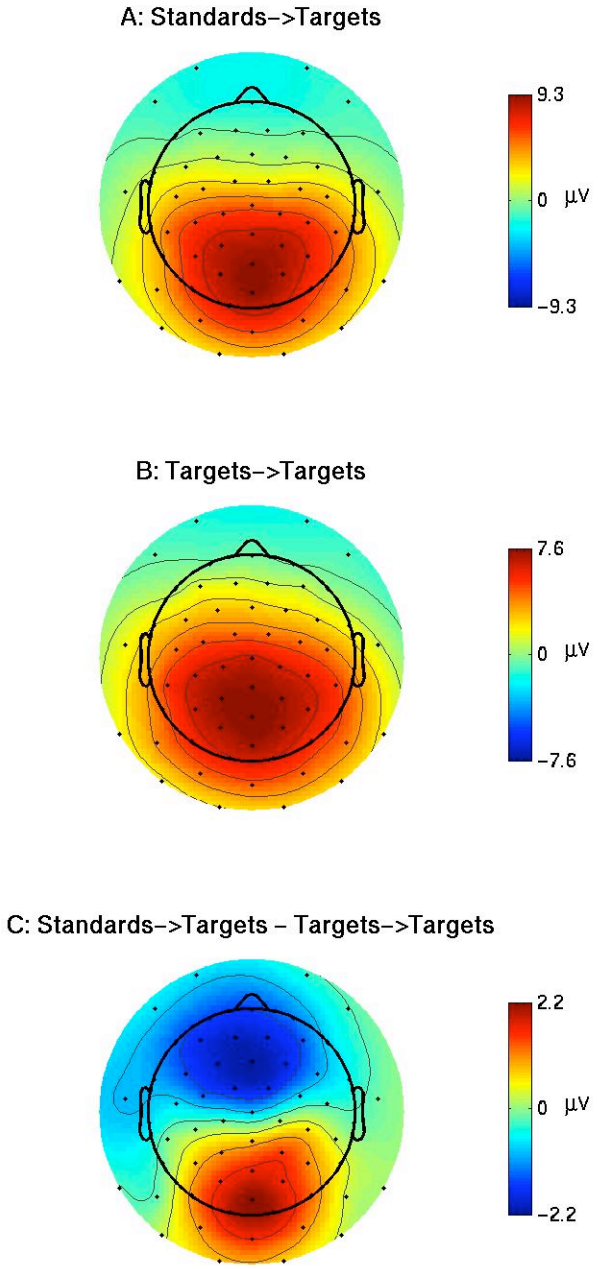

Figure IV.20 – Topographies of the grand average ERPs in Figures IV.17-18. Topographies represent the mean amplitude in the combined oddball task P3b window (350-600 ms). Electrode locations are represented by black dots. Electrodes below the head's equator are plotted progressively beyond the radius of the head (Note: Right mastoid electrode, A2, is not plotted).

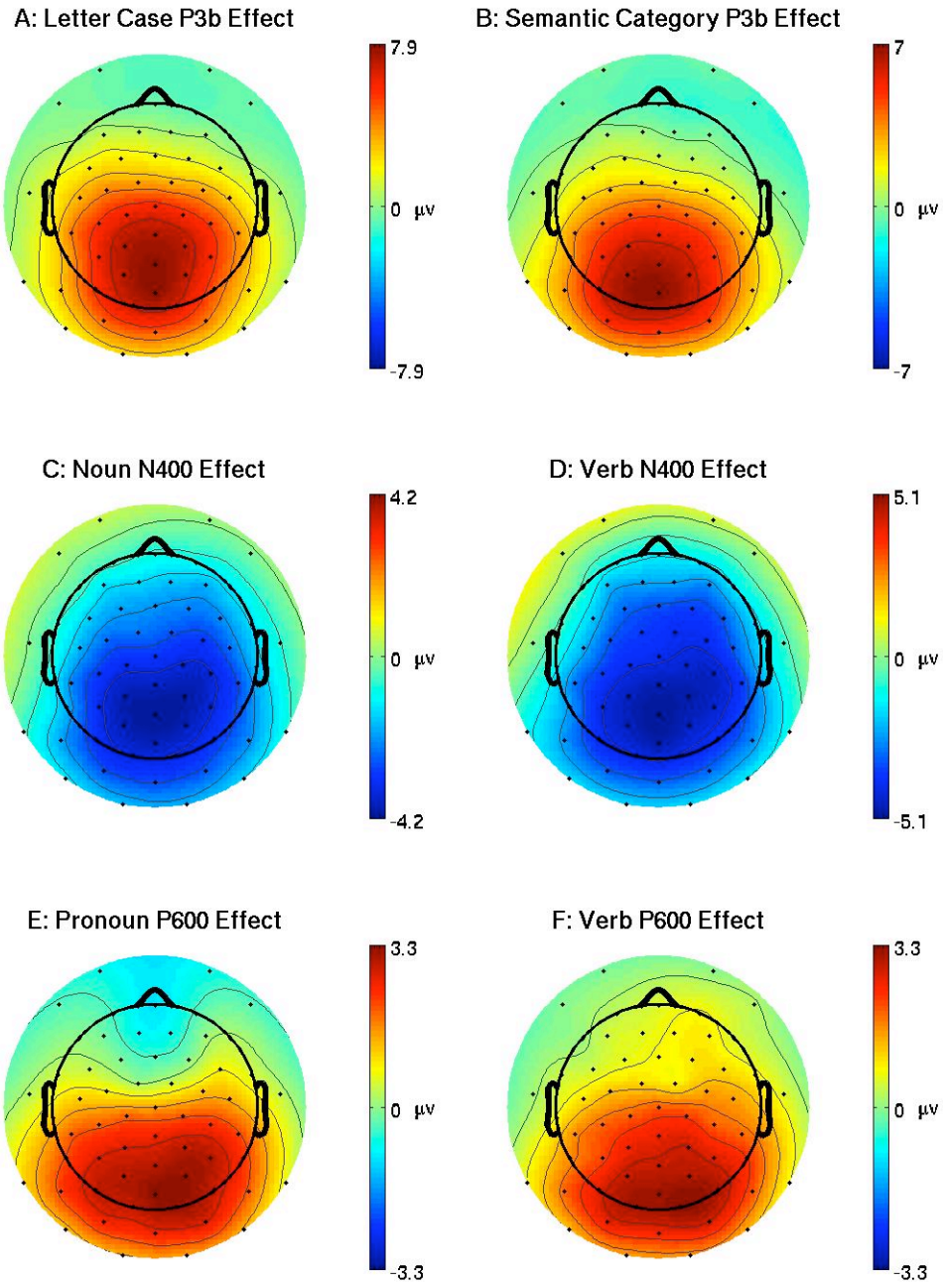

Figure IV.21 – Grand ERP effect topographies for the P3b, N400, and P600. Electrode locations are represented by black dots. Electrodes below the head's equator are plotted progressively beyond the radius of the head (Note: Right mastoid electrode, A2, is not plotted).

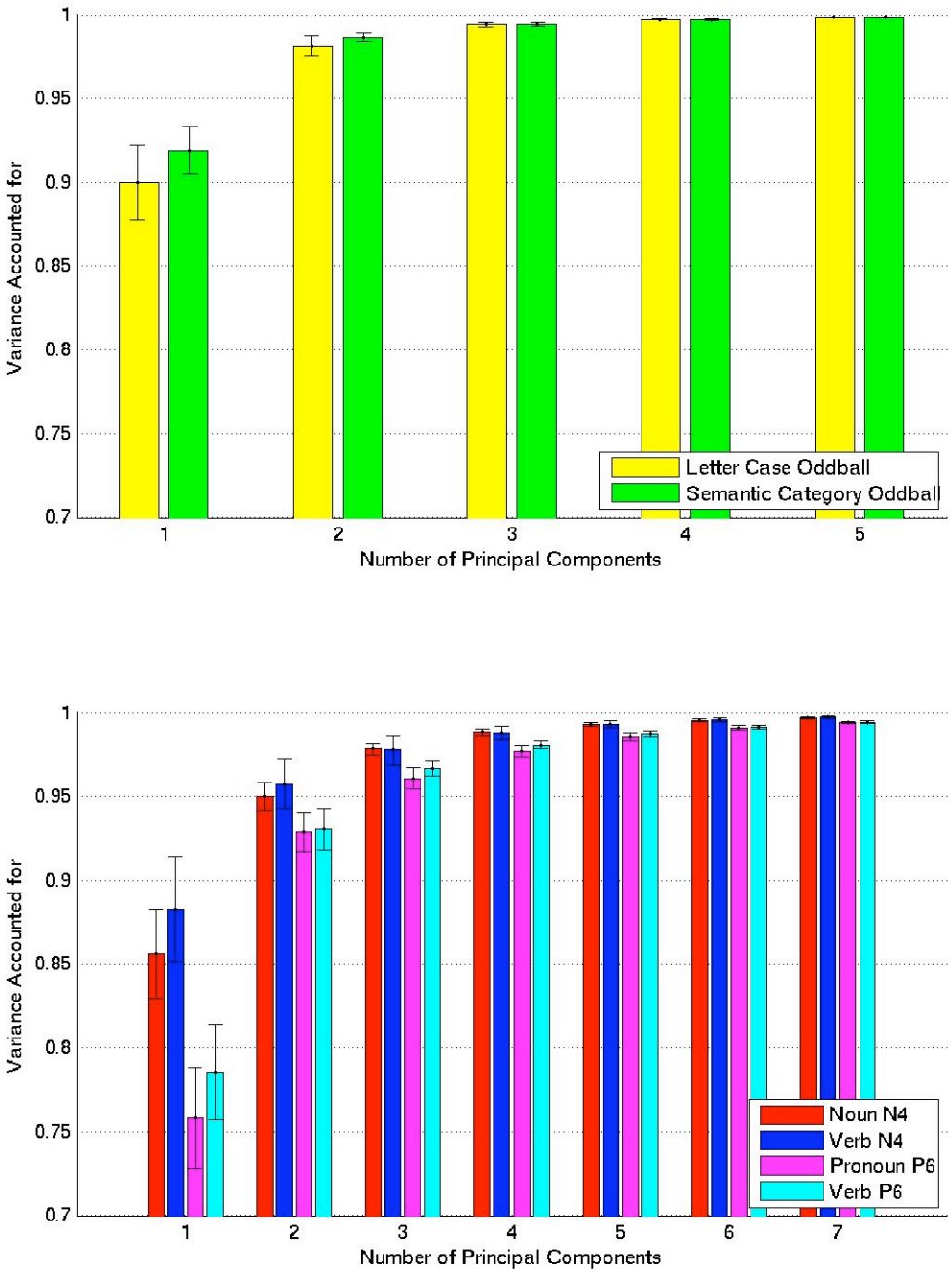

Figure IV.22- Mean cumulative percentage variance accounted for by each participant's largest five principal components of the P3b effects in the two oddball tasks **(Top)** and the largest seven principal components of the N400 and P600 effects in the sentence acceptability task **(Bottom)**. Error bars are standard error (n=16).

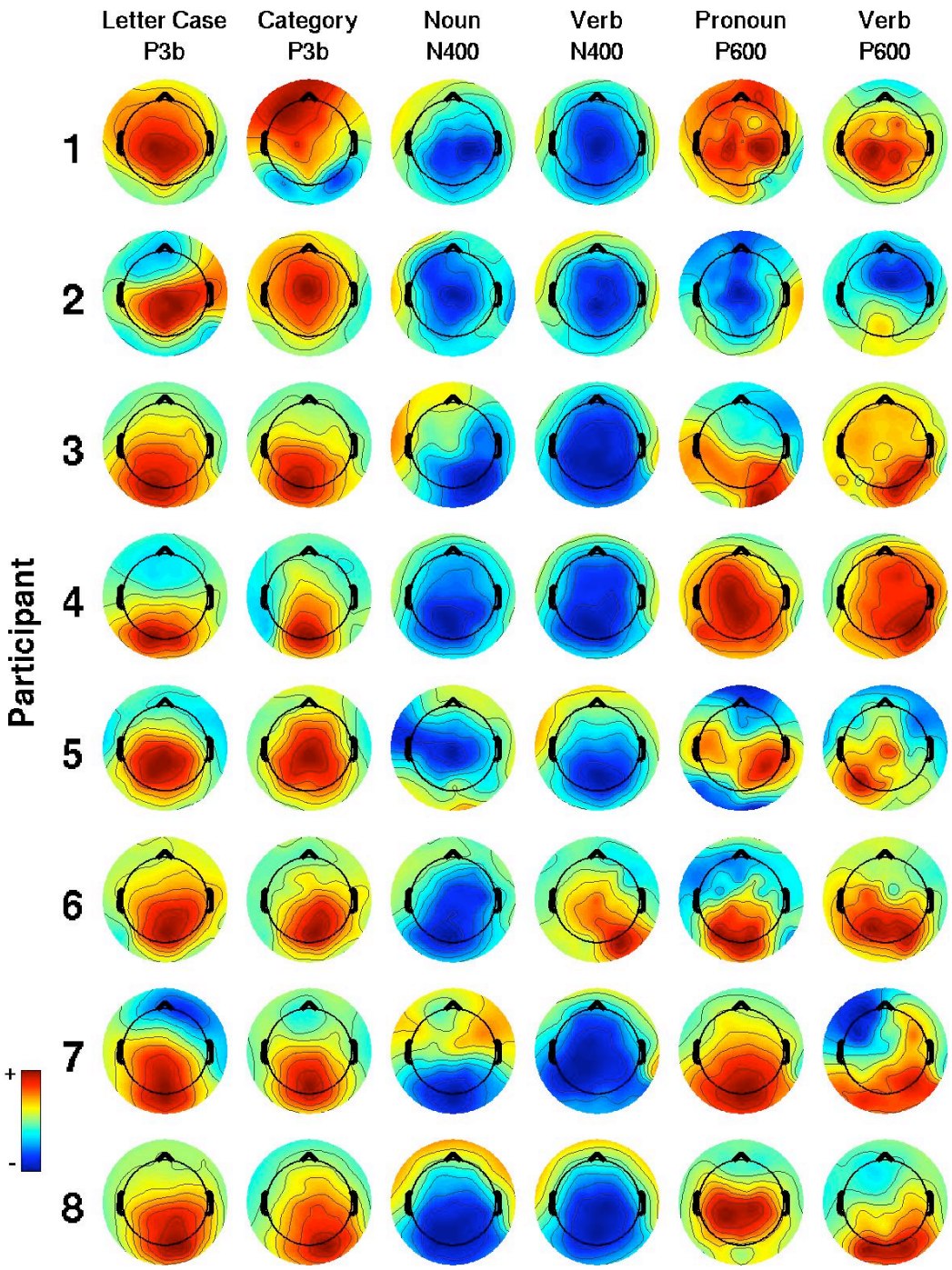

Figure IV.23- Mean P3b, N400, and P600 effect topographies of the first eight participants. Topographies are interpolated from difference wave amplitudes at the 63 non-mastoid electrodes (not shown). Voltage is represented with color and each topography is individually scaled to maximize color range. Coloring beyond the radius of the head reflect potentials recorded below the head's equator (see Figure IV.1: D).

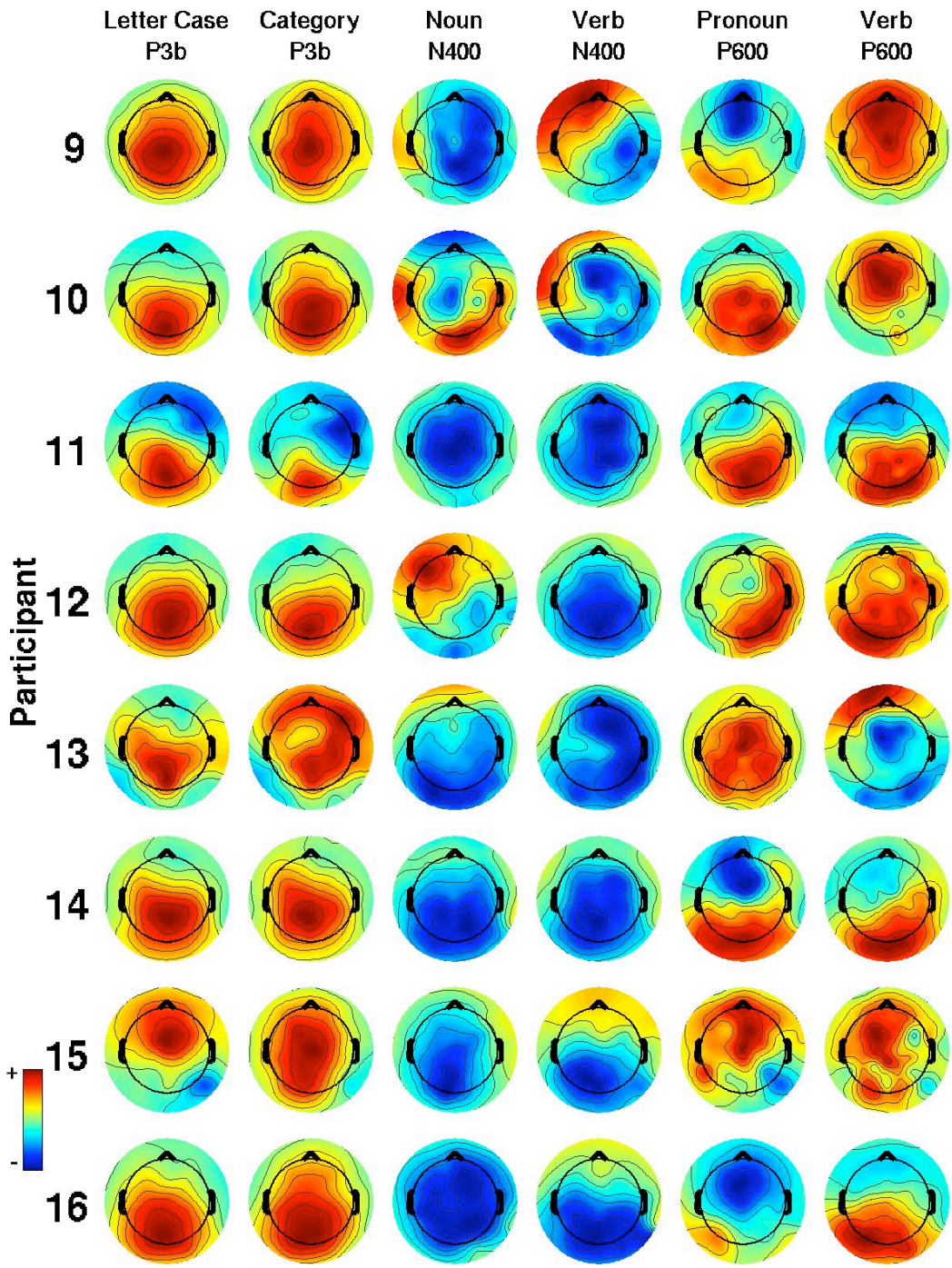

Figure IV.24- Mean P3b, N400, and P600 effect topographies of the latter eight participants. Topographies are interpolated from difference wave amplitudes at the 63 non-mastoid electrodes (not shown). Voltage is represented with color and each topography is individually scaled to maximize color range. Coloring beyond the radius of the head reflect potentials recorded below the head's equator (see Figure IV.1: D).

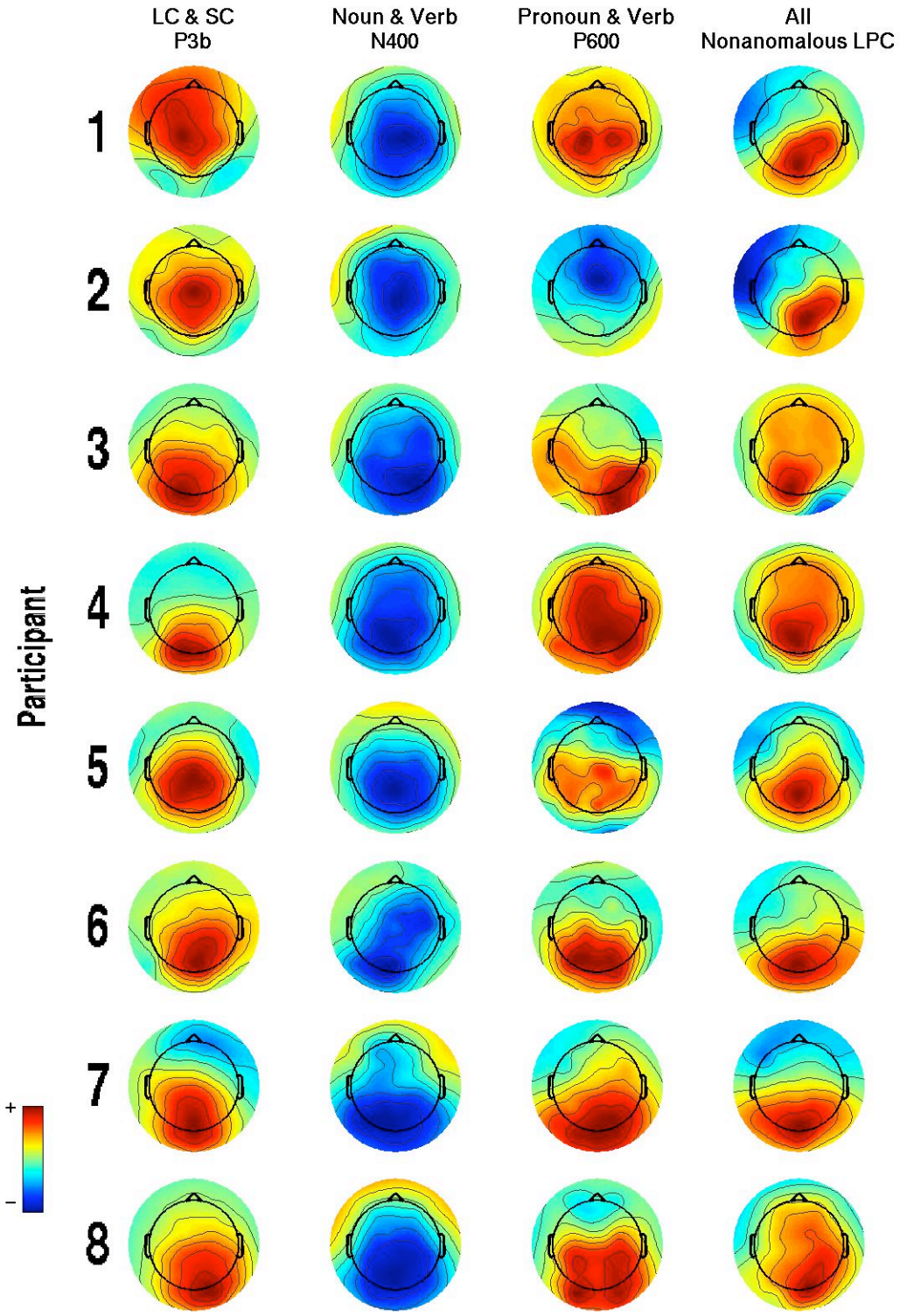

Figure IV.25- Mean P3b, N400, and P600 effect and LPC topographies of the first eight participants. See Figure IV.23 text for additional explanation. Abbreviations: LC=Letter Case, SC=Semantic Category.

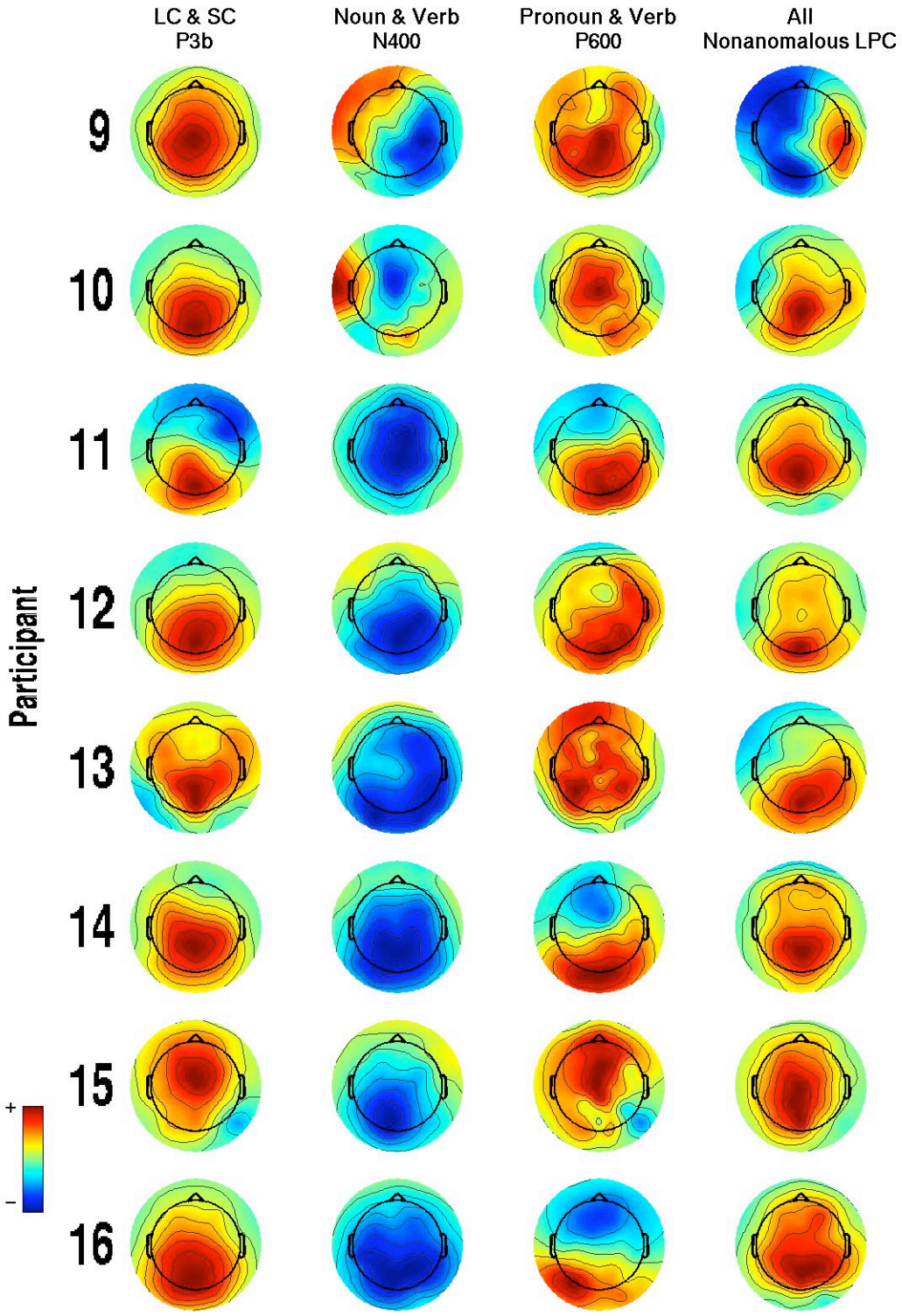

Figure IV.26- Mean P3b, N400, and P600 effect and LPC topographies of the latter eight participants. See Figure IV.23 text for additional explanation. Abbreviations: LC=Letter Case, SC=Semantic Category.

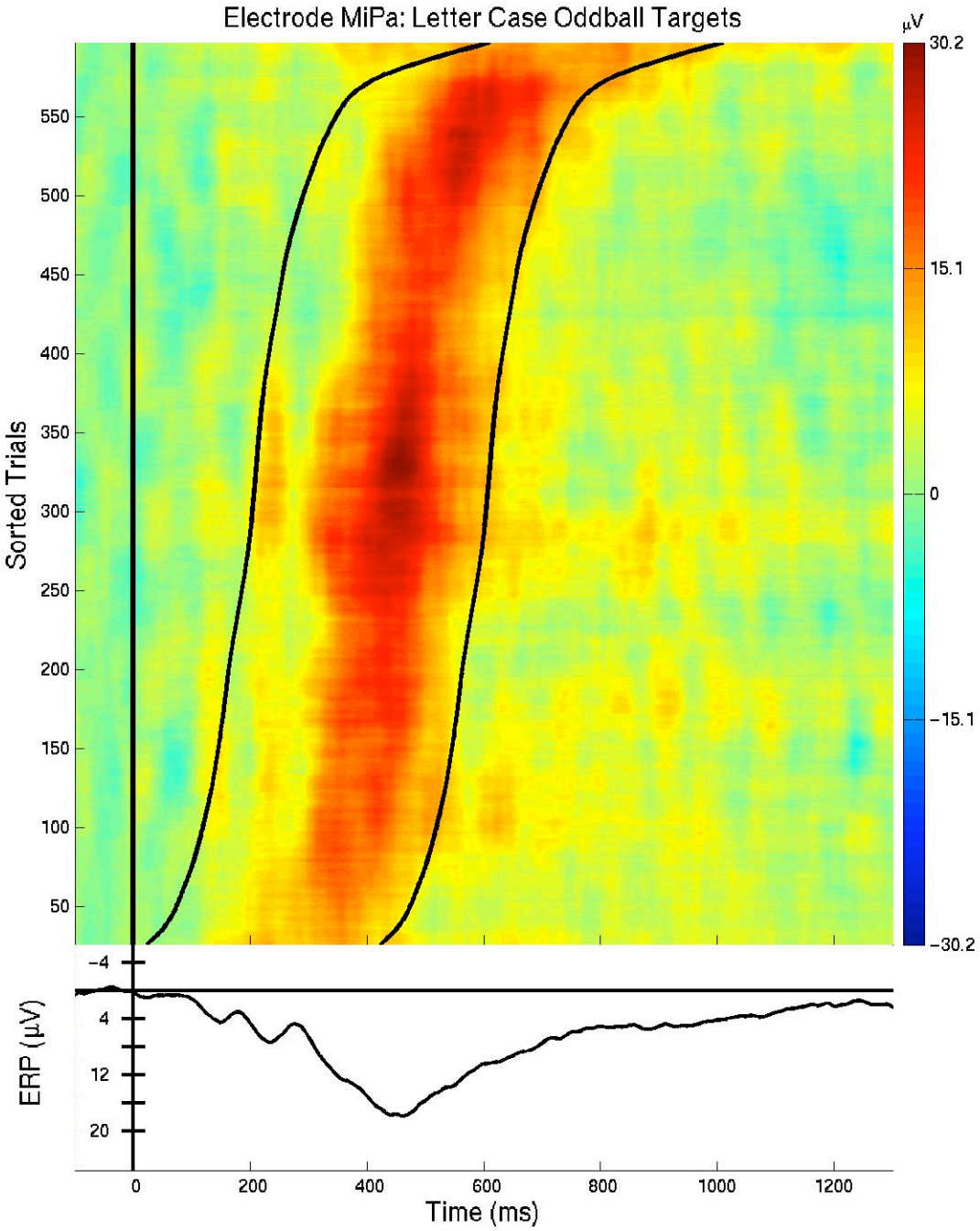

Figure IV.27 – (Top) Single trials of the EEG response to targets at electrode MiPa in the letter case oddball task. Trials are sorted by P3b latency as determined by a 400 ms long Woody filter. Sigmoidal black lines indicate the leading and trailing edges of the Woody filter window. Single trials are smoothed using a 50 trial vertical moving average. **(Bottom)** Average (ERP) of the single trials.

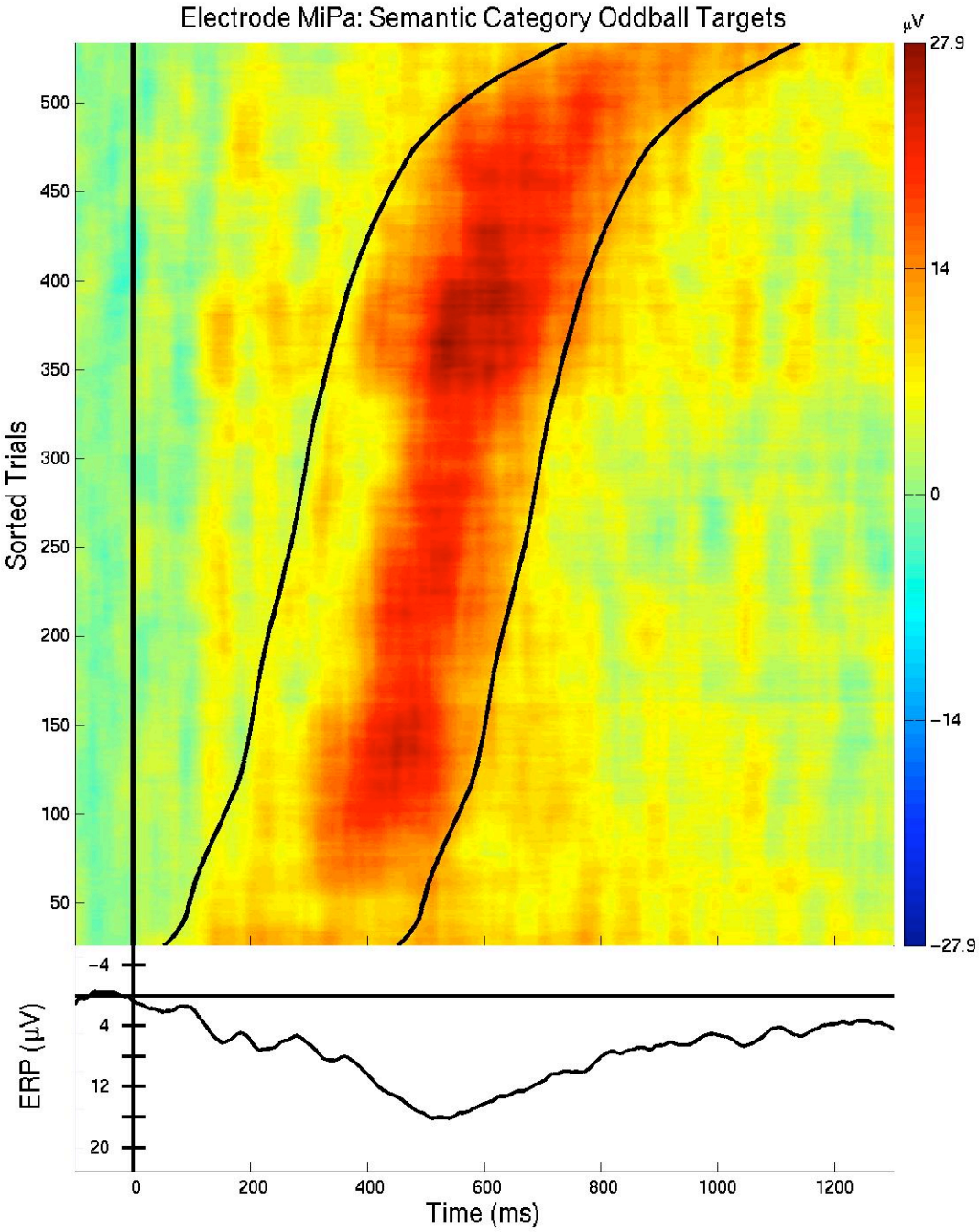

Figure IV.28 – (Top) Single trials of the EEG response to targets at electrode MiPa in the semantic category oddball task. Trials are sorted by P3b latency as determined by a 400 ms long Woody filter. Sigmoidal black lines indicate the leading and trailing edges of the Woody filter window. Single trials are smoothed using a 50 trial vertical moving average. **(Bottom)** Average (ERP) of the single trials.

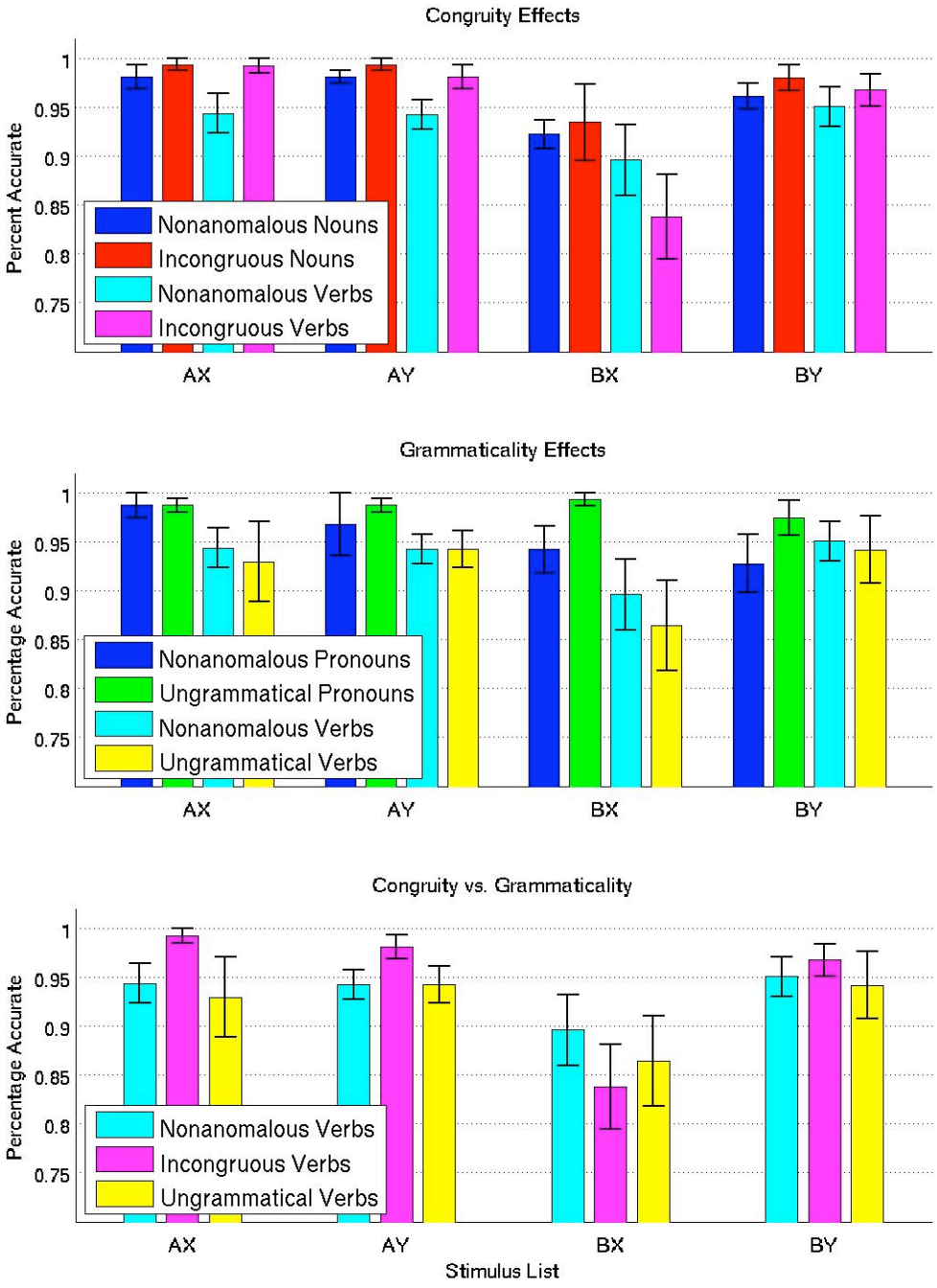

Figure IV.29 – Mean accuracy and standard error bars for incongruous and comparable nonanomalous endings **(Top)**, ungrammatical and comparable nonanomalous endings **(Middle),** and verb endings **(Bottom)**.

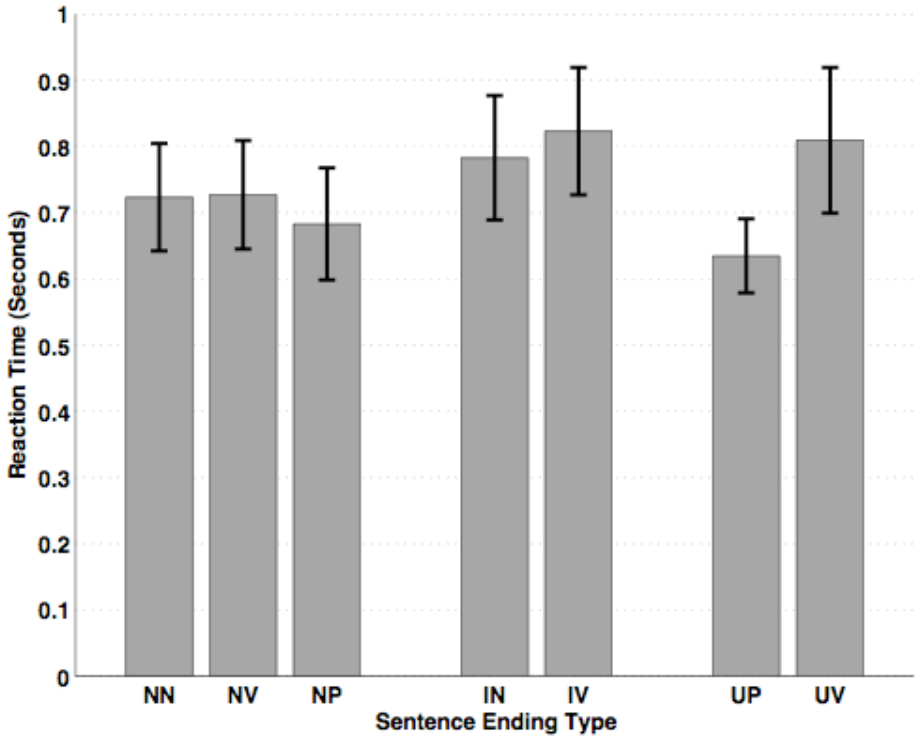

Figure IV.30 – Mean (across participants) of individual participant median reaction times in the sentence acceptability task. Reaction time is measured relative to the cue to respond, which occurs 1.4 seconds after the occurrence of the final word in each sentence. Error bars are standard error (samples=16). Abbreviations: NN=Nonanomalous Noun, NV=Nonanomalous Verb, NP=Nonanomalous Pronoun, IN=Incongruous Noun, IV=Incongruous Verb, UP=Ungrammatical Pronoun, UV=Ungrammatical Verb.

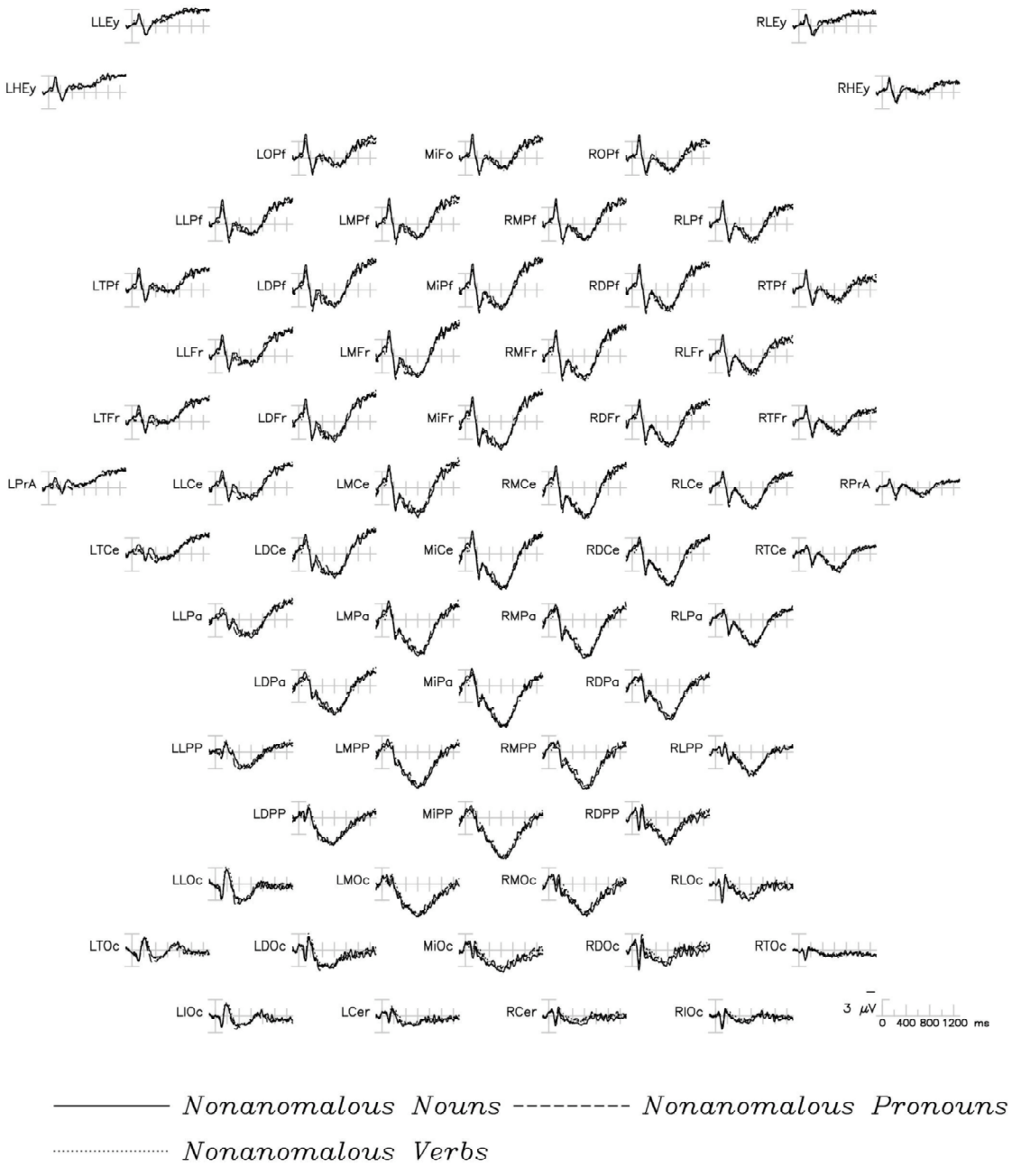

Figure IV.31 – Grand average ERPs to nonanomalous sentence final words. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

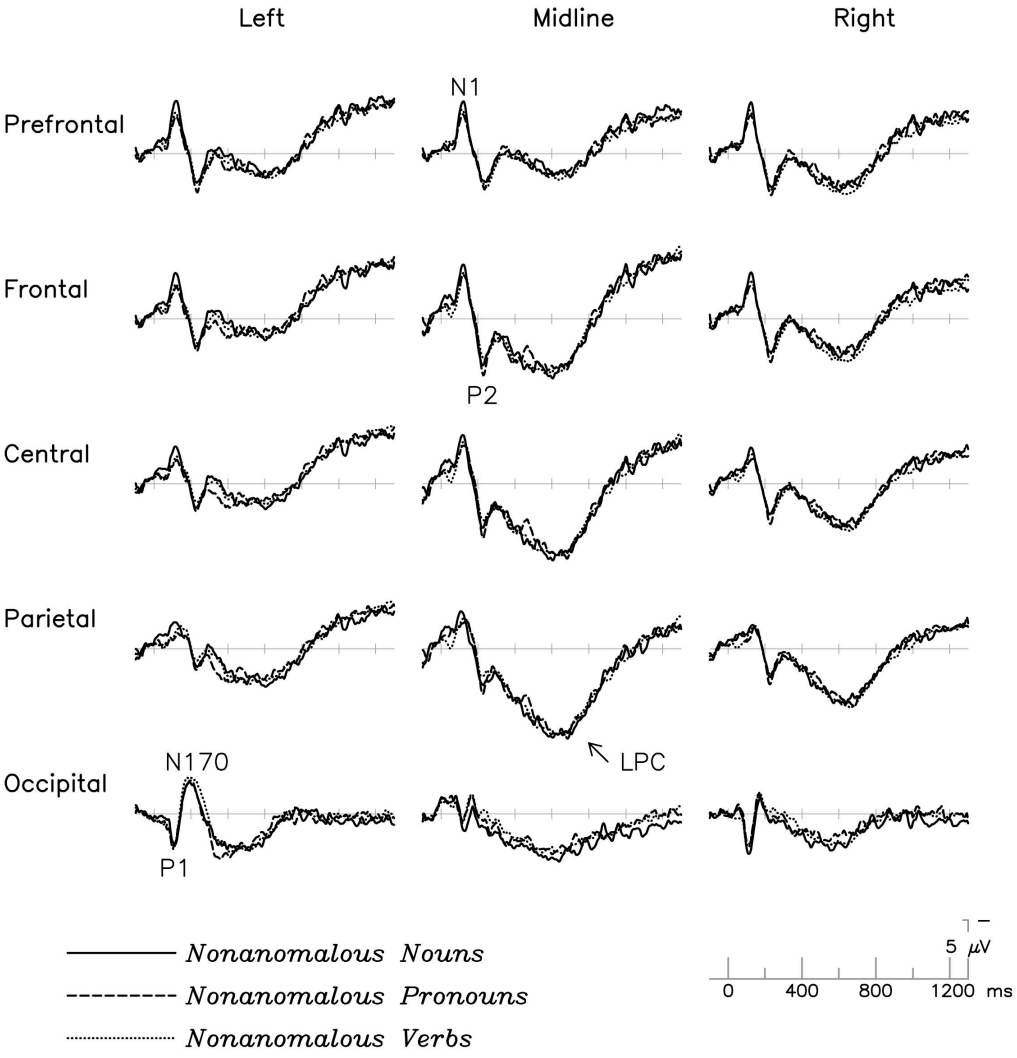

Figure IV.32 – Example grand average ERPs to nonanomalous sentence final words showing anterior N1/P2, posterior P1/N170, and LPC components.

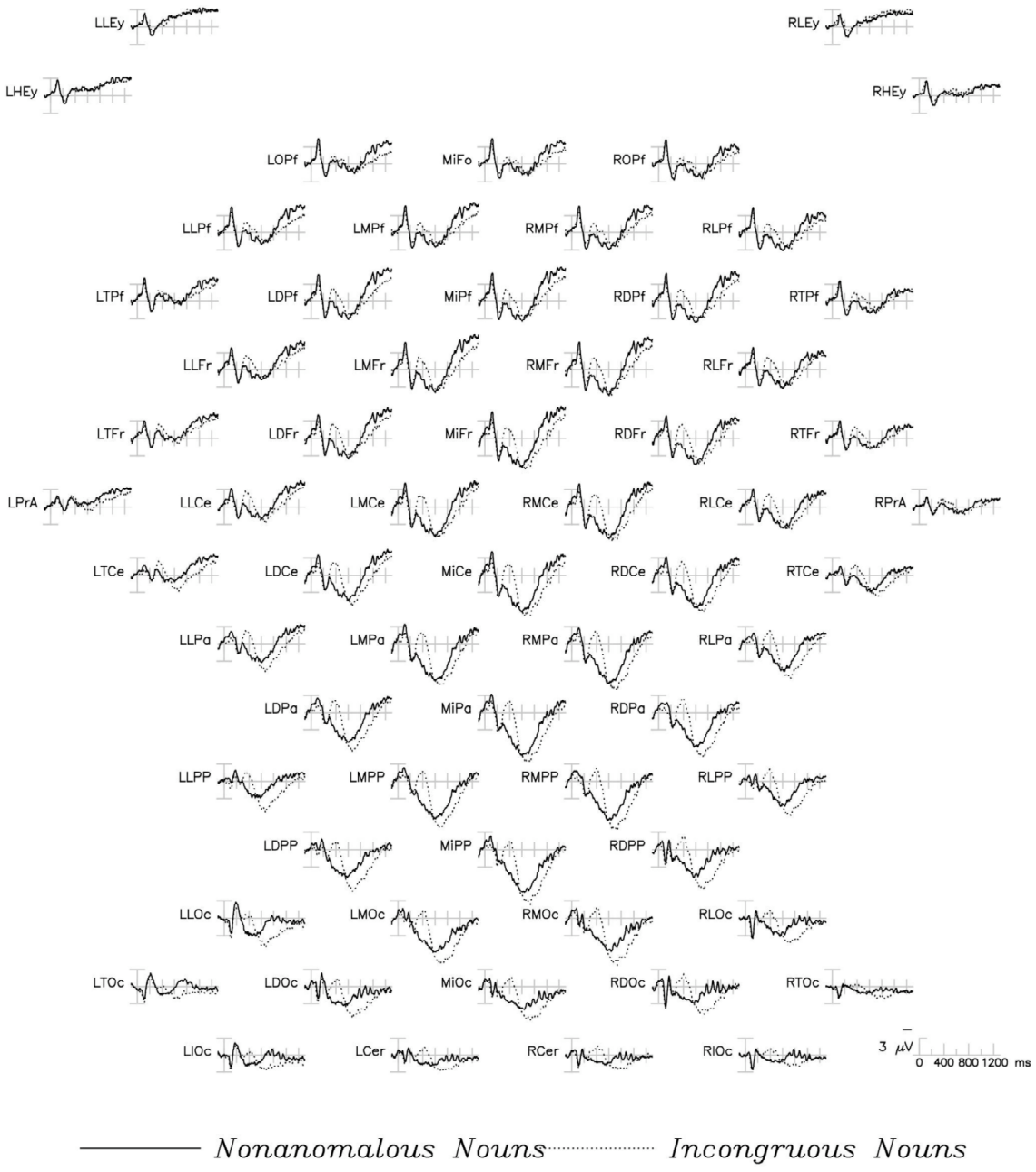

Figure IV.33 – Grand average ERPs to sentence final nouns. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

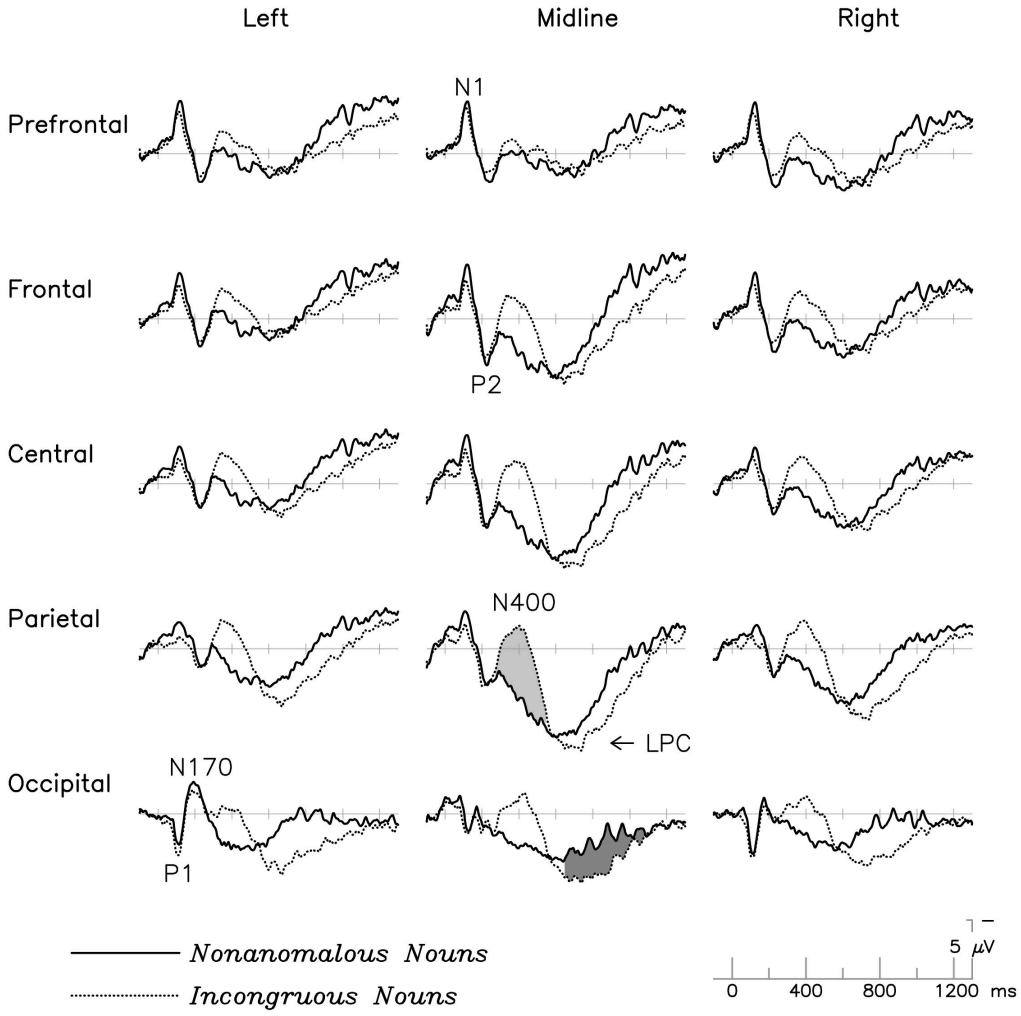

Figure IV.34 – Example grand average ERPs to sentence final nouns showing anterior N1/P2, posterior P1/N170, N400, and LPC components. Midline parietal shaded region indicates N400 window (275-550 ms). Midline occipital shaded region indicates P600 window (650-1100 ms).

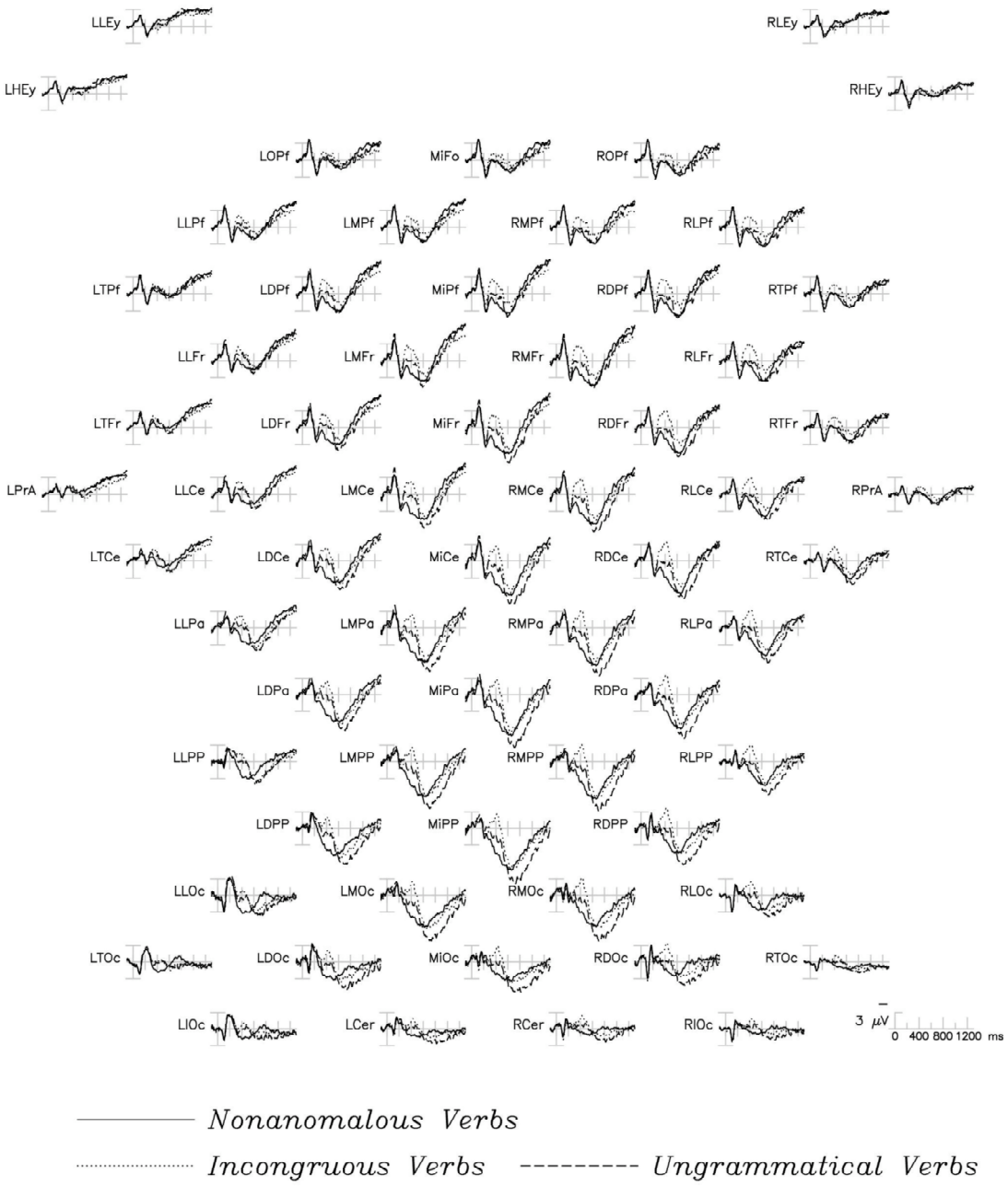

Figure IV.35 – Grand average ERPs to sentence final verbs. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

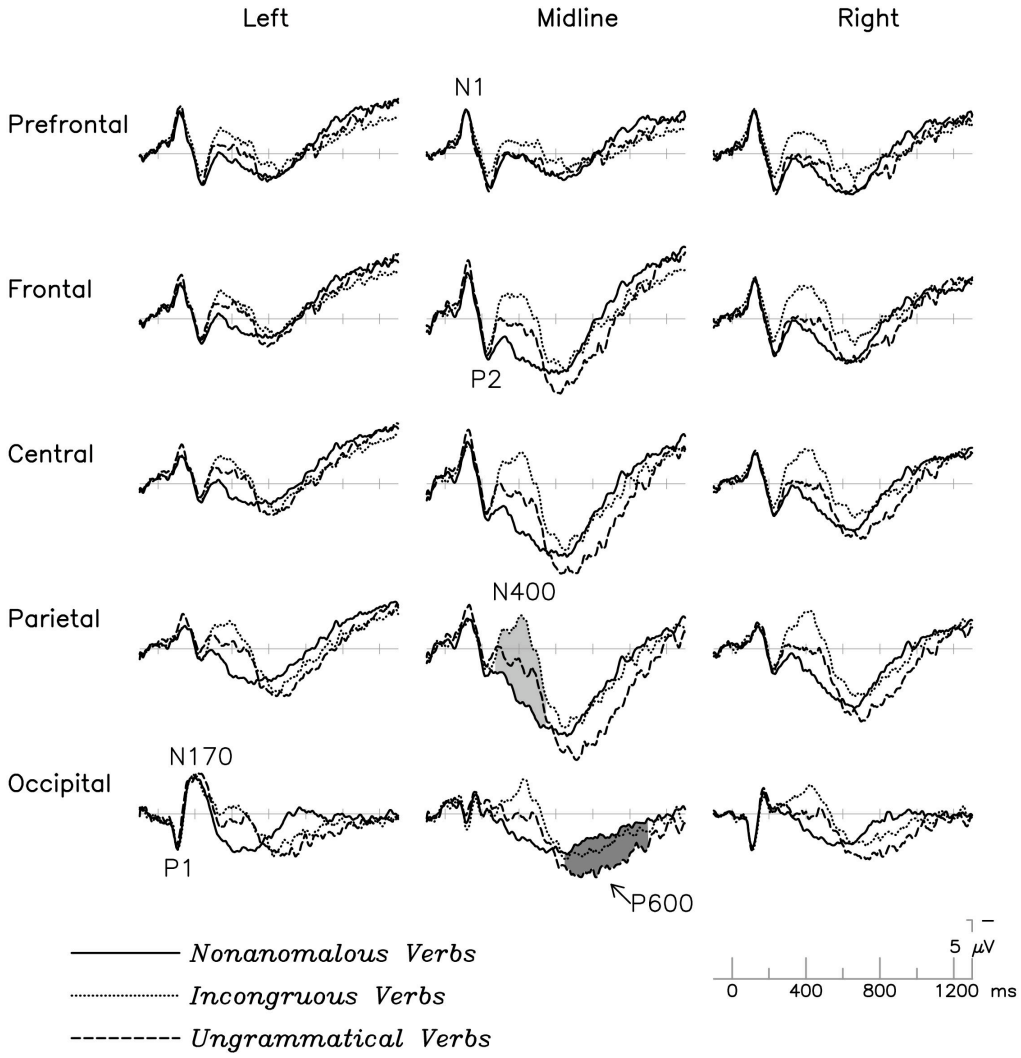

Figure IV.36 – Example grand average ERPs to sentence final verbs showing anterior N1/P2, posterior P1/N170, and N400 components, as well as the P600 effect. Midline parietal shaded region indicates N400 window (275-550 ms). Midline occipital shaded region indicates verb P600 window (650-1100) ms).

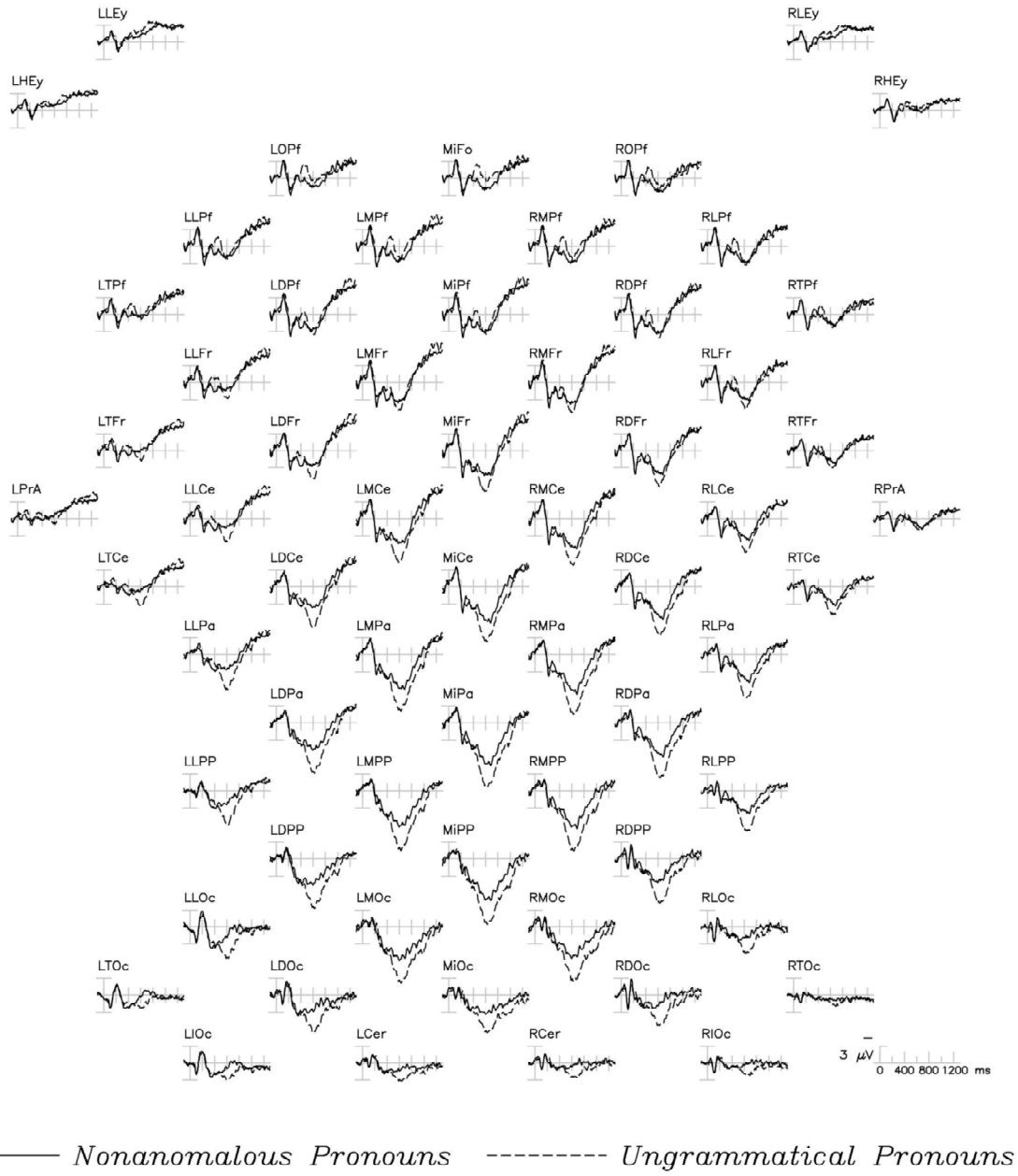

Figure IV.37 – Grand average ERPs to sentence final pronouns. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

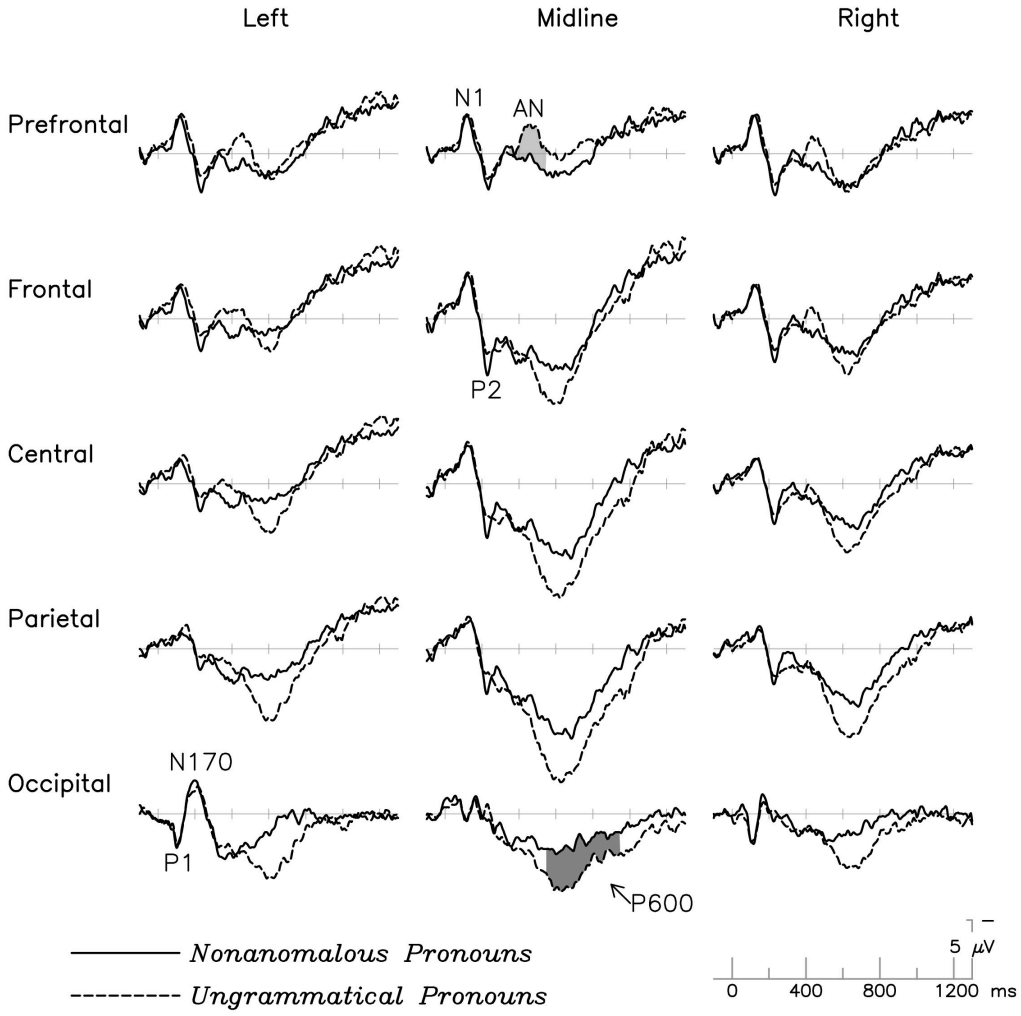

Figure IV.38 – Example grand average ERPs to sentence final pronouns showing anterior N1/P2, posterior P1/N170, and AN components, as well as the P600 effect. Midline prefrontal shaded region indicates pronoun AN window (350-550 ms). Midline occipital shaded region indicates pronoun P600 window (550-950 ms).

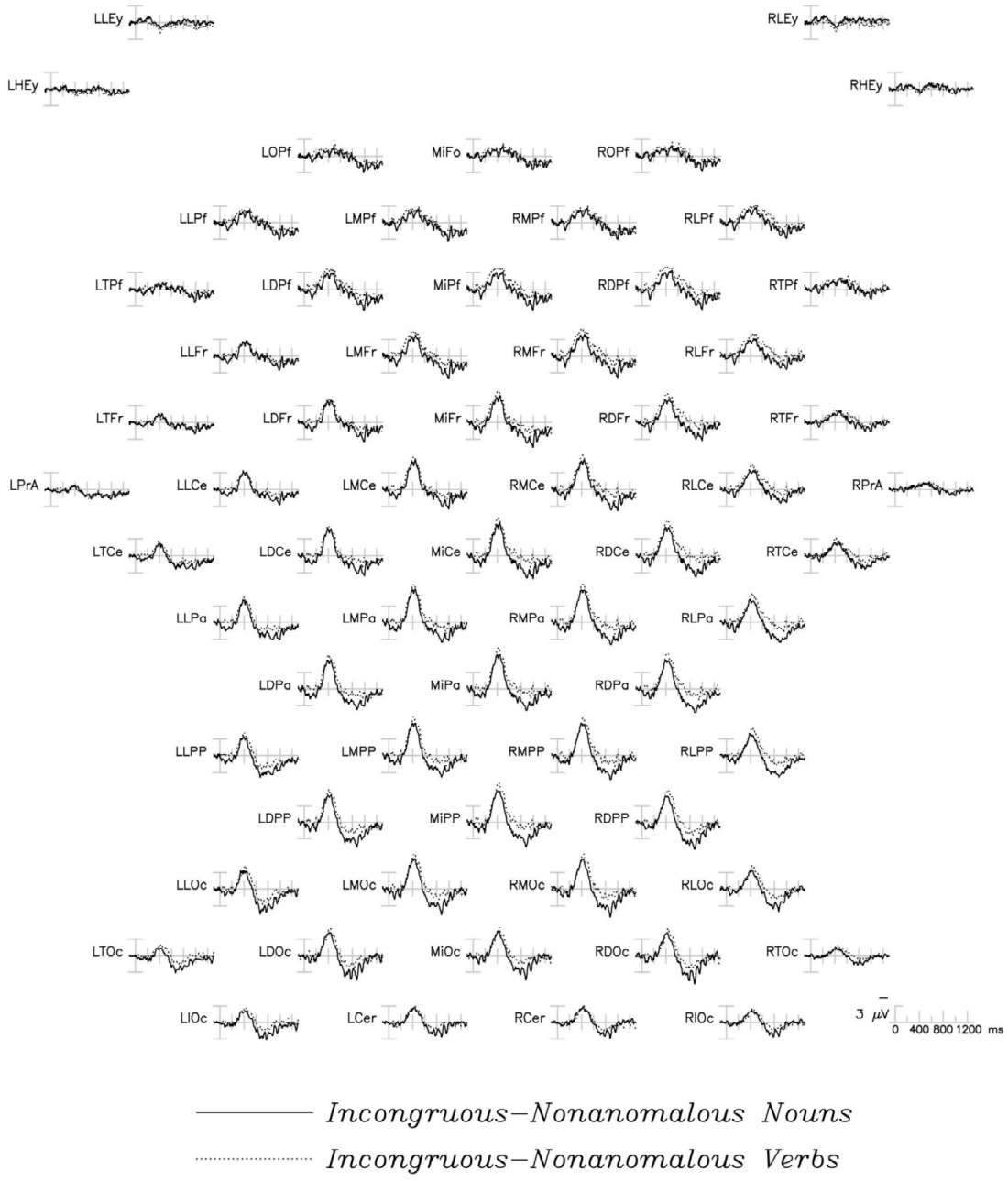

Figure IV.39 – Grand average difference waves of semantic congruity effects (ERPs to incongruous words-ERPs to nonanomalous words). Difference waves are topographically organized (anterior=towards top of page, right=towards right of page).

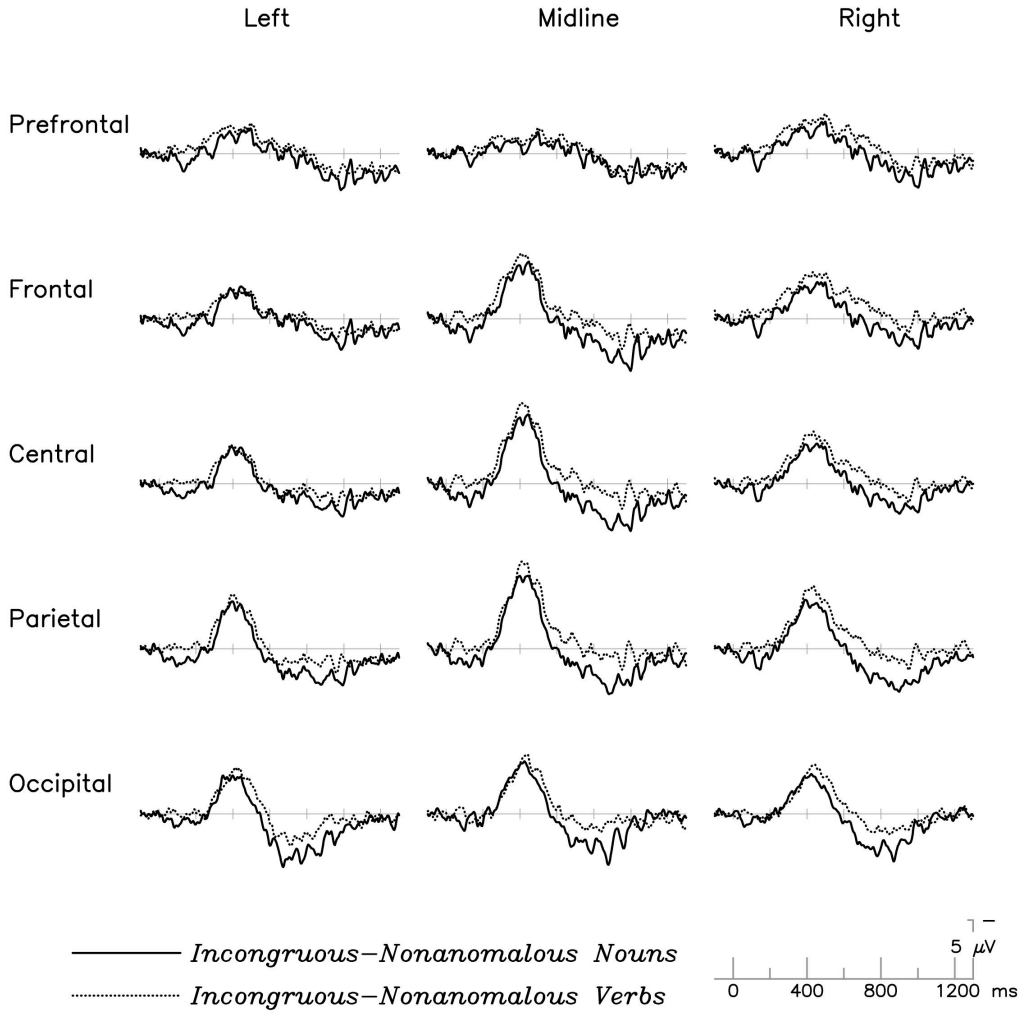

Figure IV.40 – Example grand average difference waves of semantic congruity effects (ERPs to incongruous words-ERPs to nonanomalous words).

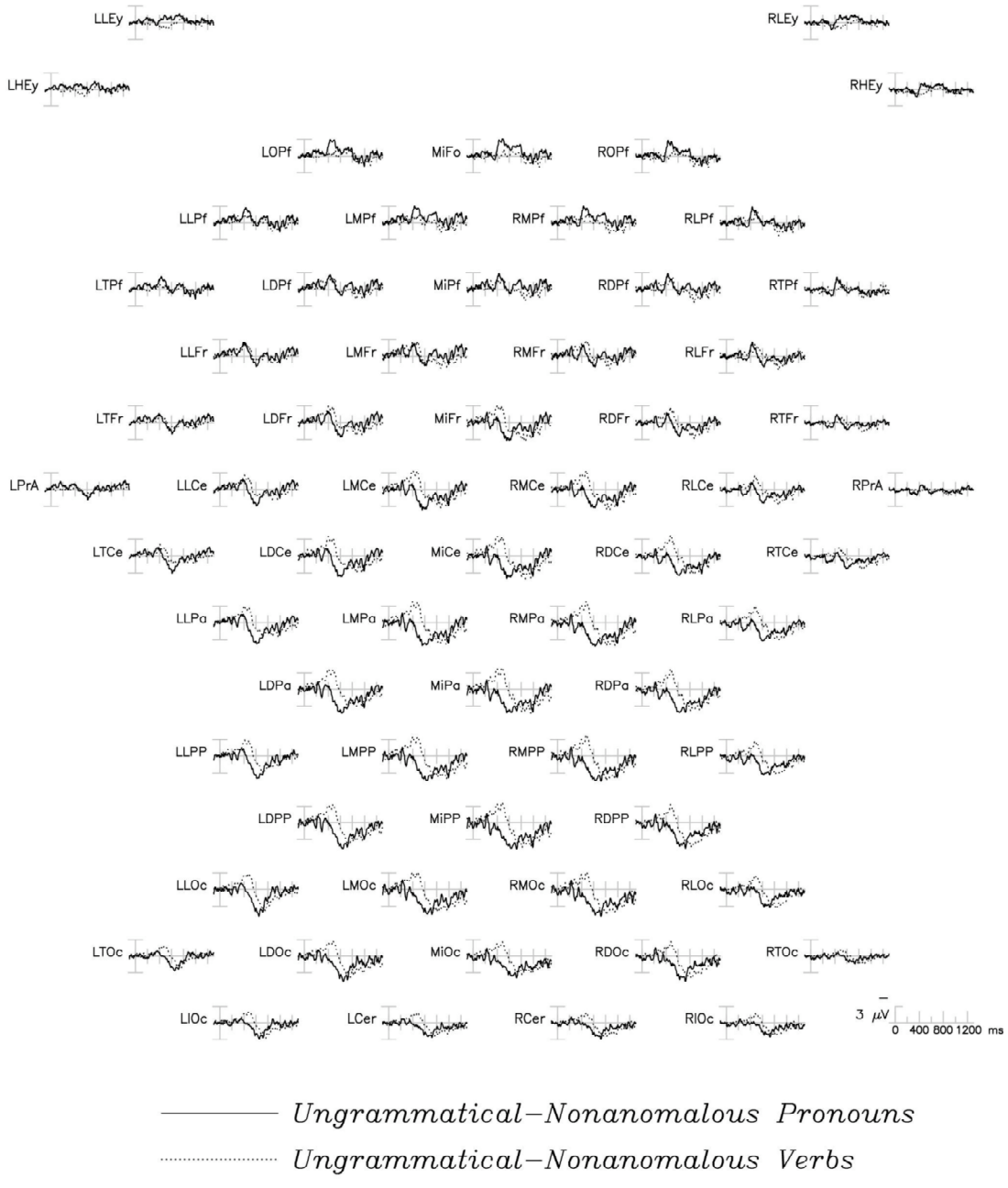

Figure IV.41 – Grand average difference waves of grammaticality effects (ERPs to ungrammatical words-ERPs to nonanomalous words). Difference waves are topographically organized (anterior=towards top of page, right=towards right of page).

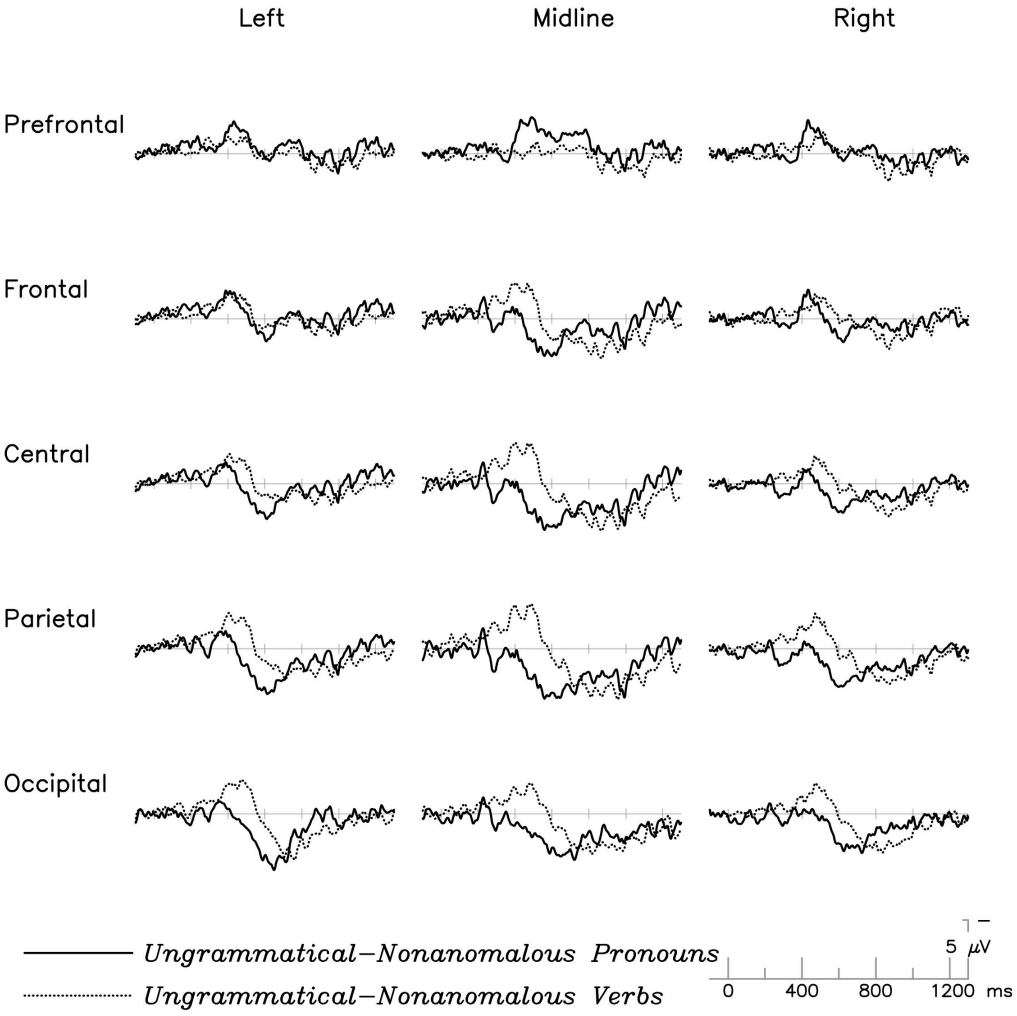

Figure IV.42 – Example grand average difference waves of grammaticality effects (ERPs to ungrammatical words-ERPs to nonanomalous words).

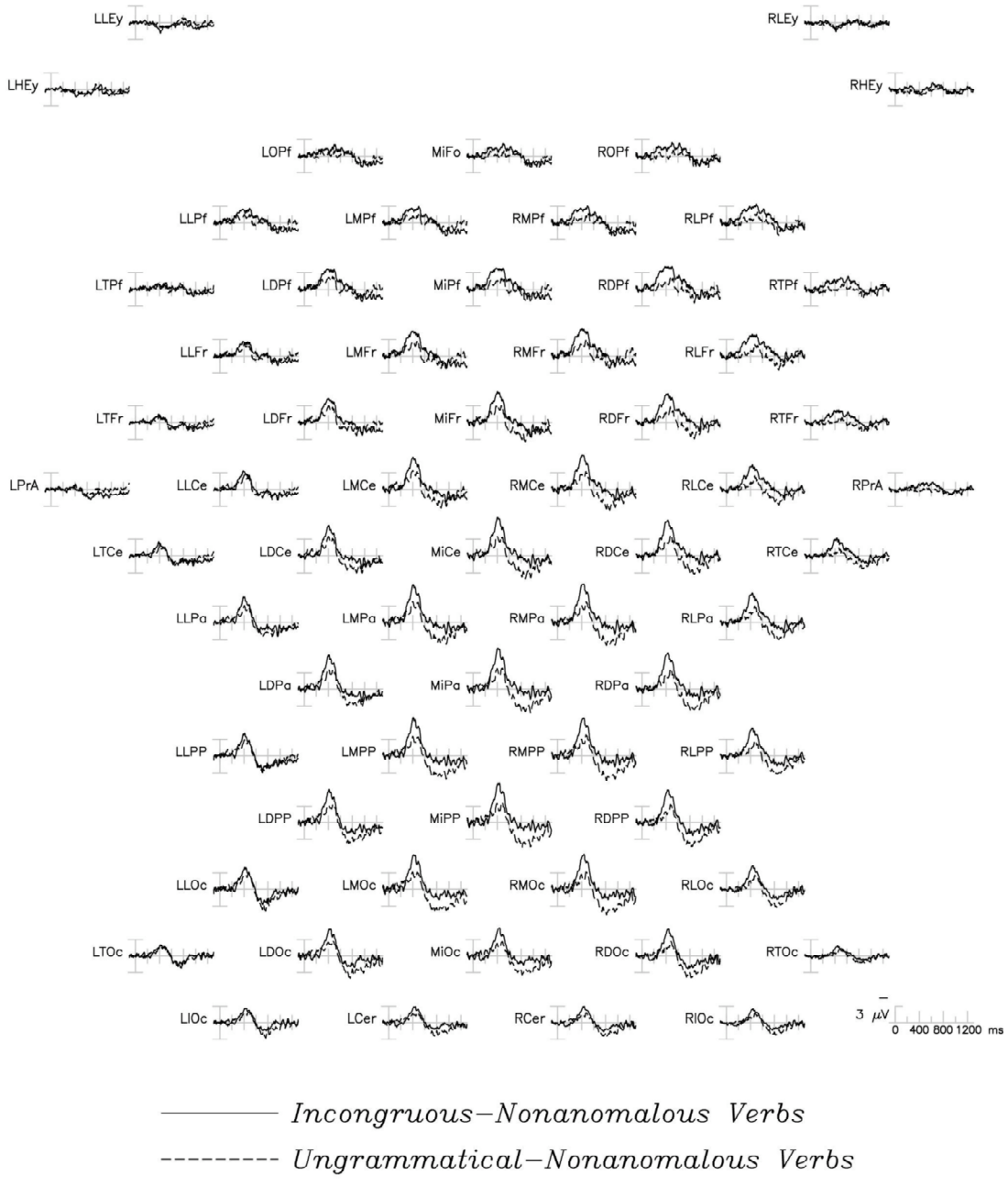

Figure IV.43 – Grand average difference waves (ERPs to anomalous verbs-ERPs to nonanomalous verbs) of verb congruity and grammaticality effects. Difference waves are topographically organized (anterior=towards top of page, right=towards right of page).

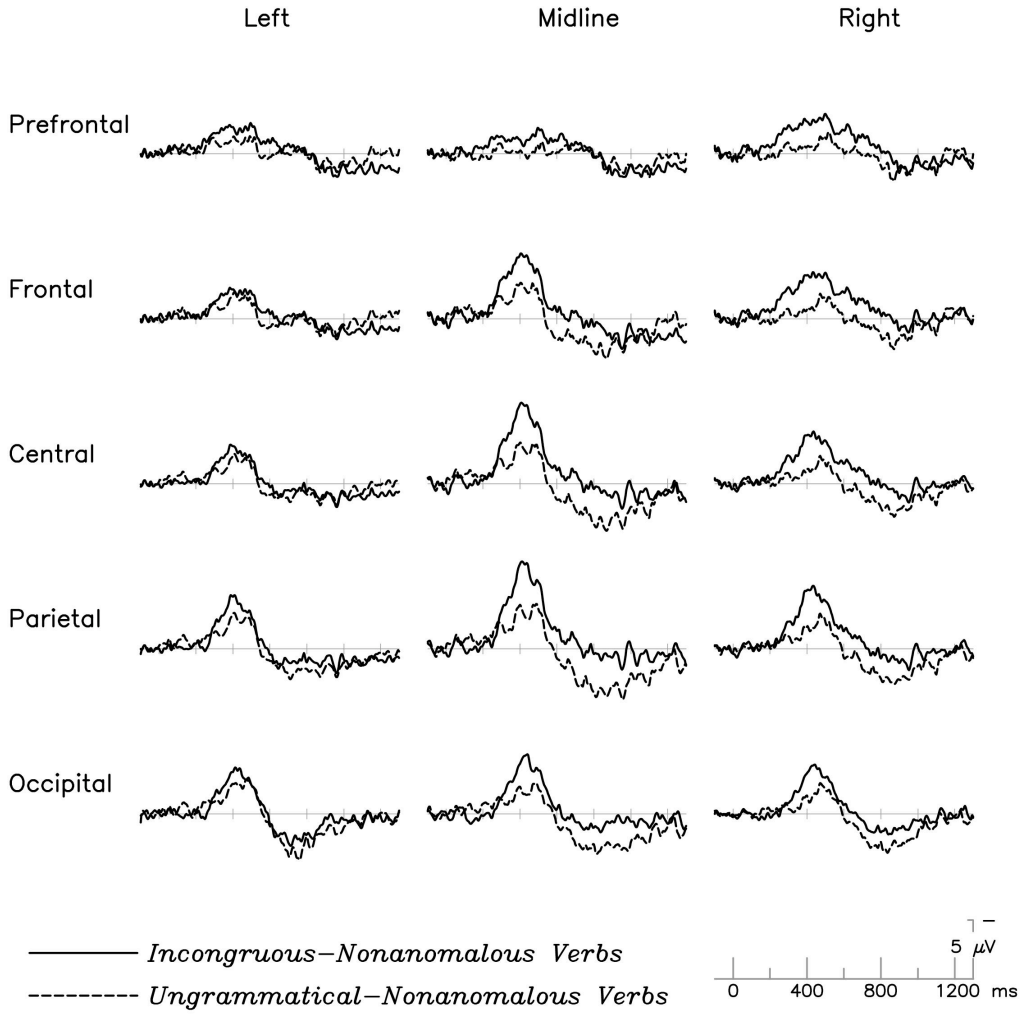

Figure IV.44 – Example grand average difference waves (ERPs to anomalous verbs-ERPs to nonanomalous verbs) of verb congruity and grammaticality effects.

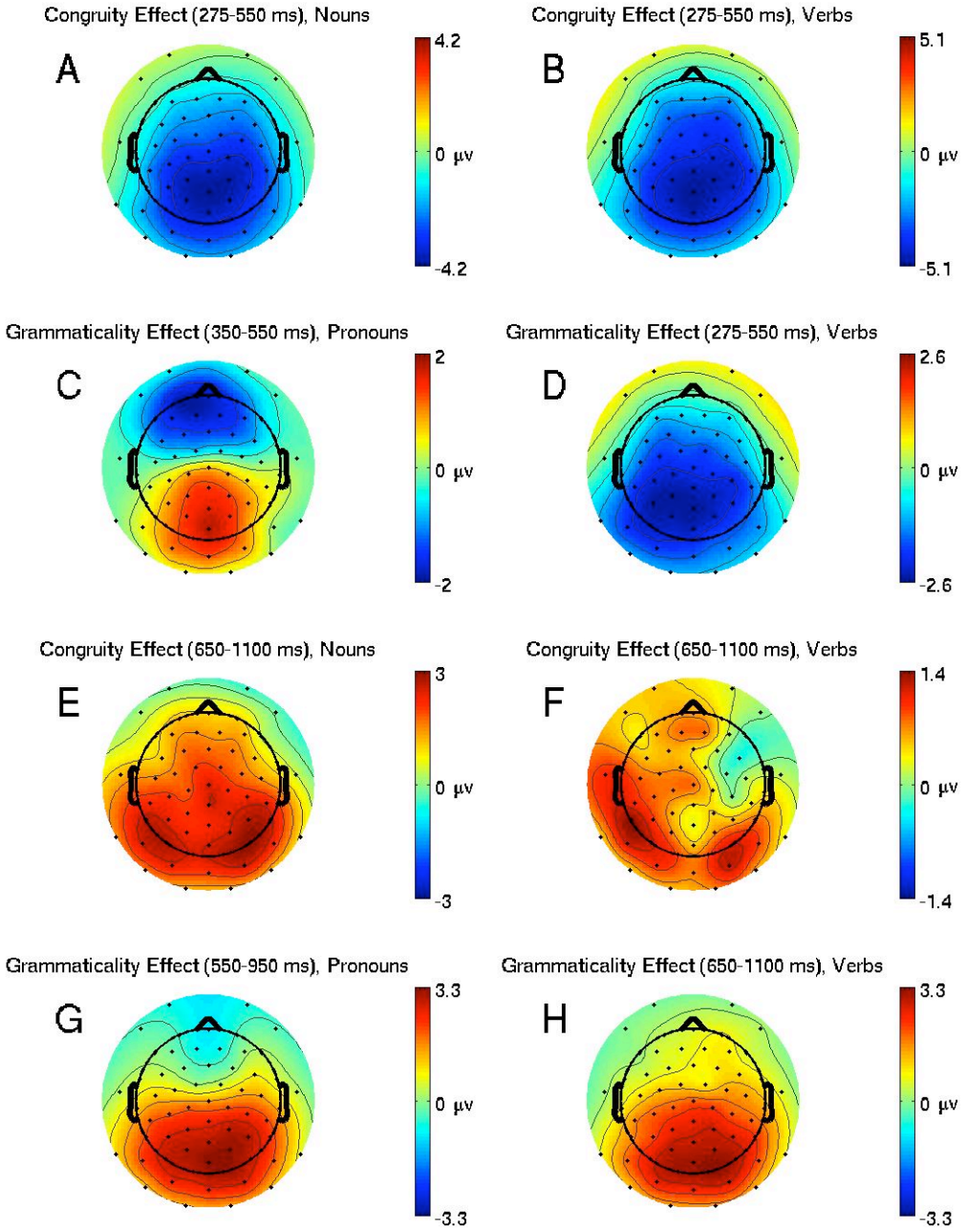

Figure IV.45 – Grand ERP topographies of semantic congruity and grammaticality effects in mid **(A-D)** and late **(E-H)** windows for each part of speech. Electrodes below the head's equator are plotted progressively beyond the radius of the head (Note: Right mastoid electrode, A2, is not plotted).

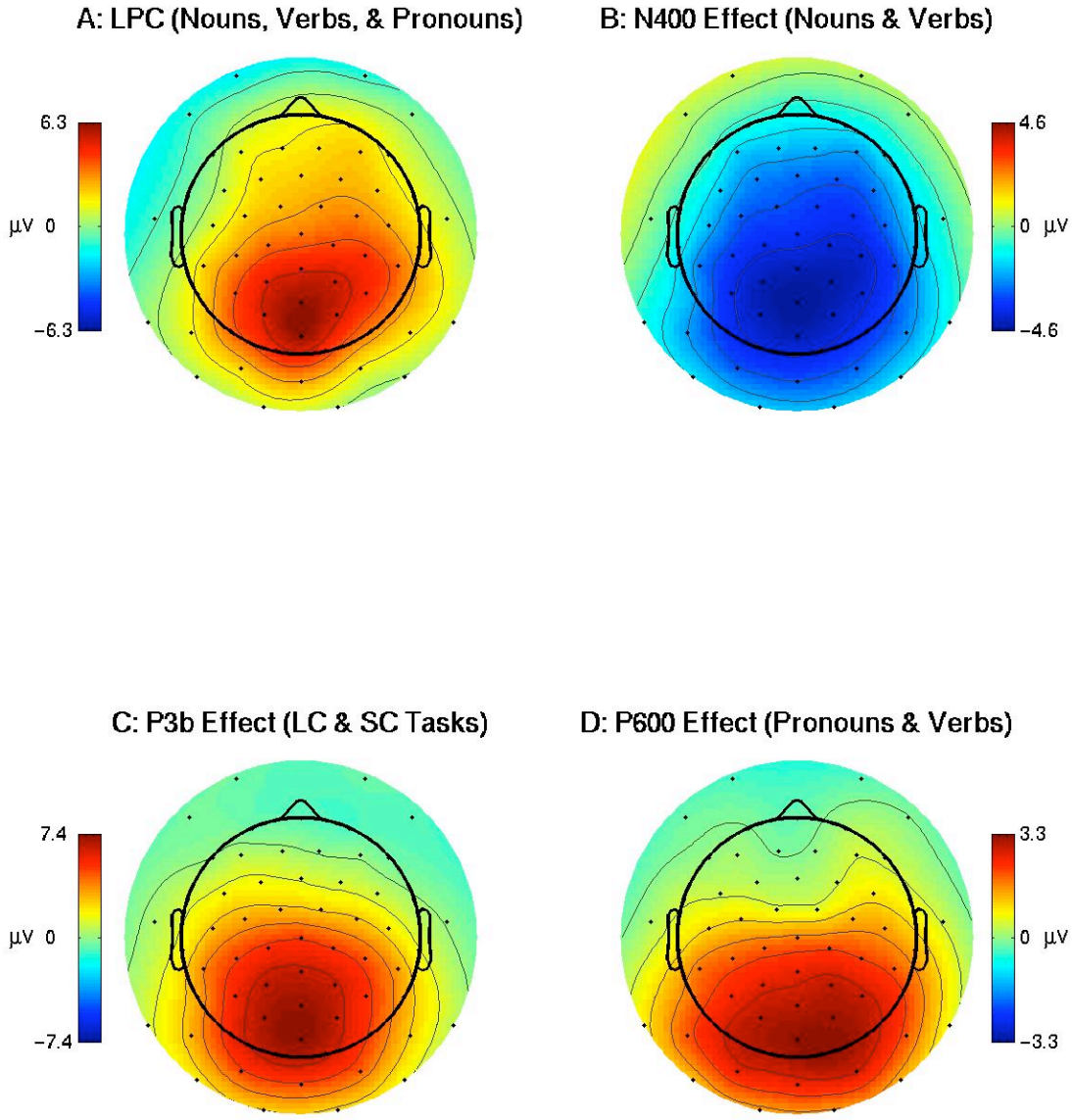

Figure IV.46 – Grand ERP topographies of the late positive component (LPC) to nonanomalous sentence endings (**A**) and the N400, P3b, and P600 effects (**B-D**). The LPC topography is derived from the mean amplitude between 450 and 800 ms in the ERPs to all nonanomalous sentence endings (each part of speech is weighted equally). N400 and P600 effects are simply the average of the N400 and P600 effects in Figure IV.44 that were computed separately for each part of speech. The P3b effect is simply the average of the P3b effects in Figure IV.20 that were computed separately for each oddball task. Electrodes below the head's equator are plotted progressively beyond the radius of the head (Note: Right mastoid electrode, A2, is not plotted).

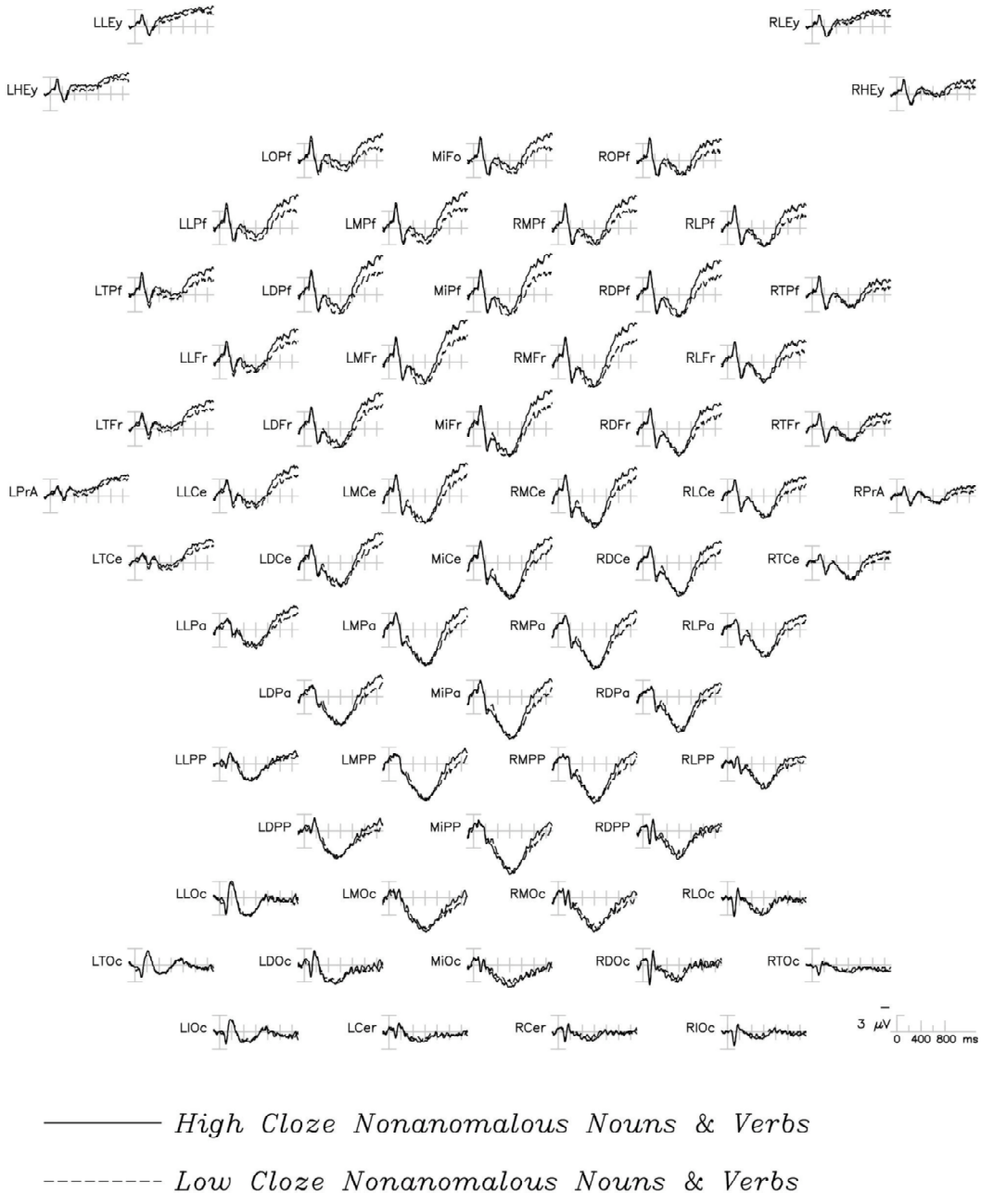

Figure IV.47 – Grand average ERPs to sentence final nonanomalous nouns and verbs. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

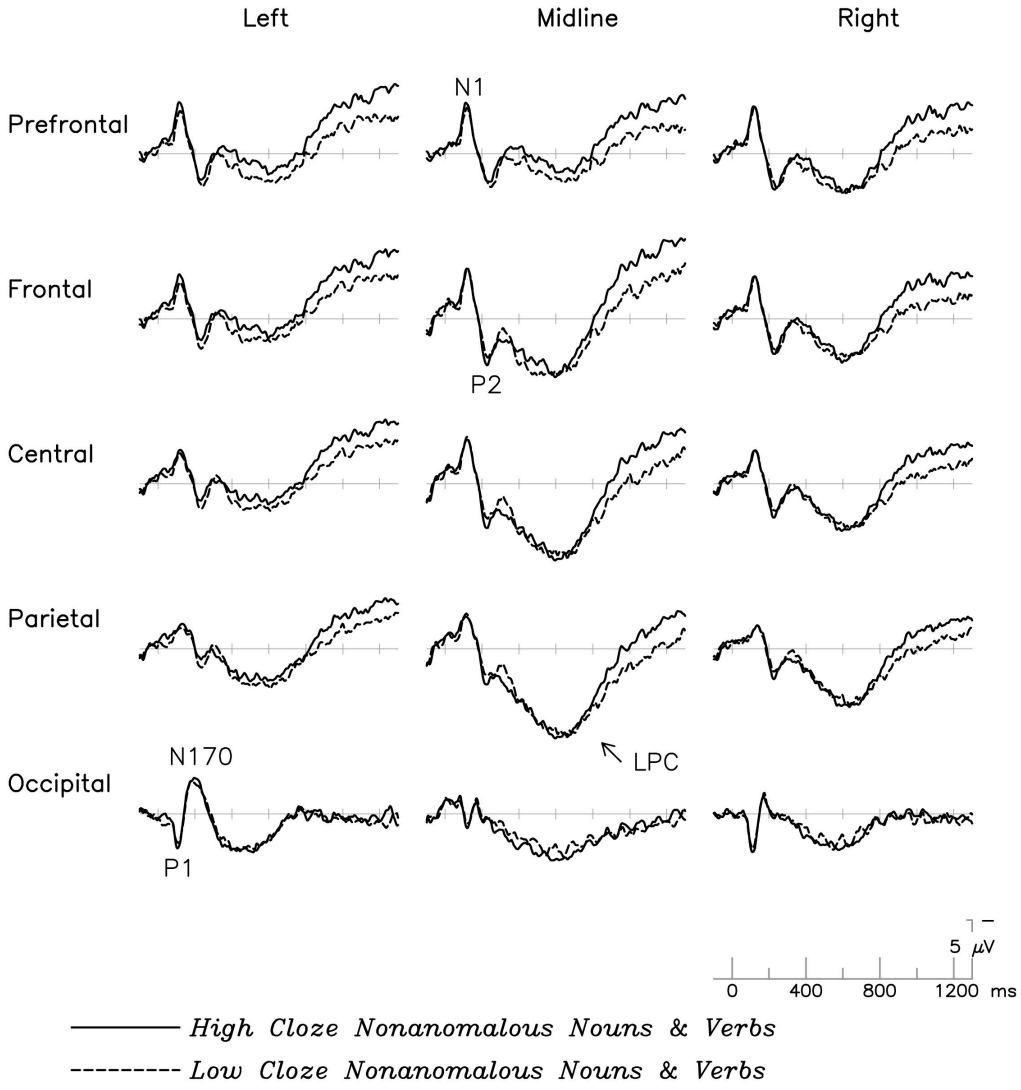

Figure IV.48 – Example grand average ERPs to sentence final nonanomalous nouns and verbs showing anterior N1/P2, posterior P1/N170, and LPC components.

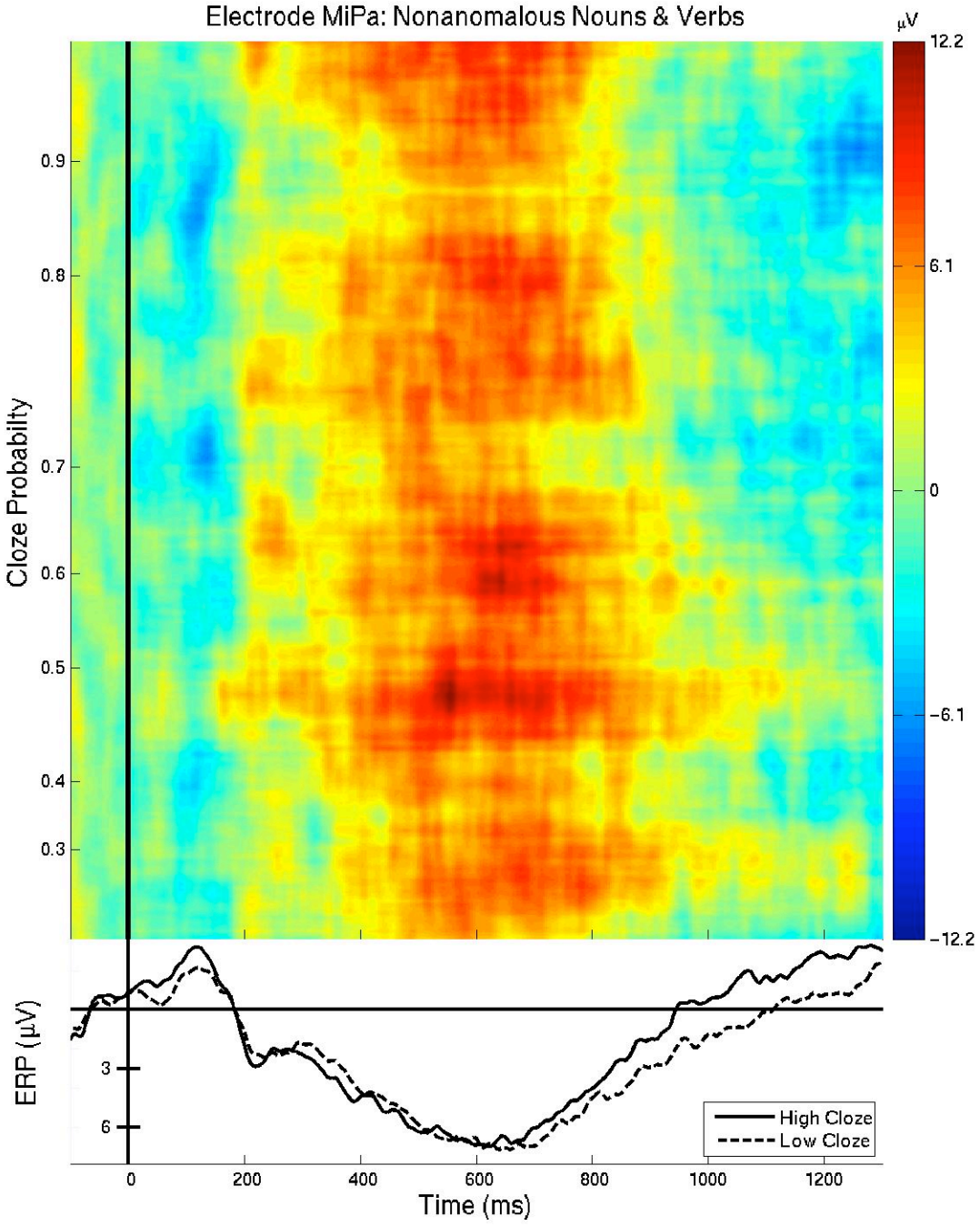

Figure IV.49 – (Top) 1381 single trials of the EEG response to nonanomalous nouns and verbs at electrode MiPa. Trials are sorted by cloze probability, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

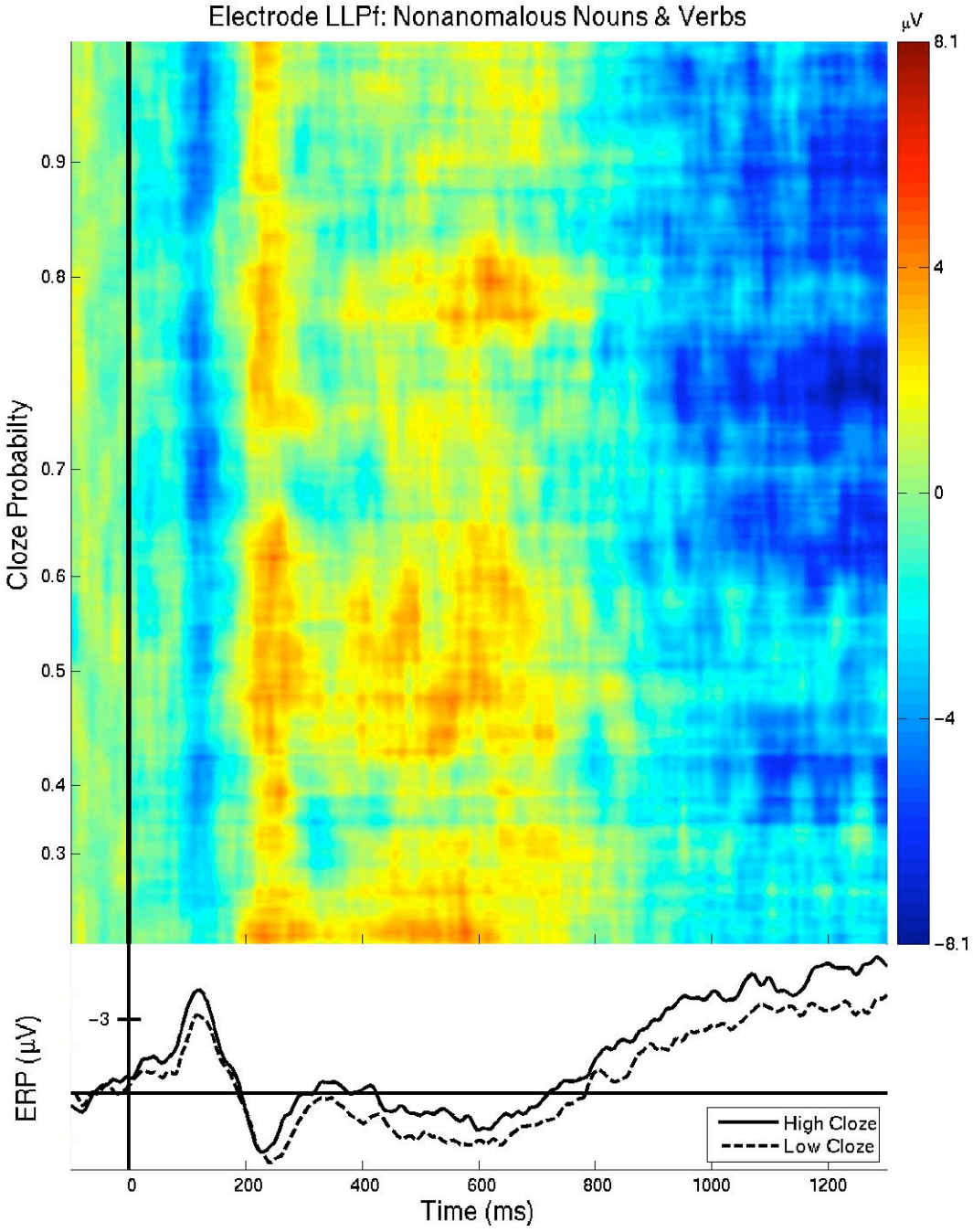

Figure IV.50 – (Top) 1381 single trials of the EEG response to nonanomalous nouns and verbs at electrode LLPf. Trials are sorted by cloze probability, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

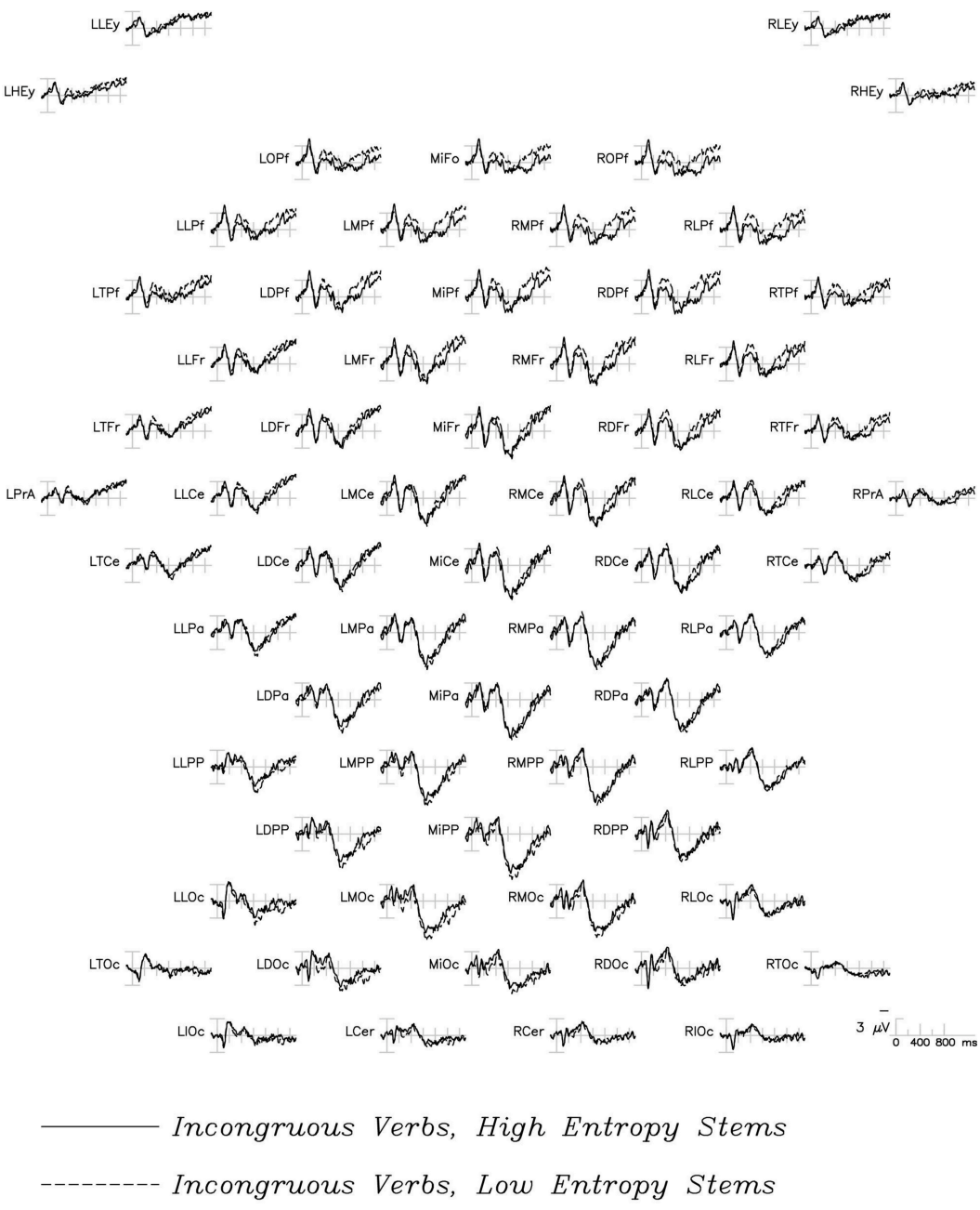

Figure IV.51 – Grand average ERPs to incongruous verb endings of high entropy (low constraint) and low entropy (high constraint) sentences. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

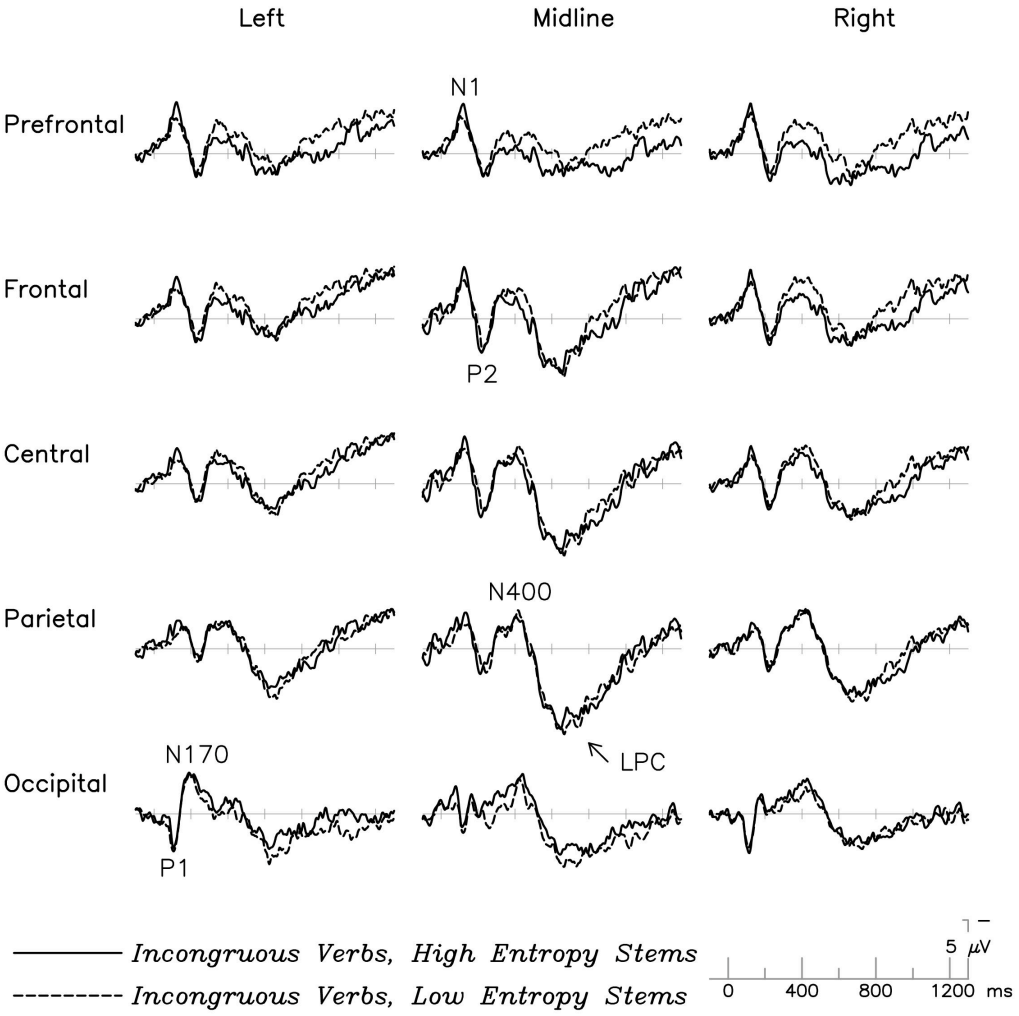

Figure IV.52 – Example grand average ERPs to incongruous verb endings of high entropy (low constraint) and low entropy (high constraint) sentences showing anterior N1/P2, posterior P1/N170, N400, and LPC components.

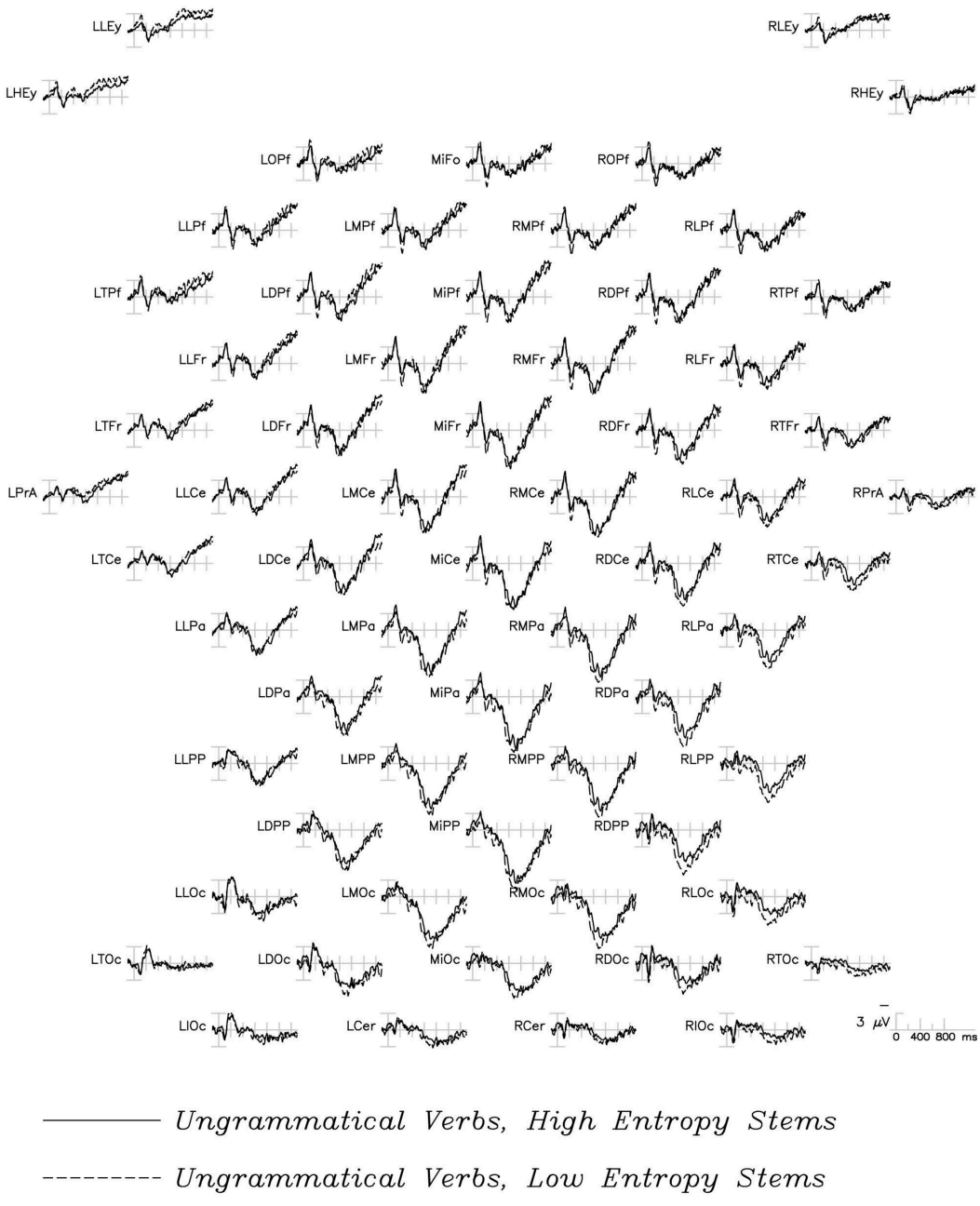

Figure IV.53 – Grand average ERPs to ungrammatical verb endings of high entropy (low constraint) and low entropy (high constraint) sentences. ERPs are topographically organized (anterior=towards top of page, right=towards right of page).

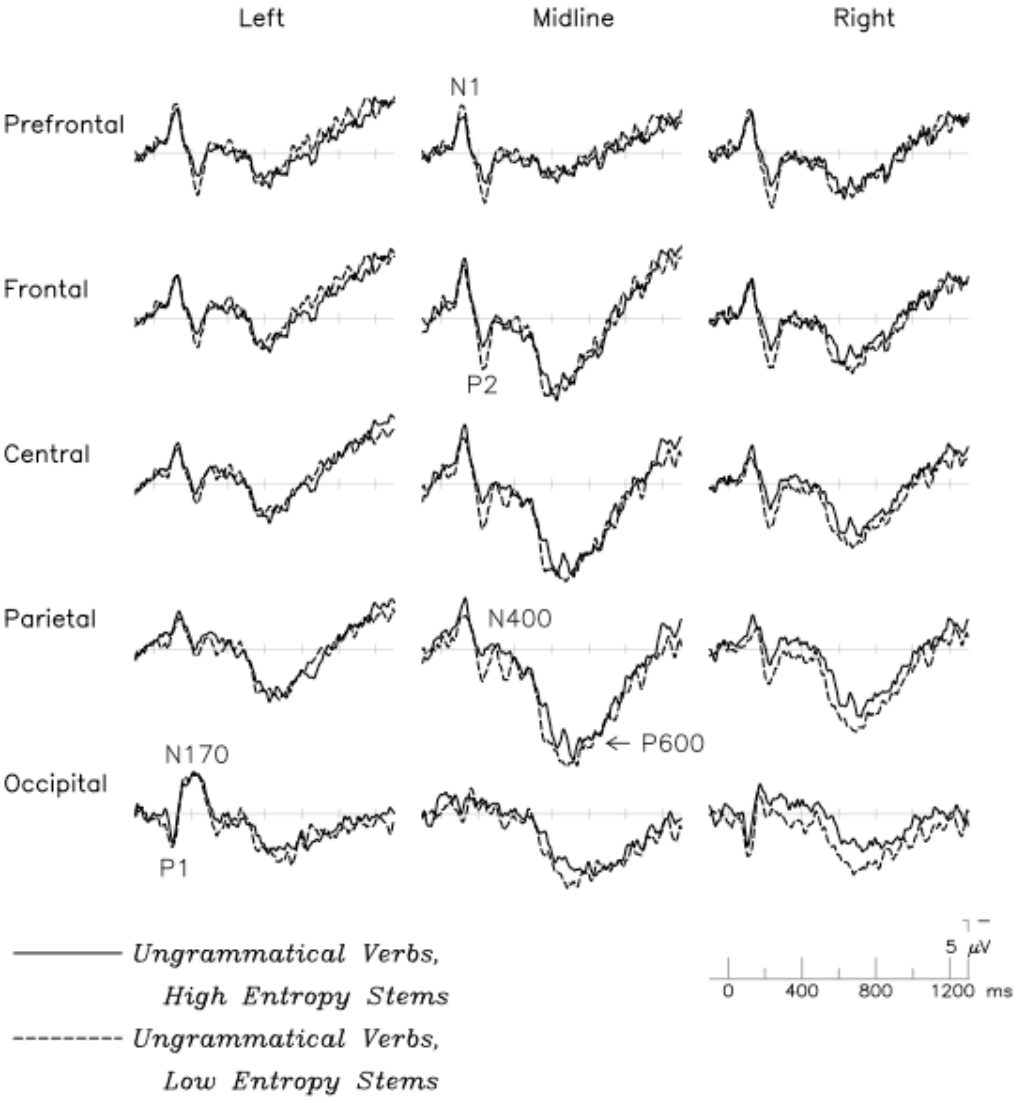

Figure IV.54 – Example grand average ERPs to ungrammatical verb endings of high entropy (low constraint) and low entropy (high constraint) sentences showing anterior N1/P2, posterior P1/N170, N400, and P600 components.

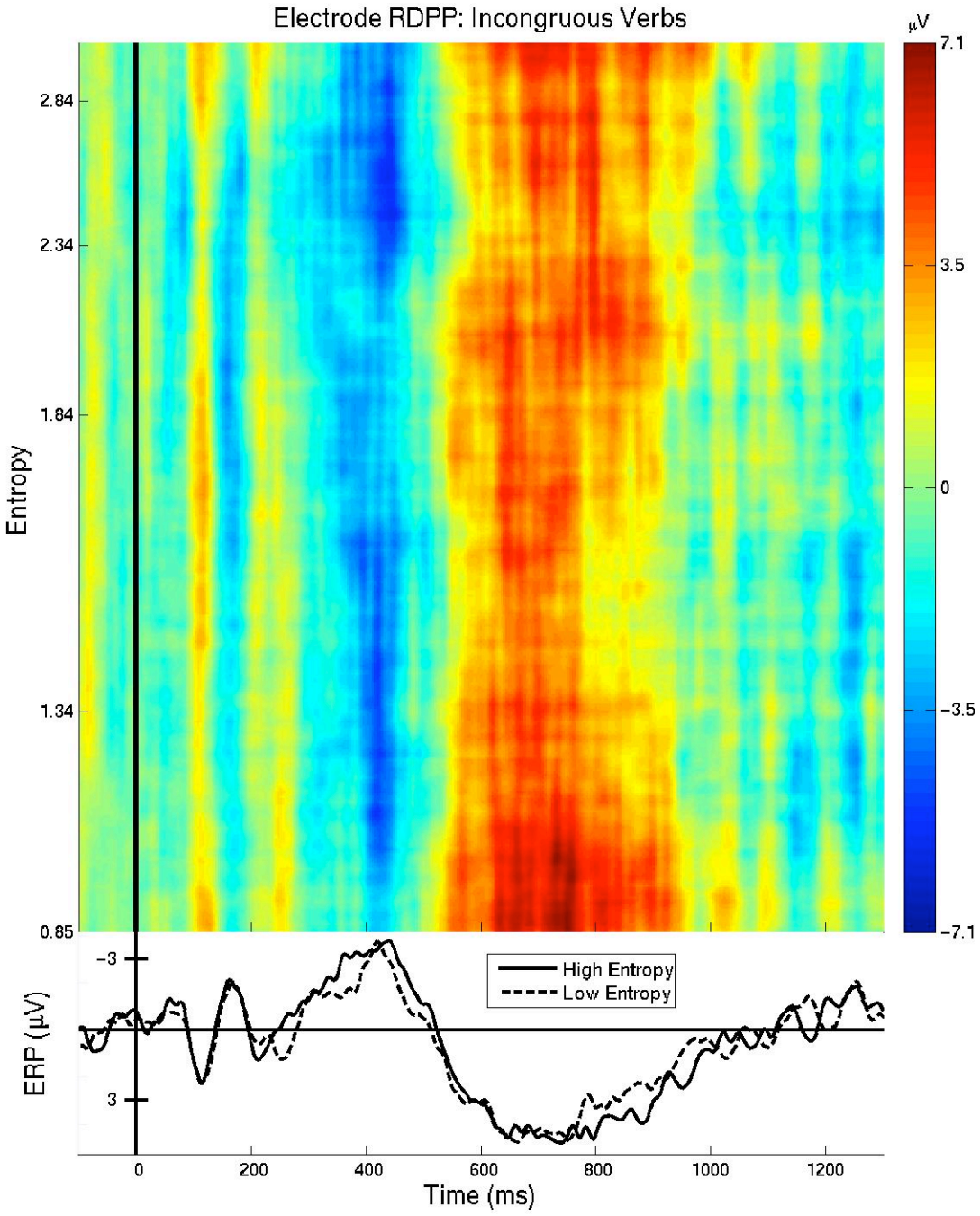

Figure IV.55 – **(Top)** 468 single trials of the EEG response to incongruous verbs at electrode RDPP. Trials are sorted by the entropy of the preceding sentence stem, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

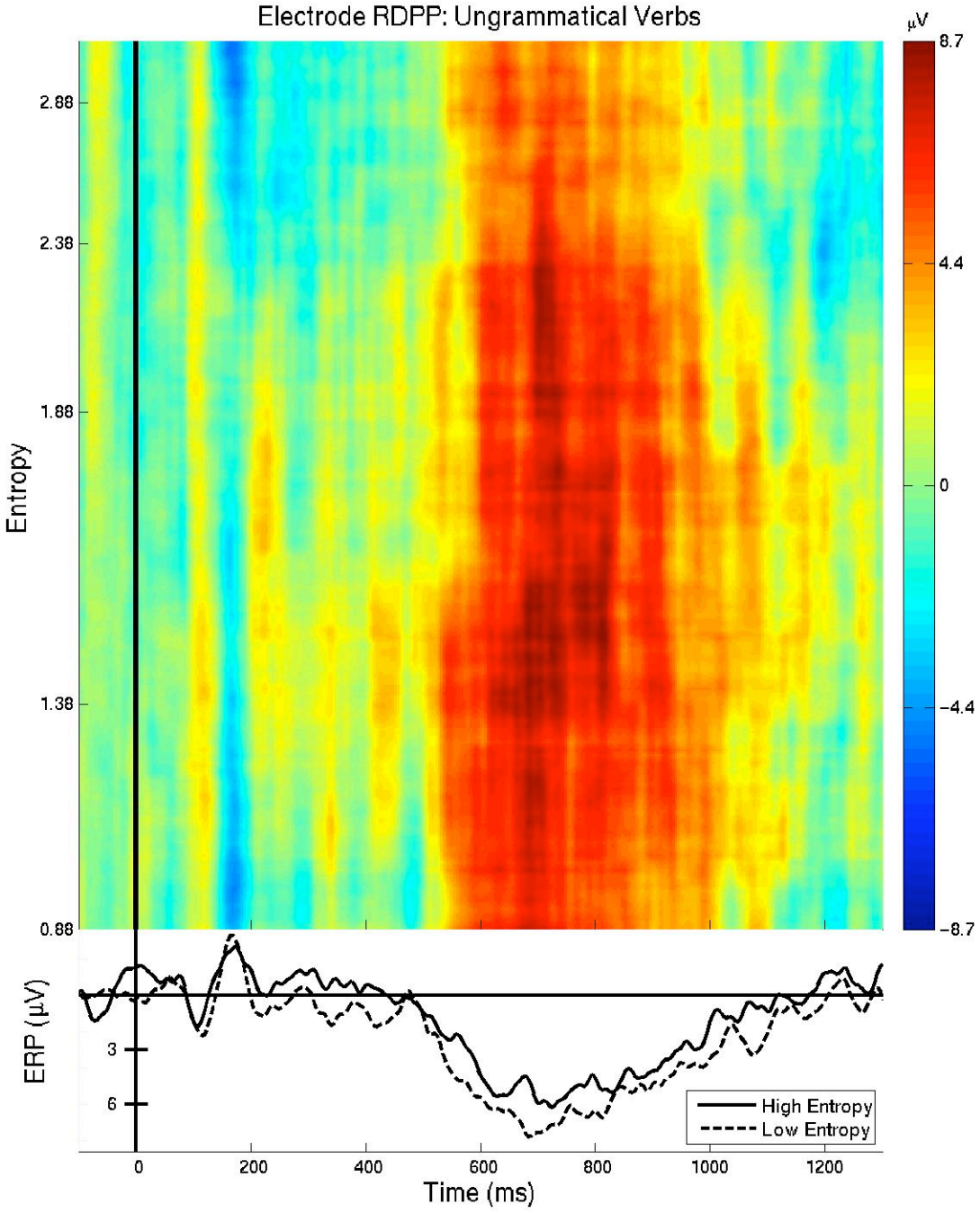

Figure IV.56 – (Top) 447 single trials of the EEG response to ungrammatical verbs at electrode RDPP. Trials are sorted by the entropy of the preceding sentence stem, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

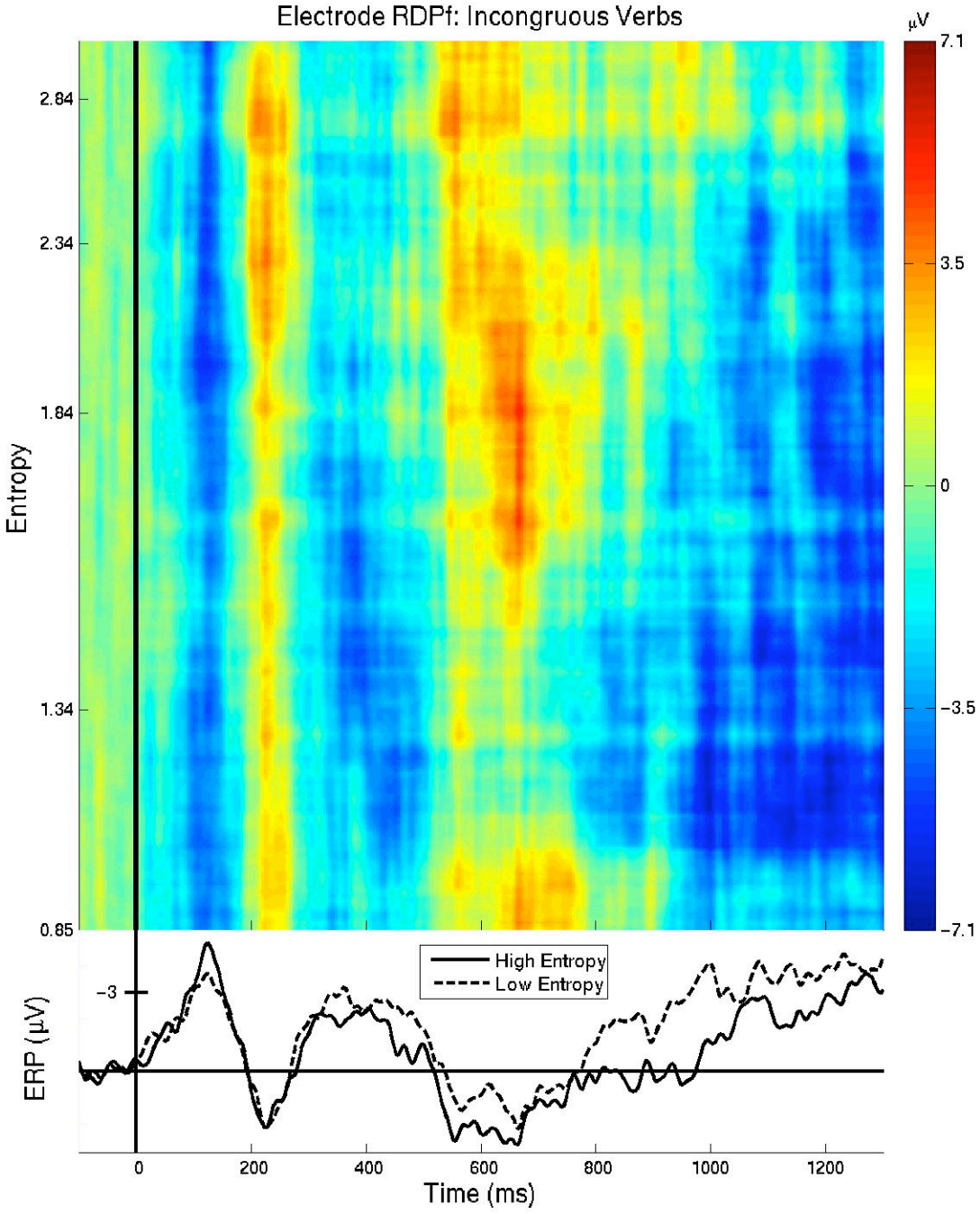

Figure IV.57 – **(Top)** 468 single trials of the EEG response to incongruous verbs at electrode RDPf. Trials are sorted by the entropy of the preceding sentence stem, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

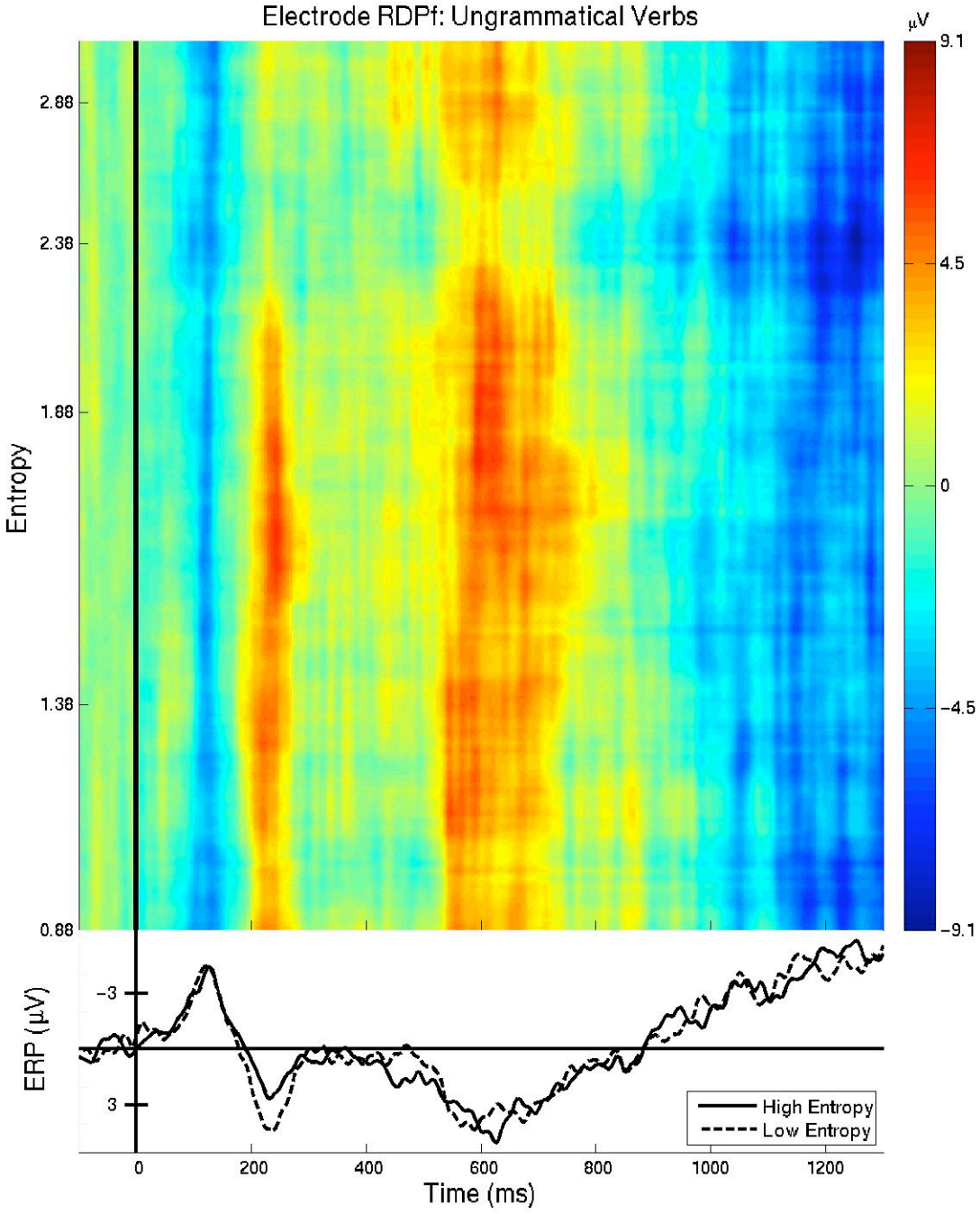

Figure IV.58 – **(Top)** 447 single trials of the EEG response to ungrammatical verbs at electrode RDPf. Trials are sorted by the entropy of the preceding sentence stem, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

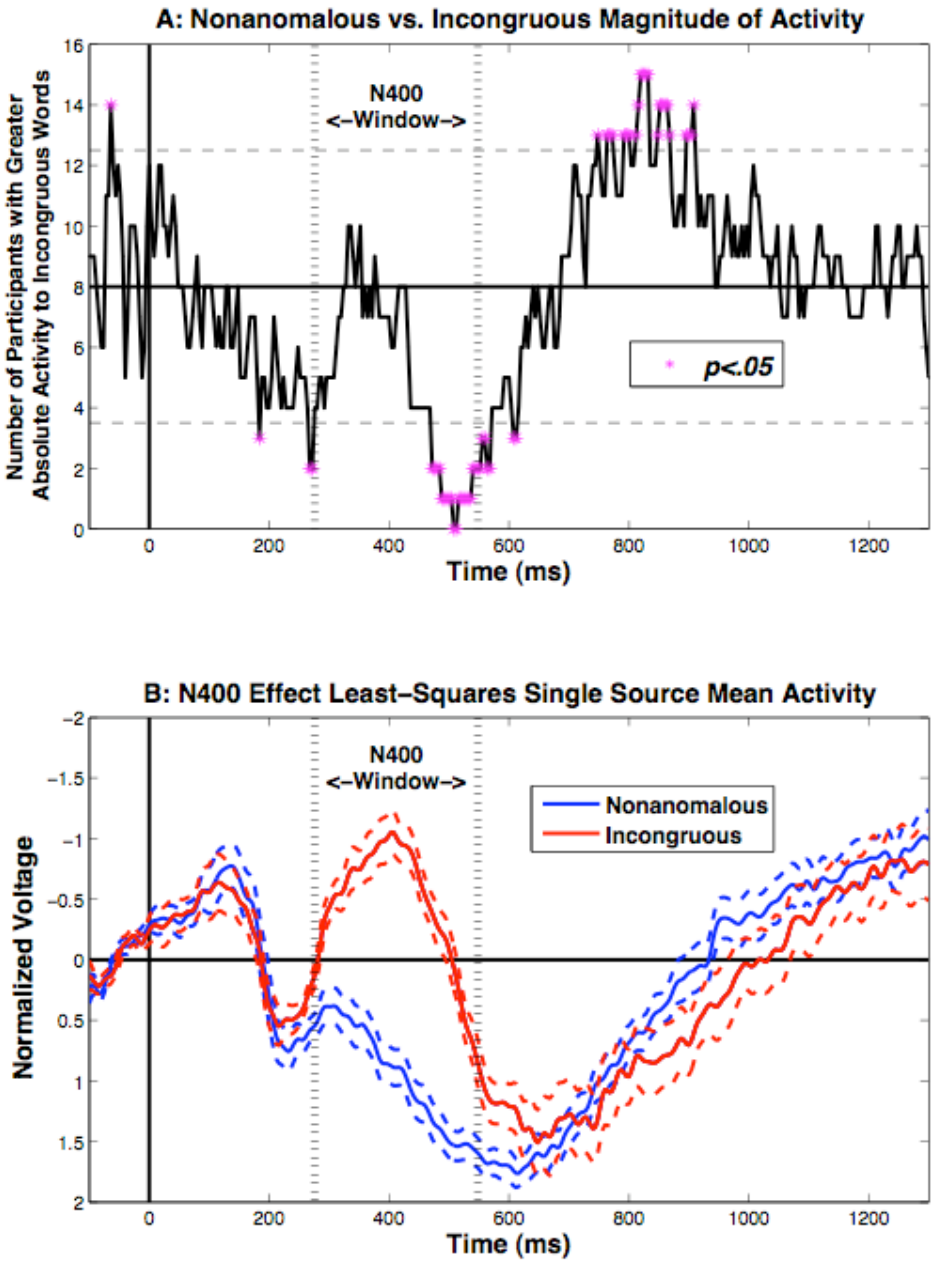

Figure IV.59 – Least-squares N400 effect source model of ERPs to all incongruous and nonanomalous noun and verb endings. (**A**) Number of participants whose N400 source shows greater activity to incongruous words. Significantly many or few numbers of participants are indicated with pink asterisks (two tailed sign test). Horizontal dashed lines indicate cut-off for statistical significance. (**B**) Mean activation of N400 source activity to incongruous and nonanomalous noun and verb endings. Colored dashed lines indicate standard error (*N*=16). To resolve the scaling ambiguity of source activations, each participant's source activity is normalized to have unit standard deviation across both conditions. Furthermore, the polarity of the activations are set such the N400 effect has a net negative scalp distribution.

IV.I Appendix 1: 64 Channel Electrode Labels and Locations

Electrode locations are given in spherical coordinates. Phi is the angle measured by an arc drawn from the top center (vertex) of the head to the electrode. Theta is the angle, measured counterclockwise around the perimeter of the head, from the right ear to the electrode.

Electrode Location Label	Label Abbreviation	Phi	Theta
Left Orbital Prefrontal	LOPf	90	114
Right Orbital Prefrontal	ROPf	90	66
Left Medial Prefrontal	LMPf	66.8	102
Right Medial Prefrontal	RMPf	66.8	78
Left Temporal Prefrontal	LTPf	90	138
Right Temporal Prefrontal	RTPf	90	42
Left Lateral Prefrontal	LLPf	77.1	126
Right Lateral Prefrontal	RLPf	77.1	54
Left Pre-Auricular	LPrA	111.3	174
Right Pre-Auricular	RPrA	111.3	366
Left Temporal Frontal	LTFr	90	162
Right Temporal Frontal	RTFr	90	18
Left Lateral Frontal	LLFr	66.8	150
Right Lateral Frontal	RLFr	66.8	30
Left Dorsal Prefrontal	LDPf	51.4	126

Right Dorsal Prefrontal	RDPf	51.4	54
Left Temporal Occipital	LTOc	133.6	210
Right Temporal Occipital	RTOc	133.6	330
Left Temporal Central	LTCe	90	186
Right Temporal Central	RTCe	90	354
Left Lateral Central	LLCe	66.8	174
Right Lateral Central	RLCe	66.8	366
Left Dorsal Frontal	LDFr	44.5	162
Right Dorsal Frontal	RDFr	44.5	18
Left Medial Frontal	LMFr	25.7	126
Right Medial Frontal	RMFr	25.7	54
Midline Frontal-orbital	MiFo	90	90
Midline Prefrontal	MiPf	44.5	90
Midline Frontal	MiFr	0	0
Right Mastoid	A2	111.3	330
Left Outer Canthus	LHEy	111.3	150
Right Outer Canthus	RHEy	111.3	30
Left Inferior Occipital	LIOc	133.6	234
Right Inferior Occipital	RIOc	133.6	306
Left Lateral Occipital	LLOc	111.3	222

Right Lateral Occipital	RLOc	111.3	318
Left Lateral Posterior Parietal	LLPP	90	210
Right Lateral Posterior Parietal	RLPP	90	330
Left Lateral Parietal	LLPa	77.1	198
Right Lateral Parietal	RLPa	77.1	342
Left Dorsal Central	LDCe	51.4	198
Right Dorsal Central	RDCe	51.4	342
Left Medial Central	LMCe	25.7	198
Right Medial Central	RMCe	25.7	342
Left Dorsal Occipital	LDOc	111.3	246
Right Dorsal Occipital	RDOc	111.3	294
Left Dorsal Posterior Parietal	LDPP	90	234
Right Dorsal Posterior Parietal	RDPP	90	306
Left Dorsal Parietal	LDPa	66.8	222
Right Dorsal Parietal	RDPa	66.8	318
Left Cervical	LCer	133.6	258
Right Cervical	RCer	133.6	282
Left Medial Occipital	LMOc	90	258
Right Medial Occipital	RMOc	90	282
Left Medial Posterior Parietal	LMPP	66.8	246

Right Medial Posterior Parietal	RMPP	66.8	294
Left Medial Parietal	LMPa	44.5	234
Right Medial Parietal	RMPa	44.5	306
Midline Central	MiCe	25.7	270
Midline Parietal	MiPa	51.4	270
Midline Posterior Parietal	MiPP	77.1	270
Midline Occipital	MiOc	111.3	270
Left Infraorbital Ridge	LLEy	133.6	112
Right Infraorbital Ridge	RLEy	133.6	68

IV.J References

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31*(4), 578-602.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences, 22*(4), 577-609; discussion 610-60.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235-260.

Chung, G., Tucker, D. M., West, P., Potts, G. F., Liotti, M., Luu, P., et al. (1996). Emotional expectancy: Brain electrical activity associated with an emotional bias in interpreting life events. *Psychophysiology, 33*(3), 218-233.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology, 88*(3), 188-209.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084*(1), 89-103.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Del Gratta, C., Della Penna, S., Ferretti, A., Franciotti, R., Pizzella, V., Tartaro, A., et al. (2002). Topographic organization of the human primary and secondary somatosensory cortices: Comparison of fMRI and MEG findings. *NeuroImage, 17*(3), 1373-1383.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*(2), 95-111.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 112*(3), 357-374.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: Wiley.

Federmeier, K. D., & Kutas, M. (1999a). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research, 8*(3), 373-392.

Federmeier, K. D., & Kutas, M. (1999b). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469-495.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123 Pt 12*, 2552-2566.

Francis, W. N., & Kucera , H. (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*(3), 183-192.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London.Series B, Biological sciences, 360*(1456), 815-836.

Galan, L., Biscay, R., Rodriguez, J. L., Perez-Abalo, M. C., & Rodriguez, R. (1997). Testing topographic differences between event-related brain potentials by using non-parametric combinations of permutation tests. *Electroencephalography and Clinical Neurophysiology, 102*(3), 240-247.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*(2), 89-106.

Gonsalvez, C. L., & Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology, 39*(3), 388-396.

Greenblatt, R. E., & Pflieger, M. E. (2004). Randomization-based hypothesis testing from event-related data. *Brain Topography, 16*(4), 225-232.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*(6), 660-676.

Haan, H., Streb, J., Bien, S., & Rosler, F. (2000). Individual cortical current density reconstructions of the semantic N400 effect: Using a generalized minimum norm model with different constraints (L1 and L2 norm). *Human Brain Mapping, 11*(3), 178-192.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883-899.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by

semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17*(3), 1101-1116.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain, 121*(6), 1133-1142.

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *The Journal of Neuroscience, 22*(7), 2936-2944.

Hemmelmann, C., Horn, M., Reiterer, S., Schack, B., Susse, T., & Weiss, S. (2004). Multivariate tests for the evaluation of high-dimensional EEG data. *Journal of Neuroscience Methods, 139*(1), 111-120.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721-742.

Homan, R. W., Herman, J., & Purdy, P. (1987). Cerebral location of international 10-20 system electrode placement. *Electroencephalography and Clinical Neurophysiology, 66*(4), 376-382.

Johnson, B. W., & Hamm, J. P. (2000). High-density mapping in an N400 paradigm: Evidence for bilateral temporal lobe generators. *Clinical Neurophysiology, 111*(3), 532-545.

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping, 14*(3), 166-185.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Kayser, J., & Tenke, C. E. (2006). Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns II: Adequacy of low-density estimates. *Clinical Neurophysiology, 117*(2), 369-380.

Kemmer, L., Coulson, S., De Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology, 41*(3), 372-384.

Kemmer, L., & Kutas, M. (2006). Salience of grammatical violations, but not verb (ir)regularity, affects the P600: An event-related potential study. *13th Annual Meeting of the Cognitive Neuroscience Society,* San Francisco.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology, 55*, 271-304.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: An event-related fMRI study of semantic processing. *NeuroImage, 17*(2), 842-850.

King, J. W., & Kutas, M. (1995). A brain potential whose latency indexes the length and frequency of words. *The Newsletter of the Center for Research in Language, 10*(2), 1-9.

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*(1), 1-36.

Kotchoubey, B., & Lang, S. (2001). Event-related potentials in an auditory semantic oddball task in humans. *Neuroscience Letters, 310*(2-3), 93-96.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 804-823.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2005). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences *Language and Cognitive Processes, 21*(5), 489-530.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293.

Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. In R. Johnson, J. W. Rohrbaugh & R. Parasuraman (Eds.), *Current Trends in Event-Related Potential Research (EEG suppl. 40)* (pp. 406-411) Elsevier Science Publishers B.V.

Kutas, M., & Dale, A. M. (1997). Electrical and magnetic readings of mental functions. In M. D. Rugg (Ed.), *Cognitive Neuroscience* (first ed., pp. 197-242). Hove East Sussex, UK: Psychology Press.

Kutas, M., & Hillyard, S. A. (1980a). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99-116.

Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83-83-143) Academic Press.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation, 11*(2), 417-441.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*(1), 47-62.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: MIT Press.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology,* (62), 203-208.

Moores, K. A., Clark, C. R., Hadfield, J. L., Brown, G. C., Taylor, D. J., Fitzgibbon, S. P., et al. (2003). Investigating the generators of the scalp recorded visuo-verbal P300 using cortically constrained source localization. *Human Brain Mapping, 18*(1), 53-77.

Munte, T. F., Heinze, H. J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia, 36*(3), 217-226.

Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms.*, 2004, from http://www.usf.edu/FreeAssociation/

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., & Nicol, J. L. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*(3), 283-317.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*(3), 105-110.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition, 56*(1), 5-13.

Pulvermuller, F. (1999). Words in the brain's language. *The Behavioral and Brain Sciences, 22*(2), 253-79.

Segalowitz, S. J., & Barnes, K. L. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology, 30*(5), 451-459.

Simos, P. G., Basile, L. F., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research, 762*(1-2), 29-39.

Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research, 15*(1), 99-103.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage, 25*(2), 539-553.

Urbach, T. P., & Kutas, M. (2002). The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology, 39*(6), 791-808.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes, 8*(4), 485-531.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19*(1), 95-112.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractioning the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*(2), 131-150.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Woody, C. D., & Nahvi, M. J. (1973). Application of optimum linear filter theory to the detection of cortical signals preceding facial movement in cat. *Experimental Brain Research, 16*(5), 455-465.

Chapter V: Identifying Reliable Independent Components

V.A Introduction

As mentioned in Chapter III, a complication with using independent component analysis (ICA)--or any other learned filter--to analyze EEG/ERP data is that there is no guarantee that the learned independent components (ICs) are reliable. This is readily apparent when using iterative, gradient descent based ICA algorithms like extended INFOMAX ICA (Lee, Girolami, & Sejnowski, 1999). Such an algorithm can learn different ICs from the same data if the algorithm is initialized differently or if the order on which the data is presented to the algorithm varies. These problems do not occur when using analytic algorithms such as second order blind identification (Belouchrani, Abed Meraim, Cardoso, & Moulines, 1997), but they too may learn different ICs from data generated by the same sources. This will occur if the amount of training data is insufficient or if some sources defy unique unmixing (e.g., Figure V.1).

A few different methods for assessing IC reliability have been proposed in the literature. These include clustering across participants (Debener et al., 2005; Debener, Makeig, Delorme, & Engel, 2005; S. Makeig et al., 2002; S. Makeig et al., 2004),

resampling (Himberg, Hyvarinen, & Esposito, 2004; Meinecke, Ziehe, Kawanabe, & Muller, 2002), and noise injection[59] (Harmeling, Meinecke, & Müller, 2004). Clustering is problematic because there is no guarantee that the IC clusters are reliable; this is especially an issue for clusters formed by hand (Debener et al., 2005; Debener et al., 2005). Furthermore, comparing ICs across participants is complicated by functionally irrelevant anatomical differences (e.g., differences in cortical folding) that can cause functionally homologous ICs to appear quite different. Resampling and noise injection methods can be too time consuming to use because they require re-running ICA multiple times on the same data set. For example, Efron and Tibshirani (1993, pg 188) recommend 1000 resamples for bootstrap based confidence intervals[60]. Using a compiled implementation of extended INFOMAX ICA, *binica* from the EEGLAB toolbox, it took over two weeks to run ICA 1000 times on a single participant's data from the visual oddball tasks reported in the previous chapter. A further problem with resampling and noise injection is that it is not clear how to determine reliability thresholds. For example, bootstrapping could provide a 95% confidence interval around an IC unmixing vector, but it does not tell you how wide of an interval to allow before rejecting that IC as unreliable.

In this chapter, I present an alternative method for identifying reliable ICs from a data set by re-running ICA on 50% splits of the data. This approach avoids problems from between-participant anatomical variation, the time demands are

[59] "Noise injection" attempts to identify reliable ICs by adding white Gaussian noise (Figure V.1: Left), which ICA cannot reliably decompose, to the data and re-running ICA to determine which ICs are affected. ICs that are most affected by the additional noise are deemed the least reliable.
[60] More specifically, Efron and Tibshirani recommend a minimum of 1000 resamples for "bootstrap corrected and accelerated confidence intervals."

manageable (one must perform ICA three times for each participant), and the reliability criterion is statistically interpretable (an alpha level). I show that the method produces results that are similar to those derived from more time intensive bootstrap resampling and show that the method depends appropriately on the amount of training data.

Given that the reliability test behaves reasonably, I use to it evaluate heuristics that have been used to select ICs for analysis: the percentage of data variance accounted for by an IC and the "dipolarity" of an IC. I find that reliable ICs do tend to account for more variance and are more frequently dipolar than unreliable ICs. However, a non-negligible percentage of ICs are exceptions to these trends, indicating a need to assess the reliability of ICs independently of these heuristics.

Finally, I demonstrate how this reliability test can be used to improve ICA results by showing how a seemingly small change in data pre-processing, removing the mean of each epoch of data, greatly enhances the reliability of extended INFOMAX ICA.

V.B Assessing Component Reliability with 50% Splits

Perhaps the simplest, most intuitive way to assess the reliability of any empirical result is to replicate the experiment. A cheap alternative to true replication is to split a data set into two comparable halves. While this diminishes the power of

the analysis, replication of the result across each data subset is evidence of its reliability.

This logic is the basis of the following five-step test of IC reliability proposed in this chapter:

1. Perform ICA on the full data set
2. Split the data set into two comparable subsets
3. Perform ICA on both subsets
4. Pair each IC from the full data set with an IC from each of the subsets forming triplets of ICs
5. If each member of an IC triple is significantly similar (i.e., homologous) to both of the other members of the triple, then the member of the triple from the full data set is deemed reliable.

This algorithm is conceptually similar to resampling based approaches for estimating reliability, such as the bootstrap and the jackknife (Efron & Tibshirani, 1993), which essentially re-perform an analysis on a large number (e.g., 1000) of possible subsets of a data set. Rather than using randomly chosen subsets of the data (the bootstrap) or methodically rotating out subsets of the data (the jackknife), using 50% splits privileges two particular resamples. This has the disadvantage of being influenced by any bias in the method used to split the data. Furthermore, it has less power than the bootstrap and jackknife, which can analyze more than 50% of the data in each

resample. In the present context, these disadvantages are offset by the minimal time demands of analyzing only two resamples.

While the proposed algorithm is conceptually simple, the details of Steps 2, 3, and 5 are far from obvious and are spelled out in the following subsections.

V.B.1 How to Split the Data

The sources of data variation of greatest interest to any EEG/ERP analysis are the experimental manipulations (e.g., targets vs. standards in an oddball task). Thus it is important to split the data such that each experimental condition is equally represented in each half. An even larger source of data variation is the experimental participant's alertness. This tends to diminish throughout the course of an experiment and can have large consequences for the participant's scalp potentials and behavior. For example, as the experiment progresses participants are more likely to blink, to relax their muscles, to make mistakes, and to elicit EEG alpha activity. To attempt to equally represent both of these factors in the two splits of the data, I propose separating the odd and even trials of each experimental condition.

V.B.2 How to Pair Independent Components

How is one to go from two sets of n ICs from comparable data sets to a set of n pairs of ICs? Identifying a potentially homologous pair of ICs first requires an IC

similarity metric. Both the topography and activation of an IC are important for ERP analyses, and these are somewhat independent of one another[61]. To determine if two ICs are homologous (Section V.B.3) it is important to compare both features, but to simply find a pair of potentially similar ICs, it is sufficient to look at only one feature. In practice, IC topographies appear to be more stable than IC activations, so I propose using the absolute value of the cosine similarity metric, r_{cos}, to compare *interpolated* IC topographies. This similarity metric deals with the IC scaling ambiguity (see Section III.E.4.a) by effectively normalizing the topographies to unit length and setting their polarities to maximize similarity. Using interpolated topographies is advantageous because they reflect not only the IC's weight to each electrode, but the relationship between electrodes.

This leads to following pairing algorithm:

1. For each IC, compute the interpolated topography values used for visualization (e.g., Figure V.2). I do this using the EEGLAB function *topoplot*, which maps the electrode weights to a disk and returns 3409 interpolated values for each map.
2. Compute the absolute value of the cosine of the angle between the interpolated IC topographies of each possible pair to quantify their similarity.

[61] The activation of IC *j* is determined by its unmixing vector, $\mathbf{w}_j$, and the scalp data. However, the topography of IC *j,* $\mathbf{a}_j$, is determined by the unmixing vectors of all the other ICs, as it must be orthogonal to them. Thus, for a given unmixing matrix, **W**, one can rather arbitrarily change an unmixing vector, $\mathbf{w}_j$, and, consequently, its activation, without affecting its topography, $\mathbf{a}_j$.

3. Pair together the two most similar ICs and remove them from further consideration.

4. Repeat Step 3 until all the ICs are paired.

This procedure "greedily" pairs up the most similar ICs, which is preferable to attempting to maximize the similarity of all pairs. Since it is likely that not all of the ICs replicate across data sets, attempting to maximize the similarity of all pairs tends to sacrifice highly similar pairs in order to avoid highly dissimilar pairs.

V.B.3 How to Test for Significant Similarity

V.B.3.a Activation and Topography Similarity Metrics

Once potentially homologous pairs of ICs have been identified, all that remains is to determine which pairs are similar enough to be considered effectively the same. As mentioned in the previous section, this requires the comparison of both IC topographies and activations. Again, to resolve the ICA scaling ambiguity I normalize the topographies of each IC to unit length and choose polarities that maximize the topography similarity of the pair. This facilitates interpretation of IC topography similarity as the same metric can be used to measure ERP topography similarity (e.g., Section IV.E.3.d). Furthermore, this solution preserves information about the magnitude of each IC's contribution to the scalp data by capturing it in the activation magnitude.

This choice of normalization leads to the following similarity metric for two IC topographies, $\mathbf{a}_i$ and $\mathbf{a}_j$:

$$\mathrm{sim}_{\mathrm{topo}}(i,j) = \left| \frac{a_i^t a_j}{\|a_i\| \|a_j\|} \right| \tag{1}$$

This is the absolute value of the cosine of the angle between the two topographies, a value that ranges from 1 (identical topographies) to 0 (maximally dissimilar topographies).

The complementary similarity metric for two IC activations, $u_i(t)$ and $u_j(t)$, is:

$$\mathrm{sim}_{\mathrm{act}}(i,j) = \max(\mathrm{RV}(i,j), \mathrm{RV}(j,i)) \tag{2}$$

where:

$$\mathrm{RV}(i,j) = \frac{\sum_t \left(\frac{u_i(t)\,\mathrm{sign}(a_i^t a_j)}{\|a_i\|} - \frac{u_j(t)}{\|a_j\|} \right)}{\sum_t \frac{u_i(t)^2}{\|a_i\|}} \tag{3}$$

which is the maximum normalized sum squared difference (i.e., residual variance) between the activations. The measure equals 0 for identical activations and grows towards infinity as the activations become increasingly dissimilar. Note, for the IC activations to be most comparable, they should be derived from the same set of scalp data, **x**(t):

$$u_i(t) = \mathbf{w}_i\mathbf{x}(t) \tag{4}$$

$$u_j(t) = \mathbf{w}_j\mathbf{x}(t) \tag{5}$$

Using residual variance, Equation 3, as a similarity metric has the advantage that the difference between activations is scaled relative to the magnitude of one of the ICs. Thus small differences between large activations will be down-weighted relative to small differences between small activations. However, the residual variance between two activations depends on which IC's activation is in the denominator. I choose to normalize by the smaller activation in Equation 2. Doing otherwise would lead ICs with large activations to appear very similar to any IC with a small activation.

It is worth noting here, how these similarity metrics compare to other metrics used in the literature. The previously mentioned resampling and noise injection methods for assessing IC reliability (Harmeling et al., 2004; Himberg et al., 2004; Meinecke et al., 2002) all use only the cosine of the angle between IC activations. This is problematic as it ignores an IC's topography and magnitude. Makeig and colleagues (S. Makeig et al., 2002; S. Makeig et al., 2004) have used automated IC clustering based on IC topographies and the log spectra of IC activity. The metric

however is hard to interpret as it first involves reducing the dimensionality of the features using principal components analysis and then sphering the combined features (i.e., it uses Mahalanobis distance).

V.B.3.b How Similar is Homologous?

While it is relatively straightforward to quantify the similarity of two ICs with metrics like the ones above, it is not clear how similar a pair of ICs need to be before they can be considered homologous. Ideally, one would like to test the null hypothesis that the two ICs are not homologous, but a priori, there seems no way to know what the distribution of non-homologous IC similarities should be. An obvious alternative would be to use a permutation test (Fisher, Lewis, & Embleton, 1993, pg. 60), which does not require assumptions about the distribution of the null hypothesis. That is, one could randomly permute the features of each IC (e.g., the topography weights) thousands of times and compute the similarity of each permuted pair. If the original ICs are more similar than $(1-\alpha)*100\%$ of the permuted pairs, then the original pair would be declared homologous with a p-value less than α. In practice, this method is too permissive. For example, I compared the topographies of the 64 ICs from one participant's visual oddball data to one another. Because the ICs all came from the same decomposition, none of them should be homologous to any of the others. However, of the 2016 possible pairs, the permutation test found that 6% of the pairs were more similar than 4000 random permutations of those pairs. This false positive

rate is over 200 times what it should be. To illustrate why the permutation test fails, the topography of an IC corresponding to blink activity is plotted in Figure V.2 along with the topographies of seven other ICs that the permutation test found to be significantly similar to the blink IC (i.e., more similar than 4000 random permutated versions of each IC). The topographies of the non-blink ICs are clearly distinct from the blink IC, especially those with posterior foci. However, the ICs all share a roughly anterior-greater-than-posterior structure, and it is this low spatial frequency similarity that the permutation test appears most sensitive to.

Another alternative to assuming the distribution of a null hypothesis are so-called "bootstrap" methods, where one approximates the distribution of a random variable from the "empirical distribution" of one's data (Efron & Tibshirani, 1993). It is this general approach that I adopt here. Specifically, to approximate the distribution of the similarity of non-homologous IC pairs, I use the similarity of all possible IC pairs from two different ICA decompositions. Presumably some of the IC pairs are homologous, but at most there will be n homologous pairs out of the n^2 possible pairs, where n is the number of electrodes. So the bias will be relatively small, and to the extent that it does bias the approximation, it will make the approximation more conservative (i.e., the test will be less likely to declare two ICs homologous). Note, for each participant, $3n^2$ pair similarities are obtained (n^2 for each of the three comparisons: "Half A" ICs to "Half B" ICs, Half A ICs to the original ICs, Half B ICs to the original ICs). Finally, to increase the stability of the empirical distribution, I combine the $3n^2$ similarities from all participants.

To illustrate, the empirical distribution of IC topography similarities from both visual oddball tasks (see Section IV.B) are plotted in Figure V.3 (top). The mode of the distribution is approximately 0, maximally dissimilar. As similarity increases, the frequency of pairs decreases until there is a small spike in frequency near maximum similarity, 1. This spike presumably reflects homologous pairs. The empirical distribution of IC activation similarities is plotted in Figure V.3 (bottom). Again, there is a small peak at the maximally similar end of the scale. The mode of the distribution is around 2.23. This is slightly greater than the expected similarity of two uncorrelated activations of equal variance, 2.

V.B.3.c Choosing a Critical Region

Given the above similarity distributions, all that remains is to select a critical region for rejecting the null hypothesis (i.e., that the a pair of ICs are as similar as a pair of ICs chosen at random) at a particular α level. Because it is necessary to test both IC features, the critical region is a segment of the two dimensional joint similarity distribution (Figure V.4). Obviously the critical region should cover the most similar corner of the joint distribution, but several shapes are possible (e.g., rectangular, a quarter circle). I use a backwards "L" shaped region (outlined in pink in Figure V.4), which favors IC pairs with one highly similar feature over pairs with two somewhat similar features.

The area of the critical region should be just big enough such that all ICs from one decomposition could be significantly similar to exactly one IC from the other

decomposition. This requires the critical region to contain 1/*n* of the empirical distribution. To accomplish this, I construct the critical region from two rectangles. The lower edge of one rectangle is set at the 90th percentile of the topography similarity distribution (.72 in Figure V.4) and the lower edge of the other is set to the 90th percentile of the activation similarity distribution
(2.2 in Figure V.4). The inner edge of each rectangle is then grown, one sample at a time, until both rectangles together contain 1/*n* of the empirical distribution.

It is important to note that in practice, the Type I error rate of this test will be higher than 1/*n*, even if the empirical distribution does accurately approximate the true distribution. This is because the ICs are first paired according to their topographies and then tested, thereby increasing the chances of rejecting the null hypothesis. While this bias makes it impossible to know what the true Type I error rate of the test is, it does not invalidate the method. This is because activation similarity is also required to reject the null hypothesis, but activations are not used to pair ICs.

V.B.4 An Example: Visual Oddball Data

To more clearly explain this conceptually simple but somewhat complicated method for assessing IC reliability, I illustrate the method here using the data from the visual oddball tasks described in the previous chapter (Section IV.B). The EEG from these tasks was pre-processed as described in Section IV.D with the exception that 1 second epochs (100 ms before to 900 ms after stimulus onset) were used to avoid any overlap between epochs.

First, for each participant, the epochs of data time-locked to targets and standards are separated and sorted according to their time of occurrence. Odd and even target and standard epochs are split to form two subsets of data, "Half A" and "Half B." ICA is applied separately to each subset and the full data set, producing three sets of 64 ICs. The similarity of all possible pairs of ICs from Half A to Half B, from Half A to the full data set, and Half B to the full data set are computed (Section V.B.3). These similarity measurements from all sixteen participants are combined to approximate the distribution of similarities of randomly paired ICs (Figures V.3-4). To get some sense of the conservativeness of the critical region of the distribution, one can compare the topography similarity critical values with the previously presented within-participant ERP effect topography similarities (Table IV.3). The critical values are less than within-participant P3b topography similarity, greater than the within-participant P6 topography similarity, and straddle within-participant N4 topography similarity.

Next, ICs from each subset are paired (Section V.B.2) with an IC from the full data set, producing triples of ICs (e.g., Figures V.5-8). If each member of the triple is significantly similar (Section V.B.3.b) to the other two members of the triple, then the IC from the full dataset is considered reliable.

Figures V.5-8 present the topographies of all the ICs from Participant 1's visual oddball data and the topographies of each IC's match from the two subsets of data. ICs are numbered in decreasing order of the mean variance of their scalp-projected activations. Thus ICs that contribute a lot to the scalp data (e.g., IC 1,

shown in Figure V.5, which accounts primarily for blink activity) have low numbers. Of the 64 ICs from the full data set, the method finds 34 to be reliable. The reliable ICs are generally the most physiologically plausible (e.g., ICs 4, 10, and 14 shown in Figures V.9-11) and have large scalp-projected activations (e.g., only one of ICs 49-64 is reliable), but there are some exceptions. For example, IC 7 can be fit with a single midline dipole (see Section V.E.2 for dipole modeling details) with a residual variance of .08 and it contributes a relatively large amount to the scalp data. However, there are no ICs with similar topographies from the data subsets. ICa 6 and ICb 18 are similar to complementary halves of IC 7, but the activations of the ICs are not very similar (Figure V.12). IC 3 is a less physiologically plausible component but its activation appears more strongly at the scalp. It also does not replicate in either of the data subsets. The ICs from the data subsets with the most similar topographies are ICa 3 and ICb 2, but their activations are rather different (Figure V.13) and their topographies suggest distinct anatomical sources.

Some of the ICs that do not meet the criteria for reliability are less clearly unreliable. For example, IC 30 replicates in Half B and pairs with a similar component from Half A, ICa 22 (Figure V.14). However, the topography of ICa 22 is not quite similar enough to ICb 34 (r_{cos}=.71) and thus it is deemed unreliable. A similar example is IC 8, which also replicates in Half A but not in Half B (Figure V.15). Both IC 30 and IC 8 are probably reliable, but there is not enough data to robustly learn them in both subsets. Resampling methods, such as the aforementioned

bootstrap and jackknife, that include more than 50% of the data in their resamples may be better at extracting these borderline reliable ICs.

V.C Comparison with Bootstrap Resampling

As mentioned previously, using 50% splits of the data has the disadvantages of being influenced by any bias in the method used to split the data and of only using half of the data in each resample. To evaluate how serious these disadvantages are, the 50% split reliability test was compared to bootstrap resampling, which has neither of these shortcomings.

This comparison was made using the data from the visual oddball tasks described in the previous chapter (Section IV.B). The EEG from these tasks was pre-processed as described in Section IV.D with the exception that 1 second epochs (100 ms before to 900 ms after stimulus onset) were used to avoid any overlap between epochs.

For each participant, 1000 resampled data sets were created by randomly selecting η epochs from the original data set with replacement (where η is the number of epochs in the original data set). Extended INFOMAX ICA was then applied to each resampled data set using the same parameters described in Section IV.D, producing 1000 additional sets of ICs for each participant. ICs from each resampled data set were uniquely paired with an IC from the original dataset using their interpolated scalp topographies (Section V.B.2). The topography and activation

similarities of each pair were computed using the metrics described in Section V.B.3.a. The median similarities of each original IC to its 1000 counterparts from the resampled data sets were used to quantify the reliability of the two features of that IC. Median similarities were used to mitigate the influence of extreme outlying values.

As described in the previous section, 50% splits of each participant's data were also made and decomposed with ICA. The ICs from the splits were paired with the original ICs and their median similarities computed, just as was done with the ICs from the bootstrap resampled data sets.

Figure V.16 shows that the two resampling methods produce strongly correlated reliability estimates. In other words, the ICs whose topographies and activations reproduce the best across the two 50% splits also tend to be the most reproducible according to the 1000 bootstrap resamples. Thus using two 50% splits reasonably approximates the results of the more time intensive bootstrap resampling.

Although reliability estimates from the two resampling techniques are similar, one might expect that bootstrap resamples would find more ICs to be reliable since they can include more than 50% of the data in each resample. To test this possibility, all ICs whose median topography and activation similarities met the significance criteria used in the previous section (e.g., topography similarity>.72 and activation similarity<1.95 or topography similarity>.87 and activation similarity<2.2) were deemed reliable. On average, 44 ICs per participant were reliable using bootstrap median feature similarities. In comparison, 43 ICs per participant were reliable using

the 50% split median feature similarities[62]. Using a one-tailed paired t-test, this difference failed to reach significance ($t(15)=1.06$, $p>.15$). Consequently, there appears to be no advantage to the more time intensive bootstrap resampling. However, with smaller or noisier data sets this might not be the case.

V.D Reliability vs. Amount of Data

Because the reliability criteria proposed here are estimated from the data, there is a danger that its performance may greatly degrade with decreasing amounts of data. In particular, since the similarity criteria are loosened to include *n* pairs of ICs (where *n* is the number of electrodes), it is possible that the algorithm will still find a large number of ICs to be reliable even with insufficient data.

To test for this possibility, I assessed the reliability of ICA decompositions of decreasing amounts of four data sets. The two data sets consisted of the data from the visual oddball (Section IV.B) and sentence acceptability experiments (Section IV.C) described in the previous chapter. A third set of data consisted of 30 electrode EEG from eight participants performing visual oddball tasks (Kemmer, unpublished). The fourth set of data was eight participants from a sentence acceptability task involving grammatical and ungrammatical sentences (Kemmer, Coulson, De Ochoa, & Kutas, 2004). For details of the latter two data sets see Groppe et al. (2004). For the first two

[62] This is a greater number of reliable ICs than found by the 50% split reliability test reported in the previous section. This is because the previous test required that three pairs of ICs be significantly similar for the original IC to be deemed reliable. Here, only the median similarity need be great enough for reliability. This was done to facilitate comparison with bootstrap resampling.

data sets (Experiments 1 & 2), ICA was performed on 100%, 66%, 50%, and 25% of the full data sets. The EEG from these tasks was pre-processed as described in Section IV.D with the exception that 1 second epochs (100 ms before to 900 ms after stimulus onset) were used to avoid any overlap between epochs.

For the 30 electrode data (Experiments 3 & 4), ICA was performed on 100%, 66%, 40%, 20%, and 2% of the full data sets. To split the data into *a*% subsets, epochs were first divided by stimulus type and sorted in order of occurrence as described in Section V.B.1. Subsequently, every *b*th epoch (where b=200/a) and the subsequent epoch were placed into the same subset. This approximately balanced the number of epochs of each stimulus type and the latencies at which the epochs were recorded across subsets. For example, for each participant in Experiment 4, this procedure produced 1, 3, 5, 10, and 100 data subsets each including 100%, 66%, 40%, 20%, and 2% of the data respectively.

Figure V.17 shows that the number of reliable ICs decreases as the amount of data decreases. Indeed, all four experiments show an approximately linear relationship between the log of the amount data per electrode[2], and the percentage of reliable ICs. Thus the reliability algorithm does depend appropriately on the amount data in the range of data set sizes explored here. Given that EEG datasets would commonly fall above the lower end of this range[63], the algorithm should perform reasonably in practice.

---

[63] One hour of EEG data, recorded at 250 Hz will provide 879, 220, 55, and 14 data points/electrode^2 for 32, 64, 128, and 256 electrode data (respectively). For comparison with Figure V.17, the $\log_{10}$ of those ratios are 2.9, 2.3, 1.7, and 1.1.

V.E Types of Reliable ICs

Given an objective method for assessing IC reliability, it is informative to try to characterize what types of EEG ICs are reliable. This knowledge would be useful for understanding what types of EEG sources ICA is likely to extract (e.g., how big does a source need to be before it can be reliably extracted?) and for evaluating heuristics for determining which ICs are worth analyzing.

In the sections below I evaluate two heuristics that have been used to select ICs for analysis, the percentage of data variance accounted for by an IC and the dipolarity of an IC. I confirm that these heuristics do tend to select reliable ICs. However, there is a non-negligible frequency of exceptions. ICA is able to reliably extract some small or non-dipolar components and not all large, physiologically plausible ICs are reliable. These results indicate a need to test IC reliability independently of these heuristics.

V.E.1 Large ICs

Research using simulated EEG/ERP data (Ghahremani, Makeig, Jung, Bell, & Sejnowski, 1996; S. Makeig, Jung, Ghahremani, & Sejnowski, 2000) suggests that the most reliable ICs should have the largest scalp activations. If this is true it would add credence to a commonly used heuristic of ignoring ICs that contribute little to the scalp data (Fogelson, Loukas, Brown, & Brown, 2004; Marco-Pallares, Grau, & Ruffini, 2005; Salustri & Kronberg, 2004) and it would suggest that ICA may not be useful for extracting small components.

To determine if relatively large ICs are indeed the most reliable, I quantified the relative size of each IC's contribution to the raw scalp data (i.e., no artifact ICs were removed) with the percentage variance accounted for metric (see Section III.D.1) for all four data sets described in Section V.D. On average, reliable ICs account for about twice the percentage of variance accounted for by unreliable ICs (Figure V.18) in all four data sets. However, around 25% of the unreliable ICs account for as much variance or more than the median percentage variance accounted for by reliable ICs. Likewise, about 25% of the reliable ICs fall below the median percentage variance accounted for by unreliable ICs. Thus percentage variance accounted for is far from a perfect predictor of IC reliability. For example, Figure V.13 shows an IC from Participant 1's oddball data that is clearly unreliable but accounts for 5% of the data variance. In contrast, IC 53, a component that reflects three bursts of muscle activity, is quite reliable but only accounts for .06% of the data variance (Figure V.19). Such EEG artifact ICs would generally be rejected from analysis and most of these small but reliable ICs probably reflect mostly muscle noise. However, the fact that ICA can reliably extract them does show that in principle it could extract small EEG components as well.

V.E.2 Dipolar ICs

Another criteria that has been used to select ICs for analysis is the "dipolarity" of an IC's topography (Debener et al., 2005; Onton, Delorme, & Makeig, 2005). Dipolarity is the extent to which an IC's topography is consistent with a single

equivalent current dipole (ECD) source or two bilaterally symmetric ECD sources. ECDs are commonly used approximations of focal (i.e., extending over a few square cm of cortex or less) EEG sources (Baillet, Mosher, & Leahy, 2001). The logic of this metric is motivated by Debener and colleagues (2005) as follows:

> Ideally, an independent component should account for synchronous activity within a connected cortical domain, and accordingly its scalp projection should match a single equivalent current dipole (or sometimes two bilaterally symmetric dipoles). In practice, many biologically plausible independent components are characterized by scalp maps fitting the projection of a single equivalent current dipole. Therefore, the goodness of fit for modeling each independent component scalp map with a single equivalent current dipole was used to quantify component quality. (pg. 312)

For dipolarity to be a useful criteria for choosing ICs to analyze, it should also be the case that dipolar ICs are reliable. To determine if this was true, the algorithm used by Debener et al. (2005) was used to automatically model all of the ICs from the experiments described in Section V.D with a single ECD. More specifically, the topography of each IC (minus the EOG and right mastoid weights) was mapped to a standard four-shell spherical head model using ideal electrode locations (Appendices 2 & 5). For each IC topography, an initial ECD location and orientation was chosen by searching a coarse grid of possible locations. Initial ECD parameters were then improved by iterative, non-linear optimization. This procedure is part of the DIPFIT 2 software (Oostenveld & Oostendorp, 2002) included with the EEGLAB toolbox (Delorme & Makeig, 2004).

The fit of an ECD model is commonly measured using the residual variance error metric, which is simply 1 minus the percentage variance accounted for metric (Section III.D.1). As done by Debener et al. (2005), ICs fit by an ECD with a residual variance of 10% or less were considered dipolar. Figure V.20 (Top) shows that most dipolar components were found to be reliable in each data set. Indeed, there is a significant relationship between dipolarity and reliability as reliable ICs are more likely to be dipolar than unreliable ICs (Figure V.20: Bottom). However, dipolarity far from guarantees reliability as a high percentage (between 38% and 22%) of dipolar ICs were not.

It is important to note that EEG source localization is a notoriously difficult problem and that the localization procedure used here was relatively crude. Better estimates of IC dipolarity could come from hand tuning localization results, using actual electrode locations instead of ideal spherical locations, and using a more accurate head model. With better estimates, the reliability of dipolar ICs may differ from what is reported here. One should also note that ICs corresponding to EEG artifacts were included in this analysis for simplicity. As the reliability of EEG artifact ICs is not necessarily of interest and EEG artifact ICs are not necessarily dipolar, this analysis should be repeated without them, although given that EEG artifact ICs probably tend to be reliable and dipolar, their inclusion probably increases the percentage of reliable dipolar ICs. Nonetheless, as they stand, these results do speak to one publicly available, published method for localizing ICs, and they illustrate the

importance of checking any criteria for IC physiological plausibility to ensure that it selects reliable ICs.

V.F Improving IC Reliability

To apply ICA to a dataset, a number of practical decisions need to be made: how to preprocess the data, which ICA algorithm to use, and what parameters of that algorithm to use. The reliability test presented in this chapter can be used to explore the different answers to these questions to improve ICA's results. While reliability is not a sufficient condition for a successful EEG decomposition, it is a necessary condition and tuning ICA to improve reliability may improve the quality of ICs in general.

For example, extended INFOMAX ICA was applied to Participant 7's data from Experiment 3 and only eight of the 30 ICs were found to be reliable (Figures V.21-22). This was shockingly low given that the other seven participants averaged 19 reliable ICs and that many of the unreliable ICs appeared physiologically plausible. Attempts to improve the reliability of Participant 7's decomposition by removing extreme epochs or by using another ICA algorithm, second order blind identification, did not improve results. Eventually it was discovered that the reliability of the ICs could be greatly improved by removing the mean of each epoch, instead of the mean of the 100 ms prestimulus baseline (for details see Groppe, Kemmer, Kutas, & Makeig, 2004), before applying ICA. Applied to the zero mean epoch data, extended

INFOMAX ICA now returned 25 reliable ICs (Figures V.23-24). Comparison of the original ICs and the more reliable ICs (Figures V.25-26) reveals that mostly the same ICs were extracted. Only three of the original ICs had no convincing counterpart in the second decomposition (Figure V.26, bottom row). Thus, removing the mean of each epoch appears to enhance the strength of the latent independent components so that they can be extracted from less data. Presumably enhancing the ability of ICA to find latent components increases its accuracy as well.

While this data set is an extreme case, removing the mean of each epoch produces more reliable ICs than removing the 100 ms prestimulus baseline for each participant in all of four experiments analyzed in Section V.D (48 data sets in total). On average, zero mean epoch data produce over 20% more reliable ICs (Figure V.27). It is not clear what causes this difference. Removing the mean of each epoch acts as a leaky high-pass filter, zeroing the DC component and dampening low frequency components. It is not clear why high pass filtering would increase IC reliability. Perhaps the low frequency EEG components (at least when viewed in approximately one second epochs) are more variable and/or have a statistical structure that is difficult for ICA to uniquely decompose (e.g., similar to the hypothetical sources in Figure V.1). On the other hand, it could be that removing the epoch mean does not aid ICA so much as removing the prestimulus baseline impairs it. Raw scalp potentials are highly variable, especially when EEG artifacts are present (as is generally the case when using ICA). Using only a small window to normalize the voltage of each epoch may enhance that variability and mask the latent sources ICA is sensitive to.

Whatever the root cause proves to be, removing epoch means is clearly an improvement over the more conventional 100 ms prestimulus baseline and indicates the utility of using reliability to explore ICA parameters. In practice, researchers have applied ICA to prestimulus baselined epoched data (Debener et al., 2005; S. Makeig et al., 2004) and to continuous data with various degrees of bandpass filtering (Debener et al., 2005; Onton et al., 2005; Tang, Sutherland, & McKinney, 2005). The reliability test presented in this chapter provides a simple, relatively time-efficient way to test these and other various alternatives to improve the robustness and possibly the accuracy of ICA's results.

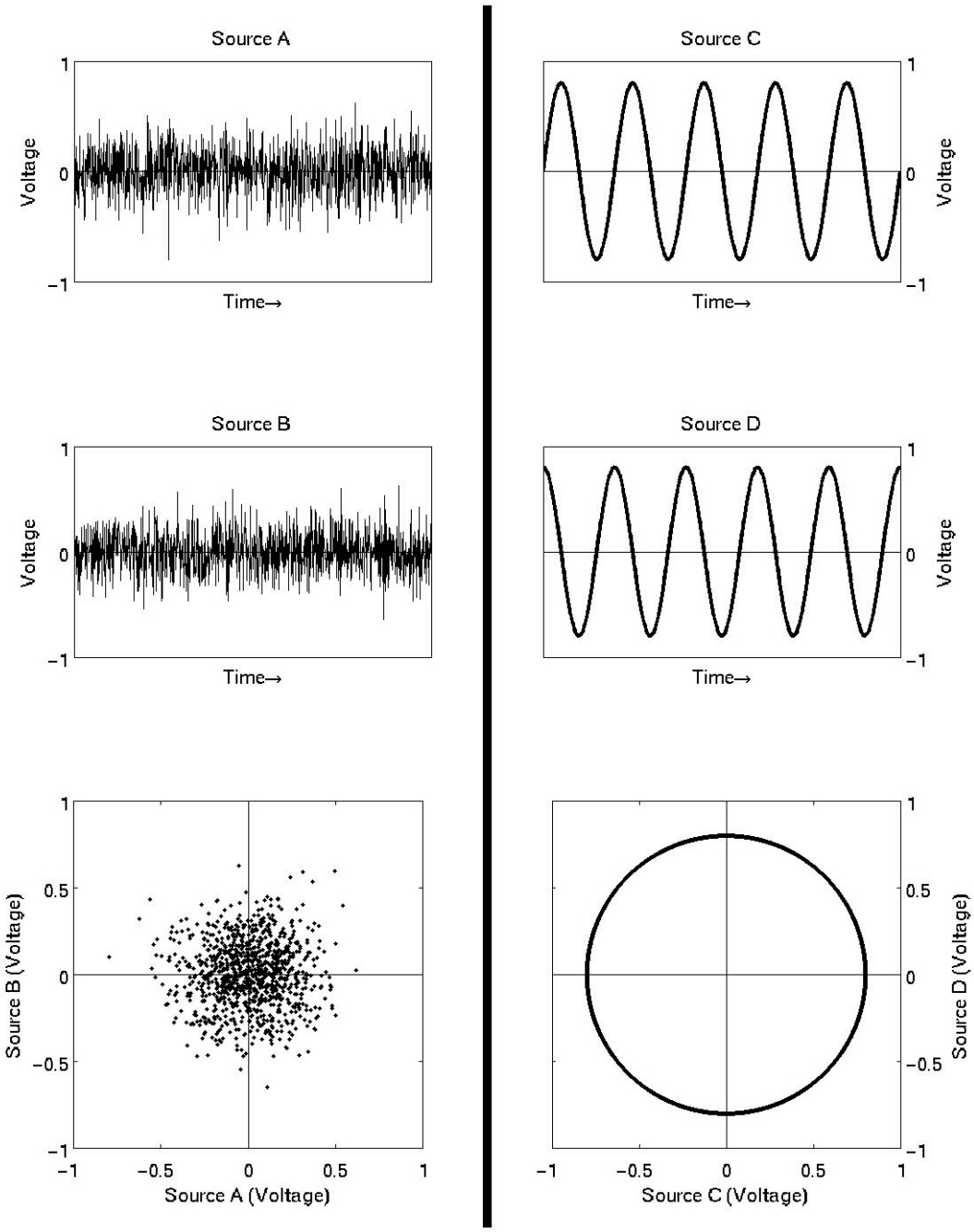

Figure V.1 – Example pairs of hypothetical scalp potential sources that ICA algorithms cannot reliably decompose: white Gaussian sources (**Left**), sin(*t*) and cos(*t*) where *t* indexes time (**Right**). An infinite number of equally independent basis are possible for such data (Meinecke, et al., 2002).

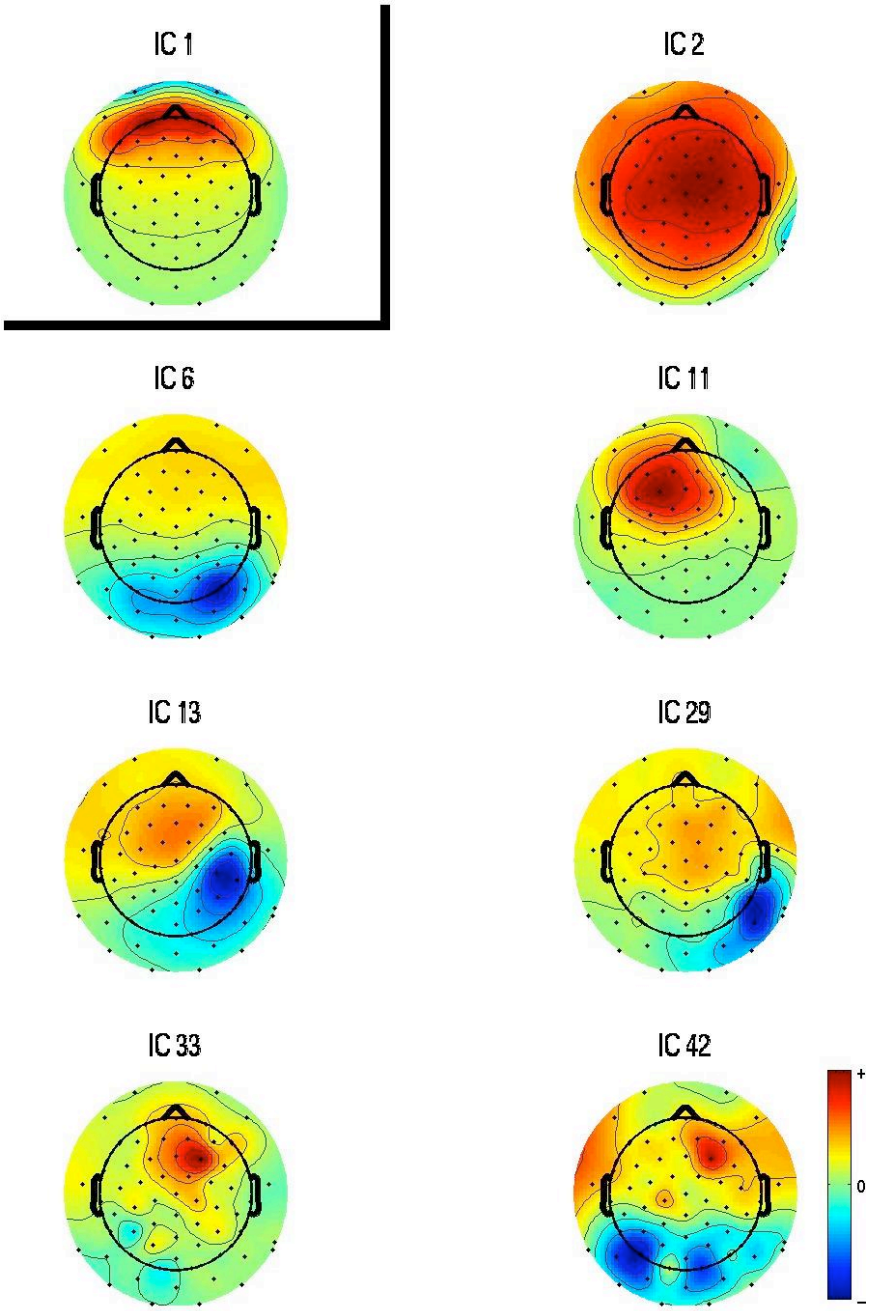

Figure V.2 – The topographies of seven ICs that are significantly similar to the topography of an IC that primarily represents blink activity, IC 1 (top left), according to a permutation test. Unitless topography weights are plotted with a color scale that is unique to each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Electrodes below the head's equator are plotted progressively beyond the radius of the head.

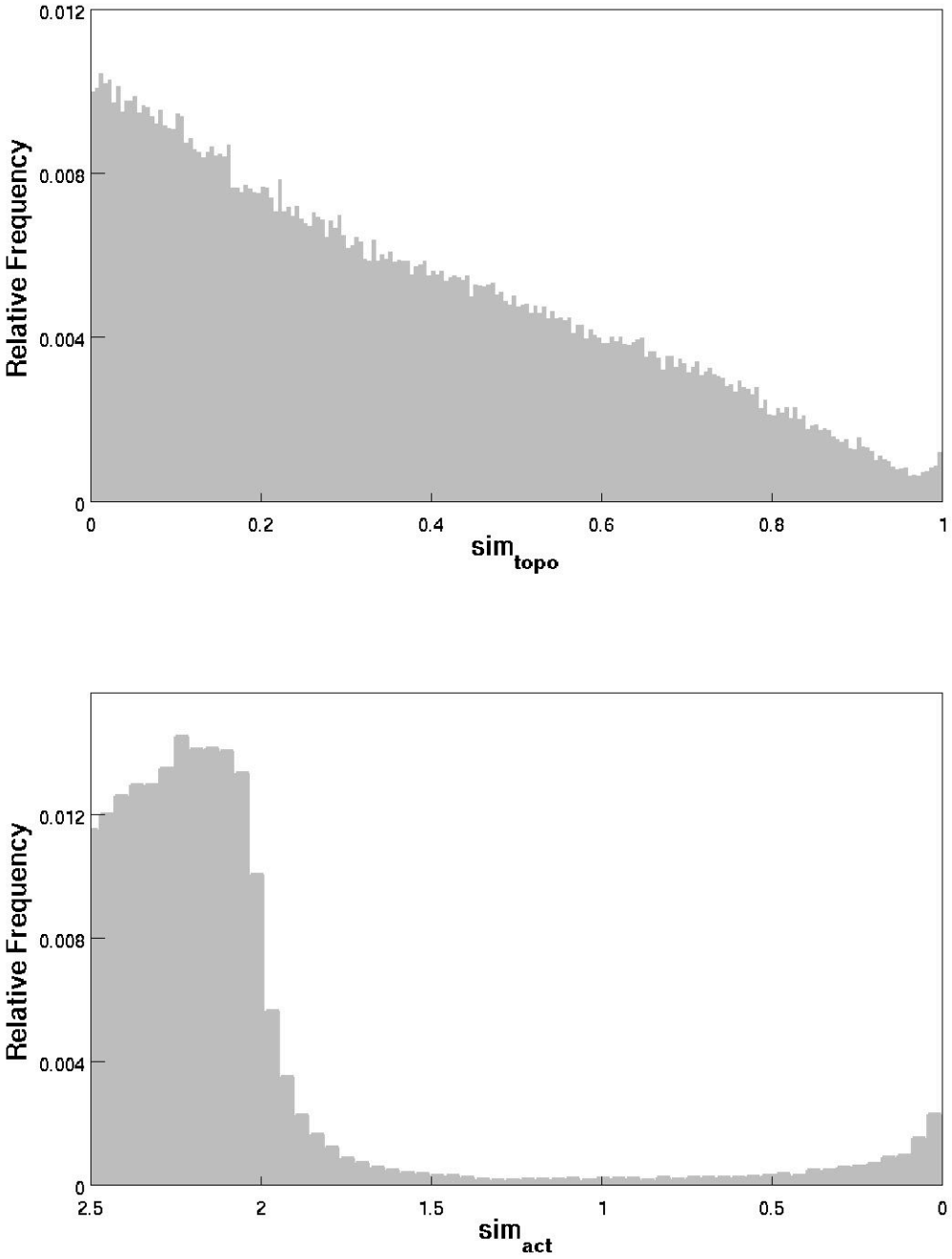

Figure V.3 – Histograms of topography (**Top**) and activation (**Bottom**) similarity for all possible IC pairs. Maximal similarity is on the right for both distributions. The activation similarity distribution tapers off towards infinity (not shown). Histograms are derived from 196,608 IC pairs (i.e., 16 participants, 3 ICA decompositions per participant, 64 ICs per decomposition).

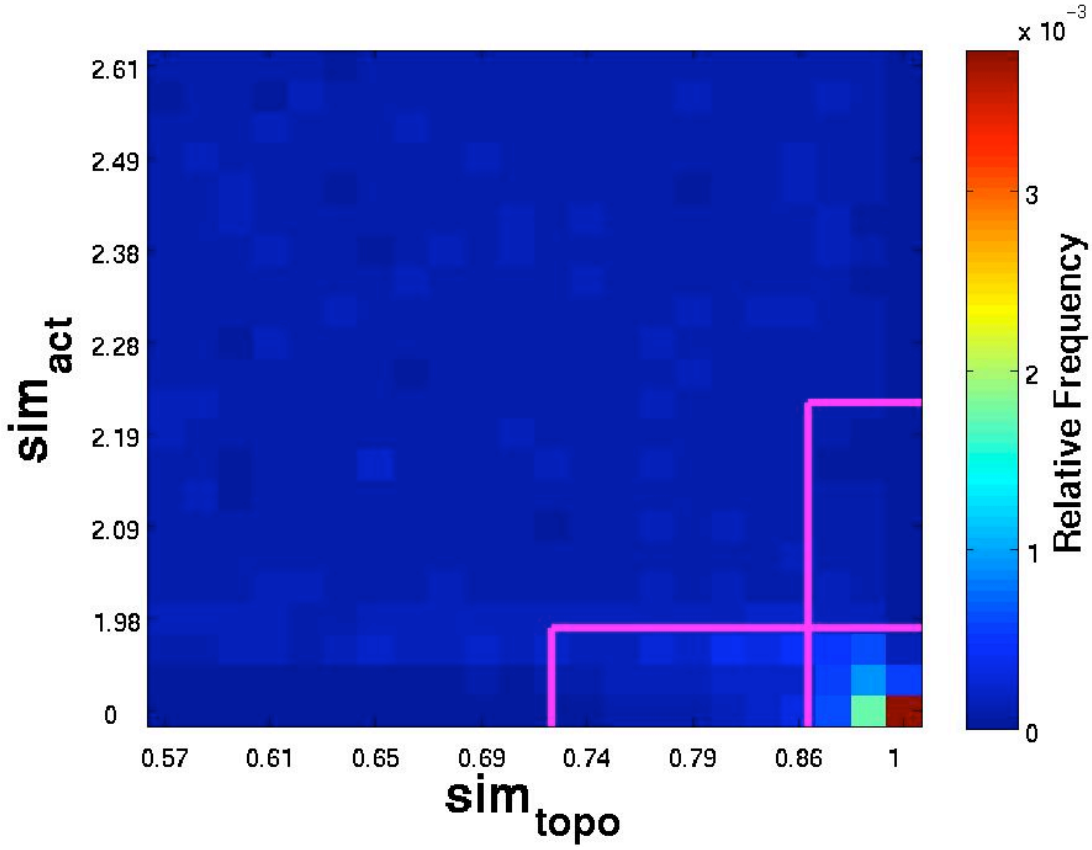

Figure V.4 – Joint distribution of IC pair topography and activation similarities from Figure V.3. To make the two similarity metrics comparable, they are binned in one percentile increments (e.g., 3% of pairs have topographies that are .86 similar or higher). Maximal similarity is represented in bottom right corner. Distribution continues to the left and up (not shown). Pink rectangles indicate backwards "L" shaped critical region that contains 1/64 of all samples. The top left edge of the vertical rectangle is at .72, 1.95. The top left edge of the vertical rectangle is .87, 2.2.

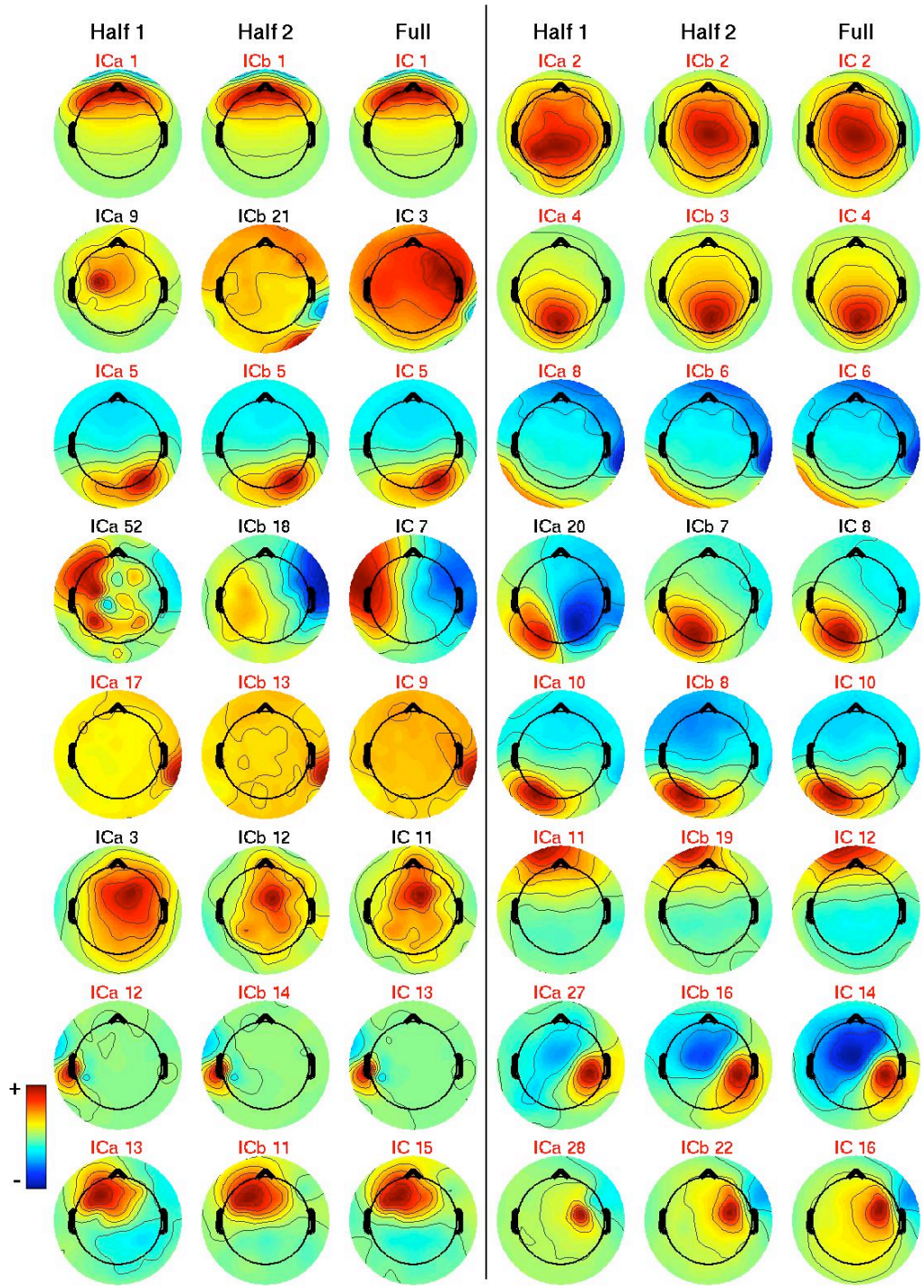

Figure V.5 – Topographies of the top 16 ICs from Participant 1's visual oddball task data (third and sixth column) and their matches from both 50% splits of the data. Unitless topography weights are plotted with a color scale that is unique to each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Triplets of reliable ICs are named in red.

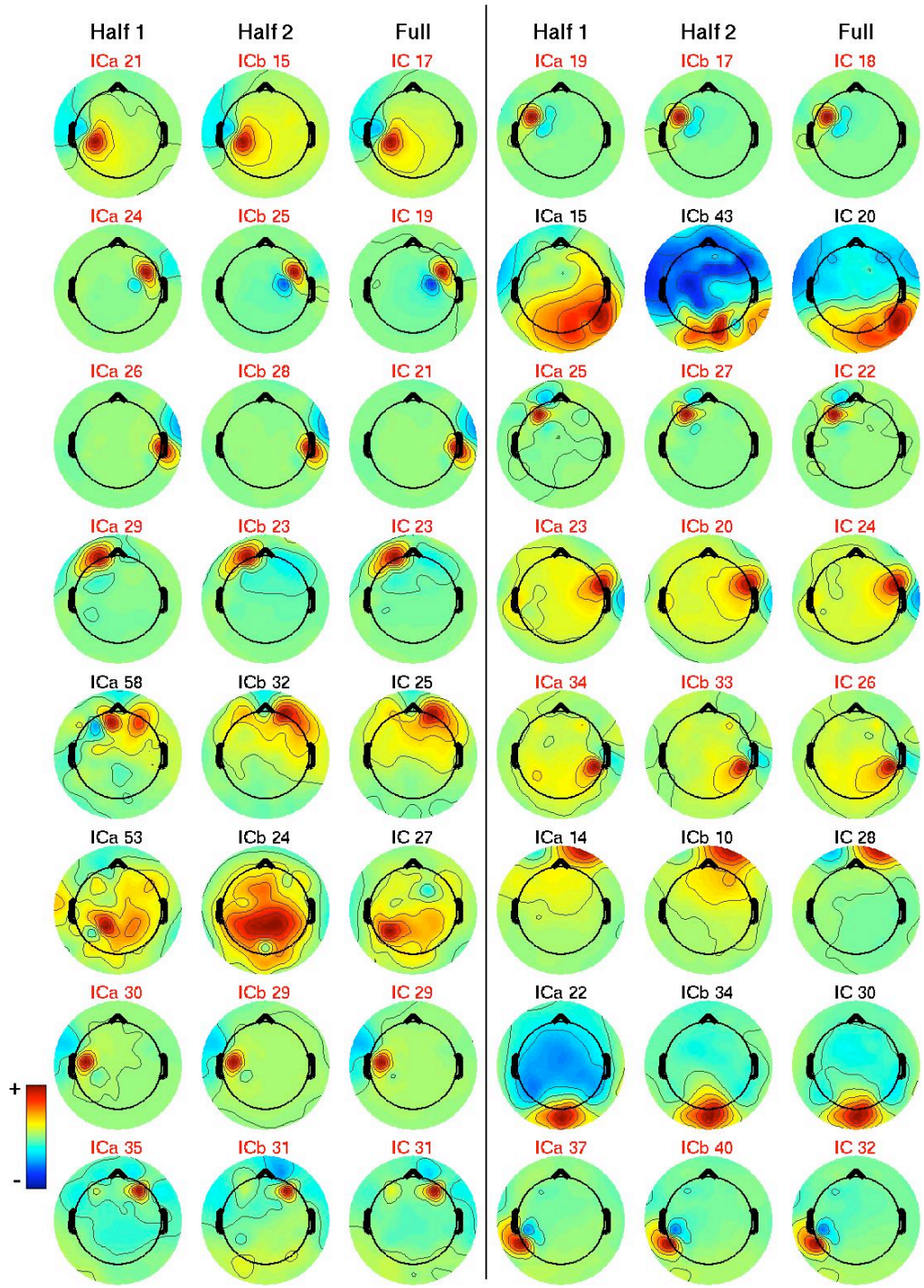

Figure V.6 – Topographies of ICs 17-32 from Participant 1's visual oddball task data (third and sixth column) and their matches from both 50% splits of the data. Unitless topography weights are plotted with a color scale that is unique to each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Triplets of reliable ICs are named in red.

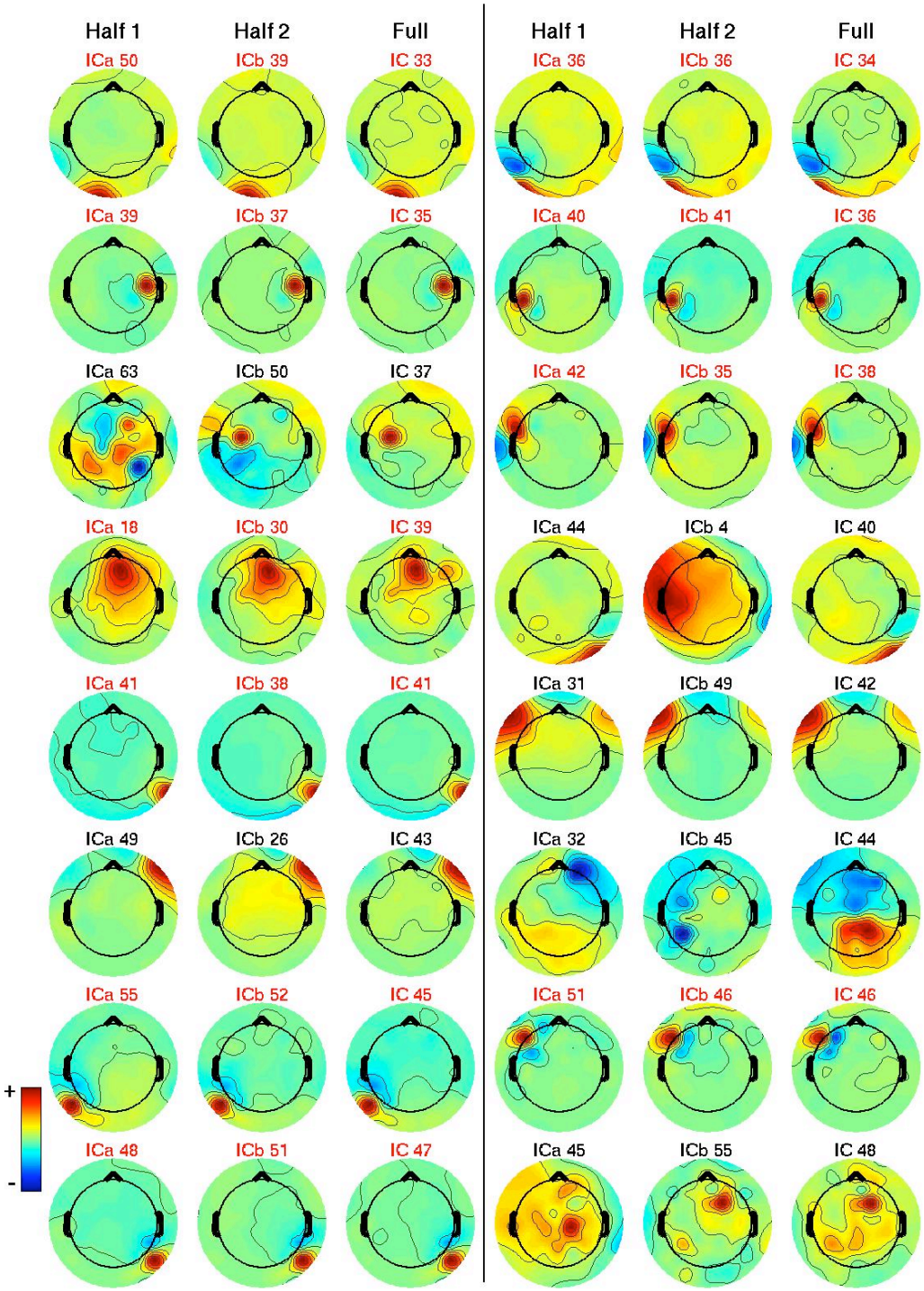

Figure V.7 – Topographies of ICs 33-48 from Participant 1's visual oddball task data (third and sixth column) and their matches from both 50% splits of the data. Unitless topography weights are plotted with a color scale that is unique to each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Triplets of reliable ICs are named in red.

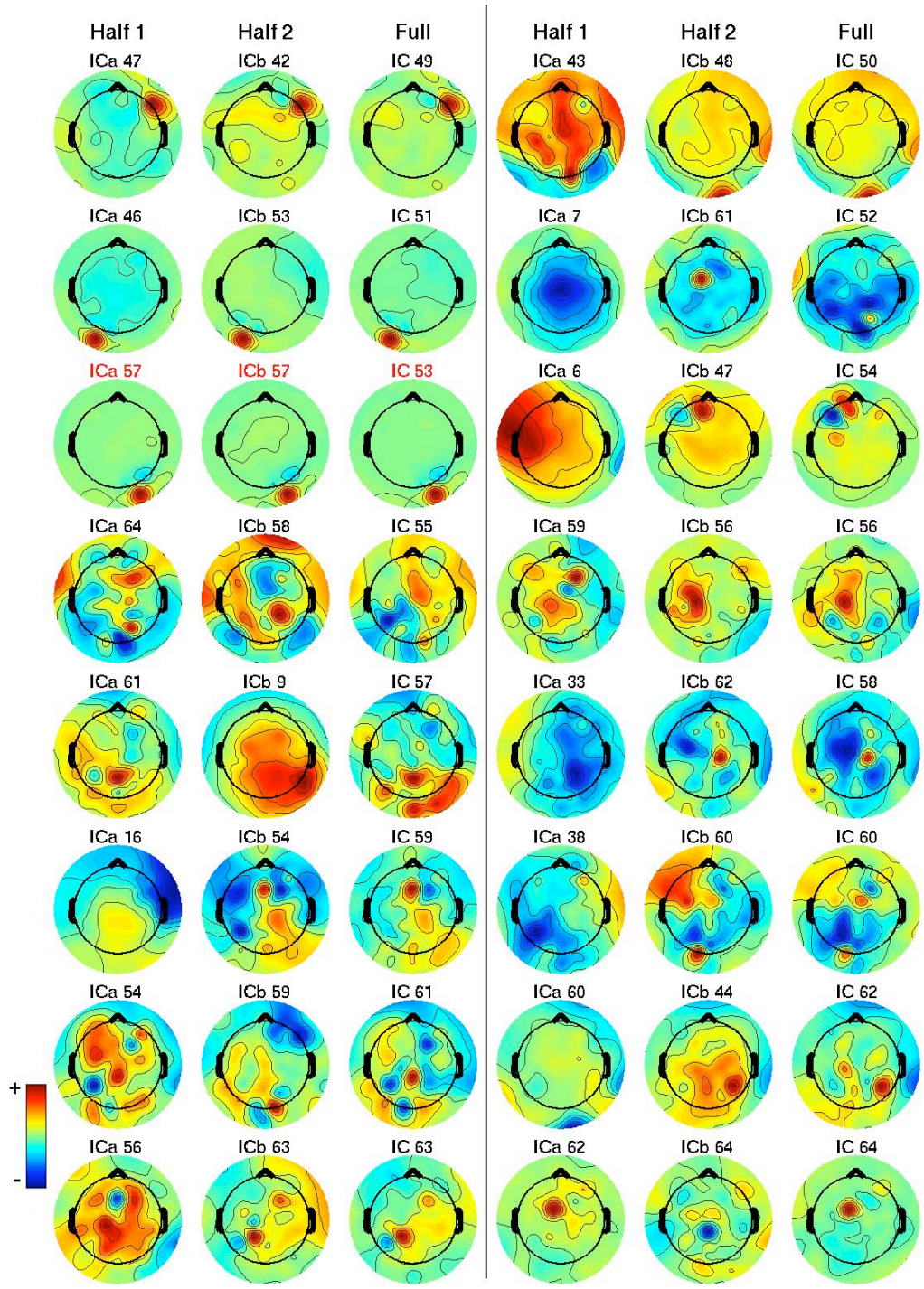

Figure V.8 – Topographies of ICs 49-64 from Participant 1's visual oddball task data (third and sixth column) and their matches from both 50% splits of the data. Unitless topography weights are plotted with a color scale that is unique to each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Triplets of reliable ICs are named in red.

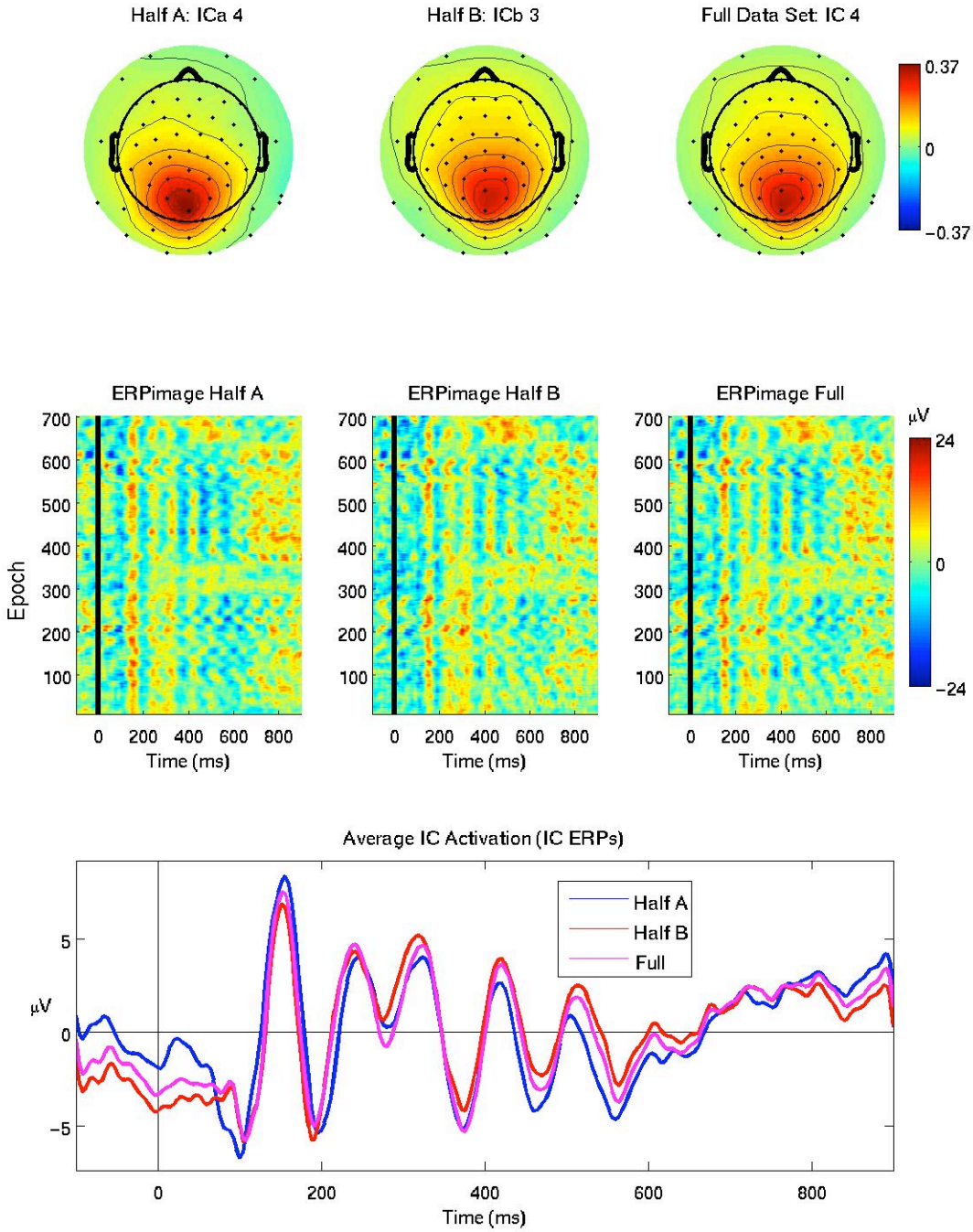

Figure V.9 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 4 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

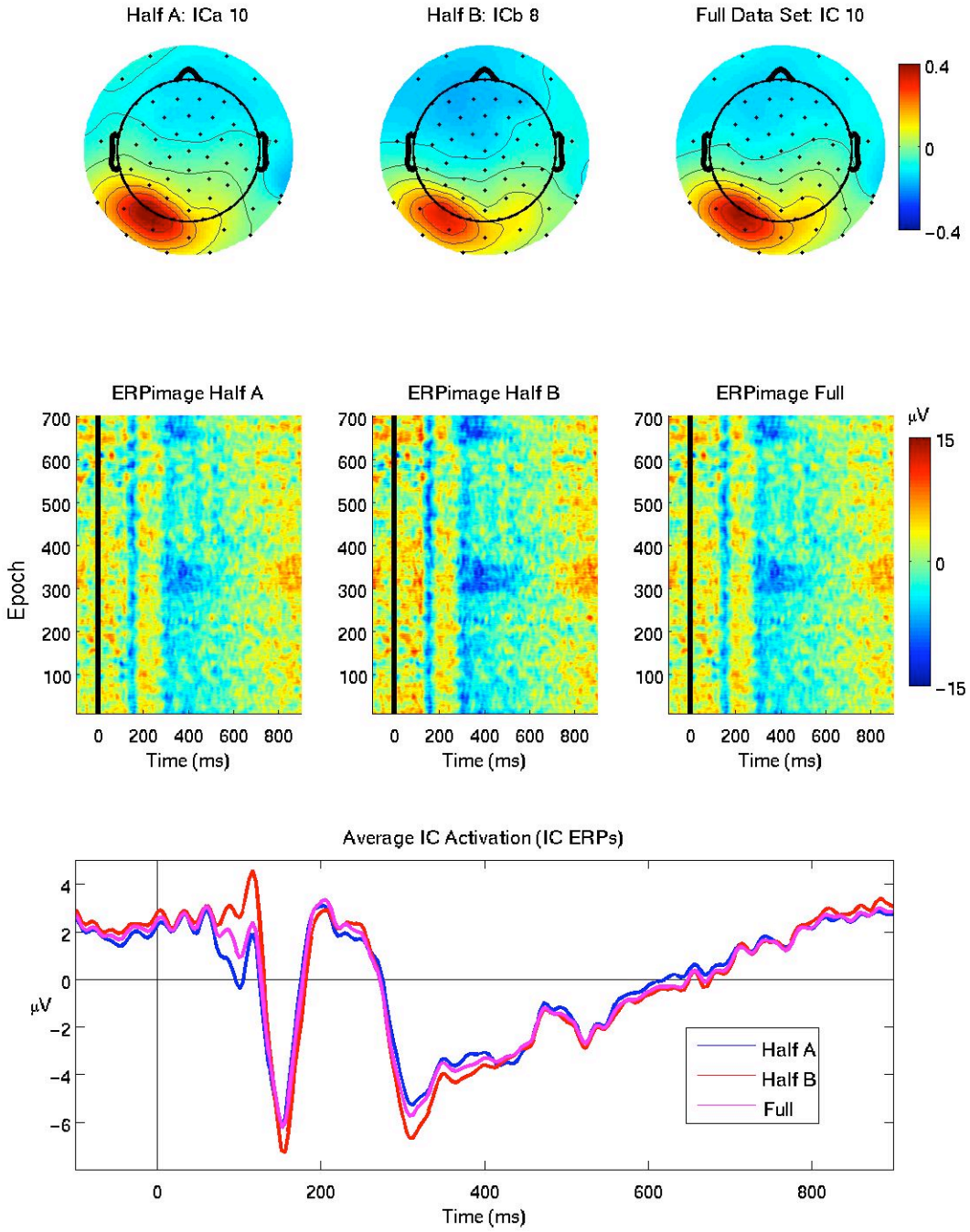

Figure V.10 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 10 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

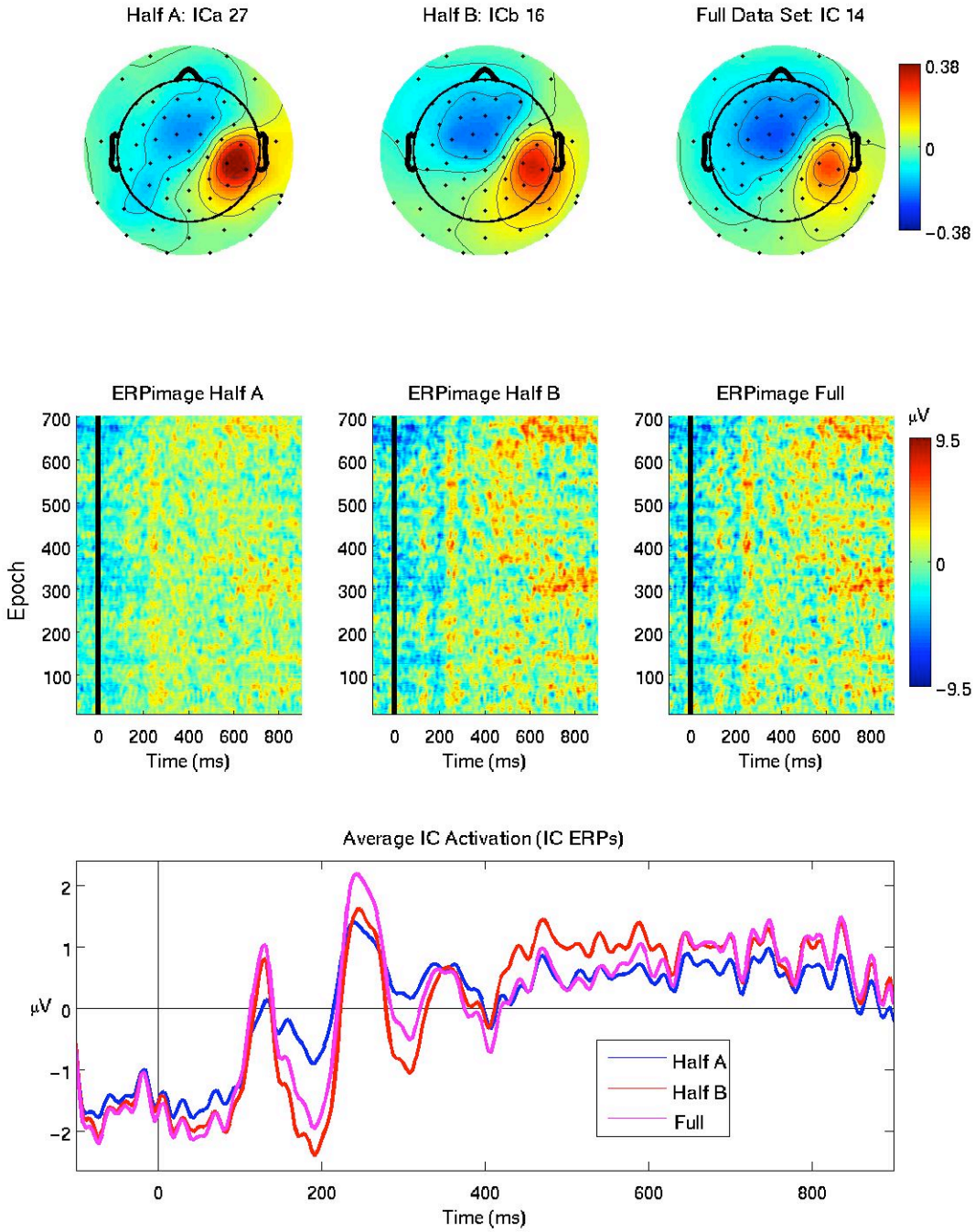

Figure V.11 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 14 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

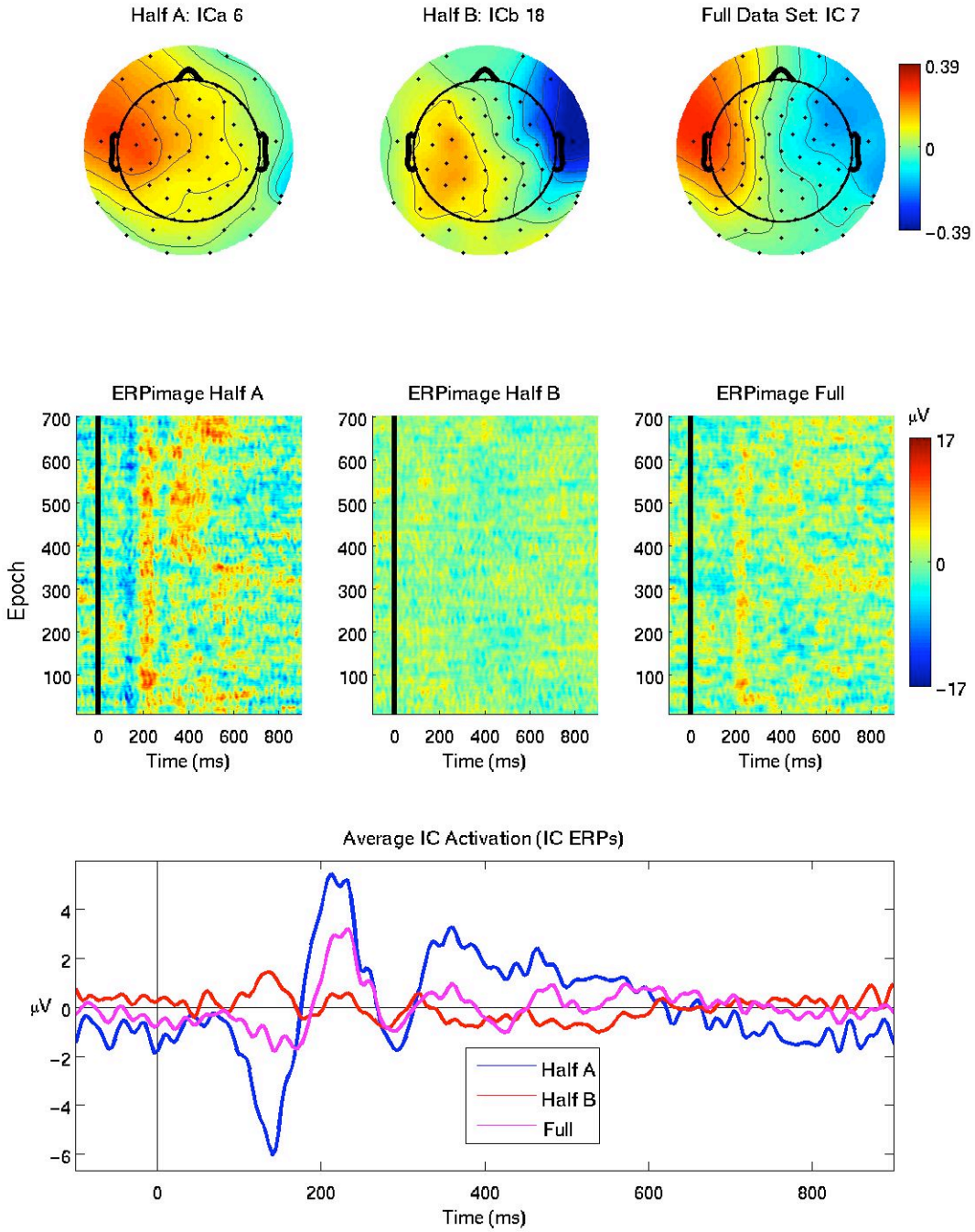

Figure V.12 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 7 from Participant 1's visual oddball data and the two ICs with similar topographies to IC 7's from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

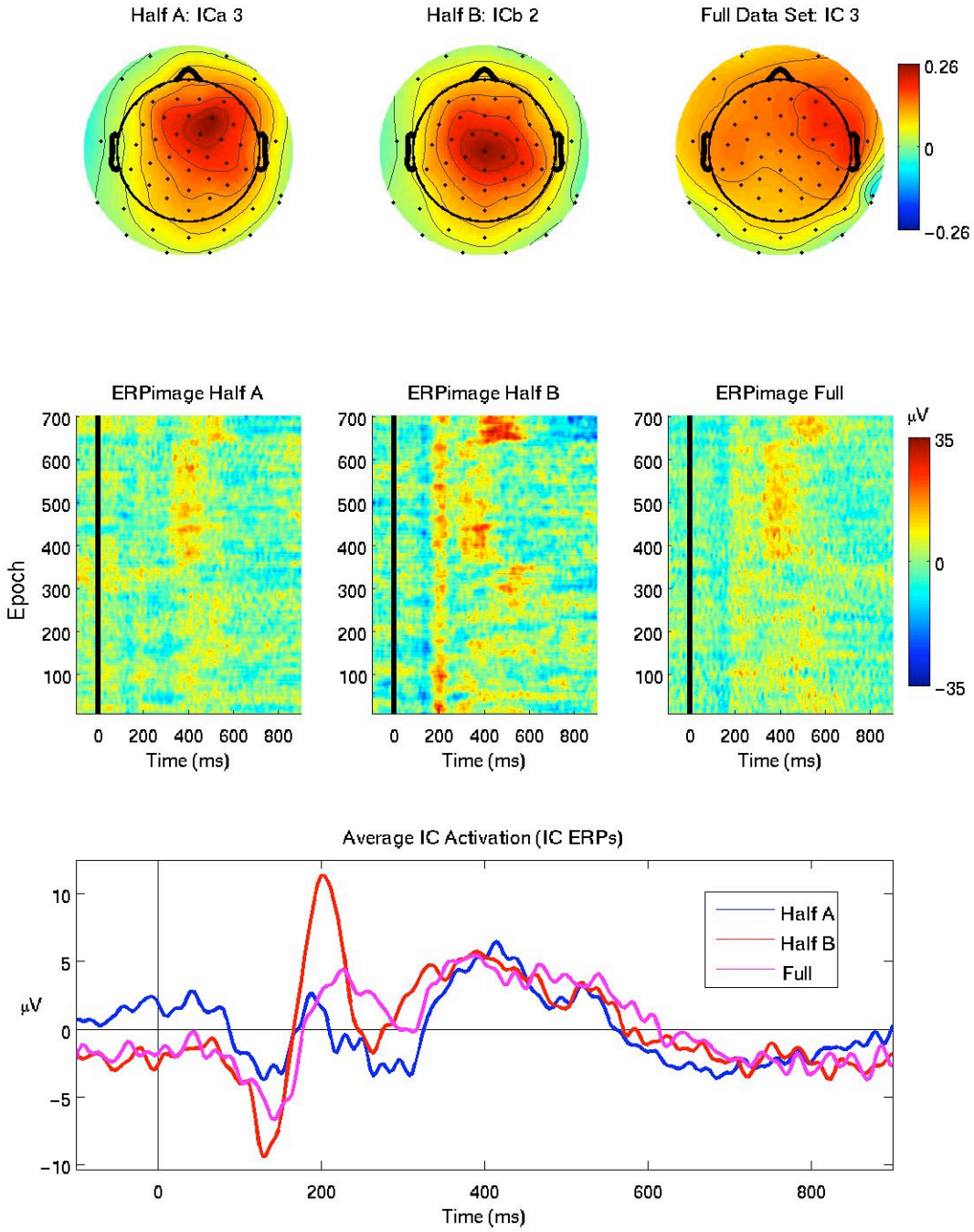

Figure V.13 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 3 from Participant 1's visual oddball data and the two ICs with the most similar topographies to IC 3's from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

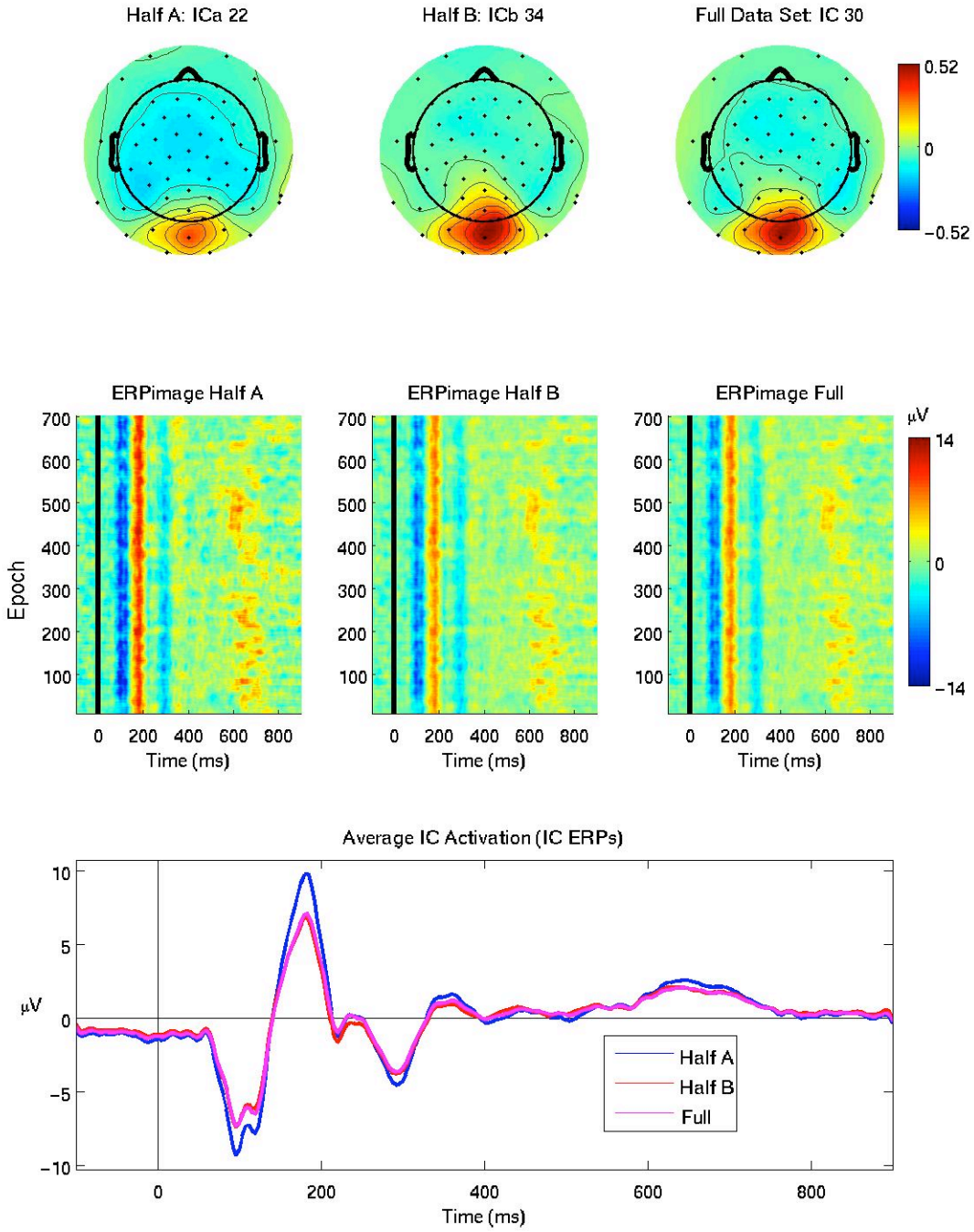

Figure V.14 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 30 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

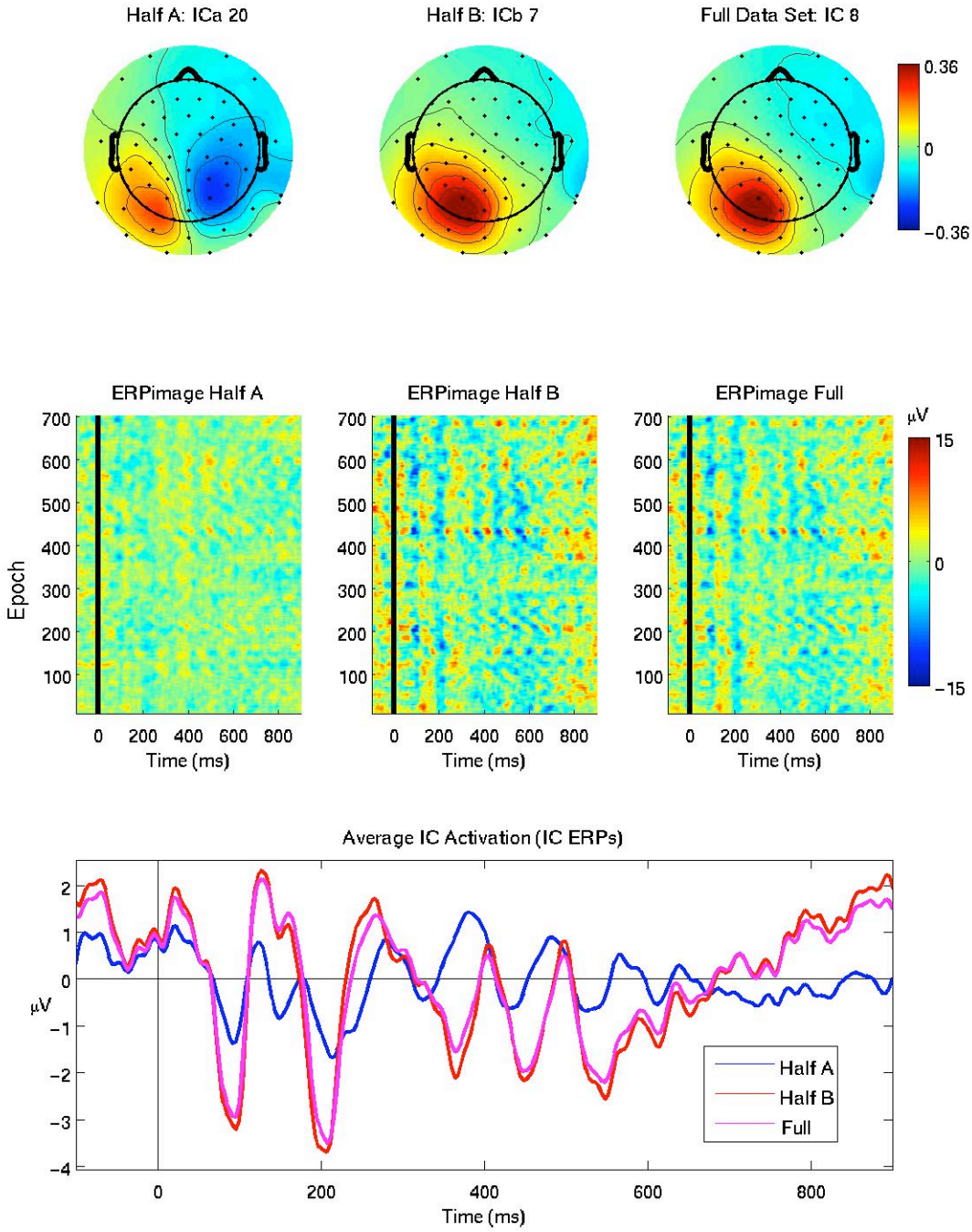

Figure V.15 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 8 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window.

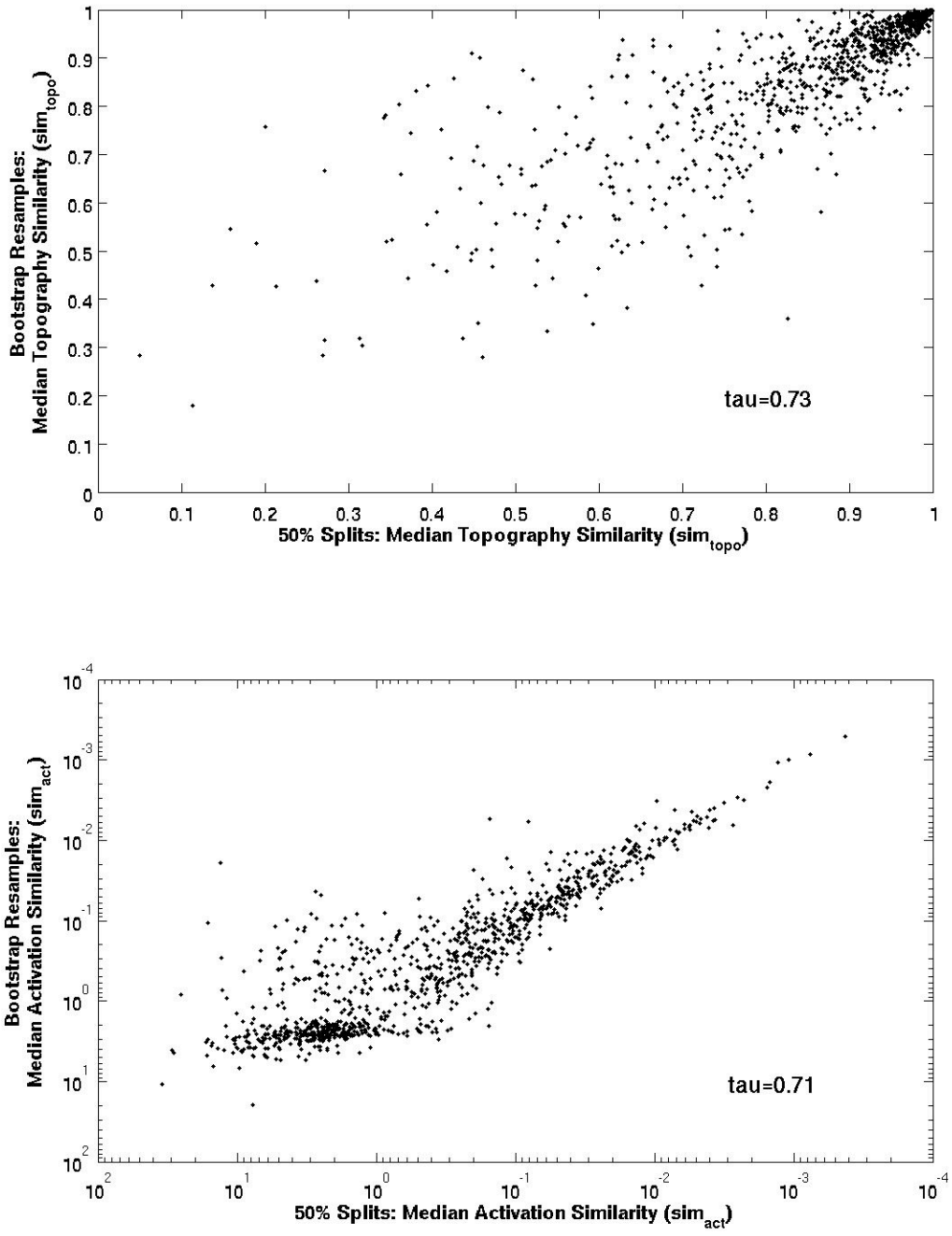

Figure V.16 – IC topography (**Top**) and activation (**Bottom**) reproducibility estimates using 50% splits and bootstrap resamples are highly correlated. Maximal similarity is the top right corner of both scatter plots. Rank correlations are reported using Kendall's tau. Activation similarities are plotted on log scales. 1024 samples per plot (e.g., 64 ICs per each of 16 participants).

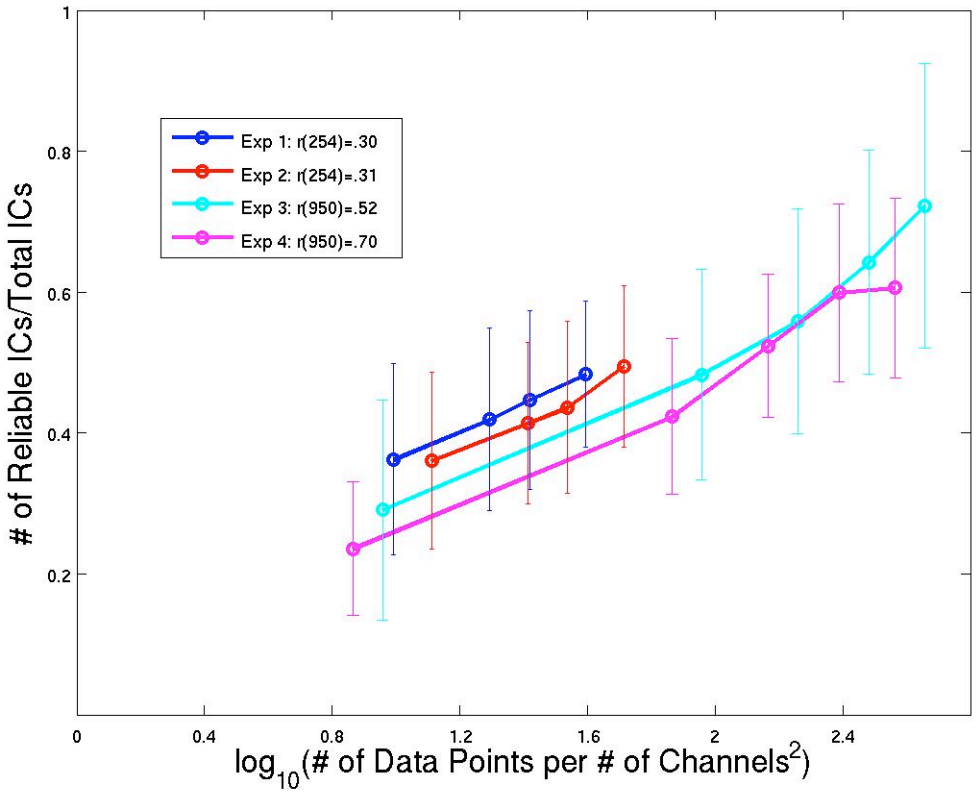

Figure V.17 – Mean percentage of reliable ICs (error bars are standard deviations) as a function of data set size. Modest to strong correlations (see legend) show that the percentage of reliable ICs tends to decrease linearly with a decrease in the log of the number of data points/electrode^2. All correlations are significant ($p \approx 0$, see legend for degrees of freedom).

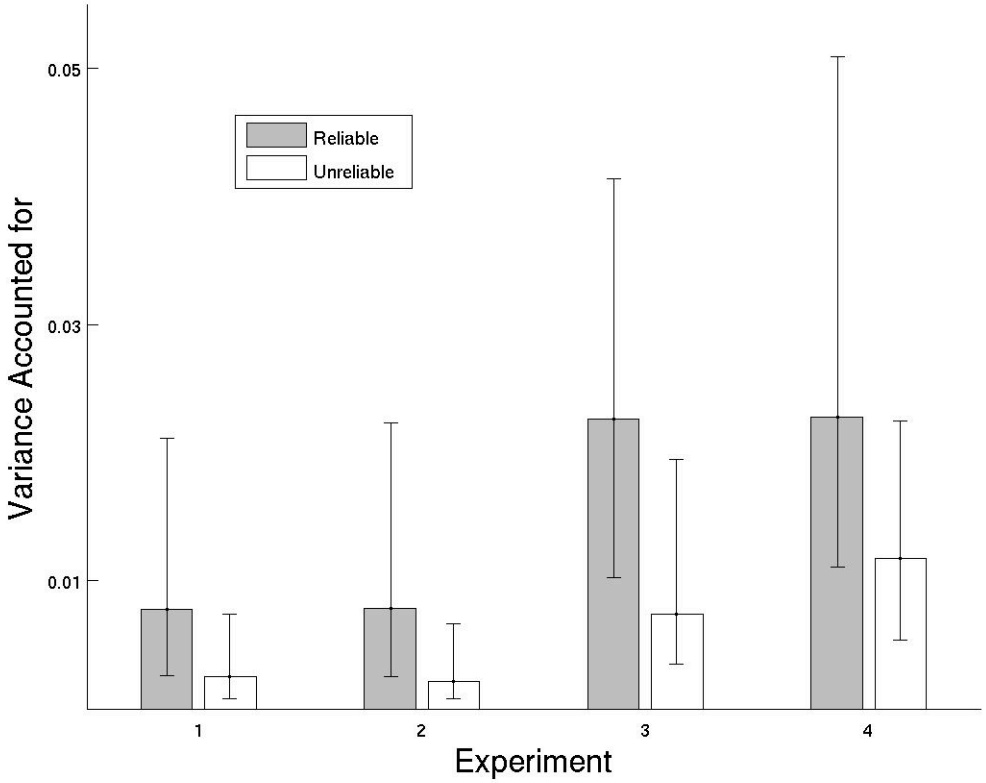

Figure V.18 – Median percentage variance accounted for (error bars indicate 25th and 75th percentiles) by reliable and unreliable ICs from all four data sets. Mann-Whitney U tests find that the reliable ICs account for more variance on average than unreliable ICs ($p<.0001$) for each experiment.

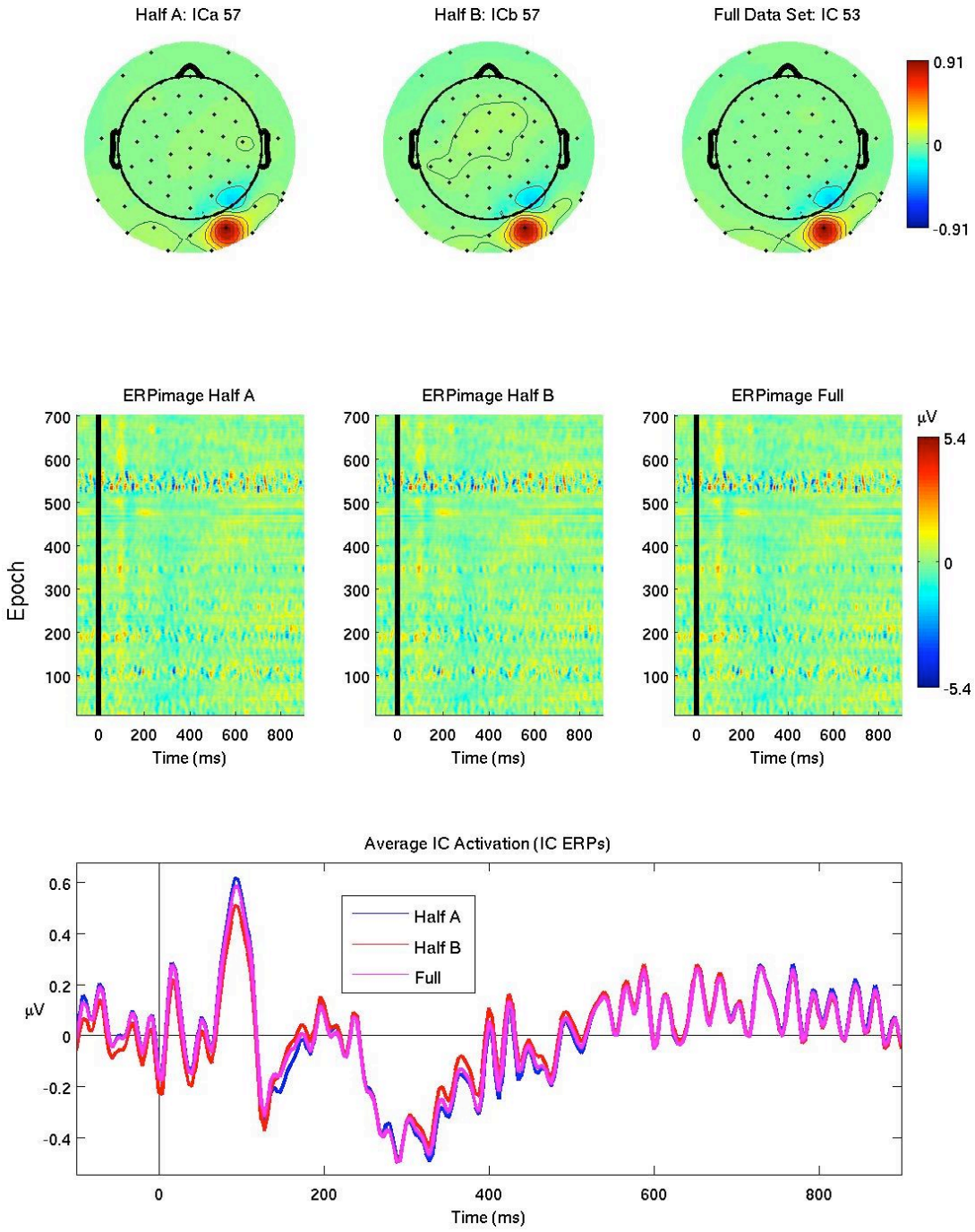

Figure V.19 – Topographies (**Top Row**), single epoch activations (**Middle Row**), and average activations across all epochs (**Bottom Row**) of IC 32 from Participant 1's visual oddball data and the two ICs it was paired with from each 50% split of the data. Topography weights are unitless and normalized to unit length. ERPimages are smoothed with a 15 epoch wide vertical moving window. Epochs are sorted in order of occurrence.

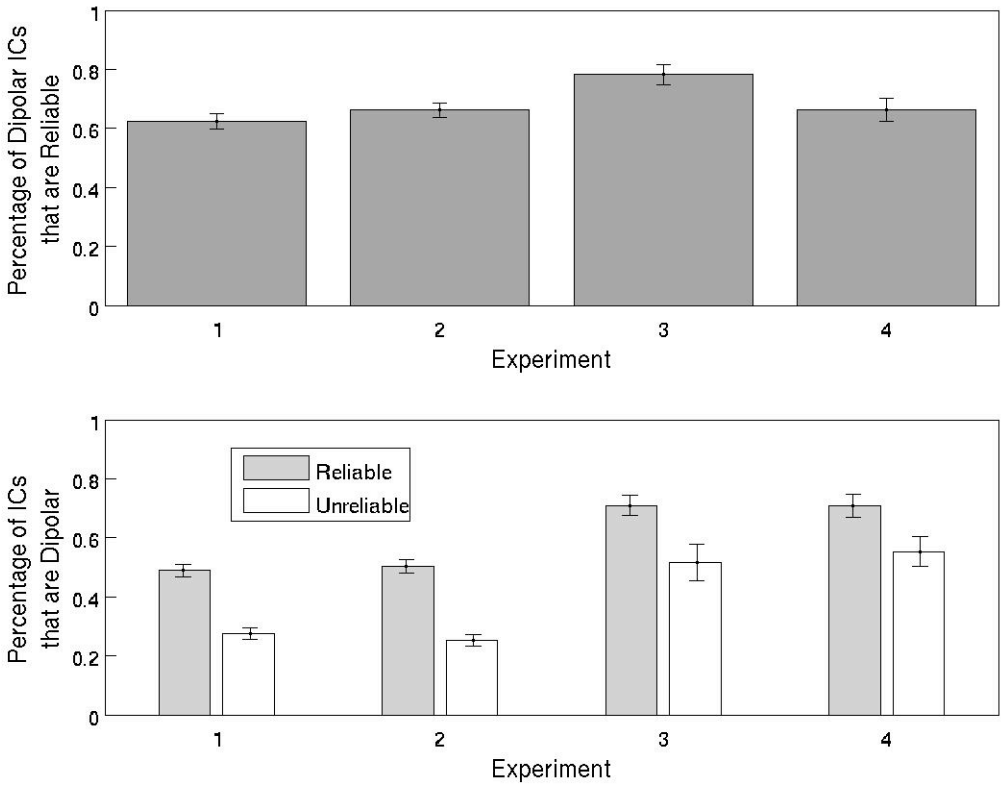

Figure V.20 – (**Top**) Percentage of dipolar ICs that were found to be reliable for four different data sets. Experiments 1 and 2 consist of 64 channel data. Experiments 3 and 4 consist of 30 channel data. (**Bottom**) For each data set, reliable ICs are more likely to be dipolar than unreliable ICs. This is borne out by a two way ANOVA with factors Reliability (reliable, unreliable) and Number of Electrodes (64, 30). Reliability ($F(1, 2520)=64.6$, $p\approx 0$) and Number of Electrodes ($F(1, 2520)=93.91$, $p\approx 0$) significantly differed without interaction ($F(1, 2520)=1.47$, $p>.22$). The fact that the 30 electrode ICs are more dipolar is expected given that there are about half as many data points to constrain the ECD model.

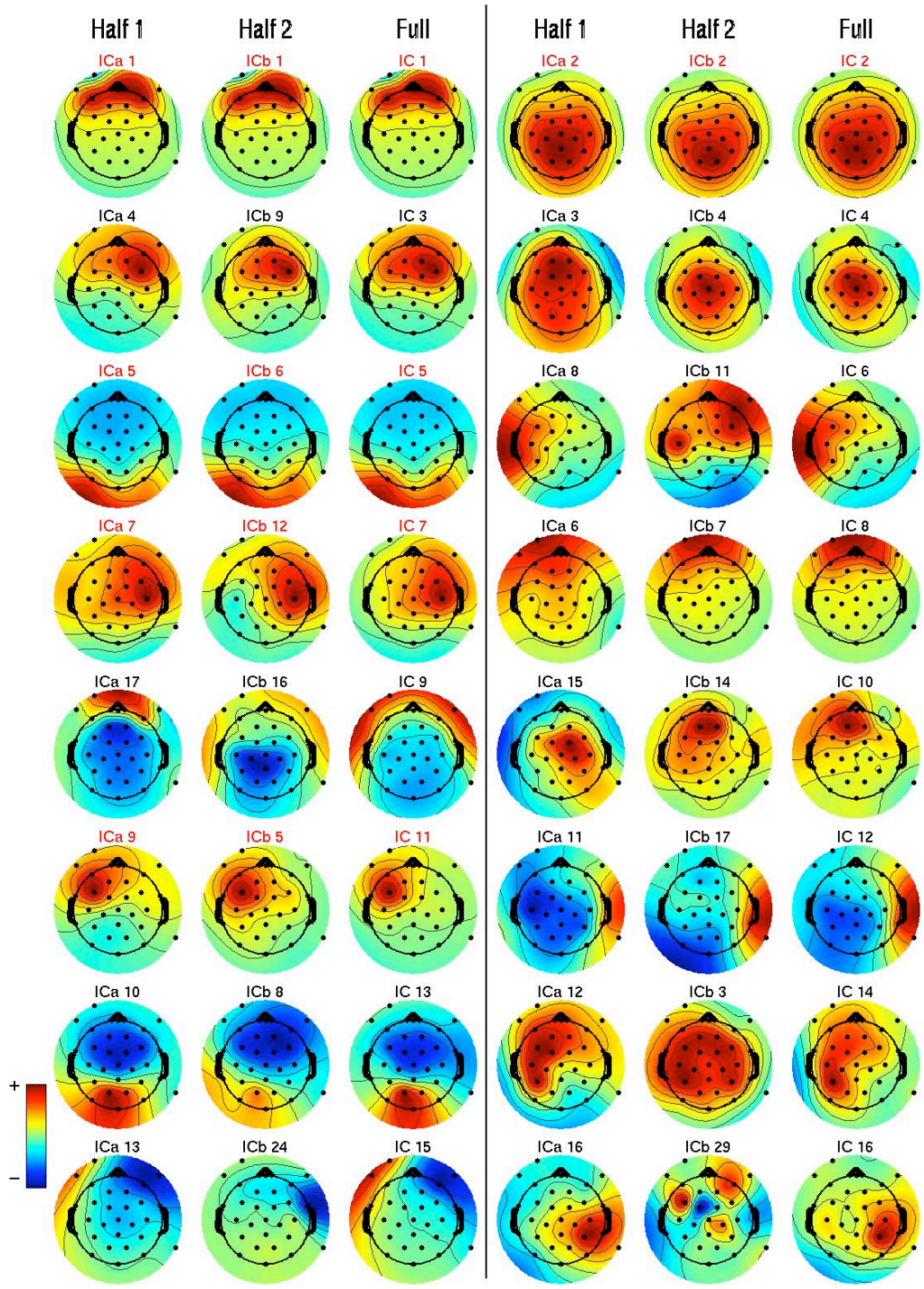

Figure V.21 – Topographies of ICs 1-16 of Participant 7's data from Experiment 3 (third and sixth column) and their matches from both 50% splits of the data. ICA was applied after the 100 ms prestimulus baseline was removed from each epoch. See Figure V.22 for further plot explanation.

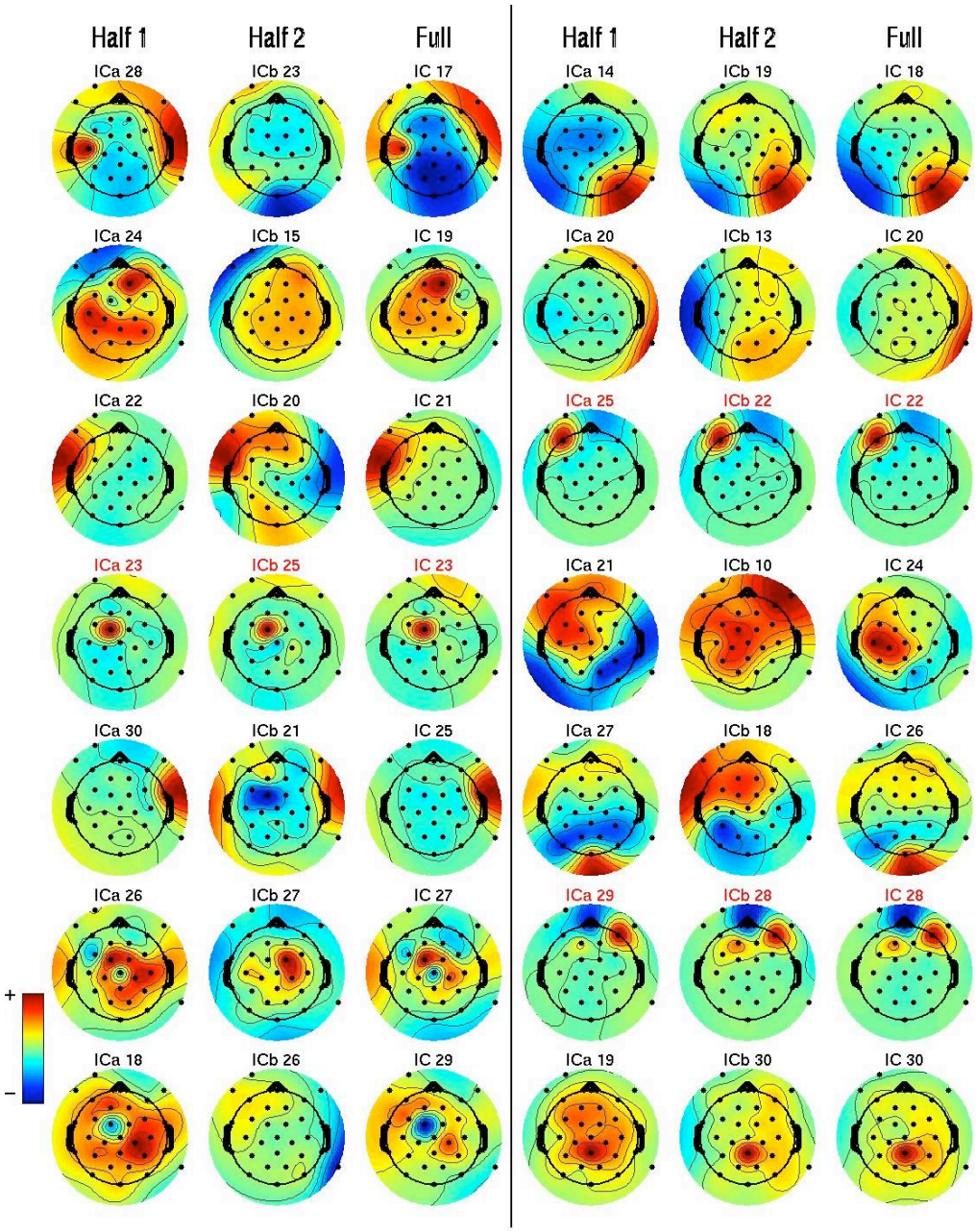

Figure V.22 – Topographies of ICs 17-30 of Participant 7's data from Experiment 3 (third and sixth column) and their matches from both 50% splits of the data. ICA was applied after the 100 ms prestimulus baseline was removed each epoch. Unitless topography weights are plotted with a color scale that is unique for each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Black dots represent electrodes. Triplets of reliable ICs are named in red.

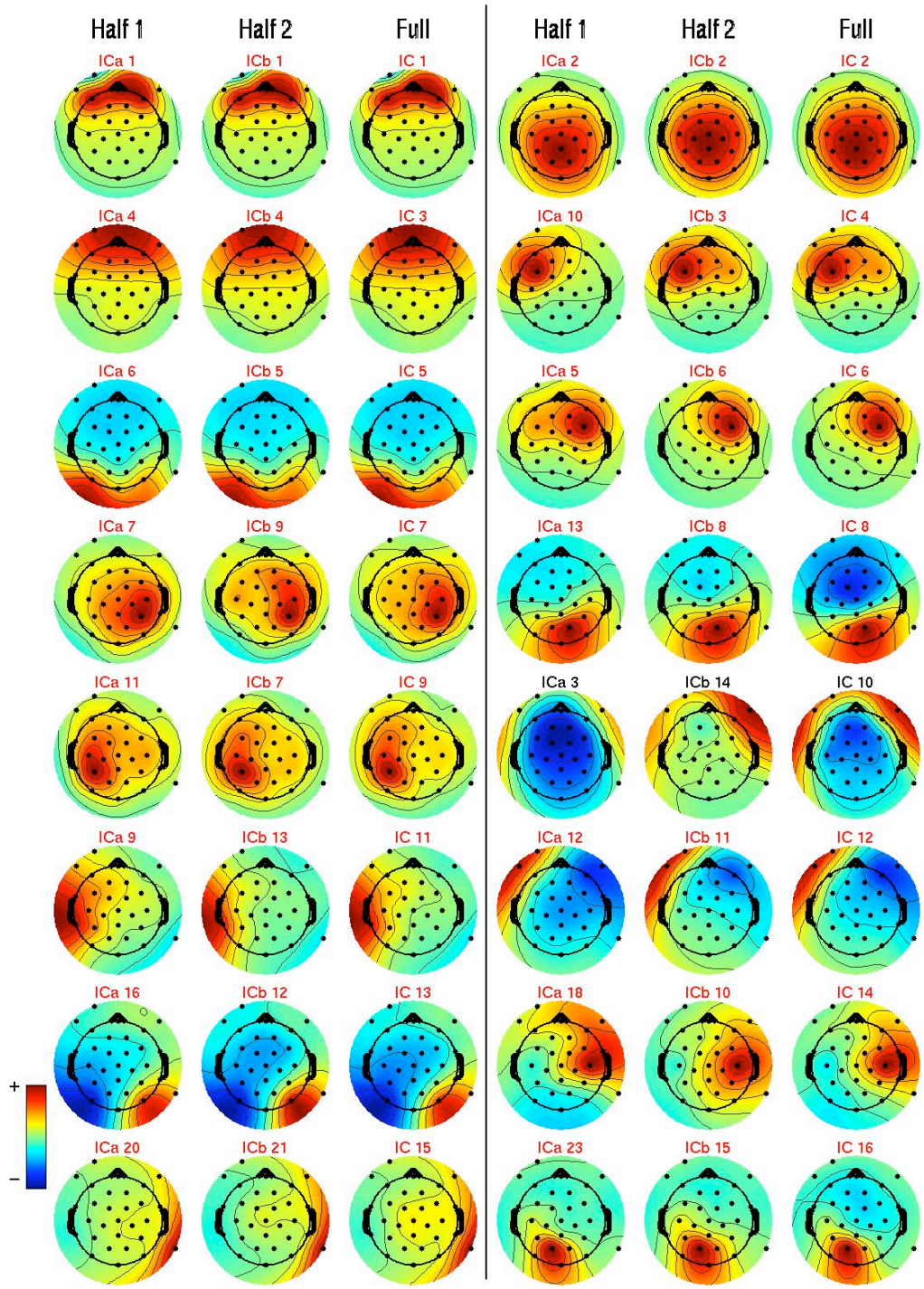

Figure V.23 – Topographies of ICs 1-16 of Participant 7's data from Experiment 3 (third and sixth column) and their matches from both 50% splits of the data. ICA was applied after the mean of each epoch was removed from each epoch. See Figure V.24 for further plot explanation.

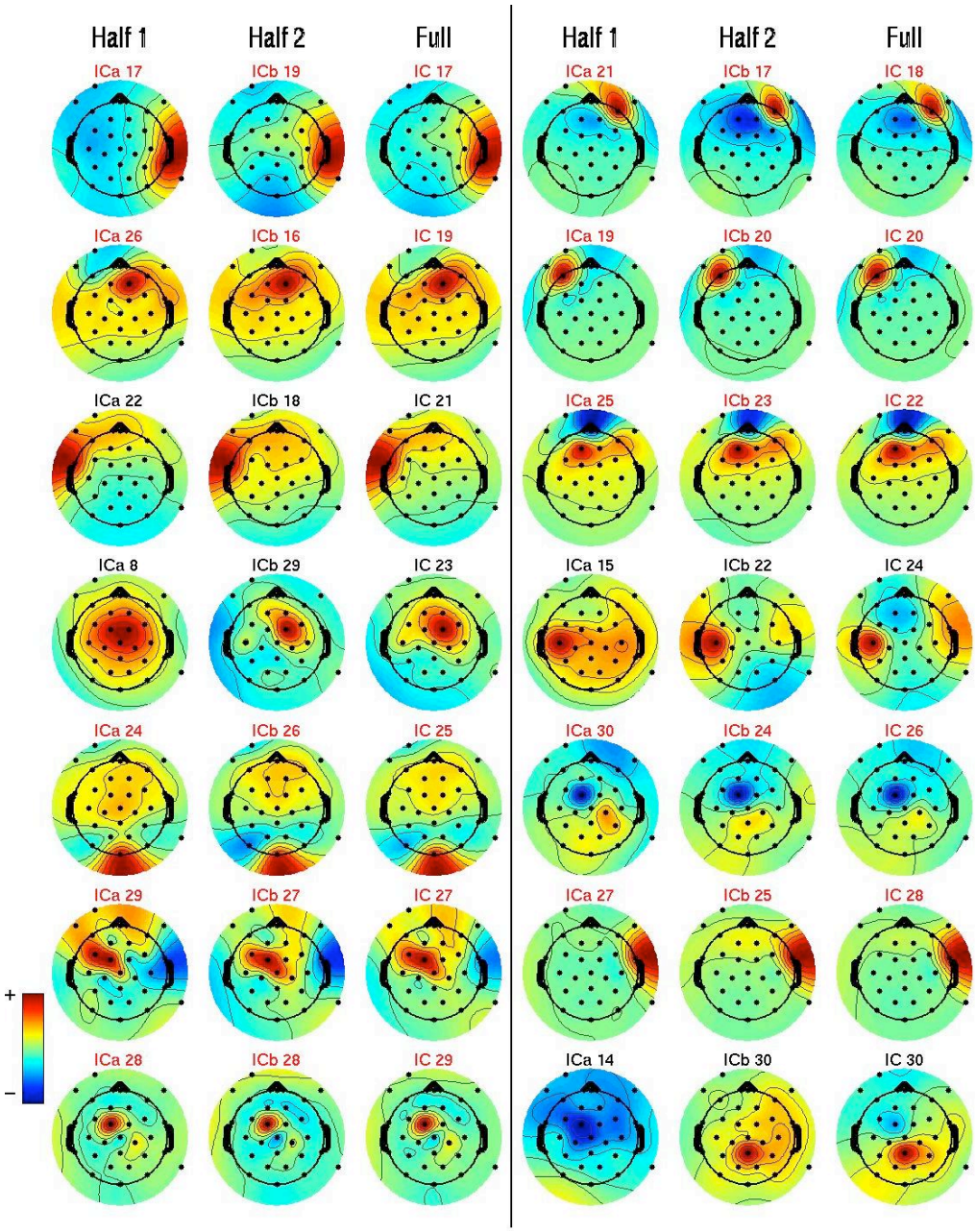

Figure V.24 – Topographies of ICs 17-30 of Participant 7's data from Experiment 3 (third and sixth column) and their matches from both 50% splits of the data. ICA was applied after the mean of each epoch was removed from each epoch. Unitless topography weights are plotted with a color scale that is unique for each topography. Topography polarities are chosen to maximize similarity with the IC from the full data set. Black dots represent electrodes. Triplets of reliable ICs are named in red.

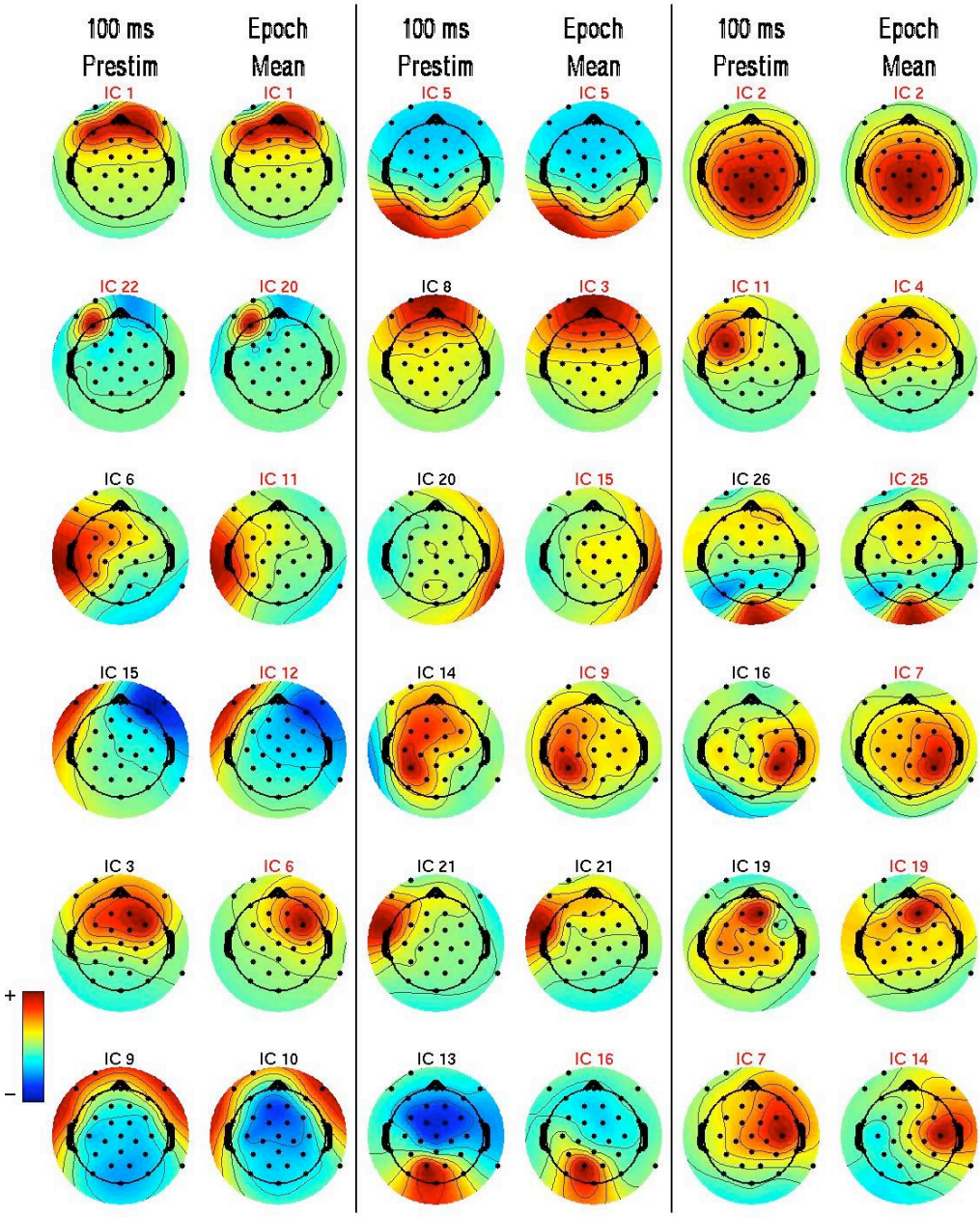

Figure V.25 – Topographies of the 18 most similar (cosine similarity metric) pairs of ICs from Figures V.21-22 and Figures V.23-24. ICs in odd columns are ICs from data with the 100 ms prestimulus baseline removed. ICs in even columns are ICs from the same data with the mean of each epoch removed. Unitless topography weights are plotted with a color scale that is unique for each topography. Topography polarities are chosen to maximize pair similarity. Black dots represent electrodes. Reliable ICs are named in red.

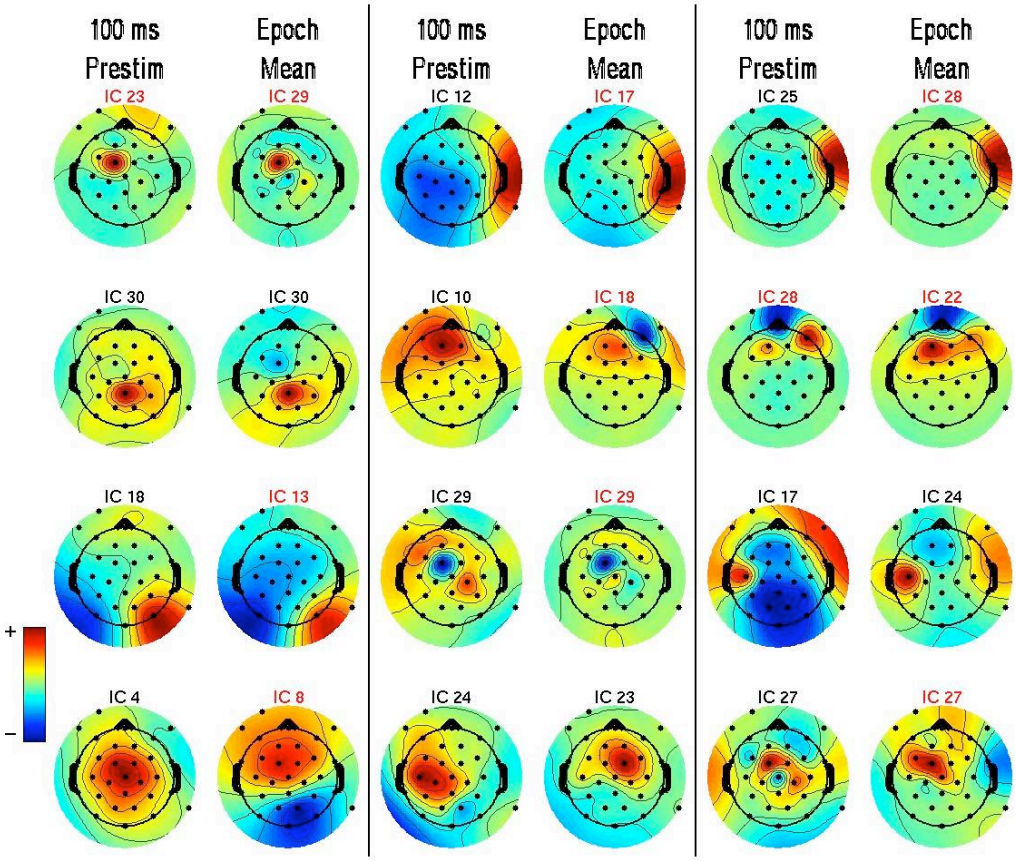

Figure V.26 – Topographies of the 12 least similar (cosine similarity metric) pairs of ICs from Figures V.21-22 and Figures V.23-24. ICs in odd columns are ICs from data with the 100 ms prestimulus baseline removed. ICs in even columns are ICs from the same data with the mean of each epoch removed. Unitless topography weights are plotted with a color scale that is unique for each topography. Topography polarities are chosen to maximize pair similarity. Black dots represent electrodes. Reliable ICs are named in red. Last row of pairs are ICs that did not appear to replicate across the two decompositions.

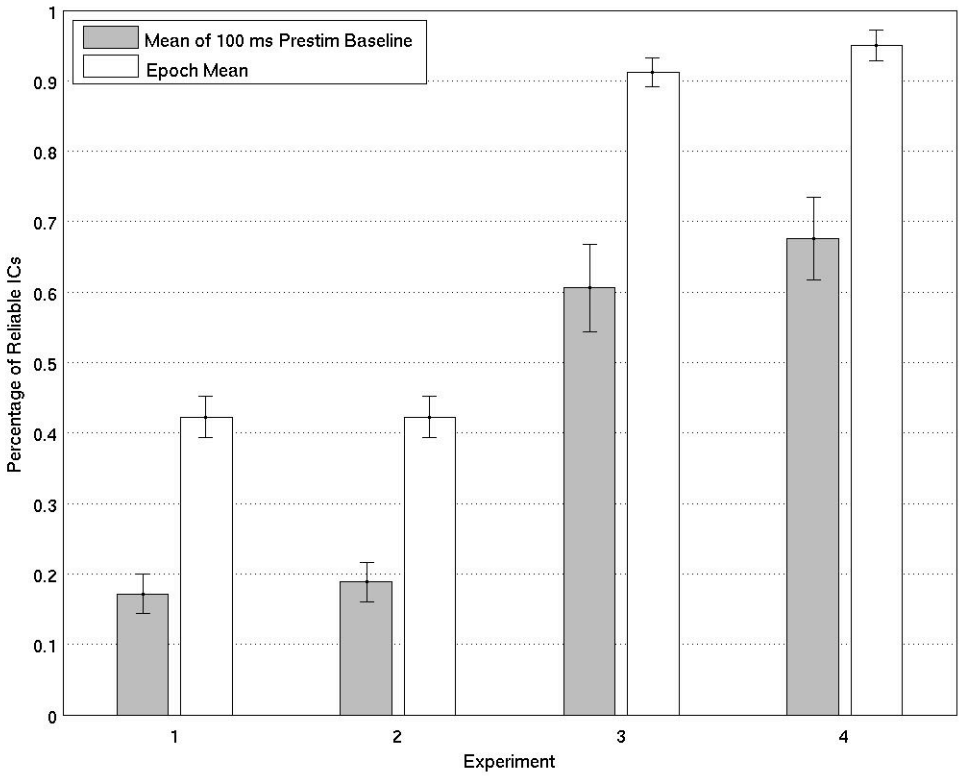

Figure V.27 – Mean fraction of ICs that are reliable per participant (error bars are standard error). Grey bars represent ICs from data that had the 100 ms prestimulus baseline removed from each epoch. White bars represent ICs from data that had the mean of each epoch removed. Experiments 1 & 2 consist of 64 electrode data. Experiments 3 & 4 consist of 30 electrode data. Zero mean epoch data produce significantly more reliable ICs than 100 ms prestimulus baselined data (Experiments 1& 2, sign test(16), $p<.00003$; Experiments 3 & 4, sign test(8), $p<.0078$).

V.G References

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine, 18*(6), 14-30.

Belouchrani, A., Abed Meraim, K., Cardoso, J. F., & Moulines, E. (1997). A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing, 45*, 434-444.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Efron, B., & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fisher, N. I., Lewis, T., & Embleton, B. J. J. (1993). *Statistical Analysis of Spherical Data*. New York: Cambridge University Press.

Fogelson, N., Loukas, C., Brown, J., & Brown, P. (2004). A common N400 EEG component reflecting contextual integration irrespective of symbolic form. *Clinical Neurophysiology, 115*(6), 1349-1358.

Ghahremani, D., Makeig, S., Jung, T.-P., Bell, A. J., & Sejnowski, T. J. (1996). *Independent component analysis of simulated EEG using a three-shell spherical head model* (Technical Report No. 96-01). La Jolla, California, USA: Institute for Neural Computation, University of California San Diego.

Groppe, D., Kemmer, L., Kutas, M., & Makeig, S. (2004). Within-subject reliability of independent component analysis of the electroencephalogram in two text processing tasks. *Cognitive Neuroscience Society Annual Meeting Program,* San Francisco. 147.

Harmeling, S., Meinecke, F., & Müller, K. R. (2004). Injecting noise for analysing the stability of ICA components. *Signal Processing, 84*(2), 255-266.

Himberg, J., Hyvarinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage, 22*(3), 1214-1222.

Kemmer, L., Coulson, S., De Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology, 41*(3), 372-384.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation, 11*(2), 417-441.

Makeig, S., Jung, T.-P., Ghahremani, D., & Sejnowski, T. J. (2000). Independent component analysis of simulated ERP data. In T. Nakada (Ed.), *Integrated Human Brain Science* (pp. 123-146). New York: Elsevier.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., & Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., & Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science, 295*(5555), 690-694.

Marco-Pallares, J., Grau, C., & Ruffini, G. (2005). Combined ICA-LORETA analysis of mismatch negativity. *NeuroImage, 25*(2), 471-477.

Meinecke, F., Ziehe, A., Kawanabe, M., & Muller, K. R. (2002). A resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Transactions on Biomedical Engineering, 49*(12), 1514-1525.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage, 27*(2), 341-356.

Oostenveld, R., & Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human Brain Mapping, 17*(3), 179-192.

Salustri, C., & Kronberg, E. (2004). Language-related brain activity revealed by independent component analysis. *Clinical Neurophysiology, 115*(2), 385-395.

Tang, A. C., Sutherland, M. T., & McKinney, C. J. (2005). Validation of SOBI components from high-density EEG. *NeuroImage, 25*(2), 539-553.

Chapter VI: Independent Component Analysis of the P3b, N400, and P600

VI.A Introduction

In order to identify subcomponents of the P3b, N400, and P600 effects, independent component analysis (ICA) was applied to the combined data from the two experiments described in Chapter IV, as well as the data from each experiment alone. This procedure produced three sets of independent components (ICs) per participant. ICs that appeared in all three sets were considered common to both experiments. Reliable ICs that appeared in less than three sets were considered unique to one of the experiments.

This method found a large number of ICs that were common to both experiments ("common ICs"), which suggest a significant, though far from total, overlap between the EEG processes generating these ERP effects. More specifically, the common ICs account for about 78%, 86%, and 71% of the grand average P3b, N400, and P600 effect variance, respectively. Identifying types of common ICs was complicated by the fact these ERP effects were generally distributed among several ICs and by individual variation. However, largely based on each individual's P3b effect, it was possible to define a somewhat homogenous class of 12 common "best

P3b ICs" from 12 of the 16 participants. This class of ICs accounted for around 50% of the P3b, N400, and P600 effects and of the LPC, and was largely responsible for the correlation between LPC and N400 effect amplitude (Section IV.G.1). The common best P3b ICs behaved like the P3b in that their average activations were typically characterized by a late positivity that was enhanced to targets/anomalous sentence endings and the peak of this positivity was generally later to stimuli that should have taken longer to discriminate. In some ways these 12 ICs did not behave like the P3b though these null effects may have been due to inadequate statistical power. In addition, the topographies of many of the best P3b ICs had a central midline focus, more typical of the P3a than of the P3b, which generally has a more posterior focus and had a more posterior focus in these data. Due to the central topographies, it appears more likely that these ICs correspond more closely to the P3a than the P3b, and index an attentional response.

An exploratory automated cluster analysis was applied as an alternative method for identifying classes of common ICs. The cluster analysis returned 10 classes of common ICs, which overall suggested a dissociation between frontal and central/posterior ICs that, on average, elicited scalp negativities and positivities (respectively) in response to deviant stimuli. The clustering results largely replicated the best P3b IC grouping and suggested nine other types of common ICs. A few of these nine other clusters also contributed to the main deviance ERP effects of interest. A bilateral pair of frontal clusters contributed to the N400 effect and the anterior negativity elicited by ungrammatical pronouns. In light of previous research on

frontal negativities elicited by text, these clusters possibly indexed working memory processes and/or an attentional response. A midline parietal and occipital pair of clusters also contributed somewhat to the P3b, N400, and P600 effects but were difficult to interpret. Finally, a left fronto-temporal cluster characterized by a negative peak at about 320 ms possibly corresponded to the N320 ERP component implicated in grapheme-to-phoneme conversion (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999) or the N340 ERP component (Mangels, Picton, & Craik, 2001), which is correlated with declarative memory encoding.

Only a few ICs unique to one experiment were identified, which most likely reflects the limited amount of data provided by each experiment alone and the fact that many ICs corresponded to EEG artifacts. Thus while a significant fraction of the P3b, N400, and P600 effects appear to be unique to one or the other the experiments, it is currently difficult to say what those unique subcomponents might be.

Final discussion of these results is postponed until Chapter VII.

VI.B Independent Component Derivation and Complexity of Deviance Effects

VI.B.1 ICA Parameters & IC Reliability

As already described in Section IV.D, extended INFOMAX ICA (Lee, Girolami, & Sejnowski, 1999) was applied to the combined data from the two

experiments described in Chapter IV (two oddball tasks and a sentence acceptability task) using the *binica* function from the EEGLAB toolbox (Delorme & Makeig, 2004). Some details are briefly recapitulated here. On average, each participant's data consisted of 122 unique time points of data per number of electrodes squared (minimum of 105) and ICA produced one set of 64 ICs for each of the 16 participants. ICs primarily accounting for EEG artifacts (e.g., blinks, muscle activity) were removed from the data and further analysis. In addition to the ICs labeled as artifacts in Chapter IV, IC source modeling (which was performed for the subsequent cluster analysis--Section VI.C.4.a) revealed four additional ICs that appeared to correspond to artifacts[64]; these were removed from the data and further analysis. It is unlikely that their removal would qualitatively change any of the results in Chapter IV as their removal has little effect on those participant's ERPs[65].

The reliability of each IC was assessed using the reliability test described in the previous chapter. 50% splits of each participant's data were constructed by separating epochs time-locked to targets, standards, well-formed sentence endings, incongruous endings, ungrammatical endings and sentence internal words. Targets and standards were separately grouped by oddball task (letter case or semantic category) and sentence endings were separately grouped by part of speech (noun, verb, and

[64] Specifically a heart beat artifact IC was identified in Participant 2, an eye movement artifact IC was identified in Participants 7 and 13, and a mechanical electrode artifact was identified in Participant 13.
[65] The difference between the ERPs of the three affected participants before and after the removal of these four ICs was quantified with the mean root mean square (RMS) metric. The four types of oddball task ERPs (ERPs to targets and standards in both oddball tasks) and the four main types of sentence acceptability task ERPs (ERPs to all incongruous, congruous, ungrammatical, and grammatical endings) were compared. The maximum RMS difference between the eight different types of ERPs across all participants was .76 microvolts (mean: .34 microvolts).

pronoun). This produced 12 sets of epochs for each participant. Epochs were sorted by their order of occurrence and odd and even epochs split to produce two halves of the entire data set. Based on these splits, the following critical values were obtained for the IC similarity test: sim_{topo}={.66, .86} and sim_{act}={1.87, 2.22}. On average 37 (SD=7) ICs per participant were found to be reliable, and 60% of these corresponded to EEG artifacts. The reliable ICs accounted for between 89% and 95% of the variance of the grand average P3b, N400, and P600 effects (Tables VI.1-3).

VI.B.2 IC Complexity of the P3b, N400, & P600 Effects

In general, the P3b, N400 and P600 effects were distributed among multiple ICs. This is visualized for one participant's data with a series of "envtopo" plots in Figures VI.1-3. Envtopo plots display the amplitude envelope of the ERPs from all electrodes (i.e., the maximum and minimum value of the ERPs across all electrodes) in black. The envelopes of the scalp-projected ERAs of the five ICs (reliable or otherwise) that together accounted for the most ERP variance in a time window of interest of any possible set of five ICs are shown in color. The topographies of these ERAs at the time point of their greatest magnitude in the time window are shown above the envelopes. The envelope of the sum of the five scalp-projected ERAs is displayed in light blue shading and provides a sense of what parts of the ERP were well accounted for or missed by those five ICs. The time windows begin 150 ms

before and end 150 ms after the peak latency of the individual's P3b, N400, and P600 effects (Sections IV.E.3.a & IV.F.3.a).

This participant's P3b is visible in Figure VI.1 (A) as the pronounced positivity in the ERP envelope that peaks at about 350 ms. No single IC accounted for the P3b. Rather, the P3b was partially the product of a midline central positivity (IC 1), occipital positivities (ICs 2, 3 & 4), a small right frontal positivity (IC 7), and a few other ICs not shown. In fact, these five ICs together accounted for only 87% of the variance of this participant's P3b effect (Figure VI.1: C). This participant's N400 and P600 effects (Figures VI.2-3) were similarly a product of multiple ICs. This was typical of the 16 participants in this study. On average, a single IC could at most account for about 50% of a participant's P3b, N400, and P600 effects. Indeed, it usually took at least five, six, and more than six ICs to account for over 90% of the variance of the P3b, N400, and P600 effects respectively (Figure VI.4), although there were some exceptions (e.g., Participant 3's IC 1 accounted for 93% of her P3b effect variance). The distributed nature of the ICA decomposition of these effects is consistent with the previously reviewed evidence that these effects are composed of subcomponents (Sections II.E) and is probably closer to reality than the minimal source models in Chapter IV (Sections IV.E.3.d & IV.F.3.e).

VI.C Independent Components Common to Experiments 1 and 2

VI.C.1 Identifying Common ICs

To identify ICs common to both experiments ("common ICs"), ICA was applied separately to each participant's data from the oddball tasks and the sentence acceptability task, thus producing three sets of 64 ICs per participant. ICA parameters were identical to those used for the combined data sets. On average, each participant's oddball task data consisted of 52 unique time points per number of electrodes squared (minimum of 45) and each participant's sentence acceptability task data consisted of 69 unique time points per number of electrodes squared (minimum of 55).

An IC from the combined data sets was considered common to both experiments if a significantly similar IC was independently derived from each data set. This was determined by the following procedure:

1. Each IC from the combined data sets, IC_{Both}, was paired with an IC derived from just Experiment 1, IC_{Ex1}, and an IC derived from just Experiment 2, IC_{Ex2}, according to their topographies (Section V.B.2). This created an IC triplet[66].

[66]One automatically formed triplet of ICs was hand corrected to produce what was believed to be a more accurate pairing (Figures VI.5-7). Specifically, Participant 8's IC_{Both} 2 was automatically paired with IC_{Ex1} 16 and this was changed to IC_{Ex1} 5. The topography of IC_{Both} 2 is slightly more similar to that of IC_{Ex1} 16 than IC_{Ex1} 5 (r_{cos} of .98 vs .97). However, the activation of IC_{Both} 2 is much more similar to that of IC_{Ex1} 5 than IC_{Ex1} 16 (e.g., ERA r_{cos} of .69 vs .38).

2. The activations and topography of IC_{Both} were compared with those of IC_{Ex1} and IC_{Ex2} to determine if they were significantly similar (Section V.B.3). The critical values derived from 50% splits of the combined data sets (Section VI.B.1.a) were used for this test.

3. The topographies of IC_{Ex1} and IC_{Ex2} were compared to determine if they were significantly similar. If the topographies were more similar than 100-100/*n* percent (where *n* is the number of ICs per data set) of the empirical distribution of topography similarities derived from 50% splits of the combined data sets (e.g., Figure V.3: Top), then the topographies were considered significantly similar.

4. If each member of the triplet was significantly similar to the other two members, then IC_{Both} is deemed to be common to both experiments.

This procedure was largely identical to that used to assess IC reliability (Chapter V) where the ICs from the two different experiments were analogous to the balanced 50% splits of the data. One difference between the procedures was that the critical values were computed from the ICs derived from balanced 50% splits of the data, not from the ICs derived from the separate experiments. This was done under the assumption that the 50% splits would more accurately reflect IC estimation error. Another difference between the procedures was that only the topographies of the ICs from the two data subsets (Experiments 1 and 2) were compared. This was done because there may have been electrical sources unique to one of the experiments and this would have made it impossible to directly compare IC activations. To illustrate, consider an

ideal case where there are three latent sources and two electrodes. One of the sources, Source A, is common to both experiments but the other two, Source B and Source C, are unique to one of the experiments. Assume that ICA is able to perfectly unmix them when applied to the data from each experiment separately. While the mixing vectors (i.e., topographies) for Source A from the two experiments will be exactly the same, the unmixing vectors for Source A will be quite different (Figure VI.8). This occurs because the Source A unmixing vector must be orthogonal to the Source B mixing vector in Experiment 1 and orthogonal to the Source C mixing vector in Experiment 2. As a consequence, applying the Source A unmixing vector derived from the data in one experiment to the data in the other will not isolate Source A's activity from the other source.

The final difference between the procedures was that when the topographies of ICs from each experiment alone (IC_{Ex1} and IC_{Ex2}) were compared, the critical similarity value was set to the (1-1/*n*)th percentile of the empirical distribution of topography similarities, where *n* is the number of ICs per participant (Step 3). This defined a critical region just large enough for every pair of ICs to be significantly similar and was done since only one IC feature was being compared (i.e., the activations of IC_{Ex1} and IC_{Ex2} were not compared). The critical value for these experiments was .89. This was somewhat less than the median similarity of each participant's letter case oddball and semantic category P3b topographies (.92) and more than the median similarity of each participant's N400 (.78) and P600 (.61) effects across parts of speech (Sections IV.E.3.e & IV.F.3.f).

On average, 19 (SD=5) of a participant's ICs from the combined oddball and sentence acceptability task data replicated when ICA was applied to the data from each task alone, and 61% of these ICs were EEG artifacts. Figures VI.9-11 illustrate the features of three example "common ICs" and their homologues from the two tasks. The topographies and ERA waveforms of these triplets were remarkably similar though the differences in ERA magnitude could be large by ERP standards. The size differences were deceptively big since the topography weights were all quite small (i.e., these differences at the scalp were a fraction of that displayed) and the IC activations could be quite large (i.e., the magnitude of the difference relative to the magnitude of the activations was significantly small).

VI.C.2 Contribution of Common ICs to Grand Average ERPs

These common ICs accounted for between 69% and 87% of the variance of the canonical P3b, N400, and P600 effects (Tables VI.1-3), which suggests a substantial degree of overlap between the latent EEG processes that produced them. However, given that about twice as many ICs were reliable (i.e., they replicated across 50% splits of the data) and that these reliable ICs accounted for more effect variance than the common ICs (Tables VI.1-3) it appears that these effects were somewhat distinct.

Figures VI.12-23 contrast various grand average ERPs from Chapter IV with the common IC and reliable IC contribution to these ERPs. The most notable difference between the two sets of ERPs in the oddball tasks was that the P3b of common ICs was about half as large as the grand average P3b (Figures VI.13 &

VI.15). About half of this loss appears to have been due to ICA instability, as the reliable ICs also failed to capture several microvolts of P3b. This difference was also apparent in the P3b difference ERPs (Figures VI.16-17, Table VI.1). In addition to these P3b differences, from 300 to 600 ms, the common ICs also missed an enhanced frontal negativity and a posterior right lateralized positivity to standards (Figures VI.12 & VI.14).

In the sentence acceptability task, what appeared to be the N400 was largely missed by the common ICs and better captured by the reliable ICs[67] (Figure VI.19). However, the N400 effect (Figure VI.22) and the LPC (Figures VI.18-21) were captured about as well by the two sets of ICs (Tables VI.2 & VI.4). In contrast, a large fraction of the P600 effect was missed by the common ICs but captured by reliable ICs (Figure VI.23, Table VI.3). The common ICs accounted for about 75% of the variance of the smaller late effects of congruity and early effects of grammaticality with the exception of the late effect of congruity on verb ERPs (Tables VI.2-3). Relative to the percentage variance accounted for by reliable ICs, this was quite good for the anterior negativity (AN) elicited by ungrammatical pronouns (Table VI.3: Column 3).

Figure VI.24 plots the topographies of the P3b, N400, and P600 effects captured by the common and reliable ICs. While the P3b effects appeared largely identical, the common IC N400 and P600 topographies lacked the medial parietal

[67] Common ICs account for 55% and reliable ICs account for 89% of the variance of the grand average ERP to incongruous endings from 275 to 475 ms.

focus of the reliable IC effects, though these differences were quite small (about half a microvolt at most).

Finally, the largest difference between the electrode data and common IC contribution to the sentence acceptability task ERPs, was that the slow, centro-frontal negativity following the LPC (possibly a readiness potential that anticipated the motor response) was absent in the common ICs (Figures VI.18-21). The negativity was moderately captured by reliable ICs.

VI.C.2.a Interim Summary I: Degree of Commonality and Distinctiveness

The contribution of independent components (ICs) common to both experiments to the grand average ERPs suggests considerable overlap in the EEG response to rare targets and anomalous sentence endings. With regard to the effects of grammaticality, this is consistent with evidence that the P600 behaves similarly to the P3b (Coulson, King, & Kutas, 1998), can be elicited by non-grammatical anomalies (Patel, Gibson, Ratner, Besson, & Holcomb, 1998), and that effects of grammaticality tend to interact with rare changes in the physical characteristics of words (Osterhout, McKinnon, Bersick, & Corey, 1996). Furthermore, it suggests that the frontal negativity elicited by ungrammatical pronouns does not simply reflect syntax specific processing (e.g., the processing of morphosyntactic information—Friederici & Kotz, 2003), which is consistent with the possibility that such frontal negativities index general working memory functions (Kluender & Kutas, 1993).

The fact that so much of the N400 effect was also captured by common ICs was surprising, as there was nothing clearly like the N400 in the grand average oddball task ERPs and a relationship between the N400 effect and oddball task potentials like the P3b is often not considered (Kutas & Federmeier, 2000; Van Petten & Luka, 2006). The apparent overlap is consistent with evidence of some common generators of these effects (Section II.G.1) and with the hypothesis that the N400 effect partially reflects a delayed P3b (ibid.). Finally, it was interesting that the LPC was largely accounted for by common ICs. This was compatible with interpretations of the LPC as related to the P3b (Chung et al., 1996), though it appeared that the LPC was more of a subset of the P3b than its homologue.

There are a couple reasons why the apparent commonality across the two experiments could have been overestimated. First of all, it is likely that some of the "significantly similar ICs" across the two experiments were false positives, given the large number of comparisons. In addition, since these ICs were derived from the EEG, they may have corresponded to EEG processes common to both experiments that have similar topographies to relatively weak ERP sources unique to one experiment. For example, midline posterior electrodes often pick up large "alpha waves," approximately 12 Hz sinusoidal EEG activity (Makeig et al., 2002). Given the potentially large number of EEG sources (Section III.B.1), it is possible that there were ERP sources unique to one of the experiments that were only weakly apparent in the EEG but had a similar topography to that of alpha processes. Since ICA tends to extract the strongest latent EEG processes (Section III.D.3), it may have extracted

alpha activity common to both experiments but not weaker ERP sources. Thus the weaker ERP ICs would have been misattributed to the alpha ICs (i.e., their variance would have been "misallocated"—Section III.C.1). While it is quite possible that these ERPs were simply composed of the largest EEG processes (e.g., Makeig et al., 2002), these two potential confounds mean that convincing evidence of common subcomponents is dependent upon corroboration from other methodologies and the identification of compelling types of common ICs. The subsequent two sections (Sections VI.C.3-4) describe two attempts to achieve the latter.

Although these results suggest some shared subcomponents, the fact that common ICs accounted for less of the P3b, N400, and P600 effects than reliable ICs derived from 50% splits of the combined experimental data suggests that there were indeed components of these effects that were unique to each experiment. In particular, it appears that the P600 was actually more distinct from the P3b than the N400, which is surprising given that there is more evidence relating the P3b to P600 than the N400 (Sections II.G.1-2). Of course, some of this difference could have reflected a greater imbalance of EEG artifact and participant alertness in the data from single experiments. It could also have reflected the small difference between the tests for reliability and commonality (i.e., in the commonality test, only the scalp topographies are compared across splits as the activations are not comparable) or the mis-pairing of ICs during the test for common components. Consequently, strong conclusions about common subcomponents are also dependent upon converging evidence from other

methodologies and the identification of compelling types of unique ICs. This is attempted in Section VI.D.

VI.C.3 The Class of "Best P3b" ICs

In light of the robustness of each participant's P3b effects relative to their N400 and P600 effects (Sections IV.E.3.b, IV.G.1, IV.G.3), probably the best way to identify a class of ICs common to these ERP phenomena is to examine how the ICs contribute to the P3b effect. Although the P3b effect was generally distributed across multiple ICs, for each participant there was usually a single IC that projected maximally to medial central or parietal electrodes and accounted for a great deal of the P3b effect (Figure VI.25: A). The topographies and oddball task ERAs of these "best P3b ICs" are presented in Figures VI.26-27. Of these 16 ICs, all but Participant 1's IC was the single IC that accounted for the most variance of that participant's P3b effect. Given the similarity of the topographies of these ICs and their prominent contribution to the P3b effect, it is possible that they largely correspond to the same EEG process across participants and were analyzed as such.

Of the 16 best P3b ICs, 12 were common to the oddball and the sentence acceptability tasks ("common best P3b ICs"), and are the focus of the following subsections. Of the four other ICs, one of these (P3 IC1) simply appeared to be unreliable. Curiously, three of the twelve common best P3b ICs (P2 IC2, P6 IC4, & P15 IC2) also failed the reliability criteria. Analysis of these three participants' ICs

found that that reliability failure was not due to poor automatic pairings or due to the slightly different criteria used for the reliability test versus the commonality test. Rather these three ICs simply appeared to be somewhat unreliable.

VI.C.3.a Common Best P3b IC Oddball Task Activity

Figure VI.28 (A-C) presents the mean topography, power spectra, and oddball task ERAs of the 12 common best P3b ICs. In the oddball tasks, the average activations of this class of ICs had a positive peak at 200 ms to both targets and standards, which was followed by a pronounced positivity to targets. The difference between targets and standards accounted for 62% and 58% of the variance of the grand average letter case and semantic category task P3b effects (respectively--Table VI.5). To determine the statistical significance of this difference, the mean activation of each IC in an individual's P3b time windows (Section IV.E.3.a) was subjected to an ANOVA with two factors: Task (letter case, semantic category) and Stimulus (target, standard). Subsequently, the number of ICs that significantly showed a particular main effect or interaction was tallied, using an alpha level of .25. The *p*-value of the tally was computed using the binomially distributed null hypothesis that significant comparisons should occur by chance with a probability of .25[68]. For example, under this null hypothesis, the probability of getting six or more of these 12 ICs to show a particular significant effect or interaction is approximately .054. Thus the overall *p*-

[68] This is a similar to a hypothesis testing procedure used in previous research to detect differences in IC spectral power (Debener et al., 2005).

value of that effect or interaction would be .054. An alpha level of .25 was used as more than half of the ICs would need to show an effect for the overall *p*-value to reach the conventional .05 level of significance.

This hypothesis testing procedure was chosen because it is unaffected by the IC scaling ambiguity, it is relatively unaffected by outliers, and because it is tailored to find effects that may be only weakly apparent in individual ICs but are representative of the entire class. In contrast, the somewhat simpler procedure of performing a single ANOVA on the normalized mean activation of each IC (or the scalp projection of that activation--e.g., Debener, Makeig et al., 2005) would be sensitive to outliers and might arrive at conclusions that are unrepresentative of most ICs in the class.

Because the ANOVA/binomial test does not indicate the direction of any significant comparisons, such comparisons were followed by one-tailed independent sample *t*-tests. As with the ANOVAs, *t*-tests were performed separately for each IC in a cluster, treating each trial as an independent sample and using an alpha level of .25. The *p*-value of the number of significant *t*-tests was again computed using the binomially distributed assumption of a false positive rate of .25. For these tests, the polarity of an IC's activation was defined as the polarity of the IC's scalp-projected activation at the electrode where its absolute magnitude was maximal. For the 12 common best P3b ICs, this maximum was unambiguous (Figure VI.26).

According to the initial ANOVA, the mean response of all 12 of the common best P3b ICs to targets and standards in each individual's P3b time windows (Section IV.E.3.a) differed (main effect of Stimulus: binomial(12,.25)=12, *p*=6e-8). Follow up

t-tests found that for all of these ICs, the mean response to targets was more positive than the mean response to standards (binomial(12,.25)=12, *p*=6e-8). The amplitude of the enhanced positivity to targets did not reliably differ across the two oddball tasks. Specifically, only three ICs showed a significant effect of oddball task on their mean activations (binomial(12,.25)=3, *p*=.61) and one showed a significant Task x Stimulus interaction (binomial(12,.25)=1, *p*=.97). However, the latency of the enhanced positivity peaked later in the presumably more difficult semantic category oddball task for all but one IC (two-tailed paired sample *t*-test, *t*(11)=3.35, *p*=.006; Figure VI.29).

While the average activations of these ICs behaved like the P3b in their sensitivity to deviance and discrimination difficulty, these ICs did not reliably show the expected effect of preceding stimulus context on the amplitude of the positivity elicited by targets. More precisely, one-tailed *t*-tests showed that targets preceded by standards elicited a significantly greater positivity than targets preceded by targets in only three of the 12 ICs (binomial(12,.25)=3, *p*=.61) and elicited the opposite, un-P3b-like effect in three of the 12 ICs (binomial(12,.25)=3, *p*=.61). These mixed results may simply reflect insufficient trials as the expected stimulus context effect on the P3b was only weakly apparent in the electrode data. Using the same statistical criteria (one-tailed independent sample *t*-test, *p*<.25), only five of these 12 participants showed a significantly larger P3b to targets preceded by standards than targets

preceded by targets at their "best P3b electrode"[69] and two showed a significant difference in the opposite direction.

Given the medial central distribution of these ICs, it is possible that they correspond to the P3a ERP component, which habituates with repeated presentations of oddball task targets (Fabiani & Friedman, 1995). To detect a possible effect of habituation, the IC responses to targets and standards in each oddball task were split into two approximately equal sized groups based on their order of occurrence (i.e., the former and latter 50% of trials—e.g., Figures VI.30-31). ANOVAs with factors Task (letter case, semantic category), Stimulus (target, standard), and Order of Occurrence (former half, latter half) were performed on the mean activation in each participant's letter case and semantic category task P3b windows. The ANOVAs found a main effect of Order of Occurrence (binomial(12,.25)=8, p=.003) and all interactions involving Order of Occurrence were not significant (binomial(12,.25)<=3, p>=.61). Follow up one-tailed t-tests found that the average response to all oddball stimuli were more positive in the former half of the experiment (binomial(12,.25)=8, p=.003) though three ICs showed the opposite effect (binomial(12,.25)=3, p=.61). To get a richer sense of this habituation effect, an ERPimage of the 12 ICs responses to all oddball stimuli was constructed by sorting responses according to their order of occurrence (Figure VI.32). From this, it appears that the habituation was primarily an early effect, occurring after the first 150 trials.

[69] The "best P3b electrode" is the participant's electrode that accounts for the most variance of that participant's P3b effect. The definition of each participant's P3b effect is given in Figure VI.25.

VI.C.3.b Common Best P3b IC Sentence Acceptability Task Activity

Figure VI.28 (D-F) presents the mean ERAs of the 12 common best P3b ICs to sentence final words in the sentence acceptability task. The mean ERAs featured an early negative peak at 150 ms, followed by a positive peak around 200 ms and a larger positivity that peaked around 600 ms, which was enhanced to anomalous endings. Between these early and late positivities, the mean ERAs to incongruous words went negative. Save for the anterior negativity (AN) elicited by ungrammatical pronouns, these 12 ICs accounted for around 50% of the grand average ERP differences between anomalous and nonanomalous words and the LPC (Table VI.5). Indeed, nine of the 12 ICs were a participant's "best LPC IC," which was the single IC that accounted for the most LPC variance.

Effects of Congruity

To determine the statistical significance of the effects of congruity on the mean activations of these ICs (Figure VI.33), the ANOVA/binomial test used in the previous section to analyze the oddball task trials was again applied and supplemented with one-tailed *t*-tests when warranted. The grand average N400 and P600 time windows[70] from Chapter IV were used to analyze these effects as the peak latencies of these effects in individual participant's ERPs were not clearly reliable (Section IV.G.1). In the N400 window, a two-way ANOVA with the factors Congruity (congruous, incongruous) and Part of Speech (noun, verb) found that eight of the ICs showed a

[70] The grand average N400 and P600 windows were 275-550 ms and 650-1100 ms, respectively.

main effect of congruity (binomial(12,.25)=8, p=.003). Follow up t-tests showed that for nine ICs the mean response to incongruous words was more negative than that to congruous words in the N400 window (binomial(12,.25)=9, p=.0004). Only one IC showed a significant difference in the opposite direction (binomial(l2,.25)=1, p=.97). Overall, the main effect of Part of Speech (binomial(12,.25)=1, p=.97) and the Part of Speech x Congruity interaction (binomial(12,.25)=1, p=.97) were not significant. On average, it appears that the effect of congruity in the N400 window was partially caused by a negativity to incongruous endings (Figure VI.28: E). However, one sample t-tests, found that only five of the ICs exhibited a significant (p<.25) negativity to incongruous words in the N400 window (binomial(12,.25)=5, p=.16). It appears that this "N400 effect" was also partially caused by a delay in the late positivity (Figure VI.29). According to a repeated measures ANOVA with factors Congruity (congruous, incongruous) and Part of Speech (noun, verb), the late positivity to incongruous words peaked later (F(1,11)=17.03, p=.002) and there was no main effect of Part of Speech (F(1,11)=.40, p=.54) although the effect of Congruity tended to be bigger for nouns (F(1,11)=4.33, p=.06). All p-values for this and all other repeated measure ANOVAs in this chapter are reported after Epsilon correction (Greenhouse-Geiser) for departures from sphericity.

As an alternative way to assess the extent to which this N400 effect was due to a difference in the polarity of activity versus a relative delay in activity, the absolute magnitude of each of the 12 ICs' ERAs to all incongruous endings was compared to that to all control congruous endings at each time point. Figure VI.34 (A) presents the

number of ICs whose absolute mean activation to incongruous endings was greater. There was no clear evidence for greater mean activation to incongruous or congruous endings within or after the N400 window. However, there was a trend for greater activation to congruous endings near the end of the N400 window, which was followed by a trend for greater activation to incongruous endings. This trend was similar to that found by the previous least squares source analysis (Section IV.G.2; Figure IV.58)

In the P600 window, ANOVAs with the factors Congruity (congruous, incongruous) and Part of Speech (noun, verb), found no significant effects (Congruity: binomial(12,.25)=5, p=.16; Part of Speech: binomial(12,.25)=3, p=.61; Congruity x Part of Speech: binomial(12,.25)=0, p=1), although there was an overall trend for the ERAs to incongruous endings to be more positive in this time range than those to comparable congruous endings (Figure VI.28: E).

Effects of Grammaticality

To assess the significance of the effects of grammaticality on the mean activity of the 12 common best P3b ICs (Figure VI.33), the ANOVA/binomial procedure was applied to their activations in the grand average N400 and P600 windows[71]. As with the congruity analysis, grand average effect windows were used as the latencies of

[71] The pronoun N400 and P600 windows were 375-550 ms and 500-950 ms, respectively. The verb N400 and P600 windows were 275-550 ms and 650-1100 ms, respectively.

these effects in individual participant's ERPs were not clearly reliable (Section IV.G.1).

In the N400 window, ANOVAs with the factors Grammaticality (grammatical, ungrammatical) and Part of Speech (pronoun, verb) found a nearly significant effect of Grammaticality in the N400 window (binomial(12,.25)=6, p=.05), which follow up t-tests showed to stem from a trend for ungrammatical words to be more positive (binomial(12,.25)=6, p=.05). There was no reliable effect of part of speech as both the main effect of Part of Speech (binomial(12,.25)=5, p=.16) and the Grammaticality x Part of Speech interaction (binomial(12,.25)=5, p=.16) were not significant, even though grammaticality tends to have different effects on verbs and pronouns in this time range (i.e., ungrammatical verb ERAs were more negative than those of grammatical verbs, but the direction of the trend was reversed for pronouns--Figure VI.28: F).

In the P600 window, a two-way ANOVA with the factors Grammaticality (grammatical, ungrammatical) and Part of Speech (pronoun, verb) found a significant effect of Grammaticality (binomial(12,.25)=7, p=.01) and the main effect of Part of Speech (binomial(12,.25)=5, p=.16) and the interaction between the two factors (binomial(12,.25)=4, p=.35) were not significant. Follow up t-tests, showed that the ERAs to ungrammatical endings were significantly more positive than those to grammatical endings (binomial(12,.25)=8, p=.003). None of the ICs showed a significant effect of grammaticality in the opposite direction.

Finally, the mean peak latencies of the late positivity to ungrammatical and nonanomalous pronoun and verb endings are presented in Figure VI.29. A repeated measures ANOVA with factors Grammaticality (grammatical, ungrammatical) and Part of Speech (pronoun, verb), found that the late positivity to ungrammatical words peaked later ($F(1,11)=12.94$, $p=.004$) and there was no main effect of Part of Speech ($F(1,11)=2.22$, $p=.16$) nor a clear interaction between the two ($F(1,11)=3.17$, $p=.10$).

Congruity vs. Grammaticality

To best contrast the effects of congruity and grammaticality, one-tailed *t*-tests were applied to the responses of these 12 ICs to incongruous and ungrammatical verbs. There was a borderline significant trend for the mean response to incongruous verbs to be more negative than that to ungrammatical verbs in the N400 window (binomial(12,.25)=6, $p=.05$), although two ICs had a significant difference in the opposite direction (binomial(12,.25)=2, $p=.84$). It does not appear that this trend can be explained by a relatively later late positivity to ungrammatical verbs, as the late positivity to ungrammatical verbs tended to peak before that to incongruous words (Figure VI.29). However, this difference was far from significant ($t(11)=.47$, $p=.65$). In the P600 window, the mean response to ungrammatical verbs was not clearly more positive (binomial(12,.25)=5, $p=.16$) than that to incongruous verbs and there was no evidence of an effect in the opposite direction (binomial(12,.25)=1, $p=.97$).

N400 Effect and LPC Amplitude Correlation

Since the amplitude of an individual's N400 effect was shown to negatively correlate with the amplitude of that individual's LPC (Section IV.G.1), a possible contribution of the 12 common best P3b ICs to this relationship was investigated by repeating the analysis on the scalp-projected contribution of these ICs to the N400 effect and LPC. The mean amplitude of each IC's contribution to the N400 effect (averaged across part of speech) and LPC (averaged across part of speech) was measured using the same electrodes and time windows as the previous analysis to make the two analyses as comparable as possible. The mean contribution of the 12 common best P3b ICs to the N400 effect and LPC was negatively correlated (*tau*=-0.48, *p*=.03), though the correlation was weaker than what was observed in the electrode data of all 16 participants (*tau*=-.63, *p*=.0003).

To get a sense for how much of the relationship between LPC and N400 effect amplitude these 12 ICs account for, the contribution of these 12 ICs was removed from these individual's ERPs and the mean amplitude of each individual's LPC and N400 effect re-measured (again using the same time windows and electrodes as used in the original analysis). Without the 12 best P3b ICs, there was no longer evidence of a relationship between LPC and N400 effect amplitude across all 16 participants (*tau*=-.13, *p*=.51).

Cloze Probability

Because of the sizeable contribution of these 12 ICs to the P3b effect (which is generally sensitive to stimulus probabilities), their activations were analyzed for possible effects of the cloze probability of nonanomalous noun and verb sentence endings[72]. The IC responses to these nonanomalous endings were split into two approximately equal sized subsets according the median cloze probability of their corresponding words[73] (Figure VI.35). ANOVAs with the factors Cloze Probability (high, low) and Part of Speech (noun, verb) found no reliable main effect of Cloze Probability in either the N400 (binomial(12,.25)=5, p=.16) or P600 (binomial(12,.25)=5, p=.16) windows, although the difference between the two was generally in the expected direction (i.e., less probable words elicited more positive activations). The interaction of the two factors was far from significant in both time windows (binomial(12,.25)=2, p=.84). To get a richer sense of a possible relationship between cloze probability and these ICs' activations, an ERPimage was constructed from their normalized activations and is presented in Figure VI.36. Although there was a clearly a great deal of variation across the range of cloze probability, there did appear to be a tendency for the late positivity to low cloze probability words to occur later than that to more predictable words. To determine the statistical significance of this trend, a Woody filter was applied separately to each IC to determine the latency of

[72] As with the scalp level analysis of cloze probability effects (Section IV.F.3.g) responses to anomalous endings were excluded from the analysis to avoid confounding the analysis with effects of anomaly. In addition, responses to nonanomalous pronouns were excluded because such closed-class words are differently affected by preceding context than open-class words (Van Petten & Kutas, 1991) and their frequencies and cloze probabilities markedly differ from that of nouns and verbs in this experiment.

[73] The mean (across the 12 ICs) of the median and semi-interquartile range (across stimulus words) of the corresponding cloze probabilities of the high and low subsets are .80 (SIQR=.08) and .44 (SIQR=.12) respectively. The mean (minimum) number of trials in the high and low subsets were 44 (41) and 43 (37) respectively, with approximately twice as many verb trials as nouns.

the late positivity. Specifically, a template of the late positivity for each IC was obtained from a 400 ms window of the IC's ERA to nonanomalous noun and verb endings. The 400 ms sample extended from 250 ms before to 150 ms after the peak of the positivity (Figure VI.29). The cosine similarity, r_{cos}, between the late positivity template and each single trial at each time point (starting at 64 ms before the onset of the stimulus) was computed. The latency of the late positivity was simply defined as the latency of maximal r_{cos}. Thus, this procedure assumed that the late positivity had a fixed, 400 ms long wave shape and polarity for each time point, but variable latency and magnitude. The rank correlation between the latency of the late positivity and the cloze probability of the eliciting words was computed separately for each IC. In general, the latency of the late positivity was only weakly correlated with the cloze probability of the eliciting word (median [min/max] Kendall's *tau*: -.01 [-.19/.10]). Lower-tailed tests of the null hypothesis *tau*=0 found that only four ICs showed a significant (p<.25) negative correlation (the expected relationship), which overall is not significant (binomial(12,.25)=4, p=.35). An upper-tailed test of the same null hypothesis found that five other ICs have a significantly positive correlation, though this is still not significant (binomial(12,.25)=5, p=.16).

Entropy

While the cloze probability based analysis did not reveal a clear relationship between the activation of these ICs and the degree to which a stimulus met expectations, it was possible that their activations were sensitive to expectation

violation. To investigate this possibility, the IC responses to incongruous and ungrammatical endings were separately split into two approximately equal sized subsets according the median entropy of their preceding sentence stems[74] (Figures VI.37, VI.39). A two-way ANOVA with the factors of Entropy and Part of Speech (noun, verb) performed on the IC responses to incongruous endings found no evidence of a main effect of Entropy (high, low) or an interaction between the two factors (binomial(12,.25)<4, p>.34) in either the N400 or P600 time window. Likewise, a two-way ANOVA with the factors Entropy (high, low) and Part of Speech (pronoun, verb) performed on the IC responses to ungrammatical endings found no evidence of a main effect of Entropy or an interaction between Entropy and Part of Speech (binomial(12,.25)<4, p>.34) in either the N400 or P600 time window. ERPimages of these data (Figures VI.38, VI.40) also revealed no clear monotonic relationship between entropy and these ICs' activations.

Habituation

In light of the general habituation of the response of these ICs in the oddball tasks, similar tests for the habituation of the late positivity to sentence final words

[74] The mean (across the 12 ICs) of the median and semi-interquartile range (across stimulus words) of the corresponding sentence stem entropies of the high and low incongruous subsets are 2.17 (SIQR=.48) and .90 (SIQR=.42) bits, respectively. The mean (minimum) number of trials in the high and low incongruous subsets were 31 (25) and 30 (24) respectively, with approximately equal numbers of nouns and verbs. The mean (across the 12 ICs) of the median and semi-interquartile range (across stimulus words) of the corresponding sentence stem entropies of the high and low ungrammatical subsets are 2.65 (SIQR=.43) and 1.26 (SIQR=.27) bits, respectively. The mean (minimum) number of trials in the high and low ungrammatical subsets were 30 (24) and 29 (24) respectively, with approximately equal numbers of pronouns and verbs.

were performed. ANOVAs with the factors Congruity (congruous, incongruous) and Order of Occurrence (former, latter) were performed on the mean activations to congruous and incongruous noun and verb endings. Part of Speech was not included as a factor in these ANOVAs due to the limited number of trials (e.g., there would be about 15 per cell), but noun and verb trials were split evenly across ANOVA bins[75]. Mean activations to these sentence endings were measured in 300 ms windows centered at the peak latency of that participant's LPC to all incongruous or congruous endings[76]. These ANOVAs found a borderline significant main effect of Order of Occurrence (binomial(12,.25)=6, p=.05) and the Congruity x Order of Occurrence interaction was not significant (binomial(12,.25)=3, p>=.61). Follow up one-tailed t-tests found that the average response to all congruous and incongruous words were more positive in the former half of the experiment (binomial(12,.25)=7, p=.01; Figure VI.41), though one IC showed the opposite effect (binomial(12,.25)=1, p=.97).

A similar analysis was carried out on the mean amplitude of these ICs' responses to grammatical and ungrammatical pronoun and verb endings using ANOVAs with factors Grammaticality (grammatical, ungrammatical) and Order of Occurrence (former, latter). Again, Part of Speech was not included as a factor in these ANOVAs due to the limited number of trials, but pronoun and verb trials were split evenly across ANOVA bins. Mean activations to these endings were measured in

[75] There were approximately an equal number of noun and verb trials in the two incongruous bins and approximately two verb trials for every noun trial in the two congruous bins.

[76] Individual congruous LPC peak latencies were determined by first deriving the ERPs to all nonanomalous noun and verb endings at all 64 electrodes. These ERPs were then low pass filtered at 7 Hz and the latency of their most positive point between 300 and 1100 ms identified. The median of these 64 peak latencies was taken as the peak latency of the congruous LPC. The same procedure was applied to incongruous endings to estimate an individual's incongruous LPC peak latency.

300 ms windows centered at the peak latency of that participant's LPC to all ungrammatical or grammatical endings[77]. These ANOVAs found a main effect of Order of Occurrence (binomial(12,.25)=8, p=.003) and the Grammaticality x Order of Occurrence interaction was not significant (binomial(12,.25)=4, p=.35). Follow up one-tailed t-tests found that pronoun and verb endings were more positive in the former half of the trials (binomial(12,.25)=8, p=.003; Figure VI.42) and no ICs showed a significant effect in the opposite direction (binomial(12,.25)=0, p=1).

An ERPimage of the 12 ICs' responses to all sentence final words is presented in Figure VI.43. A graded effect of habituation is apparent in the plot, although part of the effect appears to have been due to a brief burst of negative drift from the 175th to 200th trial.

VI.C.3.c Common Best P3b IC Variability

Although the ICs in the class of best P3b ICs were similar in some regards (e.g., their topographies and their contributions to the P3b, N400, and P600 effects), there was also a considerable amount of variation across participants and single trials. For example only seven of these 12 ICs significantly showed all three main ERA effects (i.e., target ERAs > standard ERAs in the P3b window, incongruous ERAs < congruous ERAs in the N400 window, ungrammatical ERAs > grammatical ERAs in the P600 window). This variability may indicate that the ICs in this class were not

[77] Individual, grammatical and ungrammatical LPC peak latencies were derived by a procedure analogous to that used to estimate individual congruous and incongruous LPC peak latencies (see previous footnote).

homologues of one another (in which case the failure was due to an inaccurate grouping of ICs), a failure of ICA to accurately extract subcomponents of these ERP effects (in which case the failure was due to ICA), and/or poor quality data. Alternatively, given the amount of individual variation and trial-to-trial variation in the electrode data (Section IV.H.2), it is possible that the class of ICs was accurate. Of course, without knowing the neural sources of these data, it is impossible to know to what extent any of these possibilities were true. In the absence of this knowledge, it is useful to compare IC variation with that of the electrode data. If the members of a class of ICs are more similar across participants and/or have a cleaner relationship with an independent variable than electrodes, it suggests that the ICs have successfully extracted an EEG subcomponent and, at the very least, indicates that the ICs are a useful representation of the data. On the other hand, if the converse is true, it casts doubt on the accuracy and utility of the ICs, though it clearly does not prove this is the case.

Figure VI.44 (A) presents the median topography similarity of each possible pair of the 12 common best P3b ICs (66 pairs). In general, the topographies across participants were highly similar (median r_{cos}=.95). They were significantly[78] more similar than the same participants' individual P3b effect topographies[79] (two-tailed

[78] Note that the significance of these topography and ERA/ERP similarity differences are overestimated to some extent as the similarities of the 66 comparisons are not independent (as assumed by the binomial test).

[79] An individual's P3b effect topography was measured as the mean of his/her P3b effect (averaged across both oddball tasks) in a 300 ms window centered on the peak latency of the P3b effect. The peak latency of this combined oddball task P3b effect was determined using the same procedure used to find the peak latency of the P3b effect in each oddball task (Section IV.E.3.a). The combined P3b effect

sign test, binomial(66,.5)=14, p=3e-6), N400 effect topographies[80] (two-tailed sign test, binomial(66,.5)=19, p=8e-4), and P600 effect topographies[81] (two-tailed sign test, binomial(66,.5)=3, p=1e-15). Indeed, they tended to be even more similar than an individual's P3b topographies across the two oddball tasks (r_{cos}=.92, Table IV.4).

To assess ERA activation variability, a participant's "best P3b electrode" was chosen as the foil for his/her best P3b IC. A participant's best P3b electrode was simply the electrode that alone accounted for the most variance of an individual's P3b effect (Section VI.C.3.a). ERPs and ERAs to targets and standards were averaged across both oddball tasks (Figures VI.27, VI.45) and compared from 0 to 1000 ms. ERPs and ERAs to incongruous, congruous, ungrammatical, and grammatical endings were averaged across part of speech (Figures VI.33, VI.46) and compared from 0 to 1300 ms. Figure VI.44 (B) presents the median ERA/ERP similarity of each possible between participant comparison (12 participants, 66 comparisons). Although there was clearly a great deal of between participant variation in ERA and ERP morphology, ERPs tended to be more similar across participants than ERAs. This difference was significant for ERAs/ERPs to standard and targets (two-tailed sign test, binomial(66,.5)=53, p=1e-6) and to congruous and incongruous sentence endings (two-tailed sign test, binomial(66,.5)=44, p=.009), but was not significant for ERAs/ERPs to grammatical and ungrammatical sentence endings (two-tailed sign test,

---

was used here for simplicity as qualitatively similar results are achieved when comparing P3b effect topographies from a single oddball task.

[80] An individual's N400 effect topography was measured as the mean of his/her N400 effect (averaged across nouns and verbs) in the grand average N400 effect window.

[81] An individual's P600 effect topography was measured as the mean of the pronoun P600 effect in the grand average pronoun P600 window and of the verb P600 effect in the grand average verb P600 window.

binomial(66,.5)=39, p=.17). Comparison of the mean root mean squared difference between participant ERPs and ERAs at each time point (not shown) suggested that oddball task ERAs were generally less similar between participants than their comparable ERPs throughout the time window of comparison (0-1000 ms). The same comparison for sentence acceptability task ERAs and ERPs (not shown) suggested that ERPs were more similar between participants before 600 ms, after which ERAs tended to be more similar.

Finally, to get a sense of how cleanly the best P3b IC activations followed the primary independent variables of interest in the two tasks, the separability of the IC responses to target/anomalous and standard/nonanomalous stimuli was compared to that at the best P3b electrode. If the IC response to deviance was less variable than that at the best P3b electrode, then the mean activations to deviant and non-deviant events should have been more distinct. One way to assess this is via the significance of the deviance effects as measured by independent sample t-tests (treating each single trial as an independent sample). Table VI.6 presents the median p-values of the P3b[82], N400[83], and P600[84] effects. All three effects tended to be more significant at the best P3b electrode. This difference was reliable for the P3b effect (two-tailed sign test, binomial(12,.5)=10, p=.04) but was not reliable for the N400 (two-tailed sign test,

[82] The P3b effect was measured by comparing the mean single trial amplitude to targets and standards in each individual's letter case task and oddball task P3b windows (Section IV.E.3.a). The mean amplitudes from both oddball tasks were combined into a single t-test per participant.
[83] The N400 effect was measured by comparing the mean single trial amplitude to incongruous and nonanomalous noun and verb endings in the grand average N400 window (275-550 ms). The mean amplitudes from both nouns and verbs were combined into a single t-test per participant.
[84] The P600 effect was measured by comparing the mean single trial amplitude to ungrammatical and grammatical pronoun and verb endings in the grand average P600 windows (pronouns: 500-950 ms; verbs: 650-1100 ms). The mean amplitudes from both pronouns and verbs were combined into a single t-test per participant.

binomial(12,.5)=9, p=.15) or P600 (two-tailed sign test, binomial(12,.5)=7, p=.77) effects.

VI.C.3.d Interim Summary II: Common Best P3b ICs

Common Best P3b IC: Summary

Using each individual's P3b effect, it was possible to define a "best P3b IC" in each of the 16 participants. 12 of these ICs appeared to replicate across the oddball and sentence acceptability tasks and formed a class of common ICs. These 12 ICs accounted for a bit more than half of the grand average P3b effects and generally behaved like the P3b. On average, each IC's response to targets was significantly more positive than that to standards, and this positivity peaked about 100 ms later in the presumably more difficulty semantic category task. Unlike the P3b, the response of these ICs to targets was not reliably less positive when those targets followed another target. However, this null effect may simply reflect a lack of statistical power, as the target context effect was only weakly apparent in the electrode data. Finally, the mean response of these ICs to both targets and standards appeared to quickly habituate in the oddball tasks. While habituation is more commonly considered a characteristic of the P3a rather than the P3b (Knight & Scabini, 1998), there is evidence that both habituate over the course of an experiment (Fabiani & Friedman, 1995).

These 12 ICs also accounted for about half of the variance of the deviance effects in the sentence acceptability task (with the exception of the anterior negativity to ungrammatical pronouns), appeared to account for the negative correlation between individual LPC and N400 effect amplitude, and behaved generally, but less consistently, like the P3b. In the main, the ICs were characterized by a late positivity that was larger to anomalous endings than nonanomalous (quite probable) endings, though this difference was not significant for incongruous endings. Since the average amplitude of the P3b tends to decrease with increasing task difficulty (Bentin et al., 1999; Kok, 2001; Kutas, McCarthy, & Donchin, 1977) and the incongruous endings are presumably the most difficult to detect (Section II.G.1), it is quite possible that the null effect is simply due to insufficient statistical power. The insignificant trend for the positivity to be greater to ungrammatical endings than incongruous endings was also likely to be due to a lack of power as this comparison was limited to anomalous verb endings.

Also like the P3b, the peak latency of this late positivity was generally later to incongruous than comparable congruous endings that presumably take less time to evaluate (Section II.G.1). The peak latency was also later to ungrammatical than grammatical endings, though, according to previous findings, it is not clear if ungrammatical endings should take more (Hagoort, 2003) or less time (Kuperberg et al., 2003) to detect than grammatical endings. However, the peak latency to sentence endings did deviate from the expected P3b-like pattern in other ways. Based on participant reaction times (Section IV.F.1) the late positivity to pronoun endings

should have been earlier than that to verbs, and based on previous research, the late positivity to ungrammatical verbs should have been earlier than that to incongruous verbs (Hagoort, 2003; Kuperberg et al., 2003; McElree & Griffith, 1995). To some extent these null effects could be due to noisy peak latency estimates. In addition, since participants were instructed to delay responses in either the oddball or sentence acceptability task it was difficult to know how well the peak positivity truly followed stimulus evaluation duration. Clearly immediate response versions of these tasks should be run to get better estimates of how closely the peak latency of the late positivity covaries with detection latency.

These ICs also appeared to behave unlike the P3b in that there was an insignificant trend for their response to incongruous endings to be negative in the N400 time window. While it is possible that the trend was random variation, it was also not clear that the contribution of these ICs to the N400 effect was simply due to a delayed positivity.

Finally, there was no evidence of an effect of the cloze probability of nonanomalous sentence completions or the entropy of the cloze probability distribution of possible completions and the late positivity to all sentence completions generally habituated across the course of the experiment. The null effects of cloze probability and entropy were quite possibly due to the limited range of cloze

probabilities, as all the nonanomalous sentence endings were quite predictable relative to naturalistic text[85].

Common Best P3b IC: Relation to Previous Research

There appears to be some precedent for such a class of ICs in previous ICA work. The average topography and generally linear spectra of this cluster strongly resembles those of a "frontal P3" cluster of ICs found by Debener et al. (2005) in an auditory novelty oddball paradigm and that of a "central midline" cluster found by Makeig et al. (2004) in a visual target detection task. Like the best P3b ICs, these IC clusters exhibited a late positivity following targets, which appeared to be preceded by a smaller positivity that peaked around 200 ms. Of the two clusters, the common best P3b ICs are superficially most similar to the Frontal P3 cluster which accounted for about 18-30% of the variance of the ERPs to targets[86]. However, unlike the common best P3b ICs, the Frontal P3 cluster's late positivity showed no evidence of habituating to targets. This difference may be due to the different methods of analysis[87] and/or the fact that the targets Debener et al. analyzed were novel sounds that might have seemed much more deviant in the experimental context than the targets and sentence final

[85] Philip Gough (1983) tentatively estimated that a well educated reader can accurately guess about 10% of all open class words and 40% of all closed class words, given the first seven words of a sentence (see also Gough, Alford, & Holley-Wilcox, 1981). The nonanomalous noun and verb completions used in the cloze probability analysis are generally much more predictable than this (Table IV.1: A, Figure VI.35).

[86] Debener et al. (2005) reported the variance accounted for of the grand average ERPs from 0 to 800 ms. It appears that had they used a smaller window centered around the P3b, their Frontal P3 cluster would have accounted for a greater percentage of the variance.

[87] Debener et al. analyzed the scalp-projected activations of their ICs using ANOVAs and *t*-tests. This procedure is more susceptible to outlier ICs than the procedure used here.

words used here. Given that the habituation effect was not remarkably consistent here, other factors may easily overpower it, if indeed it is real. The common best P3b ICs were like Makeig and colleagues' central midline cluster in that the peak of the late positivity appears to correlate with reaction time. However, the central midline cluster's late positivity had a maximum amplitude of about only 2 microvolts, contributed little to the grand average ERP to targets, and was more compact in time than the late positivity reported here. The grand average ERA of Makeig et al.'s "P3b cluster" was more similar to that of the common best P3b ICs, however the topographies of that cluster were generally more posterior than those of many of the common best P3b ICs. It is difficult to relate the common best P3b ICs to what I believe is the one published application of ICA to the N400 (Fogelson, Loukas, Brown, & Brown, 2004) as the authors reported the topography and ERA of ICs from only a single participant.

With respect to previous attempts to localize the sources of the P3b, N400, and P600 or comparable fMRI research, it is tempting to speculate that these ICs primarily reflect the activity of central/posterior cingulate and/or superior parietal cortex. There is some iEEG evidence of a cingulate source of the P3a (Halgren, Marinkovic, & Chauvel, 1998) but an fMRI guided source model (Bledowski et al., 2004) and a highly distributed MRI-based source model (Moores et al., 2003) of the P3b found either no evidence of a cingulate source or a cingulate source that did not contribute a large late positivity. There is some erfMRI evidence of increased cingulate activity to targets in oddball paradigms (Horovitz, Skudlarski, & Gore, 2002), however it does

not appear to correlate with individual average (ibid.) or single trial (Eichele et al., 2005) P3b amplitudes. Attempts to localize the N400 and P600 (Sections II.E.2.a, II.E.3.a) have not reported a cingulate contribution, though this could be due to erroneous model assumptions. There is some fMRI/erfMRI evidence of ungrammatical and incongruous words modulating cingulate activity. Two studies reported increased posterior cingulate activity to ungrammatical words and decreased cingulate activity to incongruous words (Kuperberg et al., 2003; Ni et al., 2000). Newman and colleagues (2001) reported an increase in central cingulate activity in response to incongruous words, but not ungrammatical words. In contrast, Friederici et al. (2001) reported no modulation of cingulate activity in response to either type of anomaly. Of these results, the best P3b IC ERAs appear most consistent with an increased hemodynamic response to ungrammatical words and a small or insignificant increased response to incongruous words, as it appears that the difference between congruous and incongruous ERAs is primarily one of polarity rather than magnitude. This difference in the polarity of activity could be due to a change in the balance of activity in superior and inferior cortical layers and/or a difference in the balance of excitatory and inhibitory synaptic potentials (Martin, 1991). If fMRI is primarily sensitive to synaptic activity (Logothetis & Pfeuffer, 2004) and does not differentiate between excitatory and inhibitory post-synaptic potentials or the cortical depth of this activity, then such a modulation would be relatively invisible to fMRI.

There are some MRI guided localization (Moores et al., 2003) results that suggest a superior parietal contribution to the P3b. Moreover, Eichele and colleagues

(2005) found that right superior parietal hemodynamic activity correlated with single trial P3b amplitude. However, attempts to localize the N400 and P600 have not found superior parietal sources (Sections II.E.2.a, II.E.3.a). Furthermore, several erfMRI studies have found no modulation of superior parietal area activity by incongruous (for review see Van Petten & Luka, 2006) or ungrammatical words (Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Kuperberg et al., 2003; Newman, Pancheva, Ozawa, Neville, & Ullman, 2001; Ni et al., Experiment 2, 2000). Thus, of the two possibilities, a central/posterior cingulate source appears more likely, and obviously, given the ambiguity if the EEG inverse problem (Section III.C.1), more than these two alternatives are possible (e.g., inferior parietal and insula sources).

Common Best P3b ICs: Interpretation

As the behavior of these 12 ICs was either consistent with or not clearly inconsistent with the expected behavior of the P3b, it is quite possible that they corresponded functionally to the conventionally measured P3b. Thus, the late positivity could have indexed the update of a model of the stimulus environment (Donchin & Coles, 1988; Friston, 2005) or an attentional response mediated by the locus coeruleus (Nieuwenhuis, Aston-Jones, & Cohen, 2005; Pineda, Foote, & Neville, 1989). It seems unlikely that the late positivity of these ICs corresponded to a "template match" (Kok, 2001; Verleger, Jaskowski, & Wauschkuhn, 1994) as the late positivity was smaller to the rather predictable nonanomalous sentence endings.

However, the behavior of the common best P3b ICs appeared equally consistent with the P3a subcomponent of the conventionally measured P3b, which is thought by many to reflect an attentional component of the nervous system's orienting response to stimuli (Friedman, Cycowicz, & Gaeta, 2001; Halgren & Marinkovic, 1995). Indeed, given the central midline focus of many of these ICs and their similarity to Debener, Makeig, and colleagues' (2005) aforementioned frontal P3 ICs that contributed significantly to the conventionally measured P3a, they corresponded somewhat more closely to the P3a than the P3b.

Further discussion of the possible functional correlates of these ICs and their implications for our understanding of language comprehension will be postponed until the final chapter. However, it is also worth considering here the possibility that these ICs did not largely correspond to a single functionally and neurally distinct process that was common to the comprehension of sentence final words and the oddball task stimuli. As mentioned previously (Section VI.C.2.a), there are several reasons why a common IC might be spurious and that compelling classes of ICs are necessary to validate apparently common ICs. In some ways this grouping of common best P3b ICs appeared quite believable: their responses to targets and standards and their topographies were quite similar across participants. Moreover, their topographies appeared to be consistent with a single focal neural source or bilaterally symmetric pair of sources (e.g., P2 IC 2). On the other hand, the response of these ICs to sentence final words was rather variable across participants, their average responses (ERAs) to major stimuli classes were less similar across participants than comparable

ERPs, and the difference between targets/anomalous endings and standards/nonanomalous endings tended to be less pronounced in their mean activations than at comparable electrodes.

This variability in IC activations makes the grouping of ICs less compelling and could have several possible causes. It could be that the class of ICs was poorly formed and that a more accurate and more consistent grouping is possible. However, the choice of each participant's most P3b-like IC was quite clear except in one case[88]. It is possible that some of the common ICs were false positives due to the large number of comparisons or that they corresponded to EEG processes that were primarily active during the sentence but did not contribute that much to the average response to sentence final words. It could be that some of the ICs in this class were not meaningful (i.e., they did not approximate a single EEG process that was functionally and neurally distinct) and were simply a way to decompose the data into somewhat independent components. Finally, of course, the variability could have reflected true individual and trial-to-trial variation or poor quality data. Surely individual variability was somewhat responsible for the IC variation given the individual variability apparent in the electrode data (Section IV.H.2). In addition, although the quality of the data could have surely been improved (given the large number of ICs corresponding to EEG artifacts), the data quality was good enough that reliable ICs were able to account for the grand average P3b, N400, and P600 effects at levels conventionally deemed acceptable for ERP source models (Section IV.E.3.d)

[88] Participant 14's IC 2 was chosen as his best P3b IC since it contributed the most to his P3b effect. However, his IC 4 had a topography more similar to the group's mean.

and the scalp level analyses of these data (Chapter IV) found the P3b and N400 effects and the LPC to be very reliable within participants.

The only of these alternatives that was explored was the possibility that better groupings of common ICs were possible. Rather than grouping ICs by their contribution to an ERP feature, the subsequent section reports the results of an automated cluster analysis that grouped ICs according to simple models of the location of their neural generators (primarily).

VI.C.4 Cluster Analysis of Common ICs

As an alternative to the best P3b IC grouping, automated cluster analysis was performed on the set of ICs common to both experiments. Automated clustering has been used in previous applications of ICA to EEG analysis (Fogelson et al., 2004; Makeig et al., 2004) and has the advantage that it returns multiple groupings of ICs, which are clearly necessary to understand the ICA decomposition of the P3b, N400, and P600 effects as they are each distributed across multiple ICs. Furthermore, it groups ICs based on their similarity rather than their contribution to an ERP effect and does not require each participant to contribute exactly one IC. Thus, cluster analysis can potentially provide more homogenous and, presumably, accurate groupings. The disadvantage of cluster analysis is that the results depend heavily upon clustering parameters (e.g., an assumed number of latent clusters) and upon a definition of IC similarity, both of which are difficult to choose a priori.

The following sections first explain the IC features and similarity metric used for clustering. They then cover the specifics of the clustering algorithm and cluster activation analysis, summarize the features and behavior of the resulting ten common IC clusters, and finally suggest interpretations of six of these clusters.

VI.C.4.a IC Source Modeling

For IC clustering the neural generators of non-artifact ICs from the combined data sets were modeled as a single or as a bilaterally symmetric pair of equivalent current dipole(s) (ECDs) using the DIPFIT 2 software (Oostenveld & Oostendorp, 2002) included with the EEGLAB toolbox (Delorme & Makeig, 2004). More specifically, the topography of each IC (not including the EOG and right mastoid weights) was re-referenced to the average reference and mapped to a standard four-shell spherical head model using ideal electrode locations (Appendix 1—Section IV.I). For each IC topography, an initial ECD location and orientation were chosen by searching a coarse grid of possible locations. Initial ECD parameters were improved by iterative, non-linear optimization. ICs with symmetric lateral foci (e.g., IC 4 in Figure VI.10) were modeled with a bilaterally symmetric pair of ECDs[89]. The initial locations and orientations of these pairs were hand chosen and iteratively improved by automatic, non-linear optimization.

[89] The anterior-posterior and dorsal-ventral coordinates of each member of the pair were constrained to be equal and the left-right coordinates were constrained to be symmetric across the midline axis. The orientation of each member of the pair could independently vary.

On average, each participant's data from the combined experiments produced 20 non-artifact ICs that could be modeled with a single or bilaterally symmetric pair of ECDs with a residual variance of .15 or less, which is a slightly greater percentage of dipolar ICs than was found in a comparable previous application of ICA (Onton, Delorme, & Makeig, 2005). However, on average only 13 of these ICs passed the reliability criteria. All but five of the 116 ICs common to both experiments were sufficiently dipolar. Specifically, one IC was not included because it could not be fit by a single ECD with a residual variance of .18 or less. Four other ICs were considered non-dipolar because their ECDs were implausibly deep (i.e., $z<-53$ in Talairach coordinates). These five ICs were excluded from the cluster analysis.

Finally, it is important to note that because ECD modeling was based on an idealized spherical head model and ideal electrode locations, the resulting models were probably not very accurate. However, the models provided a low dimensional representation of IC scalp topographies that were useful for clustering.

VI.C.4.b Clustering Algorithm and Features

Agglomerative hierarchical clustering (Duda, Hart, & Stork, 2001) using Ward's linkage criterion (Ward, 1963) was employed for the cluster analysis. Agglomerative hierarchical clustering was chosen because it provides a single grouping of ICs given a set number of clusters (as opposed to iterative algorithms like *k*-means that can return different groupings depending on initial conditions); this

property greatly facilitated the use of the cluster validation algorithm described below. Ward's linkage criterion was chosen because, like other minimum variance clustering criteria (see Duda, Hart, & Stork, 2001, Section 10.7), it is biased to return clusters of roughly equal size. This is desirable given the assumption that a certain type of IC should occur approximately once in each participant.

To accommodate the ICs fit by bilaterally symmetric ECDs, the cluster analysis was performed on the ECDs for a total of 122 clustering data points. Three features of each ECD were used in clustering. The first of these was the location of the ECD in Cartesian coordinates. The second was the location, in Cartesian coordinates, of the ECD's "scalp intersect" that was closest to the maximum absolute value of the corresponding IC topography. An ECD's scalp intersects are the two points where the line defined by the ECD location and orientation intersect the scalp of the spherical head model. Together these two features concisely represent the general location and spread of the foci of IC topographies.

The third clustering feature was the absolute *t*-scores of the first 200 ms of the ECD's corresponding IC's event-related activation[90] (ERA) to the first stimulus in each epoch (Figure VI.68). This feature was simply the ratio of ERA absolute amplitude to the standard error of that amplitude at each time point and was useful for identifying ICs whose activity was relatively phase-locked to stimulus onset (e.g., IC 4 in Figure VI.10). The absolute value of the *t*-scores were used to accommodate the polarity ambiguity of IC activations (Section III.D.2.a). In addition, because the

[90] To resolve the IC scaling ambiguity (Section III.D.2.a), IC activations were calculated after normalizing IC mixing vectors to unit length.

magnitude of the *t*-scores could vary greatly between participants, the *t*-scores of each participant's ICs were divided by the root mean squared *t*-scores of all non-artifact ICs from that participant.

To combine these three features, the 50 dimensional *t*-score feature was first reduced to five using principal components analysis. The first five principal components accounted for 93% of the feature's variance. Subsequently, to make the magnitudes of each of the three features comparable, the first two features were normalized such that their first principal components were of unit variance. The *t*-score feature was normalized such that the variance of the first principal component was .5 to down-weight its importance. Finally the three features of each ECD were concatenated to produce an 11 dimensional feature vector.

This feature space was chosen because it tended to produce clusters of ICs that projected to largely the same electrodes, thus following the conventional EEG analysis assumption that homologous EEG processes from different participants tend to project to the same electrodes. The *t*-score feature was a minimal addition, necessary for separating occipital ICs.

VI.C.4.c Cluster Validation

To determine the number of latent IC clusters, the stability-based cluster validation algorithm of Lange and colleagues (2004) was applied. The logic of the algorithm is simple: the latent number of clusters is the number that produces the most

reliable clusters. This criterion is limited in that cluster reliability is a necessary but not a sufficient condition for accurate clustering results. That being said, it is a useful analytical guide that serves to check human biases and overfitting of the data. Furthermore, the algorithm provides a sense of how likely the results are to replicate.

To estimate the reliability of *k* clusters Lange et al.'s algorithm was implemented as follows:

1. The clustering data was split in half by assigning data points from a random half of the participants to Half A and data from the remaining participants to Half B.
2. The clustering algorithm independently grouped each half of the data into *k* clusters.
3. The clusters from Half A were used to classify the data points in Half B. This was done by assigning each ECD in Half B to the nearest cluster from Half A using Ward's (1963) linkage metric.
4. The classification labels were optimally matched to the cluster labels of Half B to produce a minimal level of classification error (0-1 loss function, see Lange et al., 2004, pgs. 1303-1304)
5. The classification labels were randomly permuted and again optimally matched to the cluster labels of Half B to estimate the classification error that could be achieved by chance.
6. Step 5 was repeated 500 times.

7. Steps 1-6 were repeated 500 times.
8. The mean classification error from Step 4 was divided by the mean random classification error from Steps 5-6 to estimate the percentage improvement in classification error provided by the clustering algorithm. The lower this normalized error was, the more reliable the clustering results were.

This procedure was repeated for all values of *k* between 2 and 40. Resultant estimates of cluster classification error and random classification error are presented in Figure VI.47 (Top). Four clusters produced the most reliable results, however the clusters were overly inclusive. There was one cluster for frontal ICs, one cluster for all post-frontal left lateral ICs, one midline central/right parietal cluster, and one cluster for posterior midline and occipital right lateralized ICs. The second minima at 10 clusters produced much more reasonable groupings and was chosen as the number of clusters for analysis. Figure VI.47 (Bottom) illustrates the 10 clusters and the overall similarity structure of the clustering data with a dendrogram. The distance between each cluster node is represented by the vertical distance between it and the immediately superior or inferior nodes. For example cluster LFTe was much more distinct from cluster LPa than cluster MOc was from cluster MPa.

VI.C.4.d Cluster Mean Activation Analysis

The same procedure applied to the analysis of the mean activations of common best P3b ICs (Section VI.C.3.a) was applied to each cluster of ICs, with the difference

that an alpha level of .05 was used in all single IC hypothesis tests instead of .25. The smaller alpha level adjusted the hypothesis testing procedure such that it was more sensitive to effects that are more clearly evident in at least a few ICs, even though most of the ICs in the cluster may not show the effect. This was desirable as some of the automatically formed clusters were more heterogeneous than the hand-screened class of best P3b ICs. A second difference between the common best P3b IC analysis and the cluster activation analyses is that the same time windows used in Chapter IV were used here (i.e., the grand average P3b windows were used in place of the individual P3b windows) for simplicity.

VI.C.4.e Clusters of Common ICs

Figures VI.48-67 present the mean topographies, ERAs, and spectra of the ten IC clusters as well as the topographies of each cluster member[91]. As with the class of best P3b ICs, the mean features of multiple ICs were all computed after normalizing IC topographies to unit length and setting the polarity of the topography such that the maximum absolute channel weight was positive. The median topography similarity of all pairs of ICs in a cluster is presented above each mean topography as a measure of topography dispersion. These figures are complemented by the average scalp-projected ERAs of cluster ICs at select electrodes (Figures VI.69-98). These alternative depictions of cluster activity have the advantage of being unaffected by the

[91] One bilaterally symmetric IC contributed twice to the same cluster (cluster MCe, Participant 2, IC 2). Its features only contributed once to cluster summary statistics and analyses.

IC scaling/polarity ambiguity and are easier to compare with conventional ERP studies. Together, these figures provide a sense of how each IC cluster was affected by the main contrasts of interest (target vs. standard, congruous vs. incongruous sentence endings, grammatical vs. ungrammatical sentence endings). The effects of the minor contrasts of interest (e.g., oddball task sequence effects, cloze probability) on cluster ERAs are presented in Figures VI.114-124.

To simultaneously visualize the average response of each cluster to the main classes of stimuli, grand average envtopo plots are presented in Figures VI.99-113. These figures show which features of the grand average ERPs each cluster contributed to and the relative timing and scalp-projected magnitude of average cluster activations.

The variance of the grand average ERP effects of interest and of the LPC accounted for by each cluster is presented in Tables VI.1-4. These measures were computed across all electrodes. In addition, the pronoun anterior negativity (AN) effect of grammaticality was also measured at just frontal electrodes to better capture that effect.

Generally, the same statistical comparisons performed in Chapter IV were repeated on the IC clusters with a few additions. The sentence acceptability task analysis was supplemented by tests for effects of word concreteness (see Section IV.C.3.d), frequency (see Section IV.C.3.b), and length (see Section IV.C.3.b) on activity elicited by incongruous endings and their congruous controls. Epochs time-locked to pronoun endings were excluded from these analyses because pronoun endings were all frequent, short, and varied little in concreteness. Finally,

grammatical pronoun endings were added to the cloze probability analysis to increase statistical power and to broaden the range of cloze probabilities despite the possible confounding differences between pronouns and nouns/verbs (Section VI.C.3.b). The *p*-values for the statistical comparisons performed on all clusters are presented in Tables VI.7-21. Summary statistics of the features of the stimuli used in the cloze probability, entropy, concreteness, frequency, and length analyses are presented in Table VI.22.

These figures and tables are all referenced in the subsections that follow, which discuss the features and behavior of each cluster in turn.

VI.C.4.e.1 Left and Right Frontal Clusters (LFr & RFr)

The pair of frontal clusters (Figures VI.48-51, VI.69, VI.72) behaved similarly to one another, generally generating negativities in response to rare/anomalous stimuli. Although their contributions to the grand average P3b effects were negligible, they both elicited significantly enhanced negativities to targets in the P3b windows that peaked at about the same time (letter case task) or slightly later (semantic category task) than the P3b (Figures VI.99 & VI.102). Furthermore, the LFr cluster was significantly more negative to targets preceded by standards than to targets preceded by other targets (Figure VI.114). The RFr cluster was significantly more negative in the semantic category task but this appeared to be simply due to latency differences in the P3b window between tasks. Because of this possible confound and because the ERA peaks of these clusters did not clearly differ across tasks, the mean amplitude

analysis was repeated using a window from 325 to 750 ms (based on the mean cluster ERAs). Again, for the right cluster, the main effect of Stimulus was significant (p<2e-9) with follow up t-tests indicating that targets elicited greater negativity than standards (p<8e-8). The right cluster no longer showed a significant effect of Task (p>.40) and the interaction between Task and Stimulus (p>.09) was still not significant. In contrast, the left cluster showed main effects of Stimulus (p<1e-9), Task (p<.02), and a significant interaction (p<.02). Follow up t-tests again showed that target ERAs were more negative than standard ERAs in the letter case (p<1e-6) and semantic category (p<.003) tasks. In the semantic category task, a significant number of IC ERAs were more negative to standards instead (p<.02). Furthermore, target ERAs in the letter case task were more negative than target ERAs in the semantic category task (p<.02). No other contrasts were significant (p>.13).

In the sentence acceptability task, these same clusters contributed a good deal to the N400 effects (Table VI.2; Figure VI.107) and were the greatest contributors to the anterior negativity (AN) elicited by ungrammatical pronouns (Tables VI.3, Figure VI.103). As in the oddball tasks, anomalous stimuli generally elicited negativities (Figures VI.69-70, VI.73-74). However unlike the ERAs to standards, the ERAs to nonanomalous endings were generally positive and these positivities contribute moderately to the LPC. (Table VI.4; Figures VI.106, VI.109, & VI.111) Anomalous word negativities peaked around 500 ms for both incongruous and ungrammatical endings. Although the negativities were most pronounced to ungrammatical pronouns on average, there was no significant interaction between grammaticality and part of

speech. This negativity in the LFr cluster was more sensitive to congruity as it was greater to incongruous verbs than ungrammatical verbs (Figure VI.115). There were no effects of cloze probability, concreteness, entropy, word frequency, or word length in either cluster in the N400 window.

By the P600 window, all the N400 window mean activation effects disappeared. The only significant comparison in this later time range was an effect of verb word length (Figure VI.124).

VI.C.4.e.2 Left Fronto-Temporal Cluster (LFTe)

Though small, this cluster was quite distinct from the others (Figure VI.47: Bottom) and the topographies (Figure VI.53) and ERAs of each IC were remarkably similar. Each IC's ERA showed a negative peak at about 320 ms in both experiments (Figure VI.125). This peak was clearly bigger to stimuli in the sentence acceptability task, which also elicited a later slow negativity. The mean activations of the cluster were notably insensitive to almost all of the factors analyzed and it contributed little to the grand average P3b, N400, and P600 effects. In fact the only significant differences found were an enhanced negativity to low frequency incongruous words (relative to high frequency incongruous words) in the N400 window (Figure VI.111), and an enhanced negativity to ungrammatical verbs (relative to incongruous verbs) in the P600 window (Figure VI.103). The null effects may partially have been an artifact of the small number of ICs in this cluster, which limited the power of the binomial test.

Despite containing only four ICs, this cluster accounted for a moderate amount of the variance of the LPC to nonanomalous endings (Table VI.4).

Because of the similarity of this cluster to the N320 ERP component that has been implicated in grapheme-to-phoneme conversion (Bentin et al., 1999), the mean amplitude of the these ICs between 220 and 430 ms in the oddball tasks was re-analyzed to best measure the amplitude of the negative peak. The main effects of Task, Stimulus and their interaction were not significant (p>.18). Two-tailed t-tests were also performed on mean amplitudes in this window to test for an effect of word length. Each IC's single trials to all sentence final words were split according to median word length. The tests found no significant differences between the longer and shorter words in any of the four ICs. The same procedure repeated on just anomalous or just nonanomalous endings also failed to find any significant effects of length.

VI.C.4.e.3 Left and Right Parietal Clusters (LPa & RPa)

The lateral parietal clusters were generally characterized by slow positive going ERAs and peaks in their single trial power spectra around 10 Hz (Figures VI.54, VI.56)[92]. Neither cluster contributed much to the grand average P3b, N400, or P600 effects (Tables VI.1-3), though they contributed somewhat to the LPC (Table VI.4).

In the oddball tasks, their slow, positive ERAs were significantly enhanced by targets (Figures VI.78 & VI.81) and this was not affected by task or preceding target

[92] Though LPa and RPa do share one bilaterally symmetric IC (P13, IC20), it can clearly not account for the general similarity between the two clusters.

context (Table VI.7). Both clusters were also significantly affected by congruity. On average, this effect took the form of a brief negative wave in the N400 window, though the direction of this effect was not consistent in the LPa cluster (Table VI.8: B). Part of speech also had an effect in this window, as LPa's activation to pronouns generally went more negative than that to verbs (Table VI.10: C).

During the P600 window, congruity continued to have an effect on cluster LPa, which showed an enhanced positivity to congruous endings (Table VI.9). Concreteness and grammaticality also affected LPa's mean activity in this window but the direction of each effect was not consistent across individual ICs (Tables VI.17: B & VI.11: B).

VI.C.4.e.4 Left and Right Occipital Clusters (LOc & ROc)

The lateral occipital clusters (Figures VI.58-61, VI.84-89) were characterized by what appear to be sensory evoked potentials (Figure VI.68) followed by a slow negativity and a peak in the single trial spectra at 10 Hz. They did not contribute much to the grand average P3b, N400, or P600 effects or the grand average LPC (Tables VI.1-4). Rather they appeared to account for a lot of the grand average P1-N170-P2 complex in the oddball tasks (e.g., Figures VI.99-100) and the grand average P1-N170 in the sentence acceptability task (e.g., Figures VI.105-106). Note that the similarity between these clusters was partially due to the fact that they shared six bilaterally symmetric ICs (Figures VI.59 & VI.61).

In the oddball tasks, the early potentials consisted of a P1 (~100 ms), N170 (~150 ms), and P2 (~225 ms). The slow negativity peaked at about 675 ms and 800 ms to standards and targets, respectively (Figures VI.84, VI.87). Only the first two peaks were clearly apparent in the ERAs to sentence final words and the N170 peaked later (~175-200 ms—Figures VI.85-84, VI.88-89). The slow negativities to nonanomalous and ungrammatical endings peaked at about 700 and 800 ms, respectively, and there was no clear peak to the slow negativity following incongruous endings (Figures VI.85, VI.88).

The mean activations of both clusters to targets and standards significantly differed in the P3b windows but the direction of the difference was not consistent. A significant fraction of ICs were more positive to targets while a significant number of other ICs were more negative to targets (Table VI.7). Both clusters were more positive in the letter case task P3b window, but this appears to have been due to the different P3b window latencies rather than any true difference across tasks (Figures VI.58, VI.60). When the ANOVAs were repeated using the same windows for both tasks (either 250-500 ms or 500-1000 ms, which captured the early and late oddball mean ERA differences for these clusters), the effect of task became, at most, a trend ($p>.09$) while the main effect of Stimulus remained significant ($p<.0004$) and the interaction remained insignificant ($p>.09$). A significant fraction of the ROc cluster was more negative to targets preceded by other targets than targets preceded by standards, though this was not clearly apparent in the mean ERAs (Figure VI.114).

In the sentence acceptability task, the mean activations of the clusters were not effected by any of the variables analyzed in the N400 window. In the P600 window, LOc displayed an enhanced negativity to congruous endings and ROc showed a trend for this effect (Table VI.9). Both clusters showed a trend for an enhanced negativity to grammatical endings (Table VI.11). LOc's response to ungrammatical endings was modulated by the entropy of the preceding sentence stem, but the direction of the effect was not consistent (Table VI.15: C).

Finally, it should be noted that the marked difference in the late negativity between targets and standards was largely due to a single participant's ICs (Participant 4's IC 4 & IC 8). Without these ICs, there remained a trend for the late negativity to be greater to standards, but it was less than half the size.

VI.C.4.e.5 Midline Central Cluster (MCe)

MCe (Figure VI.62-63) was almost identical to the class of best P3b ICs. It contained two extra ICs and missed one IC from that previously analyzed grouping. MCe accounted for the most variance (over 50%) of the grand average P3b, N400, and P600 effects of any cluster (Tables VI.1-3) and accounted for a large degree of the LPC (Table VI.4). MCe's contribution to these ERP phenomena is clearly apparent in the cluster envtopo plots (Figures VI.99-103). Although there was considerable variation in the ERAs of individual ICs (Figures VI.126-127), on average, these ICs elicited a positive peak at about 200 ms that was followed by a second broader positivity, which was significantly larger in response to rare targets and anomalous

sentence endings. The later positive peak eventually died off and went negative by the end of the epoch. The ERAs to sentence final words had two additional features, an early negative peak at 150 ms and a second negativity that peaked around 425 ms.

The enhanced positivity to targets was quite significant (Table VI.7). To estimate the latency of this positivity in the oddball tasks, each MCe IC's ERA to targets in the two tasks was low-pass filtered at 7 Hz and the latency of the most positive point in the ERA between 300 and 1000 ms was chosen as the peak latency. The polarity of the ERAs was set to match the polarity of the scalp-projected ERA at its maximal channel. According to these estimates, the late positivity peaked about 120 ms earlier ($t(12)=5.69$, $p<.0002$) in the letter case oddball task (Table VI.23; Figure VI.128).

There was a significant effect of Task on mean IC amplitude, but the direction of the effect was not consistent across ICs (Table VI.7: C). There was no effect of preceding target context on the positivity's amplitude (Table VI.7: D; Figure VI.114).

In the sentence acceptability task, the mean activations to congruous words were significantly more positive than those to incongruous words during the N400 window. This was generally due to an enhanced positivity to congruous words, though for four ICs (P1 IC2, P4 IC2, P8 IC2, P15 IC2--Figure VI.127) this was partially due to a significant negativity (one-tailed *t*-test, $p<.05$) to incongruous words as well. Incongruous verbs were significantly more negative than ungrammatical verbs in this window (Figure VI.115) and pronouns were more positive than verbs (Table VI.10).

There was also an effect of word frequency in this window, but the direction of the effect was not consistent (Table VI.18).

To determine if these effects in the N400 window could have been due to differences in the latency of the late positivity, the peak latencies of the positivity in the ERAs to nonanomalous, incongruous, and ungrammatical nouns, verbs, and pronoun endings were estimated by the same procedure used to estimate the latency of the positivity to oddball task targets. Mean peak latencies are presented in Table VI.23 and Figure VI.128. An ANOVA with two repeated measures (Congruity: congruous & incongruous; Part of Speech: noun & verb) found that the positivity peaked earlier to congruous than incongruous endings ($F(1,12)=20.74$, $p<.0008$). The main effect of Part of Speech was not significant ($F(1,12)=.09$, $p>.76$), nor was the Part of Speech x Congruity interaction ($F(1,12)=2.79$, $p>.12$). A similar repeated measures ANOVA with two repeated measures (Grammaticality: grammatical & ungrammatical; Part of Speech: pronoun & verb) found that the positivity peaked earlier to grammatical endings (main effect of Grammaticality: $F(1,12)=16.7$, $p<.002$). The positivity to ungrammatical pronouns tended to peak earlier than that to ungrammatical verbs, but this difference was not significant (main effect of Part of Speech: $F(1,12)=3.82$, $p>.07$; Part of Speech x Grammaticality interaction: $F(1,12)=3.19$, $p>.09$). Finally, a paired samples *t*-test found no difference between the peak latency to incongruous and ungrammatical verb endings ($t(12)=.17$, $p>.86$).

This pattern of peak latencies cannot fully explain the amplitude results in the N400 window. The positivity did peak much later to incongruous endings, which are

much less positive in the N400 window. However, the positivity to ungrammatical and incongruous verb endings was equally late, even though incongruous verb endings were more negative in the N400 window (Figure VI.115). Furthermore, the effect of part of speech (pronouns vs. verbs) appears to have been much stronger than the effect of grammaticality on mean amplitude in the N400 window. However, the opposite was true for the peak latency of the late positivity. This conclusion was further supported by a comparison of the absolute ERA magnitudes to congruous and incongruous endings. If the contribution of these ICs to the N400 effect was purely due to a later late positivity to incongruous endings, then the absolute magnitude of the ERAs to incongruous endings should have been less that that to congruous endings in the N400 window. However, this was not clearly the case (Figure VI.34: B).

In the P600 window, both types of anomalous endings elicited enhanced positivities relative to their nonanomalous controls. Furthermore, the positivity was greater for ungrammatical verbs than incongruous verbs (Figure VI.115), greater for low frequency than high frequency endings (Figure VI.121), and greater for nonanomalous low cloze probability endings relative to high cloze probability endings (Figure VI.116). Following up on this latter effect, single trial visualization (Figure VI.129) of MCe's activity revealed no clear monotonic relationship with cloze probability and a great deal of single trial variability. However, there did appear to be a tendency for the late positivity to begin later to the most unpredictable endings and for the subsequent negativity to begin increasingly earlier with increasing cloze probability.

VI.C.4.e.6 Midline Parietal Cluster (MPa)

ICs in cluster MPa generally elicited a late positivity in response to stimuli and exhibited a marked peak around 10 Hz in the single trial spectra (Figures VI.64-65, VI.78-83). The late positivity was largely and significantly absent following standards, a difference that accounted for a considerable amount of the grand average P3b effects (Table VI.1; Figures VI.81 & VI.104). A tendency for the targets preceded by standards to elicit a greater response than targets preceded by targets (Figure VI.114) was far from significant (Table VI.7: D).

MPa's late positivity was less affected by factors in the sentence acceptability task, where it was one of the largest contributors to the LPC (Table VI.4; Figures VI.105-106). It contributed moderately to the grand average N400 effects (Table VI.2; Figure VI.107) and was significantly more positive to congruous endings in the N400 window (Table VI.8). This effect may be due to a delay in the positivity (Figure VI.64). There was an effect of grammaticality in this time range, but the direction of the effect was inconsistent across ICs (Table VI.10: B). Despite moderate contributions to the grand average P600 effects (Table VI.3), the only significant effect in the P600 window was an enhanced positivity to ungrammatical verbs relative to incongruous verbs (Table VI.12). There were no significant effects of cloze probability (Table VI.13) despite an apparent trend for the positivity to be delayed (Figure VI.116) to low cloze probability endings. Word concreteness, frequency, and length also had no effect on mean activation.

VI.C.4.e.7 Midline Occipital Cluster (MOc)

MOc ICs were similar to MPa ICs. They were characterized by a strong 10 Hz component in their single trial activations, a late positivity to targets and sentence final words that was absent to standards, and accounted for much of the LPC (Table VI.4; Figures VI.66-67, VI.96-98). MOc also contributed considerably to the grand average P3b effects (Table VI.2; Figures VI.101 & VI.104) and exhibited a trend for an enhanced response to targets preceded by other targets (Table VI.7, Figure VI.114). MOc differs from MPa in that it tended to exhibit a greater degree of stimulus-locked activity in the first 200 ms (Figure VI.68), in particular there was a positive peak at about 150 ms in both tasks.

In the N400 window, the late positivity to congruous words was enhanced relative to incongruous words (Table VI.8; Figure VI.66). This contributed a moderate amount to the grand average N400 effects (Table VI.3; Figure VI.107). The late positivity was also somewhat bigger to nouns and pronouns than to verbs (Table VI.8, Table VI.10) and was affected by word concreteness (Table VI.16: A) and length (Figures VI.123-124). However, the direction of the concreteness effect was not consistent (Table VI.16: B, C; Figures VI.119-120). With the exception of a continued effect of word length, there were no significant effects in the P600 window despite a moderate contribution to the verb P600 effect.

VI.C.4.f Interim Summary III: Clusters of Common ICs

Automated cluster analysis of the common ICs found ten classes of common components, which revealed a general dissociation between frontal components that typically exhibited negativities to rare targets and anomalous sentence endings, central/post-central components characterized by late positivities, and lateral occipital components whose early activity was relatively phase locked to the onset of stimuli. The following sections present an overview of the average behavior of these clusters and discuss in detail the five clusters that contributed the most to the P3b, N400, AN, and P600 effects, and one cluster, left fronto-temporal, that was remarkably distinct and homogenous. However, when considering the results of the cluster analysis, one should keep in mind the reliability and heterogeneity of the clusters.

Cluster Stability and Heterogeneity

The cluster analysis largely replicated the grouping of common best P3b ICs and several of the clusters resemble clusters of ICs from P3b datasets reported previously (Debener, Makeig, Delorme, & Engel, 2005; Makeig et al., 2004) [93]. However, the cluster validation procedure indicated that these specific groupings were

[93] As discussed in more detail below, the midline central (MCe) cluster resembles the "frontal P3" cluster derived by Debener and colleagues from auditory novelty oddball paradigm data and the central midline cluster derived by Makeig and colleagues from participants performing a visual target detection task. The frontal clusters somewhat resemble Makeig et al.'s frontal midline cluster (also discussed in detail below). The left and right parietal clusters are similar to the left and right "mu" clusters derived by Makeig et al., though the mean topography of the left parietal cluster is notably more posterior and lateral than that of the left mu. From that same data set, Makeig and colleagues found bilateral and central "alpha" clusters that resemble the bilateral occipital and midline parietal clusters described here. Debener et al. also report finding these latter five clusters, though they did not report the features of these clusters as they were not central to their analysis of the P3b or novelty P3.

not remarkably stable. Using ten clusters from half the data set, only 70% of the remaining ICs could be correctly classified, which is modest considering that comparable random labeling could correctly classify 30% of the ICs. Furthermore, the minima in classification error used to select the clusters was relatively shallow. Different IC features (e.g., spectra, ERAs) or weightings of features would return somewhat different clusters. This is problematic as there was not a principled reason for choosing the exact set of IC feature weightings used here; they were chosen because they were simple and returned reasonable results. That being said, even large changes in feature weights (e.g., clustering on just ECD locations or scalp intersects) did not result in wildly different clusters (not shown). The left fronto-temporal (LFTe) cluster and the groupings of the medial parietal and occipital ICs appeared to be the most unstable, but the others appeared rather robust.

Interpreting the IC clusters is also complicated by the heterogeneity of most clusters. Overall, clusters primarily contained ICs that maximally projected to nearby electrodes. However, there remained notable variation among topographies and ERAs of the ICs in a cluster, especially in the larger clusters.

As discussed previously (Section VI.C3.d) this variation is not surprising, given the individual variability apparent in the scalp data and was probably not due to excessively noisy data. In addition, previous ICA decompositions of the P3b have also found substantial variability across participants. For example, Makeig and colleagues (2004) decomposed the P3b from a 31 electrode dataset into nine primarily automatically-derived clusters of independent components and on average each cluster

contained ICs from 10 out of the 15 participants. Debener, Makeig, and colleagues (2005) reported two types of hand-derived IC clusters that contributed to the P3b (measured with 64 electrodes) and appeared to replicate across two different experiments. On average each cluster contained at least one IC from 92% of the participants, but there was a great deal of variability of the source models of the ICs in the "Parietal P3" cluster[94].

Nevertheless, the cluster heterogeneity makes the groupings less convincing and precludes a tidy, functional interpretation of most clusters. It is possible that some of the clusters included outliers (e.g., P9 IC1 in Cluster LFr & P2 IC2 in Cluster MCe). These were left in the analyses due to the difficulty of incorporating a way to identify outliers into the cluster validity procedure and because they were few in number and the activation analysis was insulated against them. Finally, one should note the large number of statistical comparisons, the limited range of the lexical features (Table VI.22), and the fact that only the factors of part of speech and congruity were controlled for in the tests for effects of lexical features (Tables VI.13-VI.21). All of these factors increased the probability of false positives and decreased the chances of detecting real differences.

In light of these possible confounds, it is clear that the clustering results should be considered exploratory and need to be replicated. Given the large number of existent P3b, N400, and P600 datasets, the replicability of these results should be inexpensive to assess.

---

[94] Variability in the source models indicates some degree of variability in the topography of the ICs. Debener and colleagues did not report any other measures of the within-cluster variability of IC features.

Overview of Oddball Task Effects

During the oddball tasks, mean IC activity elicited by targets differed from that elicited by standards in every cluster save LFTe. While the difference between targets and standards was widespread, only the midline clusters contributed greatly to the P3b effect with MCe contributing the most. Their contributions appeared to be approximately simultaneous.

Only two clusters were modulated by target sequence, LFr ($p<.0003$) and ROc ($p<.02$). Finally, the few significant effects of task (in RFr, MCe, LOc, and ROc) appeared to be simply due to the different latencies of the window of analysis or were inconsistent across individual ICs. There were no interactions between oddball task and stimulus (targets & standards). Supplementary analysis on the frontal clusters using a window tailored to their mean ERAs did reveal what appeared to be a true effect of task in LFr. LFr IC activity to targets was more negative in the letter case task than the semantic category task and in the latter, some ICs were more positive to targets than standards.

Overview of Sentence Acceptability Task Anomaly Effects

During the sentence acceptability task, the differences between activity elicited by congruous and incongruous words was widespread in the N400 window, apparent in all clusters except for LFTe, ROc, and LOc. MCe was by far the greatest

contributor to the canonical N400 effect, but the frontal and other midline clusters also accounted for moderate amounts of the variance of the N400 effect. Two clusters, LFr and MPa were also significantly affected by grammaticality in the N400 time range and the two frontal clusters accounted for a great deal of the AN grammaticality effect.

There were fewer significant effects in the P600 window. Only clusters LPa and MCe were affected by grammaticality. As with the canonical N400 effect, MCe accounted for much more of the canonical P600 effect than any other cluster. The clusters affected by grammaticality, as well as LOc, were also affected by congruity in this time range. Effects of anomaly did not interact with part of speech in either time range. The LPC that occurs in the P600 window was mostly accounted for by the midline clusters and MCe's contribution generally preceded that of MPa and MOc (Figures VI.105-106, VI.106-109, VI.11-112).

Incongruous and ungrammatical verbs differed somewhat in both time windows. In the N400 window, LFr showed a greater negativity to incongruity and MCe showed a greater positivity to ungrammaticality. In the P600 window, ungrammatical verbs elicited greater activity in MCe, MPa, and LFTe.

Overview of Sentence Acceptability Task: Other Effects

There were a few effects of part of speech in the N400 window. MOc was more positive to nouns and pronouns than to verbs. MCe was more positive and LPa more negative to pronouns than to verbs. Finally, the more subtle effects of cloze probability, entropy, concreteness, word frequency, and word length were rarely

significant and the few significant comparisons were only weakly so (p>.02) and/or were inconsistent across individual cluster members. The one exception was an enhanced positivity for long words in MOc (Figures VI.99-100), which was rather significant in both time windows (p<.0006).

Cluster MCe

The cluster that accounted for most of the P3b, N400, and P600 effects was cluster MCe, which consisted of nearly the same ICs in the previously analyzed grouping of common best P3b ICs (Section VI.C.3). In the oddball tasks, the behavior of MCe was qualitatively identical to that of the common best P3b ICs, although possible effects of cluster habituation were not tested for. Given the overlap between the two groupings, there is surely a trend for habituation, though the significance of the effect might be quite different given the higher alpha level for individual IC analyses. In the sentence acceptability task there were a few significant effects that were not significant in the 12 common best P3b ICs. Specifically, in the N400 window, the response to incongruous endings was significantly negative, the response to incongruous verbs was more negative than the response to ungrammatical verbs, and the response to verbs was generally more negative than the response to pronouns. In the P600 window, the late positivity was now significantly enhanced by incongruity, it was larger to ungrammatical verbs than incongruous verbs, and it was slightly smaller to highly probable than somewhat probable endings.

The fact that these effects are significant in MCe but not in the common best P3b ICs is most likely due to the higher alpha level used when testing individual ICs rather than the slight differences in the ICs in each group. Moreover, it does not change their status as borderline reliable effects. In general, I think the behavior and feature of the MCe ICs suggest the same interpretation as the common best P3b ICs and adds credence to that previous grouping.

Left and Right Frontal Clusters

The left and right frontal (LFr & RFr) clusters were generally characterized by a negativity in response to targets or anomalous words. In both clusters this negativity was enhanced relative to that elicited by standards in the oddball task and it contributed to the frontal negativity elicited by oddballs, though this was a negligible fraction of the grand average P3b effect. The frontal negativity may correspond to the "slow wave" ERP component that typically follows the canonical auditory oddball P3b (Squires, Squires, & Hillyard, 1975), although in this case, the latency of the frontal positivity was approximately simultaneous with that of the P3b.

In contrast to these negativities, the clusters tended to elicit positive potentials to nonanomalous words, a difference that accounted for over 50% of the AN effect (at frontal electrodes) and generally over 20% of the N400 effect. These differences were both significant in LFr, but only the effect of congruity was significant in RFr. LFr proved more sensitive to other variables as well. It was more sensitive to congruity

than grammaticality, it was affected by preceding target context, and it showed a late effect of word length (but only for verbs).

These frontal clusters find some corroboration in previous work on the P3b. The clusters are similar to the frontal midline cluster reported by Makeig and colleagues (2004) that elicited a 1-3 microvolt negativity to simple visual targets. Nothing like these clusters were reported by Debener et al.'s (2005) ICA decomposition of the auditory P3b; this may reflect the auditory nature of the paradigm or differences in clustering (their clusters were formed by hand). A careful attempt to localize a visuo-verbal P3b (Moores et al., 2003) using a distributed source model derived from individual MRIs generally found no frontal negativities, though a frontal negative slow wave contribution to their data may have been strongly diminished by their application of a high pass spatial filter (a spherical spline Laplacian) given that this potential appears to be diffusely generated by frontal cortex (Baudena, Halgren, Heit, & Clarke, 1995). An fMRI guided attempt to localize a grand average visual P3b (Bledowski et al., 2004)localized the frontal negative component of the P3b to the left and right insula, but unlike LFr and RFr, the insula source activity was clearly biphasic. Event-related fMRI studies of oddball paradigms frequently find greater frontal and pre-frontal activity in response to targets especially in the middle and inferior frontal gyri (Eichele et al., 2005; Horovitz et al., 2002; A. A. Stevens, Skudlarski, Gatenby, & Gore, 2000; M. C. Stevens, Calhoun, & Kiehl, 2005). Furthermore there is evidence that this activity positively correlates with P3b

amplitude (Eichele et al., 2005) and target-to-target interval (Horovitz et al., 2002; M. C. Stevens et al., 2005).

The left frontal cluster corresponds a bit better to attempts to identify the source of the N400 and event-related fMRI (erfMRI) studies of semantic and grammatical anomalies. In an attempt to localize the N400m, magnetic analog of the canonical N400, Halgren and colleagues (2002) found a widespread left frontal contribution that began around 300 ms. Previous work using the same localization method found a left ventro-lateral prefrontal contribution to the effect of word repetition on the N400m (Dale et al., 2000; Dhond, Buckner, Dale, Marinkovic, & Halgren, 2001). Effects of congruity and grammaticality on relatively superior left frontal lobe activity is common in the erfMRI literature (Kiehl, Laurens, & Liddle, 2002; Kuperberg et al., 2003; Newman et al., 2001). Furthermore, some of these studies show enhanced activity for semantic relative to syntactic violations in some left frontal regions (Kuperberg et al., 2003; Newman et al., 2001). Without detailed source localization of the LFr ICs it is impossible to evaluate how well they correspond to any of this previous research. Moreover, it appears that the effects of anomaly on these clusters was not due to differences in the magnitude of activity so much as the polarity of activity; such a difference might not be apparent in fMRI (Section VI.C.3.d). Nevertheless, the potential correspondence is worth noting.

Given the existence of several enhanced frontal negativities to words in sentences that should be particularly taxing to working memory (Kleunder & Kutas, 1993; Martin-Loeches, Munoz, Casado, Melcon, & Fernandez-Frias, 2005; Vos,

Gunter, Kolk, & Mulder, 2001) and the effect of participant working memory span on these effects (Vos et al., 2001), it is likely that these frontal clusters reflect an engagement of working memory. Some researchers have suggested a left frontal "semantic working memory system" used for the search of semantic memory or to hold semantic information "in mind" (Gabrieli, Poldrack, & Desmond, 1998; Kuperberg et al., 2003). This interpretation would be consistent with the greater sensitivity of the frontal clusters to semantic anomalies, but it is inconsistent with the fact that cluster activity was, if anything, greater in the letter case oddball task than in the semantic categorization oddball task. This inconsistent behavior may reflect the functional heterogeneity of the cluster.

An alternative interpretation of the frontal clusters is that they reflect an orienting response to rare targets and anomalous stimuli. This interpretation would be consistent with a large body of evidence of a frontal lobe (esp. dorsolateral prefrontal cortex) contribution to ERPs elicited by deviant stimuli (for review see Friedman, Cycowicz, & Gaeta, 2001). This might explain why targets preceded by other targets elicited a smaller response. Moreover, assuming that a deviance response is proportional to the speed of deviance detection, this could explain why targets in the presumably easier letter case oddball task elicited a larger negativity in the LFr cluster. However, it would not explain the greater negativity to semantically incongruous verbs than ungrammatical verbs, as ungrammatical verbs should be easier to detect and it would suggest that nonanomalous final words do not elicit an orienting response even though they require a behavioral response.

Finally, another possible explanation is that these frontal ICs reflect "imagistic encoding" that is used for comprehending concrete words and pictures. This function has been used to explain the fact that concrete words, in the absence of supportive context (e.g., a congruous sentence stem), elicit larger N400s than abstract words, and this difference is largest at anterior electrodes (Holcomb, Kounios, Anderson, & West, 1999; Kounios & Holcomb, 1994; West & Holcomb, 2000). Although the frontal clusters showed no effects of concreteness in the concreteness analysis, this could well have been caused by the limited range of word concreteness; the difference between the mean concreteness ratings of the median splits used in this study was less than half that used in the previously cited studies that showed concreteness effects. "Imagistic processing" might explain why the letter case task elicited as much or more activity in these clusters as the semantic category task. However, it does not explain the effects of rarity/anomaly or target sequence nor the differential activation to incongruous and ungrammatical verbs in LFr (Table IV.1: A). A more general working memory or attention function is a more parsimonious fit to the data.

Midline Parietal and Occipital Clusters

The topographies of these ICs, their enhanced response to targets, and the fact that they accounted for a sizeable portion of the P3b effect suggests a correspondence between these clusters and the canonical P3b. However, their average responses to targets were not modulated by preceding stimulus context and the magnitude of the ERAs to targets were not clearly affected by the difficulty of the oddball task.

Moreover, the behavior of these clusters in the sentence acceptability task seems inconsistent with the P3b. In general, the amplitude of average late positivity elicited by sentence final words is little affected by anomaly despite the low probability of such stimuli in naturally occurring language. In addition, the latency of the late positivity was little affected by anomaly or part of speech, especially in comparison to the late positivity in the ERAs of Cluster MCe ICs and participant reaction time trends (Section IV.F.1.b). However, these reaction time patterns may be misleading since participant responses were delayed. According to previous research, incongruous endings could take anywhere from 50 to 300 ms longer to evaluate than congruous endings (Baumgaertner, Weiller, & Buchel, 2002; Cardillo, Aydelott, Matthews, & Devlin, 2004; Kiehl et al., 2002; Kuperberg et al., 2003) and about 100-200 ms longer to evaluate than ungrammatical endings (Hagoort, 2003; Kuperberg et al., 2003). The latency of the LP of these clusters could be consistent with reaction time differences on the lower end of these differences. Clearly, speeded response versions of these experiments should be done to get a better sense as to the extent to which the latencies of cluster activity correlated with the time required to evaluate these stimuli.

The mean topographies, ERA, and spectra of the midline parietal cluster (MPa) are very similar to the "central alpha cluster" of ICs found in previous ICA decompositions of P3b paradigm data (Debener et al., 2005; Makeig et al., 2004). Those studies did not report anything like the midline occipital cluster (MOc). This

was probably due to the limited electrode coverage in the earlier study (i.e., 29 scalp electrodes) and possibly due to differences in clustering criteria in the later[95].

The topographies of MPa and MOc ICs generally suggest symmetric, near-midline parietal and occipital generators. However, it is difficult to connect them to the source localization and erfMRI literature because it was not clear what variables they were sensitive to. The fact that MOc was modulated by word length is interesting, but it could be spurious due to the large number of comparisons. This seems likely given that MOc ERAs to pronouns were more positive than those to verbs, despite the longer length of verbs. Future work looking at the effects of word length in the oddball task and in the response to sentence internal words should clarify this.

For the moment, it is tempting to believe that these ICs reflect some type of working memory update (Donchin & Coles, 1988) or response related processing (Nieuwenhuis et al., 2005) since their response to standards is so small relative to that elicited by other types of stimuli and standards are the one type of stimuli in these comparisons that required no response. The fact that the late positivity did not closely follow inferred speed of processing in the sentence acceptability task may have simply been caused by a poor signal to noise ratio due to insufficient participants, the fact that responses were delayed, or inaccurately constructed clusters. However, given that these clusters seemed particularly unstable (the assignment of many ICs to MPa, MOc,

[95] Debener and colleagues did not report the features of their central alpha cluster as it was not central to their analysis of the P3b and novelty P3. Thus it is currently impossible to say just how closely it resembles MPa, MOc, or their combination.

and ROc was highly dependent on the clustering features) it is not worth expending much inferential effort on them until they are shown to be reliable.

Left Fronto-Temporal Clusters

Although this cluster was not significantly affected by rarity or anomaly, the homogeneity of the cluster and the fact that it was the one left lateralized cluster without a right lateralized counterpart in a language comprehension experiment makes it worthy of comment. The topographies and negative peak of the ICs in this cluster resemble that of the N320 ERP component described by Bentin and colleagues (1999) in visuo-verbal oddball tasks. The N320 was greater to pronounceable letter strings relative to non-pronounceable letter strings (e.g., "rtgdfs") and it was diminished or absent in an oddball task that required only a letter size judgment[96]. Based on these factors and its topography and latency, Bentin and colleagues proposed that the N320 reflects grapheme to phoneme translation. This cluster was also similar to the N340 component reported by Mangels et al. (2001) in a visual word memorization task. N340 amplitude appeared to reflect depth of processing, as it was greater to words that were later recognized than to forgotten words and it was decreased by divided attention.

The lack of an effect of oddball task on the LFTe ICs is inconsistent with the behavior of the N320, though it is possible that the four participants who contributed

[96] Bentin et al. do not report the grand average ERPs at left fronto-temporal electrodes in the size judgment task, so it is difficult to know whether or not a small N320 was present. In the grand average ERP topographies from that task (their Figure 8), there is no sign of an N320.

to this cluster automatically read all words in the oddball tasks, even when they were only required to evaluate the case of the letters. The lack of an effect of word length on LFTe also contradicts a grapheme to phoneme translation function. However, it is highly likely that the range of word lengths used was too small to reveal an effect (the mean lengths of the median splits only differed by 2 characters on average). The small contribution of the cluster to the LPC, whose amplitude sometimes correlates with subsequent stimulus memory (Section II.C.1), weakly suggests that LFTe corresponds to the N340, but the LPC contribution was small and it is not clear that the N340 and N320 are distinct components (i.e., decreased N340 amplitude may reflect a lack of grapheme to phoneme translation to shallowly processed text).

The effect of word frequency on the LFTe response to incongruous endings and the late effect of grammaticality on the LFTe response to verb endings may simply be spurious results of the large number of statistical comparisons. The spuriousness of the frequency effect is suggested by the fact that there was no difference between verbs and pronouns in the N400 window, despite the much greater frequency of pronouns. Nevertheless, the frequency effect appears somewhat consistent with evidence that MEG word frequency effects in sentence comprehension localize primarily to the left anterior temporal pole and left orbital cortex (Halgren et al., 2002). Moreover, the fact that frequency effects were significant only for unpredictable sentence endings is in line with previous studies of frequency effects on the EEG response to nonanomalous words (Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Van Petten & Kutas, 1991). These studies found that effects of frequency on

mean amplitude in the N400 window only appeared when a word is relatively unpredictable (Dambacher et al., 2006) or presumably relatively unpredictable (Van Petten & Kutas, 1991)[97].

If this class of IC is accurate, the frequency effect for anomalous open-class words is robust, and the null effect of cloze probability is real, it would indicate an intriguing neural distinction between the general (i.e., prior) probability of a word and the probability of a word based on immediate context (i.e., its conditional probability). At this point however, it is impossible to say what function the LFTe ICs reflect, but the similarity of the ICs to one another and their distinctiveness from the other ICs suggests that they were accurate and are amenable to future study. Hopefully future analyses of the response of these ICs to sentence internal words will better reveal which variables this cluster was and was not sensitive to.

VI.D Independent Components Unique to One Experiment

Reliable ICs derived from a single experiment were deemed unique to that experiment if a significantly similar IC was not derived from either the other experiment or the combined data. The reliability of ICs derived from a single experiment was assessed using the reliability test described in the previous chapter.

[97] Van Petten and Kutas found that frequency effects in the N400 window were only significant for nonanomalous open-class words that occurred early in sentences. They did not estimate the probability of these words, but presumably words early in a sentence are less predictable than words late in the sentence.

For the data from the oddball tasks, 50% splits of each participant's data were constructed by separating epochs time-locked to targets and standards for each task, producing four sets of epochs. Epochs were sorted by their order of occurrence and odd and even epochs split to produce two halves of the entire data set. Based on these splits, the following critical values were obtained for the IC similarity test: sim_{topo}={.70, .84} and sim_{act}={1.97, 2.21}. On average 29 (SD=9) ICs per participant were found to be reliable, and 66% of these corresponded to EEG artifacts[98]. The reliable ICs accounted for 84% and 86% of the variance of the grand average letter case and semantic category task P3b effects (respectively). Of the 160 reliable, non-artifact ICs, 73 were unique to the oddball tasks.

For the sentence acceptability task data, epochs time-locked to well-formed sentence endings, incongruous endings, ungrammatical endings and sentence internal words were separated. Sentence endings were separately grouped by part of speech (noun, verb, and pronoun) to produced eight sets of epochs for each participant. Epochs were sorted by their order of occurrence and odd and even epochs split to produce two halves of the entire data set. Based on these splits, the following critical values were obtained for the IC similarity test: sim_{topo}={.69, .84} and sim_{act}={1.97, 2.21}. On average 31 (SD=7) ICs per participant were found to be reliable, and 63% of these corresponded to EEG artifacts. The reliable ICs accounted for between 87%

[98] The same artifact tests used on the combined datasets were used to automatically detect ICs corresponding to muscle noise or mechanical electrode artifacts. In addition to these tests, eight and eleven ICs from Experiment 1 and 2 (respectively) were hand labeled as muscle noise. All other types of IC artifacts were labeled by hand.

and 88% of the variance of the grand average N400 and P600 effects. Of the 182 reliable, non-artifact ICs, 89 were unique to the sentence acceptability task.

Interpreting the ICs unique to either the oddball tasks or the sentence acceptability task was complicated by their limited number and heterogeneity. Of the reliable, non-artifact ICs unique to one experiment, on average only 4.4 (SD=2.2) ICs per participant could be convincingly modeled with a single equivalent current dipole or a bilaterally symmetric pair of ECDs using the same algorithm and criteria described in Section VI.C.4.a. To some extent, the limited number of reliable and dipolar ICs unique to each experiment surely was caused by the diminished amount of data from which ICA could learn. This is evident from the fact that non-artifact reliable ICs from each experiment alone, accounted for less grand average ERP effect variance than non-artifact reliable ICs from the combined experimental data (Section VI.B.1).

Given the small number of reliable, dipolar ICs unique to one experiment, the clustering procedure applied to the common ICs would only have been able to return a few, very crude clusters. Instead of clustering, the "top" P3b, N400, and P600 effect ICs unique to one experiment are presented in Figures VI.130-137. These ICs give a sense of what might have been unique to each task.

To select the top ICs of each effect, the root mean square (RMS) amplitude of each IC's scalp-projected contribution to the effect (e.g., the letter case target-standard

difference ERP) in the effect window was computed[99]. The oddball task ICs whose contribution to either P3b effect was above the 90th percentile of that of all ICs (uncommon, common, or unreliable) are presented in Figures VI.130-131. The sentence acceptability task ICs whose contribution to the N400 or P600 effect (computed across parts of speech) was above the 86th percentile of that of all ICs are presented in Figures VI.132-137.

Of the unique oddball task ICs, two types of ICs were apparent. There were a handful of P3bish ICs (Figures VI.130-131: F-I) that elicited enhanced positivities to targets primarily at medial and parietal electrodes. In addition, there were a few ICs (Figures VI.130: A, C, D, K) with inferior lateral foci. Two of these had a vertex positive peak (Figures VI.131: C, D) and the others had a vertex negative peak (Figures VI.131: A, K) at about 200 ms. These ICs were reminiscent of Participant 2's common IC 2 (Figures VI.63, VI.126-127), which exhibited a vertex positive peak at about 200 ms.

Of the top N400 effect ICs, there were several posterior lateral ICs (Figures VI.132-134: C, E, F, L) that were generally negative in the N400 window (though sometimes more negative for congruous endings). There were a few midline parietal ICs characterized by positivities that were reduced and/or delayed to incongruous endings (Figures VI.132-134: M, O, T, N). There were three frontal negativity ICs (Figures VI.132-134: B, I, Q) and two ICs (Figures VI.132-134: A, S) that may have corresponded to readiness potentials. Some of these same general categories of ICs

[99] The letter case P3b effect window was 325-575 ms. The semantic category P3b effect window was 425-675 ms. The N400 effect window was 275-550 ms. The P600 effect window was 550-1000 ms.

(and many of the same ICs) were also found in the top P600 ICs. Here there were four frontal negativities (Figures VI.135-137: C, N, O, Q), the same two "readiness-potential-like" ICs (Figures VI.135-137: A,S), and three ICs (Figures VI.135-137: K, L, M, T) that elicited broadly distributed, parietal positivities to ungrammatical endings. In addition, there were two right lateral parietal positivity ICs (Figures VI.135-137: B,G) that were greater to ungrammatical endings.

VI.D.1 Interim Summary IV: Independent Components Unique to one Experiment

Identifying the types of components unique to each task was difficult due to the limited number of task-unique reliable ICs. The lack of reliable ICs was surely partially due to insufficient data as the reliability tests required running ICA on quarters of the entire data set. This is informative given that each quarter of the data contained approximately 30 data points per number of electrodes squared which is comparable to the amount of data used in some previous applications of ICA (Debener et al., 2005; Salustri & Kronberg, 2004). While it is possible that cleaner data would produce more reliable decompositions, this result suggests that more than 30 data points per number of electrodes squared is needed to model "late cognitive" ERP components at the level of accuracy conventionally used for equivalent current dipole models.

Of the ICs unique to the oddball tasks there was a handful of P3b-like ICs that were similar to Makeig and colleagues' "P3b" cluster of ICs (2004) derived from a visual target detection task and somewhat similar to Debener and colleagues' "Parietal P3" cluster of auditory novelty oddball ICs (2005). The ERA of these previously reported clusters are comparable to the Midline Central cluster of common ICs reported here, but their topographies are much broader. These ICs may reflect a greater degree of coordination among the parietal areas of some participants in the oddball tasks. The other possible class of IC unique to the oddball tasks were four ICs with strongly lateralized, inferior foci and peaks in their ERAs at about 200 ms. However, the polarity of the peaks was not consistent across the ICs and these ICs are similar to some members of the lateral occipital clusters of common ICs. Consequently, it is not clear that these four ICs reflect a single process that is distinct from anything in the sentence acceptability task.

In the sentence acceptability task, there were some hints as to the unique N400 and P600 ICs. Specifically there were less than a handful of posterior lateral ICs with marked negativities to incongruous words and/or enhanced positivities to ungrammatical words. In addition, there were two ICs that might reflect readiness potentials. However, most of the ICs unique to the sentence acceptability task were not markedly distinct from the common ICs.

VI.E Summary

ICA of the combined data from the oddball and sentence acceptability tasks found that the P3b, N400, and P600 effects were generally distributed among a handful of ICs, some of which appeared to replicate when ICA was applied to the oddball task or sentence acceptability task separately. These "common ICs" suggested a great deal of overlap between the EEG processes that generated these three effects though the overlap was far from total. Surprisingly, common ICs accounted for more of the N400 effect than the P600 effect. The common ICs also accounted for a great deal of the late positive component (LPC) ERP component, which follows the final words of sentences, though the LPC appears to be more of a subset of the P3b than its homologue.

Understanding in more detail what types of EEG processes might have been shared by these effects was complicated by individual variability and the fact that these three ERP effects were generally distributed across multiple ICs. However, largely based on each individual's P3b effect, a somewhat homogenous class of 12 common "best P3b ICs" from 12 of the 16 participants was defined. This class of ICs accounted for around 50% of the P3b, N400, and P600 effects and of the LPC, and was largely responsible for the correlation between LPC and N400 effect amplitude (Section IV.G.1). The common best P3b ICs behaved like the P3b in that their average activations were typically characterized by a late positivity that was enhanced to targets/anomalous sentence endings and the peak of this positivity was generally later to stimuli that should have taken longer to discriminate. In some ways these 12

ICs did not behave like the P3b (e.g., the late positivity was not smaller to targets preceded by other targets relative to targets preceded by standards, their average response to incongruous sentence endings tended to go negative in the N400 window), though these null effects may have been due to inadequate statistical power. In addition, the topographies of many of the best P3b ICs had a central midline focus, more typical of the P3a than of the P3b, which generally has a more posterior focus and had a more posterior focus in these data. Due to the central topographies, it appears more likely that these ICs correspond more closely to the P3a than the P3b, and index an attentional response.

Automated cluster analysis was applied as an alternative method for identifying classes of common ICs. The cluster analysis returned 10 classes of common ICs, which overall suggested a dissociation between frontal and central/posterior ICs that, on average, elicited scalp negativities and positivities (respectively) in response to deviant stimuli. Although the clustering results were not remarkably stable (i.e., it was not clear exactly how many latent clusters of ICs there were and the results were not remarkably reliable), they largely replicated the best P3b IC grouping and suggested nine other types of common ICs. A few of these nine other clusters also contributed to the main deviance ERP effects of interest. A bilateral pair of frontal clusters contributed to the N400 effect and the anterior negativity elicited by ungrammatical pronouns. In light of previous research on frontal negativities elicited by text, these clusters possibly indexed working memory processes and/or an attentional response. A midline parietal and occipital pair of clusters also contributed

somewhat to the P3b, N400, and P600 effects but were difficult to interpret. Finally, a left fronto-temporal cluster characterized by a negative peak at about 320 ms possibly corresponded to the N320 ERP component implicated in grapheme-to-phoneme conversion (Bentin et al., 1999) or the N340 ERP component (Mangels et al., 2001), which is correlated with declarative memory encoding.

The fact that much of the P3b, N400, and P600 effects was left unexplained by the common ICs suggested that there were indeed some EEG processes unique to one of the two tasks. However, due to the limited number of candidate "unique ICs" it was not possible to characterize what was unique to each task. The small number of unique ICs most likely was due to the limited amount of data provided by each experiment alone and the fact that many ICs corresponded to EEG artifacts.

Table VI.1 – Variance of grand average P3b effects (target-standard ERPs) from the two oddball tasks accounted for by each cluster of ICs. Bottom row is the difference between the 12th and 13th rows. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

Oddball Effects: Percentage variance accounted for

Cluster	P3b Letter Case (325-575 ms)	P3b Semantic Category (425-675 ms)
LFr	-0.06	0.03
RFr	-0.12	-0.11
LFTe	0.01	0.01
LPa	0.03	0.08
RPa	0.07	0.07
LOc	0.08	0.08
ROc	0.07	0.13
MCe	0.65	0.57
MPa	0.29	0.37
MOc	0.25	0.32
All Common ICs	0.74	0.81
All Reliable ICs	0.92	0.95
All Reliable-All Common	0.18	0.14

Table VI.2 – Variance of grand average congruity (incongruous-congruous ending ERPs) effects accounted for by each cluster of ICs. Bottom row is the difference between the 12th and 13th rows. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

Congruity Effects: Percentage variance accounted for

Cluster	Noun (275-550 ms)	Verb (275-550 ms)	Noun (650-1100 ms)	Verb (650-1100 ms)
LFr	0.1	0.3	0.24	0.18
RFr	0.24	0.34	0.03	-0.66
LFTe	0.1	0.08	-0.12	-0.05
LPa	0.09	0.06	-0.15	-0.19
RPa	0.1	0.04	0.02	0.13
LOc	-0.06	0.04	0.03	-0.12
ROc	-0.11	0.04	0.19	-0.07
MCe	0.74	0.51	0.55	0.22
MPa	0.14	0.29	0.15	-0.41
MOc	0.15	0.1	0.26	0.29
All Common ICs	0.84	0.87	0.77	0.19
All Reliable ICs	0.89	0.94	0.94	0.63
All Reliable-All Common	0.05	0.07	0.17	0.44

Table VI.3 – Variance of grand average grammaticality (ungrammatical-grammatical ending ERPs) effects accounted for by each cluster of ICs. Bottom row is the difference between the 12th and 13th rows. Third column values were derived only from frontal electrodes (electrodes anterior to MiFr). Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

Grammaticality Effects: Percentage variance accounted for

Cluster	Pronoun (375-550 ms)	Pronoun (375-550 ms), Frontal Electrodes	Verb (275-550 ms)	Pronoun (500-950 ms)	Verb (650-1100 ms)
LFr	0.05	0.66	0.22	-0.08	-0.2
RFr	0.05	0.54	0.16	-0.13	-0.16
LFTe	-0.12	0.06	0.11	-0.11	0.1
LPa	0.09	-0.01	0	0.11	0.03
RPa	0.05	0.04	0.01	0.15	0.07
LOc	-0.18	-0.25	-0.02	-0.09	-0.05
ROc	0.05	0.01	0.02	0.1	0.07
MCe	-0.13	-1.25	0.45	0.6	0.61
MPa	0.03	0.15	0.29	0.11	0.23
MOc	0.04	-0.1	0.11	0.08	0.27
All Common ICs	0.75	0.79	0.77	0.73	0.69
All Reliable ICs	0.86	0.83	0.90	0.92	0.90
All Reliable-All Common	0.11	0.04	0.13	0.19	0.21

Table VI.4 – Variance of the grand average late positive component (LPC) to sentence final words accounted for by each cluster of ICs. Bottom row is the difference between the 12th and 13th rows. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

Late positive component: Percentage variance accounted for

Cluster	Congruous (450-800 ms)	Incongruous (525-900 ms)	Grammatical (450-800 ms)	Ungrammatical (450-900 ms)
LFr	0.22	0.13	0.23	0.08
RFr	0.21	-0.01	0.22	0.02
LFTe	0.18	0.10	0.19	0.12
LPa	0.08	0.03	0.06	0.07
RPa	0.09	0.12	0.08	0.10
LOc	0.02	0.01	0.04	-0.02
ROc	0.02	0.07	0.05	0.04
MCe	0.49	0.55	0.41	0.59
MPa	0.57	0.51	0.57	0.45
MOc	0.40	0.53	0.42	0.39
All Common ICs	0.90	0.90	0.90	0.86
All Reliable ICs	0.95	0.96	0.95	0.94
All Common-All Reliable	0.05	0.05	0.04	0.09

Table VI.5 – Percentage of the variance accounted of various grand average ERP effects and components by the class of 12 common best P3b ICs. Grand averages are derived from all 16 participants' ERPs. P3b, N400, P600, and AN effect time windows are identical to those used in Chapter IV. Abbreviations: LPC=Late positive component, AN=Anterior Negativity.

Common Best P3b ICs

Grand Average ERP Effect or Component	Percentage Variance Accounted for
Letter Case Task P3b Effect (325-575 ms)	0.62
Semantic Category Task P3b Effect (425-675 ms)	0.58
N400 Effect (Nouns & Verbs: 275-550 ms)	0.57
P600 Effect (Pronouns: 500-950 ms; Verbs: 650-1100 ms)	0.62
LPC (Congruous Endings: 450-800 ms)	0.57
LPC (Incongruous Endings: 525-900 ms)	0.56
LPC (Grammatical Endings: 450-800 ms)	0.53
LPC (Ungrammatical Endings: 450-900 ms)	0.61
AN Effect (Pronouns: 375-550 ms)	0.01
N400-like Effect of Grammaticality (Verbs: 275-550 ms)	0.48
P600-like Effect of Congruity (Nouns & Verbs: 650-1100 ms)	0.53

Table VI.6 – Median (SIQR) one-tailed independent samples *t*-test *p*-values of the three primary deviance effects for the 12 common best P3b ICs and their comparable best P3b electrode (Section VI.C.3.a). Star indicates that the *p*-values for the best P3b ICs are significantly larger than their comparable electrodes (two-tailed sign test, binomial(12,.5)=10, *p*=.04).

Median Deviance Effect Significance

	Common Best P3b ICs	Common Best P3b Electrode
P3b*	1e-5 (9e-4)	4e-11 (3e-8)
N400	.06 (.10)	.003 (.02)
P600	.08 (.18)	.04 (.08)

Table VI.7 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), LC=Letter Case, SC=Semantic Category, ST=Targets preceded by Standards, TT=Targets preceded by Targets.

Oddball Tasks Analysis

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Stimulus	**$1x10^{-9}$**	**$8x10^{-8}$**	1	**.03**	**$2x10^{-5}$**	**$2x10^{-9}$**	**$2x10^{-12}$**	**$3x10^{-13}$**	**$5x10^{-8}$**	**$5x10^{-12}$**
Task	.13	**.01**	1	.51	.33	**.01**	**$4x10^{-7}$**	**.02**	.48	.17
Stimulus x Task	.49	.40	1	.15	1	.08	.28	.49	.49	.54

B) Stimulus *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Targ. > Stand.	.14	.40	NA	**.03**	**$2x10^{-5}$**	**.001**	**$4x10^{-6}$**	**$1x10^{-7}$**	**$1x10^{-6}$**	**$6x10^{-12}$**
Stand. > Targ.	**$1x10^{-6}$**	**$8x10^{-8}$**	NA	.51	1	**.001**	**$4x10^{-4}$**	1	.49	1

C) Task *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
LC > SC	NA	**.01**	NA	NA	NA	**$3x10^{-6}$**	**$5x10^{-11}$**	.14	NA	NA
SC > LC	NA	1	NA	NA	NA	.40	.66	.14	NA	NA

D) Oddball Sequence Effect *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
ST > TT	1	.40	1	1	.06	1	.28	.49	.49	.17
TT > ST	**.0003**	.40	1	.51	1	.09	**.02**	.14	.49	1

Table VI.8 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) Con.=Congruous, Incon.=Incongruous.

Congruity Analysis: N400 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Congruity	**.02**	**.001**	.19	**.03**	**.006**	1	.66	**$1x10^{-6}$**	**.02**	**.006**
Part of Speech	1	1	1	.51	.34	.40	.66	.49	.49	**.04**
Congruity x Part of Speech	1	.40	.19	1	1	.40	.28	1	.49	.54

B) Congruity *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Con. > Incon.	**.02**	**.0001**	NA	.15	**.006**	NA	NA	**$2x10^{-5}$**	**.003**	**.04**
Incon. > Con.	1	1	NA	.51	1	NA	NA	.49	.49	.54

C) Part of Speech *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Noun > Verb	NA	NA	NA	NA	NA	NA	NA	NA	NA	**.006**
Verb > Noun	NA	NA	NA	NA	NA	NA	NA	NA	NA	1

Table VI.9 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) Con.=Congruous, Incon.=Incongruous.

Congruity Analysis: P600 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Congruity	.14	1	.19	**.03**	1	**.01**	.09	**.02**	.49	.53
Part of Speech	1	1	1	.51	1	1	1	.49	1	.17
Congruity x Part of Speech	.49	.40	1	1	.34	.09	.28	1	.49	1

B) Congruity *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Con. > Incon.	NA	NA	NA	**.03**	NA	1	NA	.49	NA	NA
Incon. > Con.	NA	NA	NA	.51	NA	**.01**	NA	**.003**	NA	NA

Table VI.10 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) Gram.=Grammatical, Ungram.=Ungrammatical.

Grammaticality Analysis: N400 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Grammaticality	**.003**	.40	.19	.15	1	.40	.08	.49	**.02**	1
Part of Speech	.14	.09	1	**.004**	.34	.40	.08	**.02**	.49	**.04**
Grammaticality x Part of Speech	.14	.40	.19	.15	.34	1	.28	.14	.49	1

B) Grammaticality *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Gram. > Ungram.	**$2x10^{-5}$**	NA	NA	NA	NA	NA	NA	NA	.14	NA
Ungram. > Gram.	1	NA	NA	NA	NA	NA	NA	NA	.49	NA

C) Part of Speech *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Pronoun > Verb	NA	NA	NA	.51	NA	NA	NA	**.02**	NA	**.006**
Verb > Pronoun	NA	NA	NA	**.03**	NA	NA	NA	1	NA	1

Table VI.11 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) Gram.=Grammatical, Ungram.=Ungrammatical.

Grammaticality Analysis: P600 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Grammaticality	.49	1	1	**.03**	.34	.09	.08	**.0003**	.14	1
Part of Speech	.14	1	1	.15	1	.40	.66	1	.49	.54
Grammaticality x Part of Speech	.49	1	1	.15	1	1	.28	.49	1	.54

B) Grammaticality *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Gram. > Ungram.	NA	NA	NA	.15	NA	NA	NA	1	NA	NA
Ungram. > Gram.	NA	NA	NA	.15	NA	NA	NA	**$2x10^{-5}$**	NA	NA

Table VI.12 – *p*-values from one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: Incon.=Incongruous, Ungram.=Ungrammatical.

Congruity vs. Grammaticality Analysis

A) Anomalous Verb Endings: *t*-Test *p*-Values, N400 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Incon. > Ungram.	.49	1	.19	1	1	1	1	1	.49	1
Ungram. > Incon.	**.02**	.40	1	.51	1	1	1	**.02**	.14	1

B) Anomalous Verb Endings: *t*-Test *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Incon. > Ungram.	.49	1	**.01**	.51	1	.40	1	1	1	.54
Ungram. > Incon.	.14	.40	1	1	.34	.40	.66	**.02**	**.02**	.54

Table VI.13 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) CP=Cloze Probability, HCP=High Cloze Probability, LCP=Low Cloze Probability.

Cloze Probability Analysis

A) Nonanomalous Endings: ANOVA *p*-Values, N400 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Cloze Probability	.49	.40	1	1	1	.40	.66	.14	.49	.54
Part of Speech	.14	.40	1	.51	1	.40	.28	.49	.49	.54
CP x Part of Speech	.49	.40	1	1	1	1	.66	1	.14	1

B) Nonanomalous Endings: ANOVA *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Cloze Probability	.49	.40	1	.51	1	1	.66	**.02**	1	.17
Part of Speech	.14	1	1	1	.34	.40	.66	1	.49	1
CP x Part of Speech	.49	.09	1	.15	1	.40	.28	1	.49	.17

C) Nonanomalous Endings: Cloze Probability *t*-Test *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HCP > LCP	NA	NA	NA	NA	NA	NA	NA	1	NA	NA
LCP > HCP	NA	NA	NA	NA	NA	NA	NA	**.02**	NA	NA

Table VI.14 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Effects of part of speech were not followed up here due to lack of relevance. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

Entropy Analysis: Incongruous Endings

A) ANOVA *p*-Values, N400 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Entropy	.49	.40	1	1	1	1	.66	.49	.49	.17
Part of Speech	1	.40	.19	.51	1	.40	.28	1	**.02**	.17
Entropy x Part of Speech	1	.40	1	1	1	1	1	1	.49	1

B) ANOVA *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Entropy	.49	.40	1	.51	1	.40	.66	.14	1	1
Part of Speech	1	1	1	.51	1	.09	**.02**	1	.14	.54
Entropy x Part of Speech	1	1	1	1	1	1	.66	1	.14	1

Table VI.15 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Effects of part of speech were not followed up here due to lack of relevance. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test) HE=High Entropy, LE=Low Entropy.

Entropy Analysis: Ungrammatical Endings

A) ANOVA *p*-Values, N400 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Entropy	.49	1	.19	1	1	1	1	1	.49	.54
Part of Speech	.49	.09	.19	.15	.34	.40	.08	**.02**	.49	.17
Entropy x Part of Speech	1	.40	1	.51	.34	.40	1	.14	1	1

B) ANOVA *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Entropy	.49	.40	.19	1	1	**.02**	.08	1	.49	1
Part of Speech	.14	.40	1	.51	1	.09	.08	.14	1	.54
Entropy x Part of Speech	1	.40	1	.15	1	.40	.28	.49	1	1

C) Entropy *t*-Test *p*-Values, P600 Window

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HE > LE	NA	NA	NA	NA	NA	.40	NA	NA	NA	NA
LE > HE	NA	NA	NA	NA	NA	.09	NA	NA	NA	NA

Table VI.16 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Conc.=Concreteness, PS=Part of Speech, HC=High Concreteness, LC=Low Concreteness.

Concreteness Analysis:

Congruous & Incongruous Endings, N400 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Concreteness	1	1	1	1	1	1	1	1	.49	.54
Conc. x Part of Speech	1	1	1	1	1	.49	.66	.14	1	**.04**
Conc. x Congruity	1	.09	.19	1	1	.40	.66	.14	1	.54
Conc. x PS x Congruity	1	1	1	.51	1	.40	1	1	1	1

B) Noun Endings: Concreteness *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HC > LC	NA	NA	NA	NA	NA	NA	NA	NA	NA	.17
LC > HC	NA	NA	NA	NA	NA	NA	NA	NA	NA	.17

C) Verb Endings: Concreteness *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HC > LC	NA	NA	NA	NA	NA	NA	NA	NA	NA	1
LC > HC	NA	NA	NA	NA	NA	NA	NA	NA	NA	.54

Table VI.17 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Conc.=Concreteness, PS=Part of Speech, HC=High Concreteness, LC=Low Concreteness.

Concreteness Analysis:

Congruous & Incongruous Endings, P600 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Concreteness	.49	1	1	**.004**	1	.40	1	.14	1	.54
Conc. x Part of Speech	.49	1	1	1	1	1	1	.49	1	.17
Conc. x Congruity	1	1	1	1	.34	.40	1	.49	.49	.54
Conc. x PS x Congruity	1	1	1	.15	1	.40	1	.14	1	1

B) Concreteness *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HC > LC	NA	NA	NA	.15	NA	NA	NA	NA	NA	NA
LC > HC	NA	NA	NA	.15	NA	NA	NA	NA	NA	NA

Table VI.18 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Freq.=Frequency, PS=Part of Speech, HF=High Frequency, LF=Low Frequency

Frequency Analysis:

Congruous & Incongruous Endings, N400 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Frequency	.49	1	1	.51	1	1	1	1	1	.54
Freq. x Part of Speech	.14	1	1	1	1	.49	.66	**.02**	1	.54
Freq. x Congruity	.14	1	**.01**	.51	1	1	1	.14	1	1
Freq. x PS x Congruity	.49	1	1	1	.34	1	1	.14	1	.54

B) Noun Endings: Frequency *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HF > LF	NA	NA	NA	NA	NA	NA	NA	1	NA	NA
LF > HF	NA	NA	NA	NA	NA	NA	NA	.14	NA	NA

C) Verb Endings: Frequency *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HF > LF	NA	NA	NA	NA	NA	NA	NA	.49	NA	NA
LF > HF	NA	NA	NA	NA	NA	NA	NA	1	NA	NA

D) Congruous Endings: Frequency *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HF > LF	NA	NA	1	NA	NA	NA	NA	NA	NA	NA
LF > HF	NA	NA	.19	NA	NA	NA	NA	NA	NA	NA

E) Incongruous Endings: Frequency *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HF > LF	NA	NA	**.01**	NA	NA	NA	NA	NA	NA	NA
LF > HF	NA	NA	1	NA	NA	NA	NA	NA	NA	NA

Table VI.19 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Freq.=Frequency, PS=Part of Speech, HF=High Frequency, LF=Low Frequency

Frequency Analysis:

Congruous & Incongruous Endings, P600 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Frequency	1	1	1	.51	1	1	.66	**.02**	1	1
Freq. x Part of Speech	1	1	1	.51	1	1	1	.49	.14	1
Freq. x Congruity	.14	1	.19	.51	1	1	1	.49	.14	1
Freq. x PS x Congruity	1	1	1	1	1	1	1	.49	1	1

B) Frequency *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
HF > LF	NA	NA	NA	NA	NA	NA	NA	.4867	NA	NA
LF > HF	NA	NA	NA	NA	NA	NA	NA	**.0031**	NA	NA

Table VI.20 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Len.=Length, PS=Part of Speech.

Word Length Analysis:

Congruous & Incongruous Endings, N400 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Length	1	1	1	.51	1	1	.28	1	.14	**.04**
Len. x Part of Speech	.49	1	1	.15	1	.40	.08	1	1	.54
Len. x Congruity	1	.40	1	1	1	.40	.66	.14	.49	1
Len. x PS x Congruity	.49	1	1	1	1	1	.66	1	1	1

B) Length *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Long > Short	NA	NA	NA	NA	NA	NA	NA	NA	NA	**.0006**
Short > Long	NA	NA	NA	NA	NA	NA	NA	NA	NA	1

Table VI.21 – *p*-values from cluster ANOVAs and follow up one-tailed *t*-tests (see Section VI.B.3.c). Significant *p*-values are printed in bold. Other effects of part of speech and congruity are not reported here due to redundancy. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital. Other abbreviations: NA=Not Applicable (i.e., ANOVA did not warrant *t*-test), Len.=Length, PS=Part of Speech.

Word Length Analysis:
Congruous & Incongruous Endings, P600 Window

A) ANOVA *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Length	.49	1	1	.51	1	1	.66	.49	.14	**.04**
Len. x Part of Speech	**.003**	.40	1	.51	1	.40	.66	.49	1	1
Len. x Congruity	.49	1	.19	1	.34	1	1	1	.49	.54
Len. x PS x Congruity	1	1	.19	.51	1	1	.66	1	1	1

B) Length *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Long > Short	NA	NA	NA	NA	NA	NA	NA	NA	NA	**.0006**
Short > Long	NA	NA	NA	NA	NA	NA	NA	NA	NA	1

C) Nouns: Length *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Long > Short	.49	NA	NA	NA	NA	NA	NA	NA	NA	NA
Short > Long	1	NA	NA	NA	NA	NA	NA	NA	NA	NA

D) Verbs: Length *t*-Test *p*-Values

	LFr	RFr	LFTe	LPa	RPa	LOc	ROc	MCe	MPa	MOc
Long > Short	.49	NA	NA	NA	NA	NA	NA	NA	NA	NA
Short > Long	**.003**	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table VI.22 – Central tendencies and measures of dispersion (in parenthesis) of stimulus features used to create median split ERAs. Central tendency and dispersion are measured with mean and standard deviation with the exception of frequency, for which median and semi-interquartile range are used.

Central Tendencies of Median Splits

	Lower 50%	Upper 50%
Cloze Probability (Nonanomalous Endings)	0.34 (0.21)	0.80 (0.14)
Entropy (Incongruous Endings)	0.9 (0.5)	2.4 (0.7)
Entropy (Ungrammatical Endings)	1.2 (0.5)	2.7 (0.6)
Concreteness (Congruous/Incongruous Endings)	3.9 (0.7)	5.5 (0.7)
Frequency (Congruous/Incongruous Endings)	18 (14)	131 (85)
Length (Congruous/Incongruous Endings)	3.9 (0.6)	6.2 (1.3)

Table VI.23 – Mean onset and peak latency of the late positivity in the event-related activations of ICs in the Midline Central cluster. Standard error (*n*=13) are in parentheses.

Midline Central Cluster Late Positivities: Peak Latency

	Latency (in ms)
Letter Case Oddball Task: Targets	453 (22)
Semantic Category Oddball Task: Targets	577 (31)
Nonanomalous Noun Endings	504 (31)
Nonanomalous Verb Endings	529 (35)
Nonanomalous Pronoun Endings	540 (32)
Incongruous Noun Endings	699 (34)
Incongruous Verb Endings	660 (42)
Ungrammatical Pronoun Endings	574 (36)
Ungrammatical Verb Endings	667 (34)

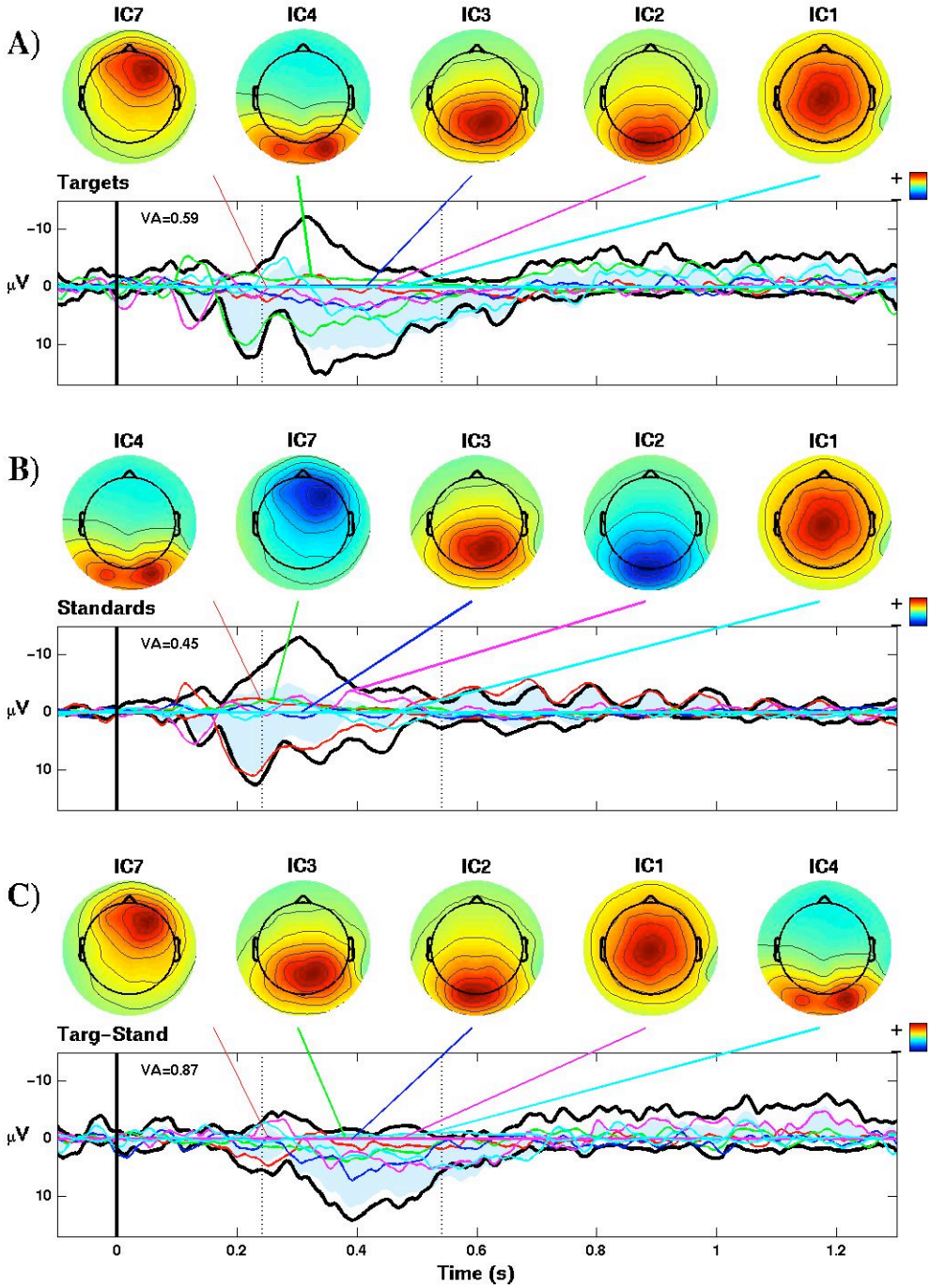

Figure VI.1 – Envtopo plots (see Section VI.B.2) of letter case oddball task ERPs (**A**,**B**) and difference wave (**C**) from Participant 16. Featured ICs are the five that account for the most P3b effect variance of any set of five ICs (reliable or otherwise) in this participant's letter case task P3b window (242-542 ms), marked with vertical dashed lines. The percentage variance accounted for (VA) by the featured ICs in the P3b window is given in the upper left corner of each plot.

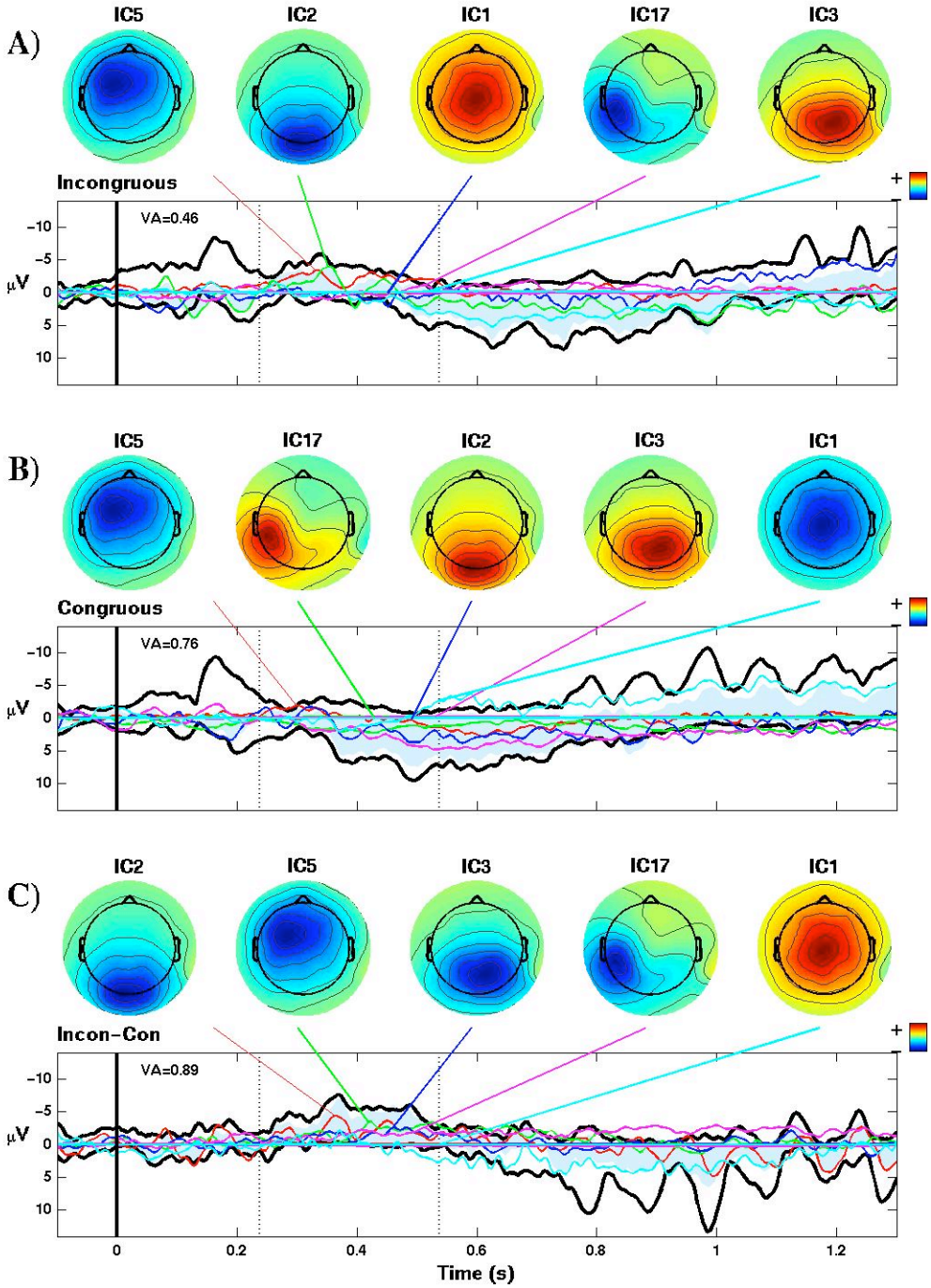

Figure VI.2 – Envtopo plots (see Section VI.B.2) of Participant 16's ERPs to incongruous (**A**) and congruous (**B**) noun and verb sentence endings and their difference wave (**C**). Featured ICs are the five that account for the most N400 effect variance of any set of five ICs (reliable or otherwise) in this participant's N400 window (238-538 ms), marked with vertical dashed lines. The percentage variance accounted for (VA) by the featured ICs in the N400 window is given in the upper left corner of each plot.

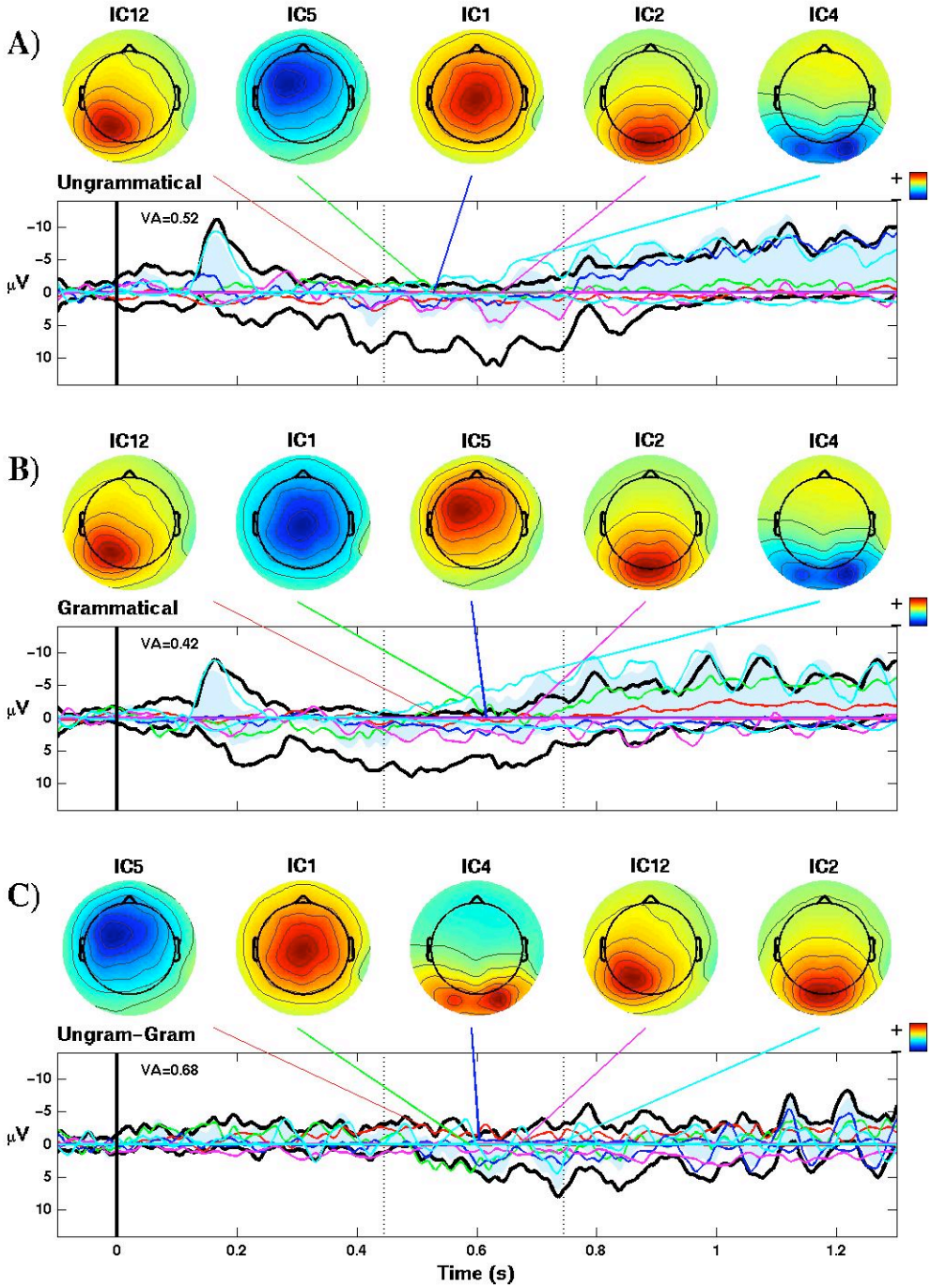

Figure VI.3 – Envtopo plots (see Section VI.B.2) of Participant 16's ERPs to ungrammatical (**A**) and grammatical (**B**) pronoun and verb sentence endings and their difference wave (**C**). Featured ICs are the five that account for the most P600 effect variance of any set of five ICs (reliable or otherwise) in this participant's P600 window (446-746 ms), marked with vertical dashed lines. The percentage variance accounted for (VA) by the featured ICs in the P600 window is given in the upper left corner of each plot.

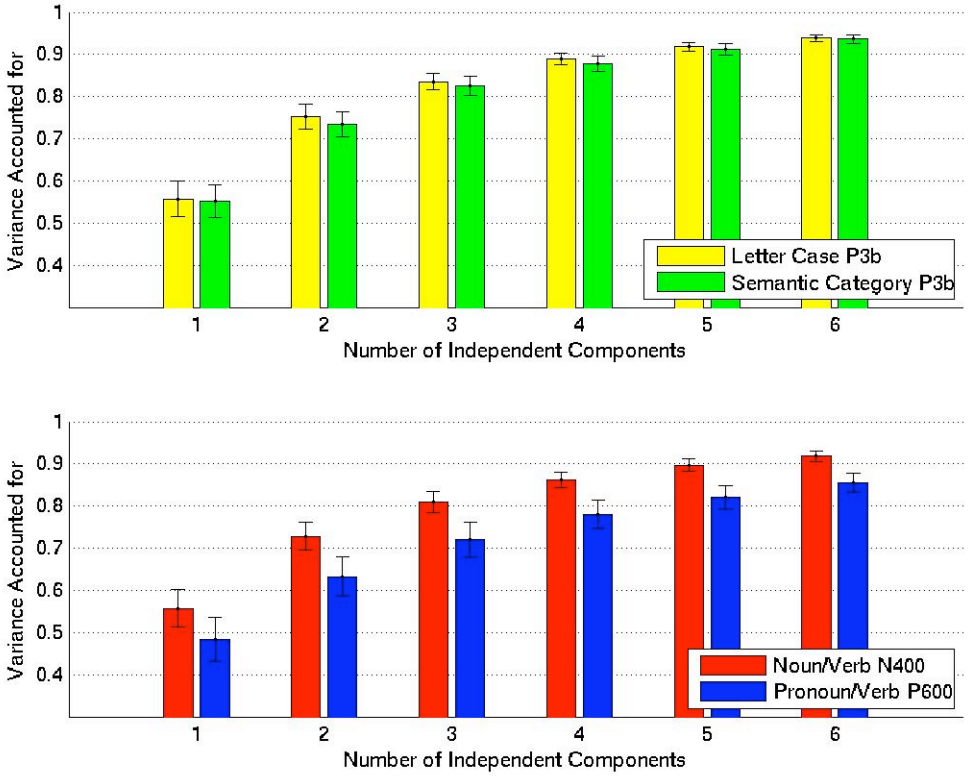

Figure VI.4 – Mean individual participant P3b, N400, and P600 effect percentage variance accounted for by increasing numbers of "best" independent components (ICs). Error bars are standard error. Best ICs account for the most variance of any set of ICs (reliable or otherwise) of that size. To minimize estimation error, the N400 and P600 effects for each part of speech were combined. Individual effects were defined by a 300 ms window around each effect peak in individual difference ERPs (e.g., ERPs to incongruous endings minus ERPs to congruous endings; see Sections IV.E.3.a & IV.F.3.a). Using grand average windows would most likely decrease the maximum amount of effect variance a given number of ICs can account for.

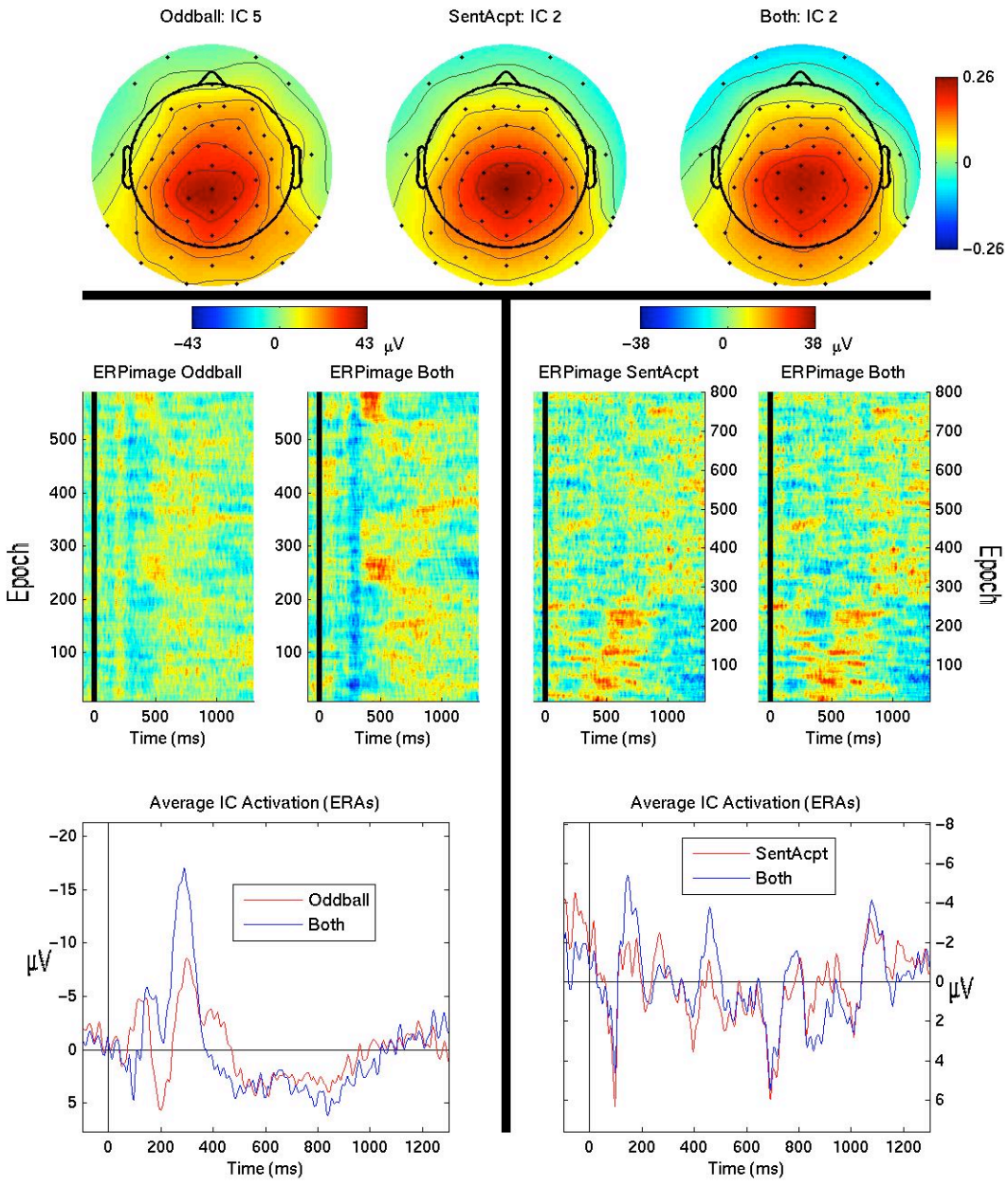

Figure VI.5 – Features of an IC (from Participant 8) derived from both experiments and its homologue derived from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization. IC 5 from the oddball experiment was hand paired with IC 2 from both experiments. This pairing was used in place of the automatic pairings presented in the subsequent two figures.

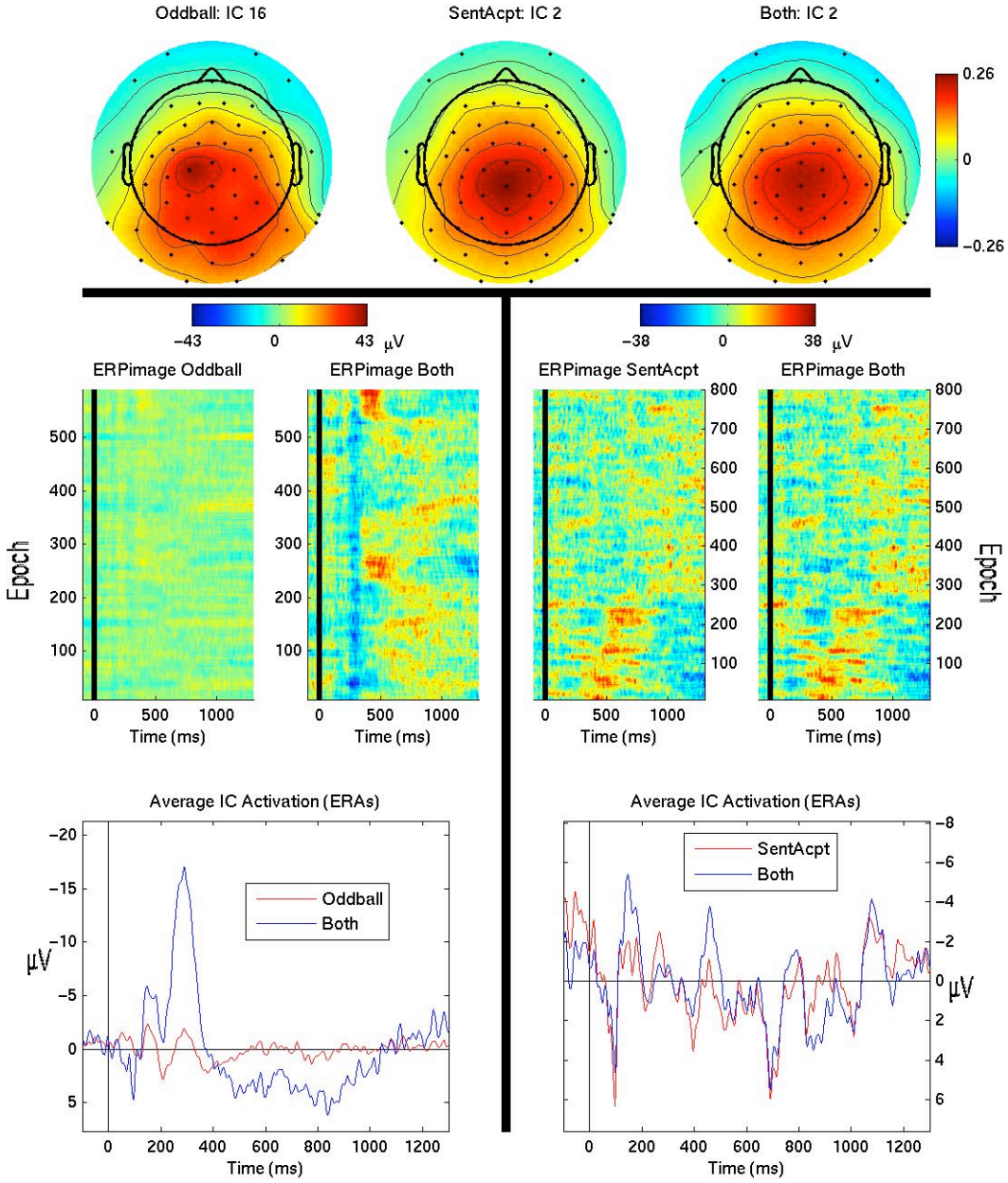

Figure VI.6 – Features of an IC (from Participant 8) derived from both experiments and its automatically selected pair from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization. The parings in Figure VI.5 were used in place of these automatic pairings.

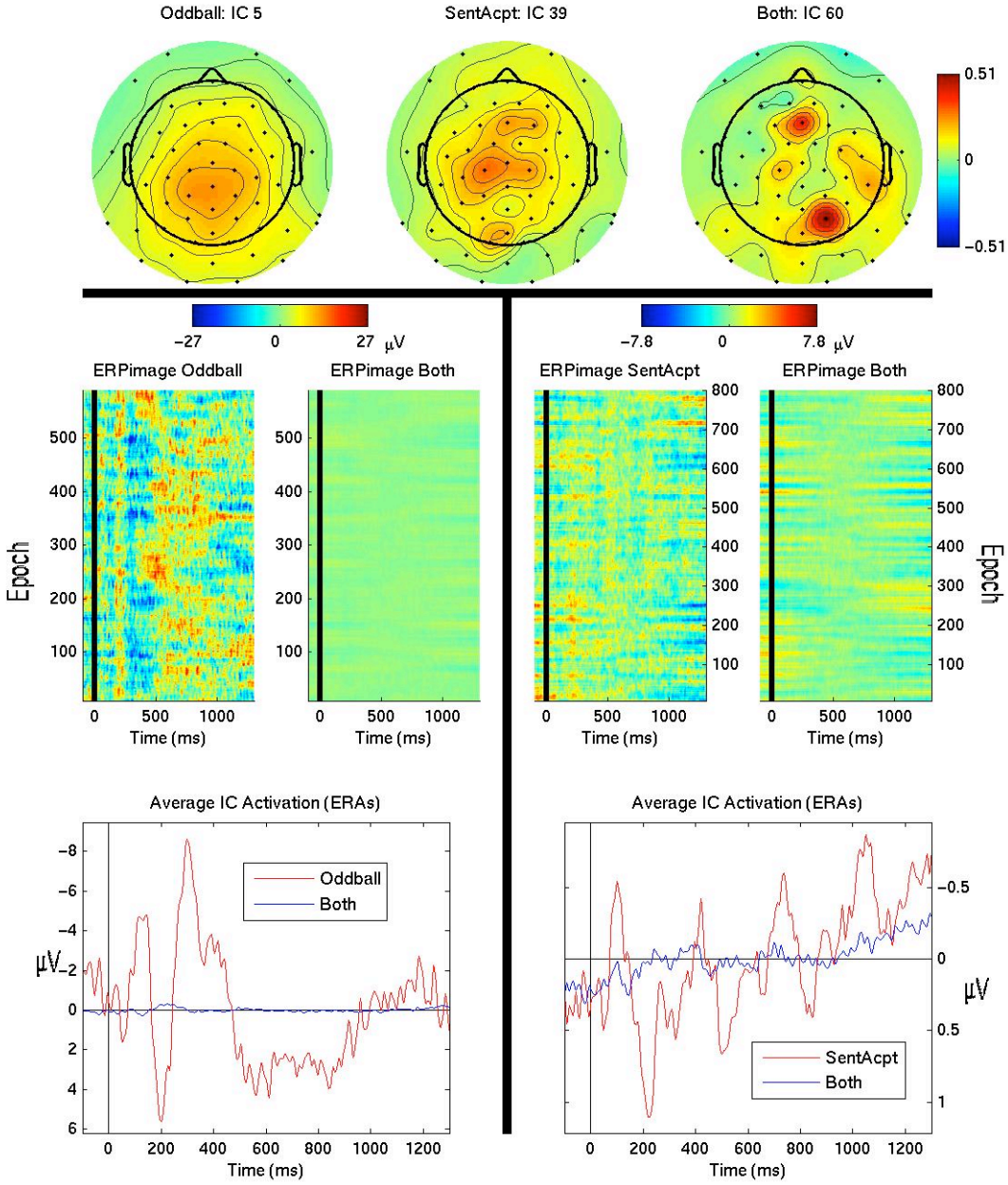

Figure VI.7 – Features of an IC (from Participant 8) derived from both experiments and its automatically selected pair from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization. The parings in Figure VI.5 were used instead of these automatic pairings.

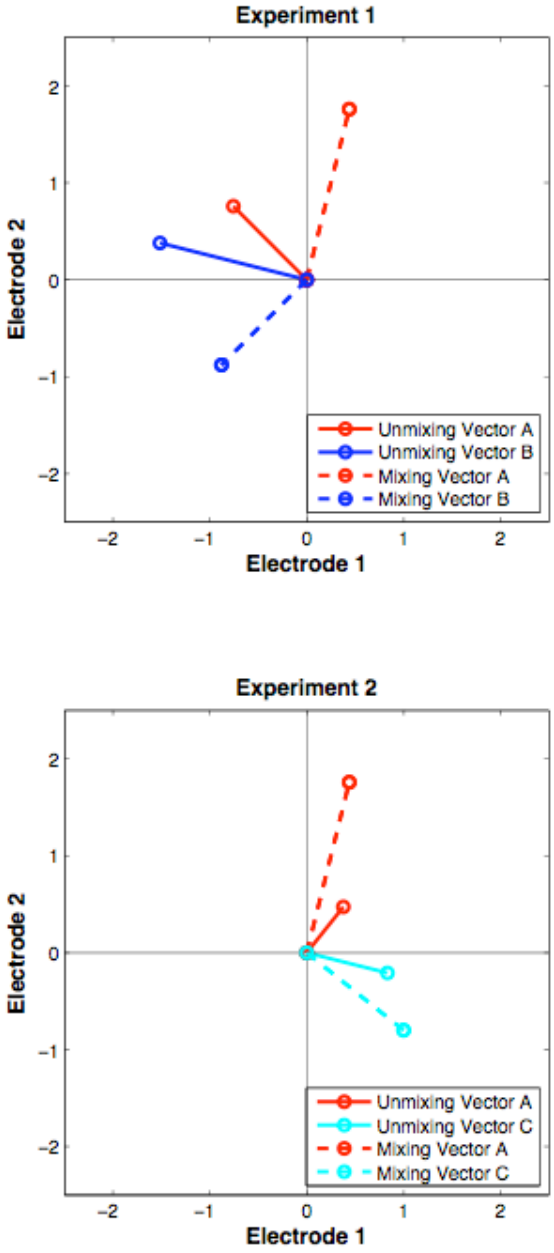

Figure VI.8 – Hypothetical ideal ICA unmixing and mixing vectors of two data sets from two different experiments. Axes represent unitless weights. The data for Experiment 1 is generated by Source A and Source B. The data for Experiment 2 is generated by Source A and Source C. Source A's mixing vector is identical but its unmixing differs across experiments.

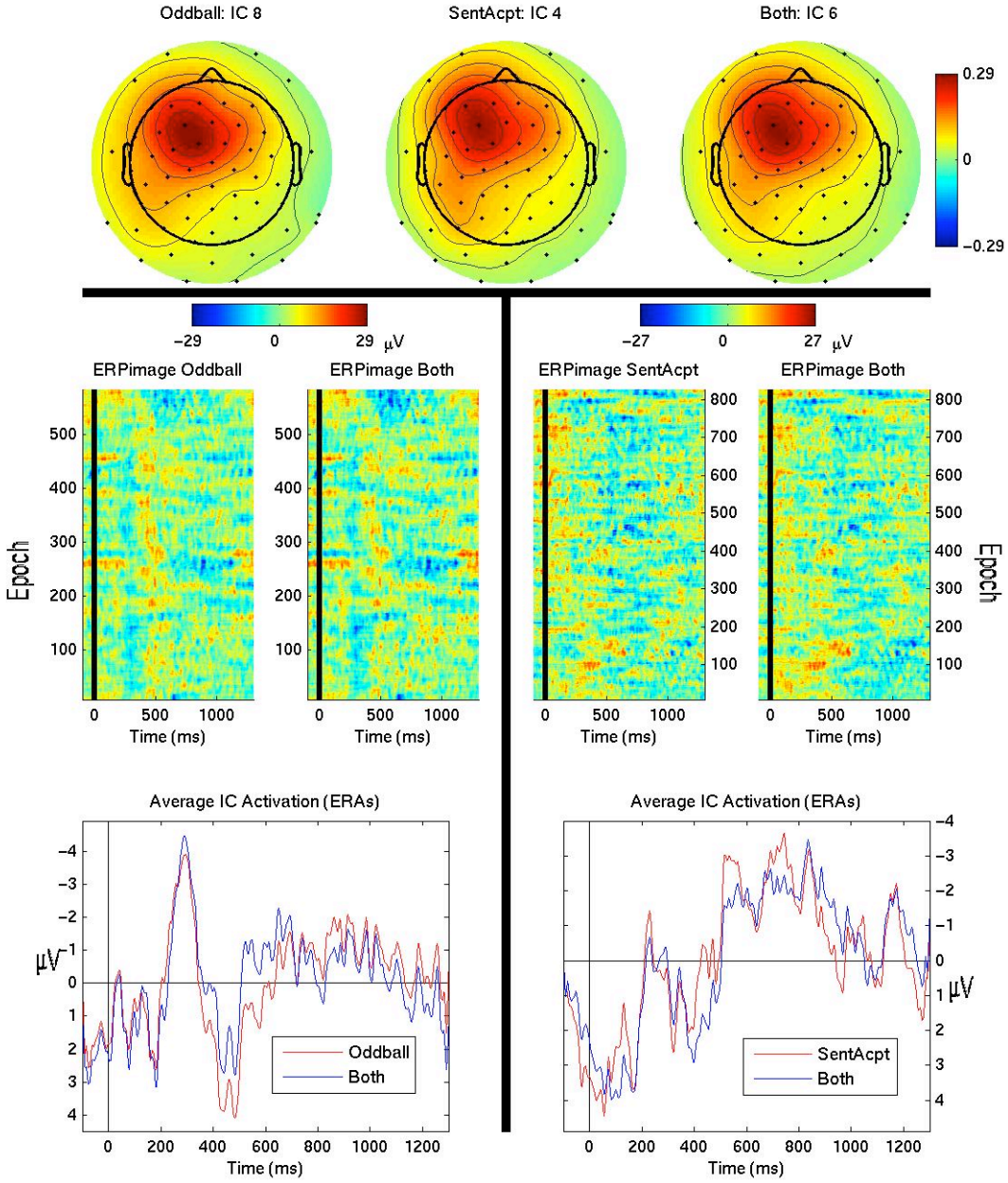

Figure VI.9 – Features of an IC (from Participant 4) derived from both experiments and its homologue derived from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization.

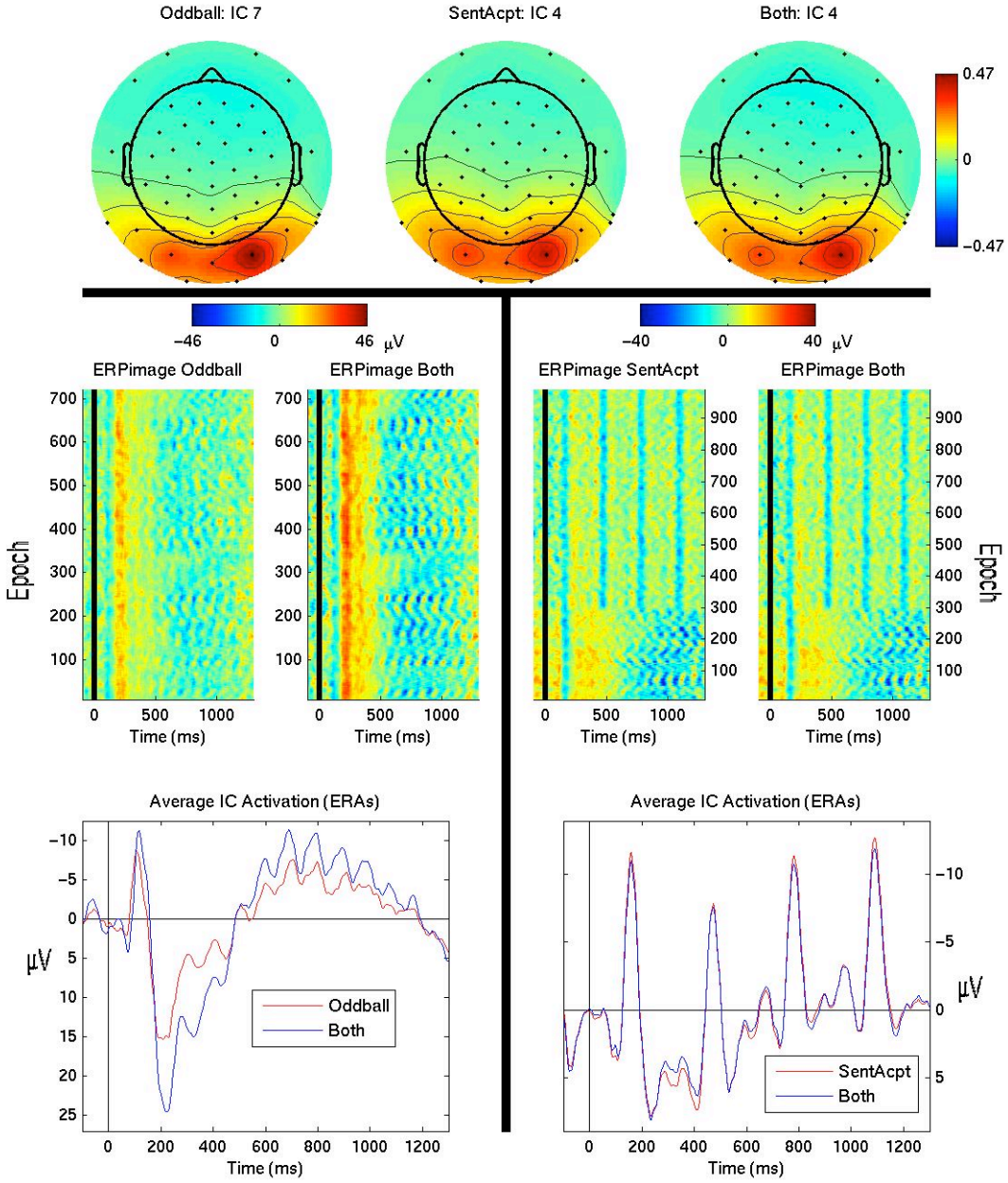

Figure VI.10 – Features of an IC (from Participant 16) derived from both experiments and its homologue derived from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization.

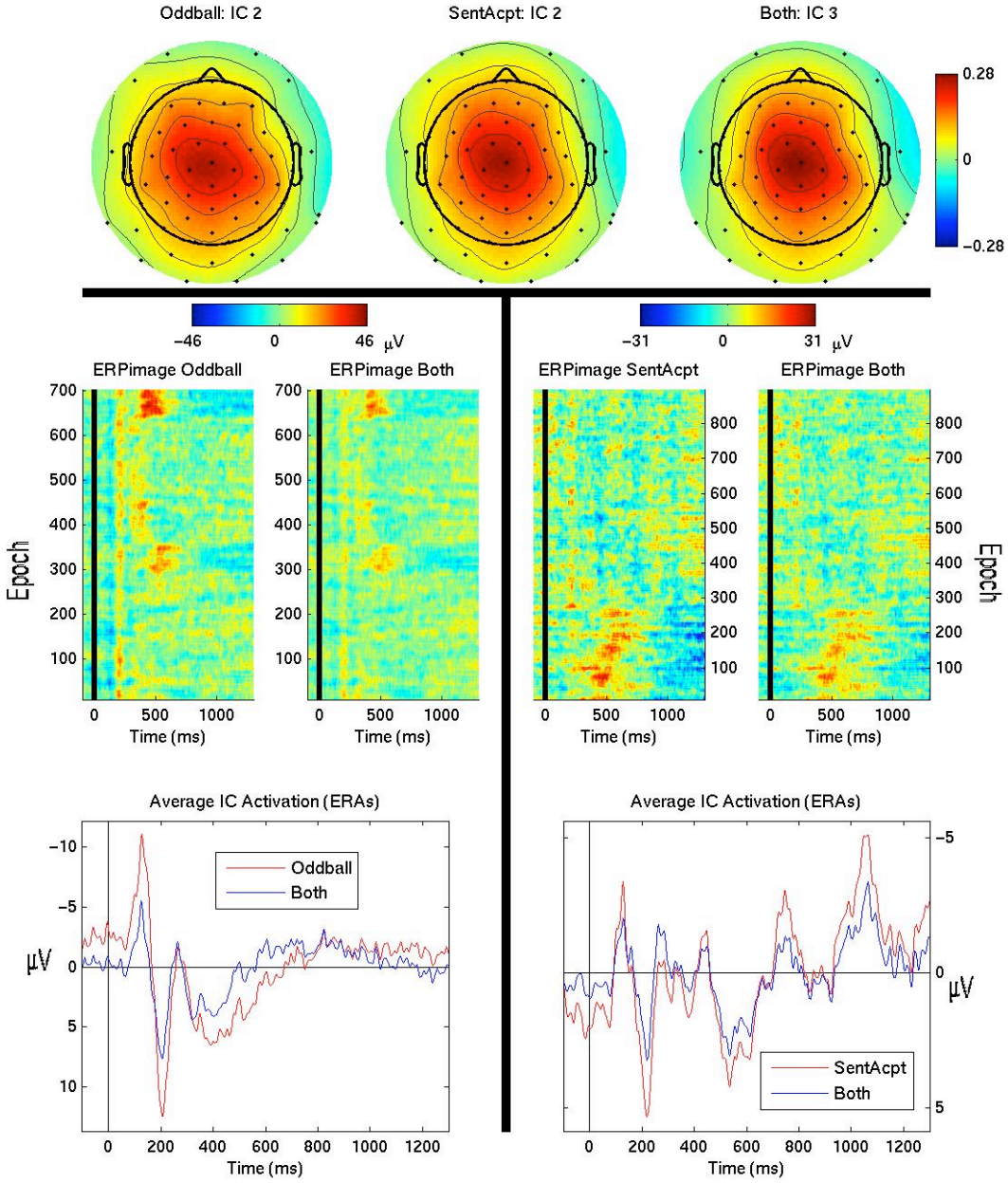

Figure VI.11 – Features of an IC (from Participant 1) derived from both experiments and its homologue derived from only the oddball and sentence acceptability (SentAcpt) experiments. Topographies (**Top**) are normalized to unit length. IC activations (**Bottom Left** & **Right**) are computed after topography normalization.

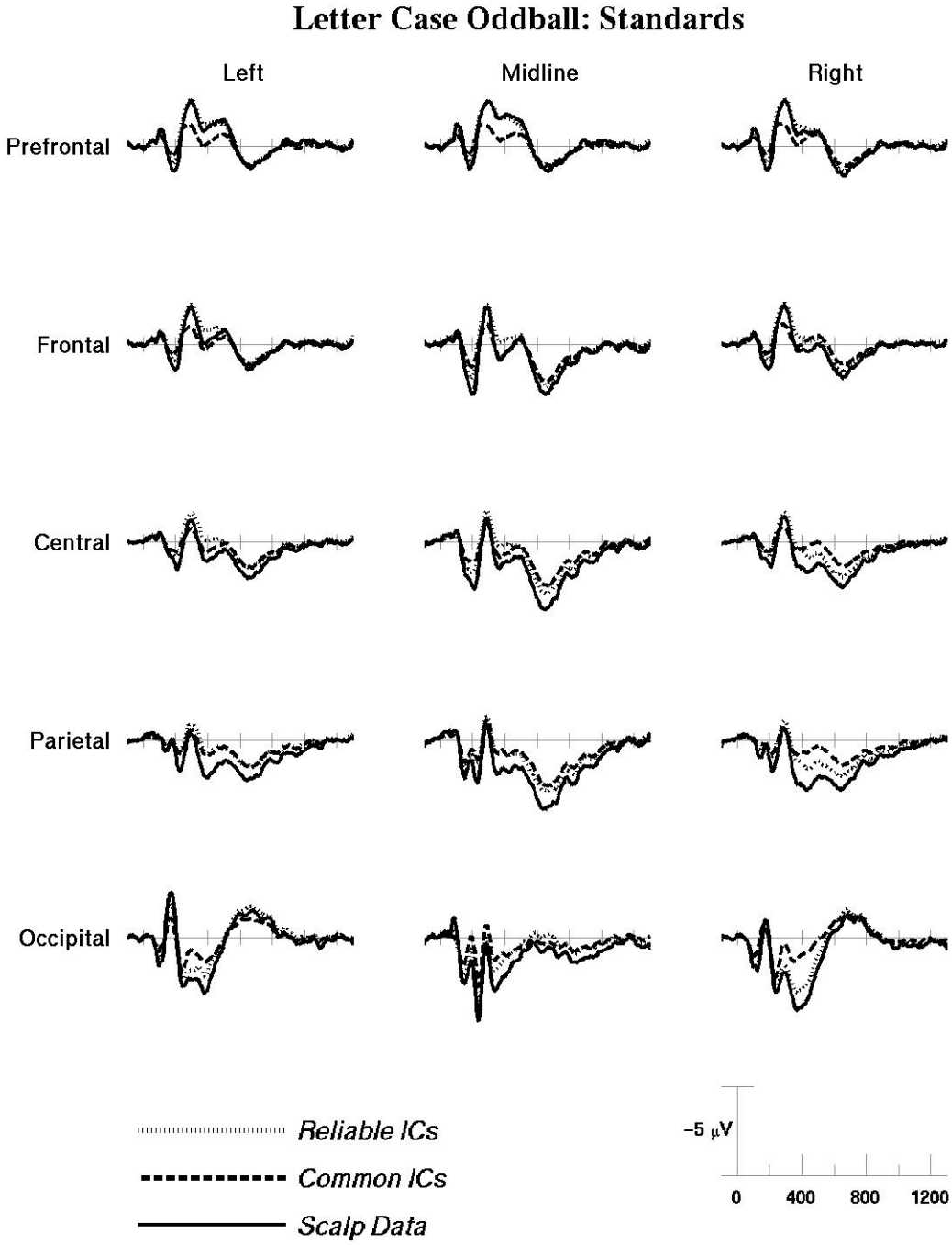

Figure VI.12 – Example grand average ERPs to standards in the letter case oddball task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

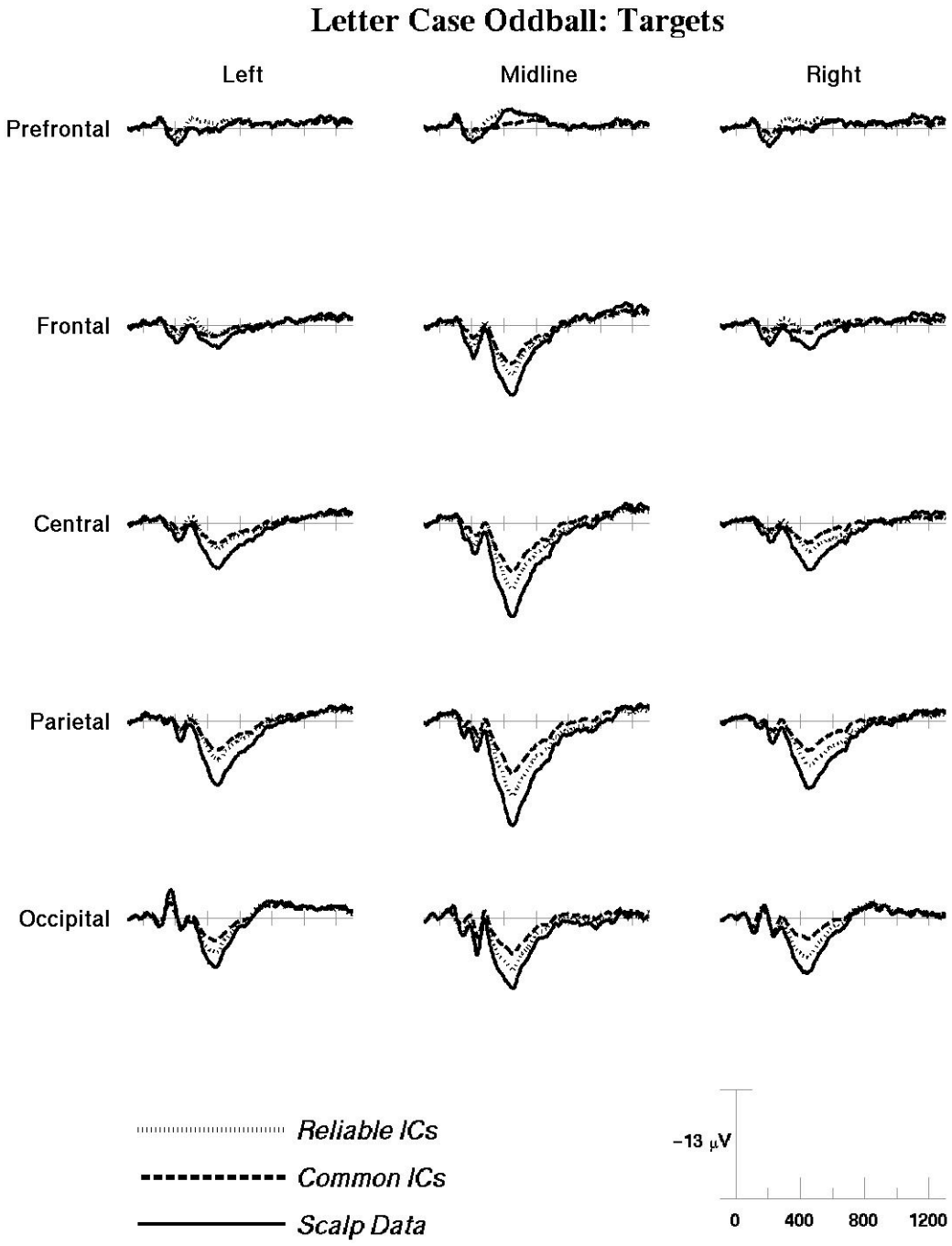

Figure VI.13 – Example grand average ERPs to targets in the letter case oddball task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

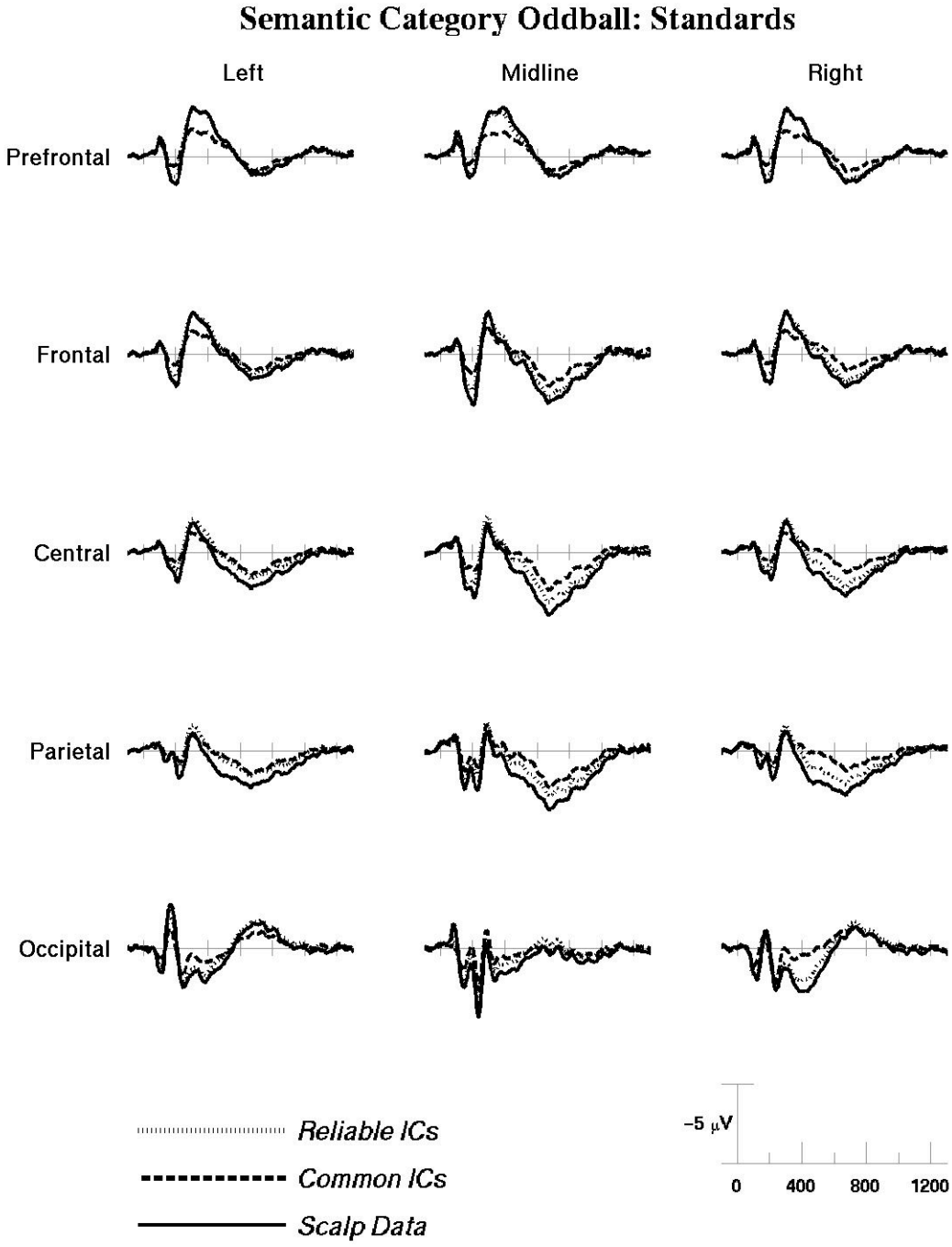

Figure VI.14 – Example grand average ERPs to standards in the semantic category oddball task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

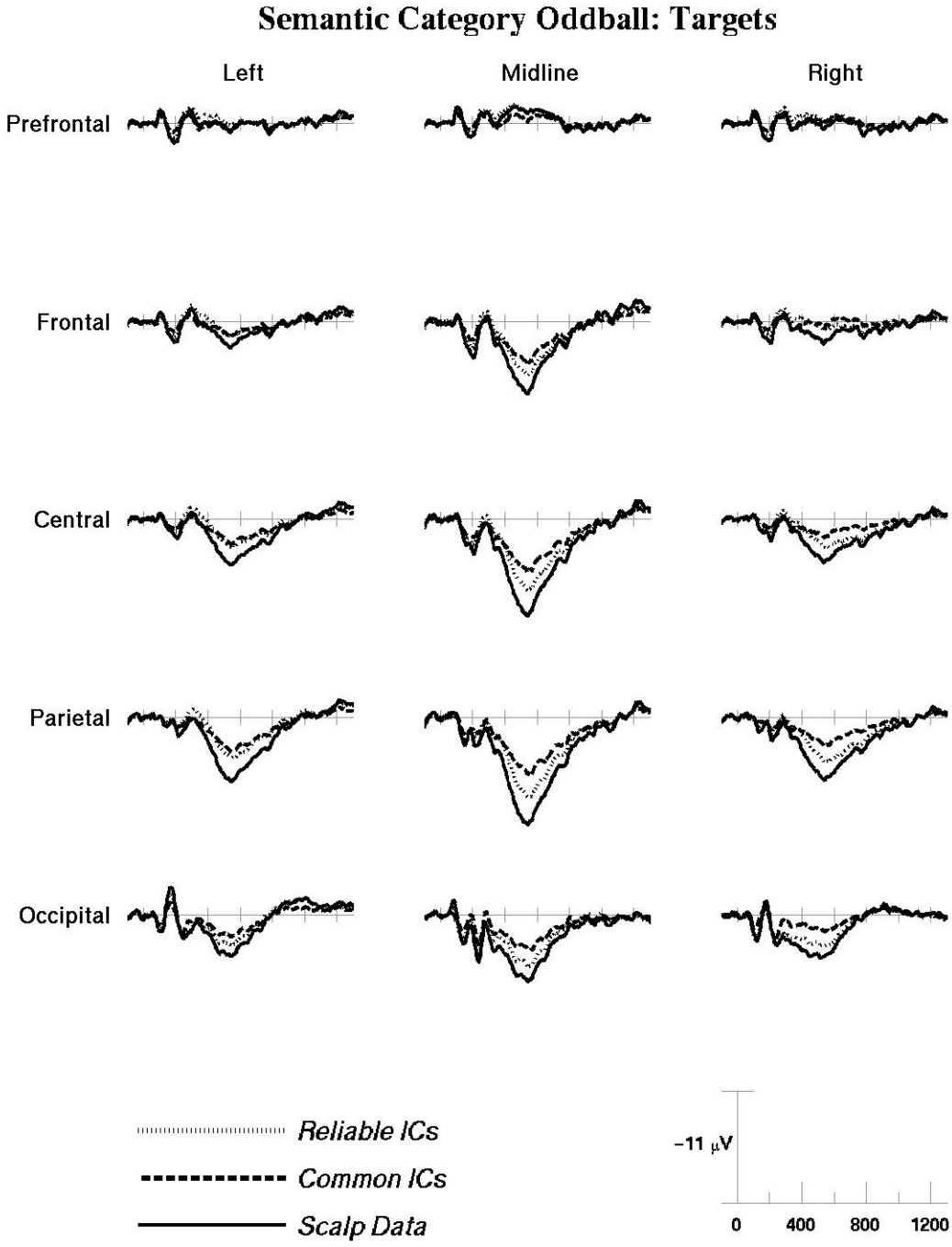

Figure VI.15 – Example grand average ERPs to targets in the semantic category oddball task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

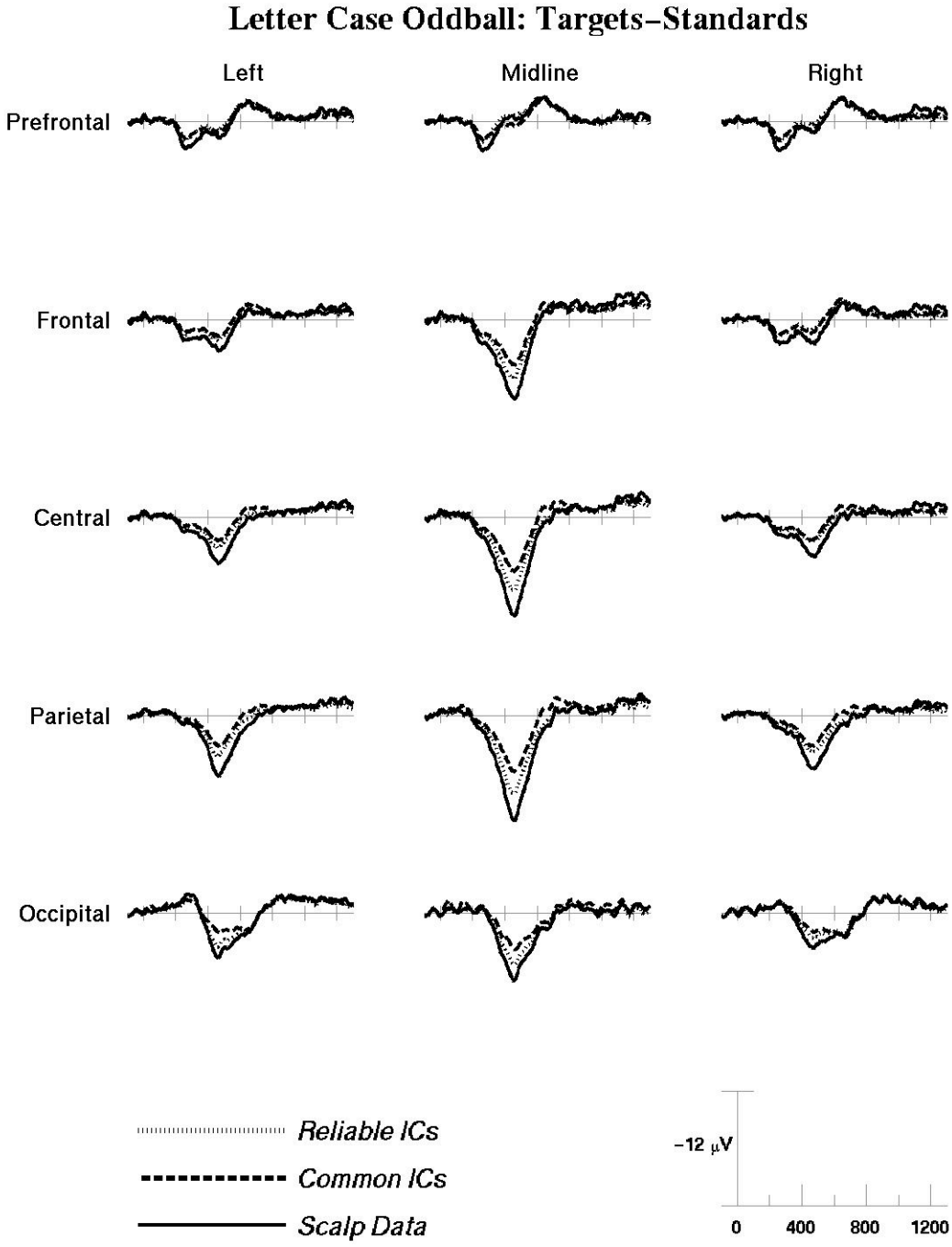

Figure VI.16 – Example grand average difference ERPs illustrating the P3b effect in the letter case oddball task. Also plotted are the contributions to these difference ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

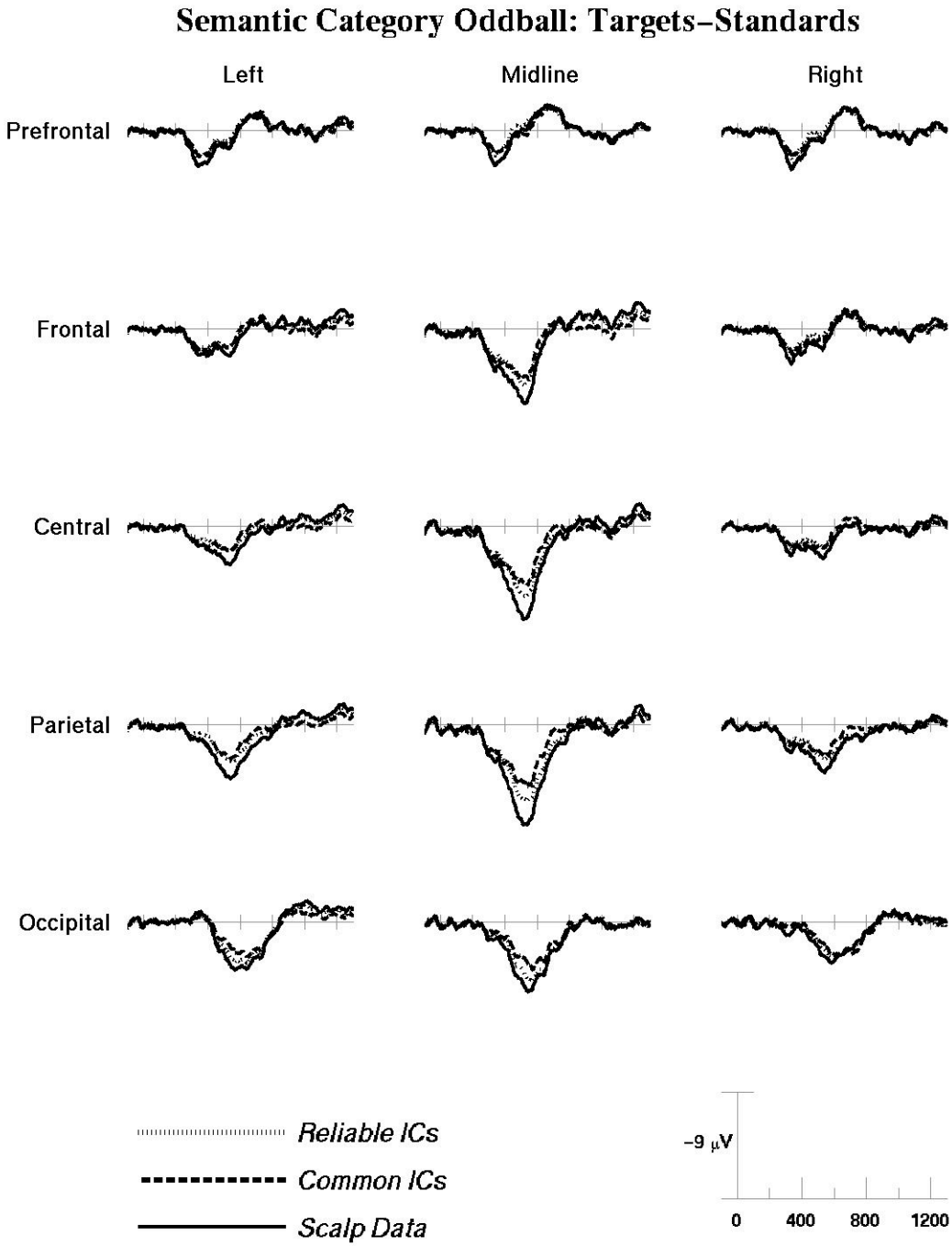

Figure VI.17 – Example grand average difference ERPs illustrating the P3b effect in the semantic category oddball task. Also plotted are the contributions to these difference ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

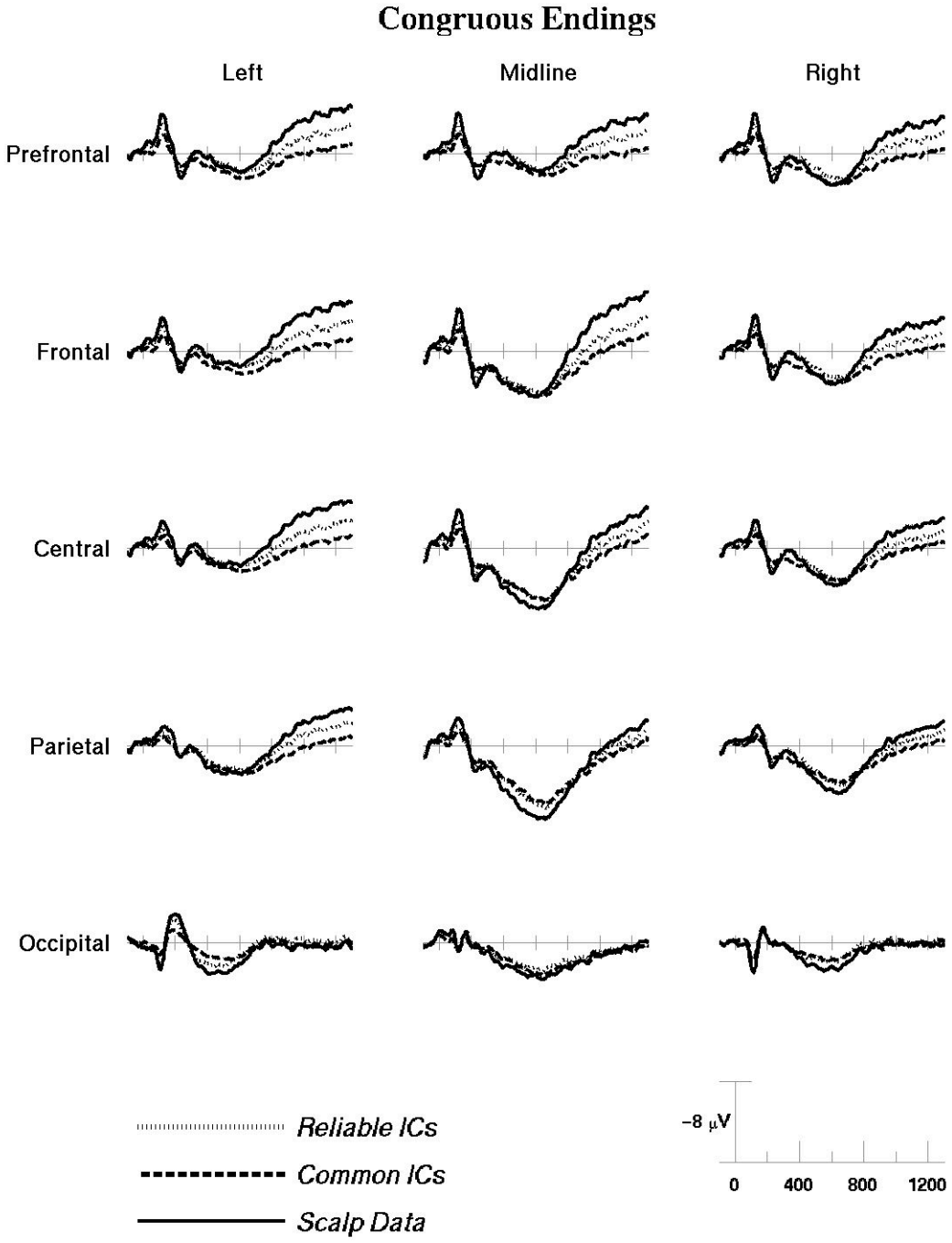

Figure VI.18 – Example grand average ERPs to congruous (noun and verb) endings in the sentence acceptability task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

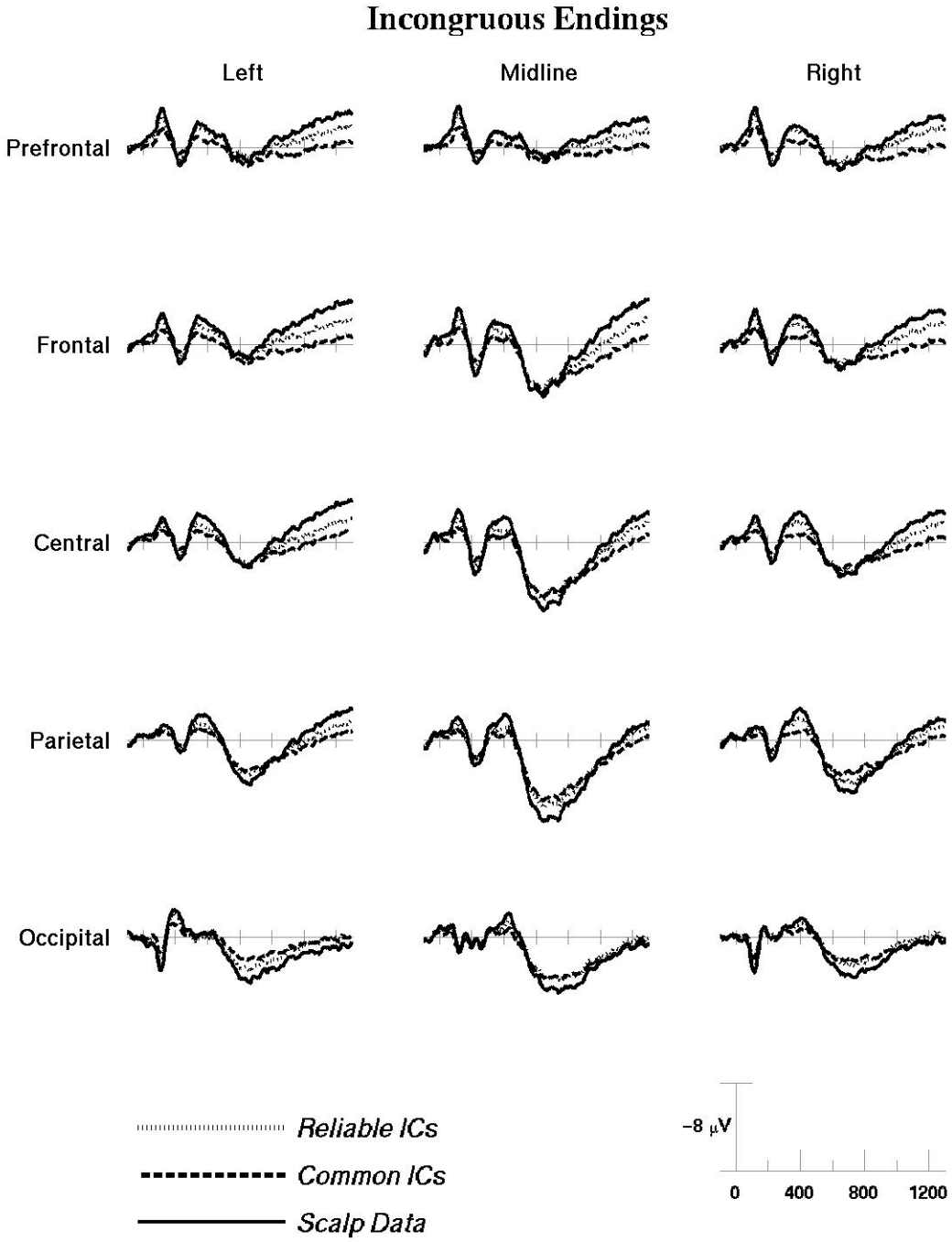

Figure VI.19 – Example grand average ERPs to incongruous (noun and verb) endings in the sentence acceptability task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

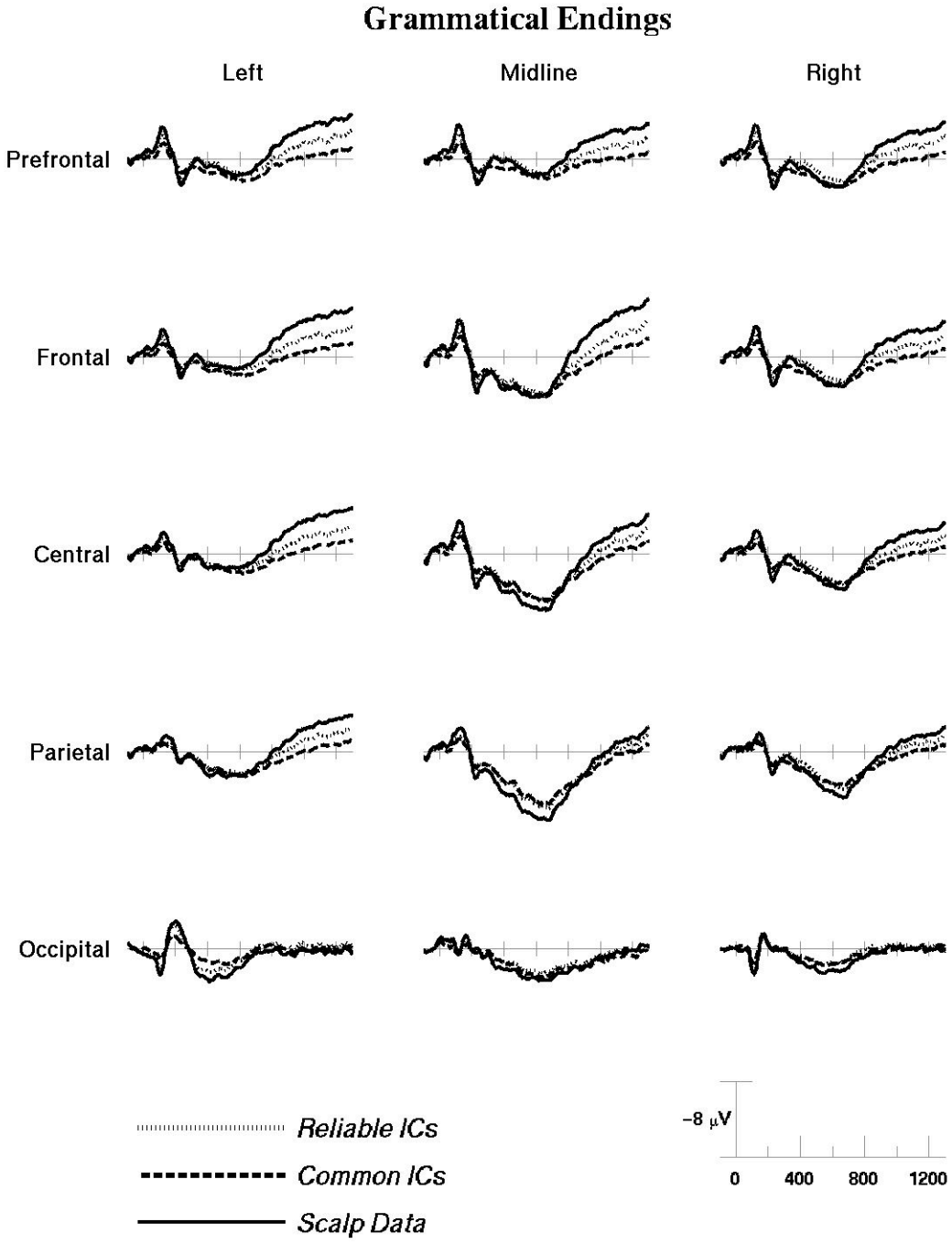

Figure VI.20 – Example grand average ERPs to grammatical (pronoun and verb) endings in the sentence acceptability task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

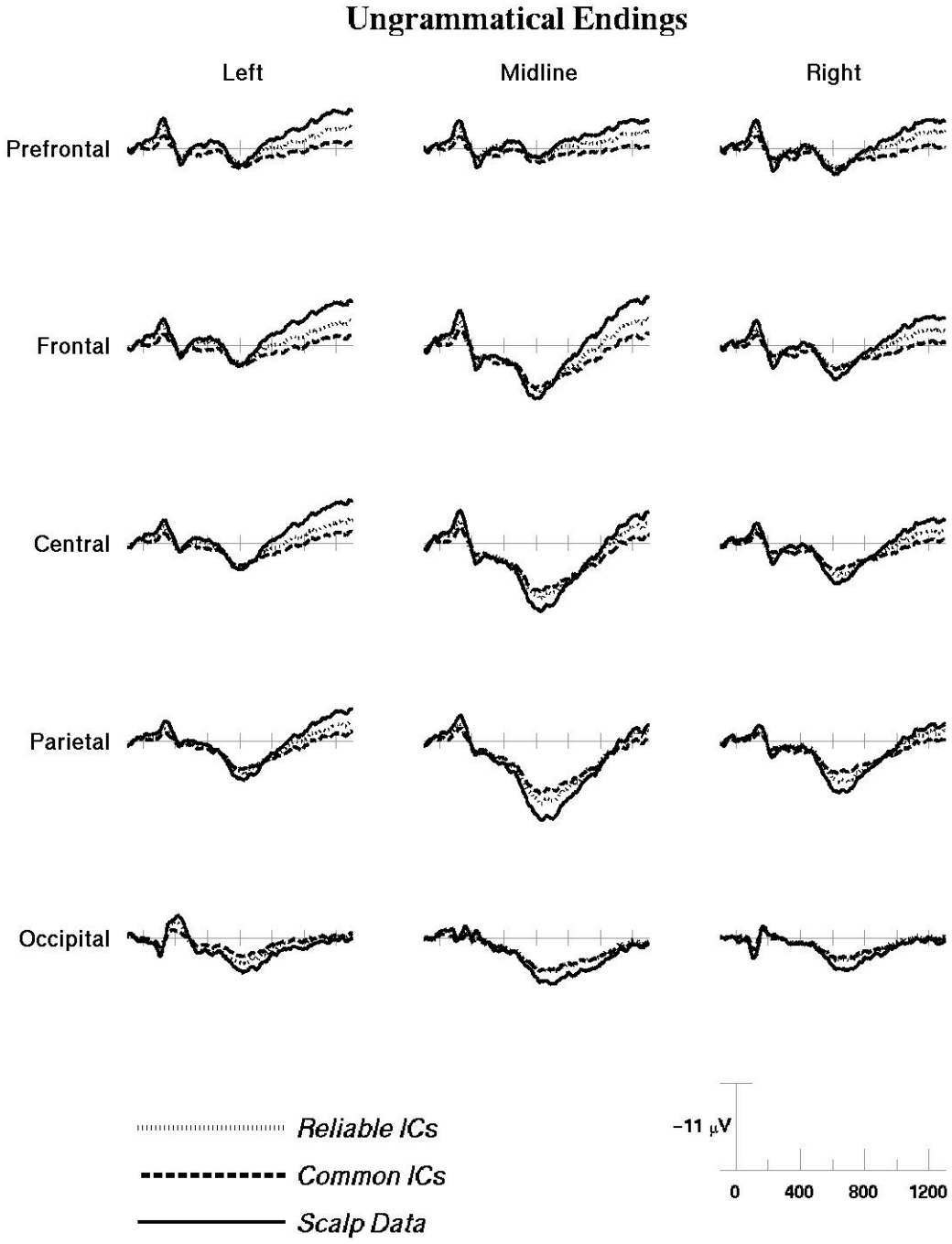

Figure VI.21 – Example grand average ERPs to ungrammatical (pronoun and verb) endings in the sentence acceptability task. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks.

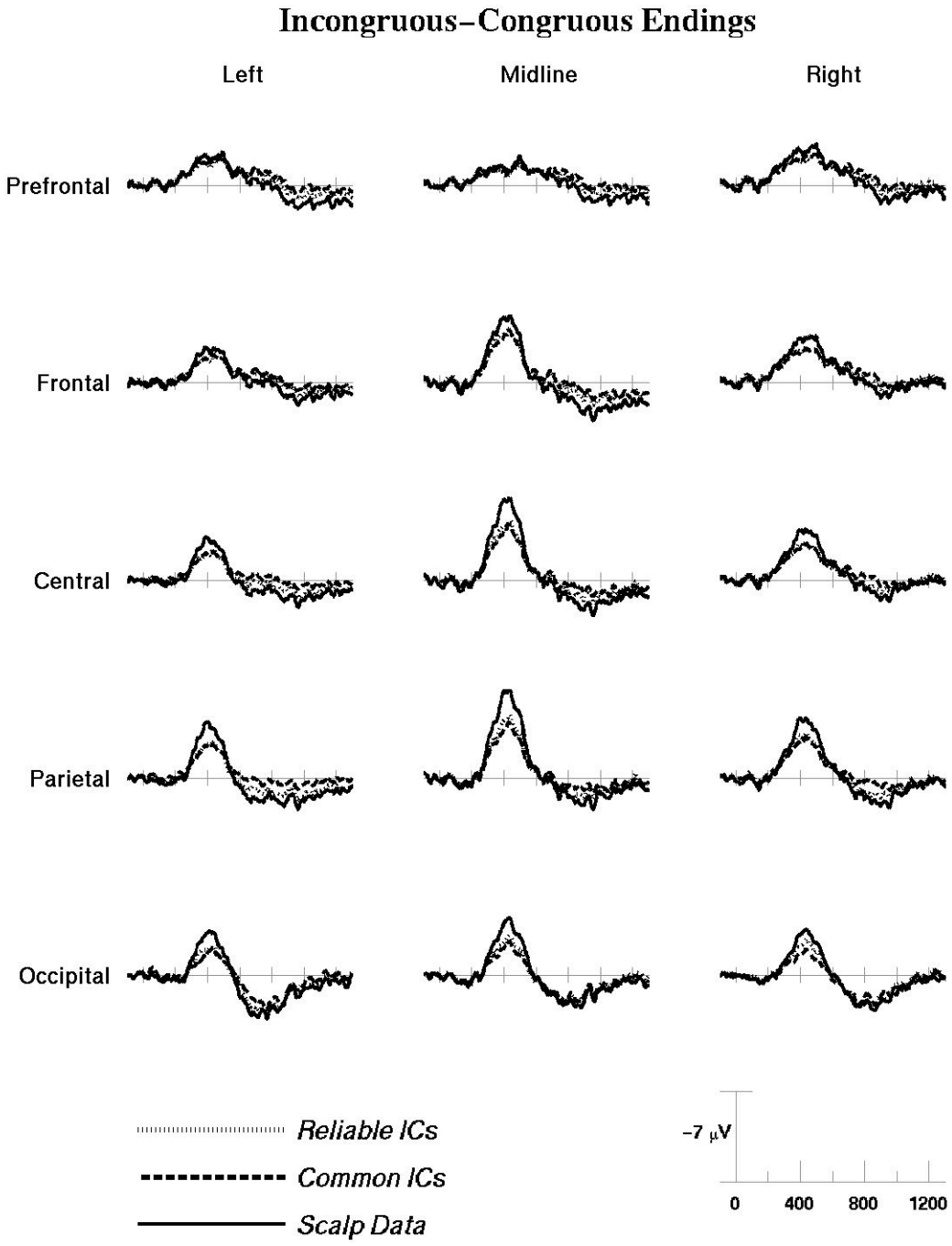

Figure VI.22 – Example grand average difference ERPs illustrating the N400 effect. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks. Difference ERPs are computed from both noun and verb ending ERPs.

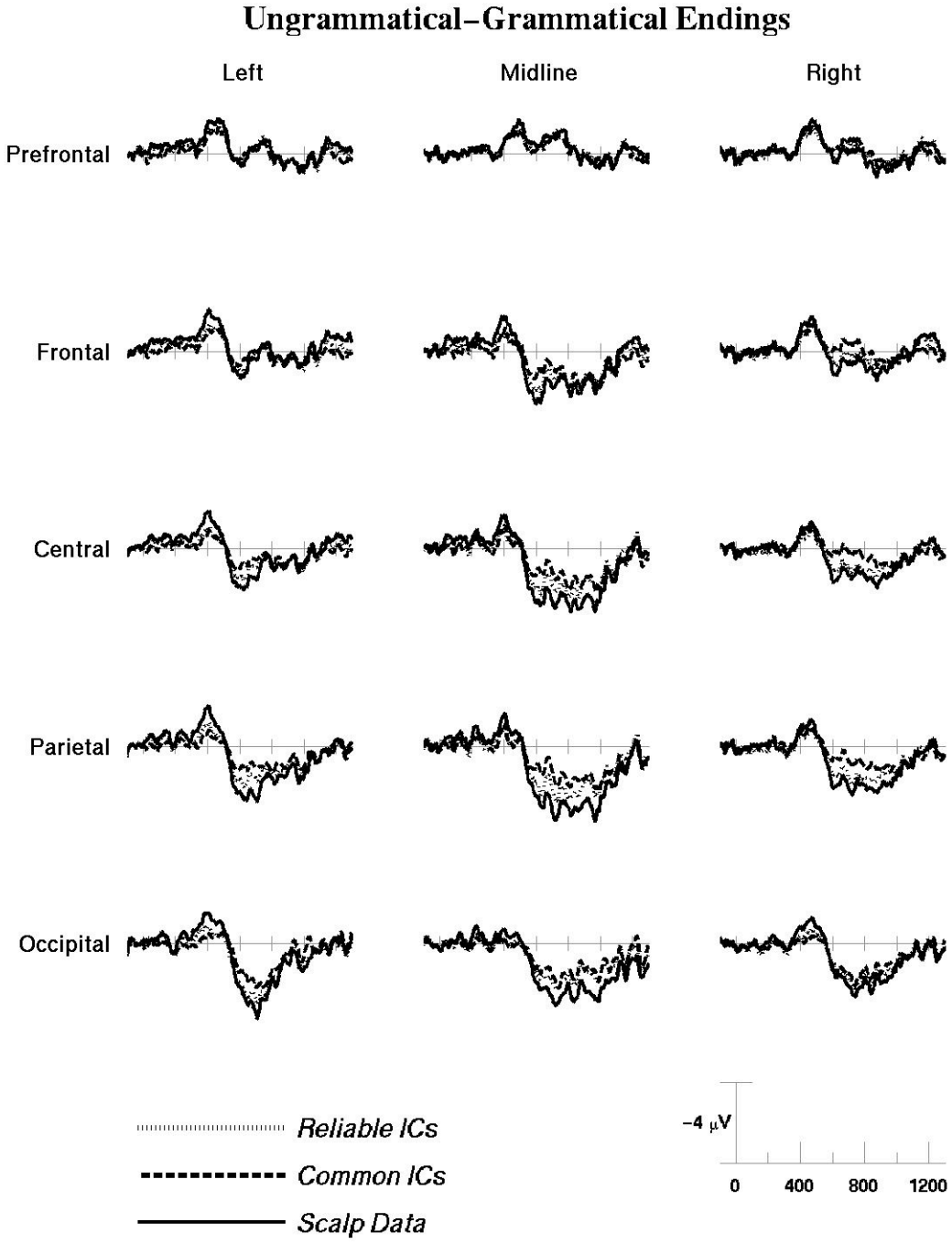

Figure VI.23 – Example grand average difference ERPs illustrating the P600 effect. Also plotted are the contributions to these ERPs of the independent components (ICs) common to the oddball and sentence acceptability tasks and of the reliable ICs derived from both tasks. Difference ERPs are computed from both pronoun and verb ending ERPs.

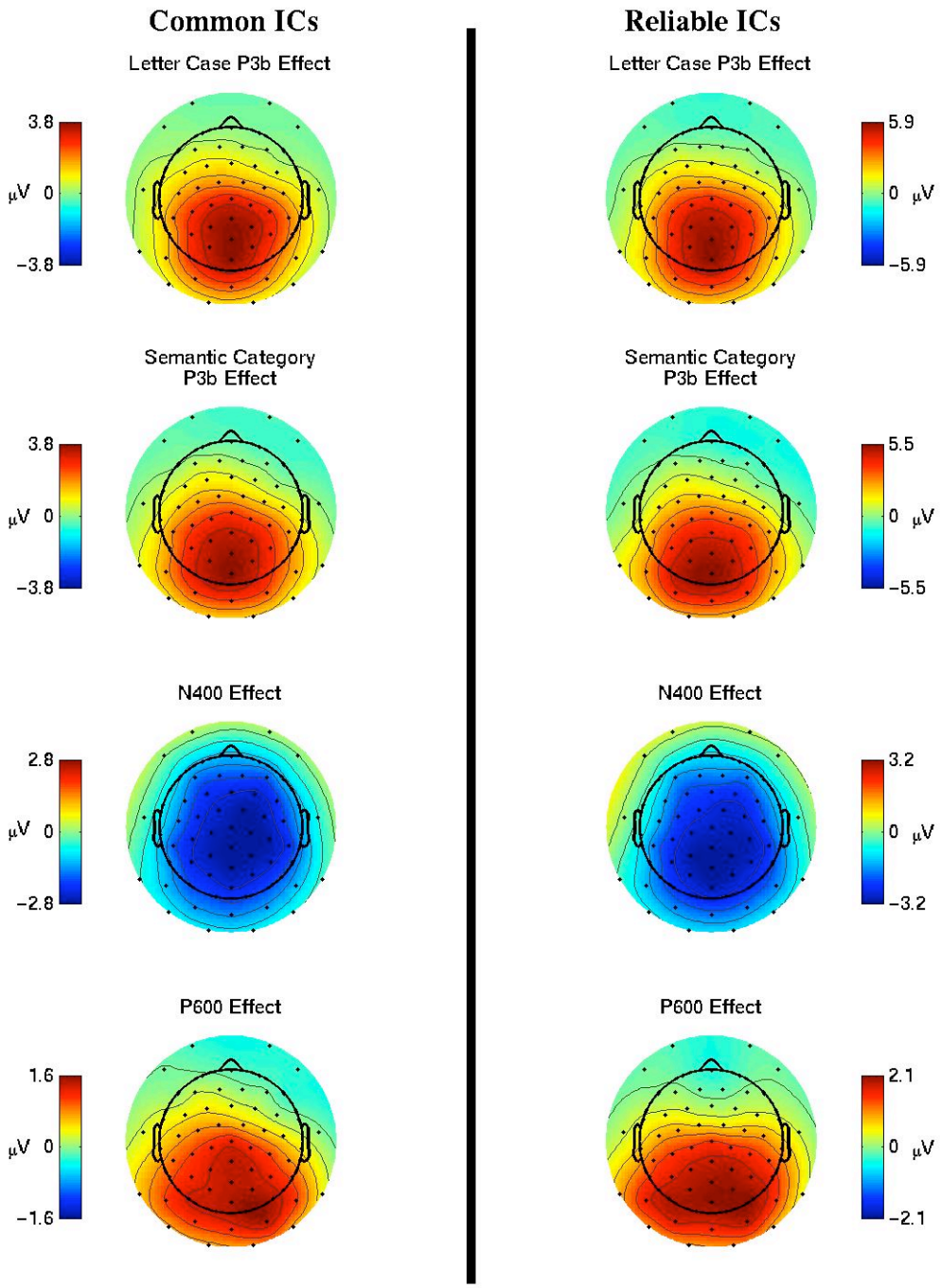

Figure VI.24 – Topographies of the grand average P3b, N400 (nouns and verbs), and P600 (pronouns and verbs) effects captured by ICs common to both experiments (**Left**) and by reliable ICs (**Right**). Effect windows are 325-575 ms (letter case task P3b), 425-675 ms (semantic category task), 275-550 ms (N400), and 550-1000 ms (P600). (Note: Right mastoid electrode, A2, is not plotted).

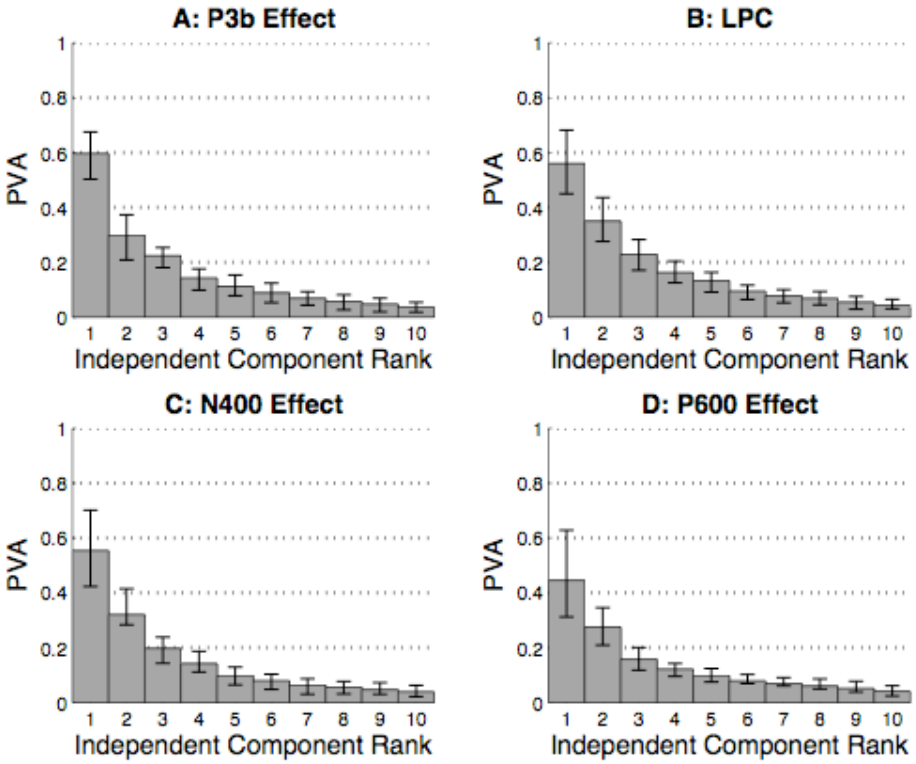

Figure VI.25 – Median (across all 16 participants) ERP effect or component percentage variance accounted for (PVA) by the 10 ICs that individually accounted for the most variance of that effect or component. For each participant, ICs were ordered from greatest to least before averaging. Error bars indicate 25th and 75th percentiles. The P3b effect was measured by taking the average P3b effect from the two oddball tasks and deriving individual P3b time windows that spanned from 150 ms before to 150 ms after the peak of the P3b effect. The LPC was measured by averaging the ERPs to nonanomalous noun, pronoun, and verb endings (each part of speech weighted equally) and then deriving individual LPC time windows that spanned from 150 ms before to 150 ms after the peak of the LPC. The N400 effect was measured by subtracting the ERPs to all incongruous endings from the ERPs to all comparable congruous endings; the segments of these difference ERPs in the grand average N400 window (275 to 550 ms) were the N400 effect. The P600 effect was measured by subtracting the ERPs to all ungrammatical endings from the ERPs to all comparable grammatical endings; the segments of these difference ERPs in the grand average P600 window (550 to 1000 ms) were the P600 effect. The derivation of individual P3b effect and LPC windows is described in Section IV.G.1. Grand average windows were used for the N400 and P600 effects since the peak latencies of these effects in individual participants were not clearly reliable (ibid.).

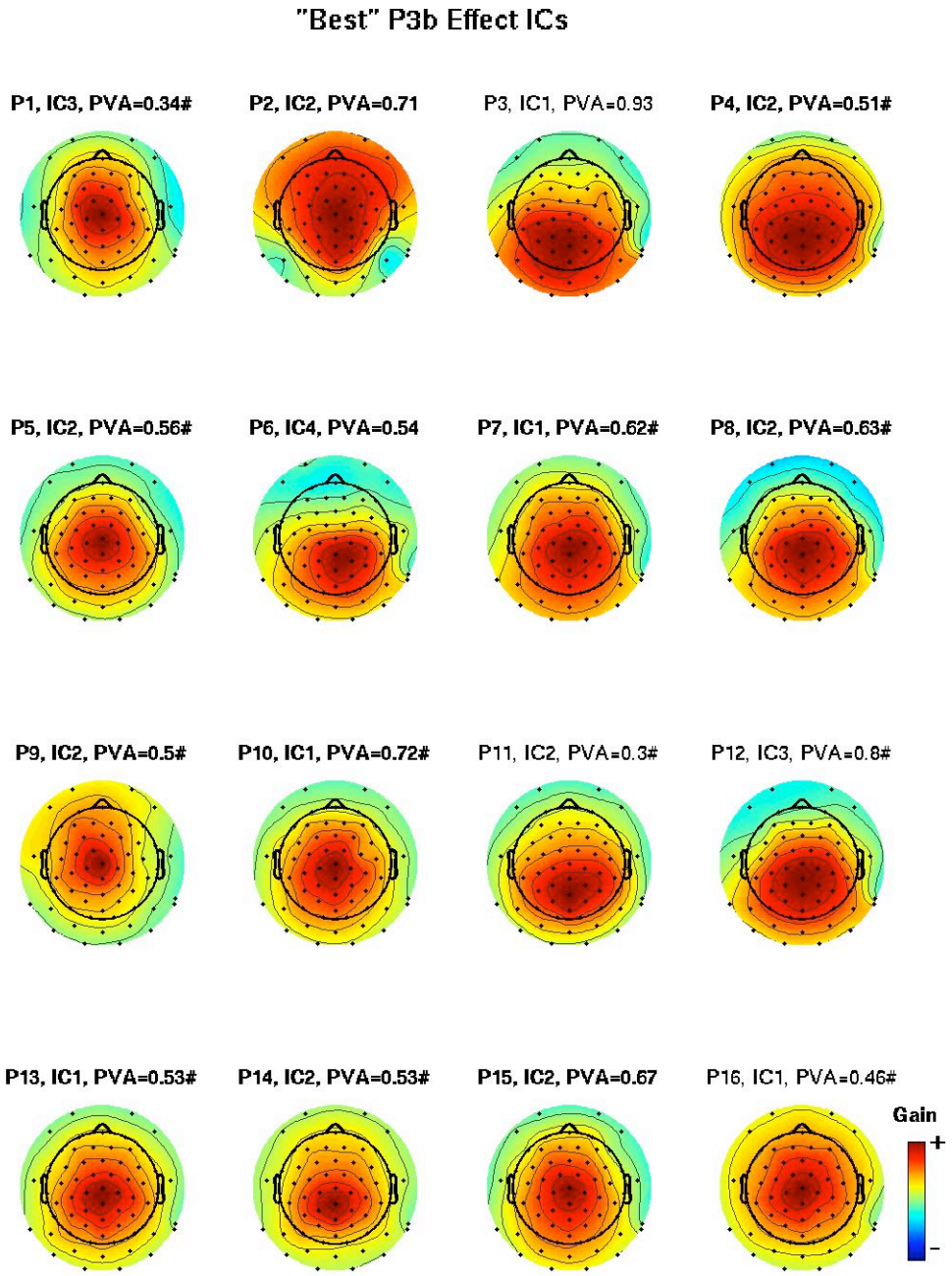

Figure VI.26 – Topographies of each participant's hand-picked single most P3b-like IC. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute gain, -=-maximum absolute gain). The polarity of each IC has been set such that the maximum absolute gain weight is positive. Bold headings indicate that the IC is common to both the oddball and sentence acceptability tasks. # indicates that the IC is reliable. Abbreviations: P=Participant, IC=Independent Component, PVA=percentage variance accounted for (of the P3b effect averaged across both oddball tasks).

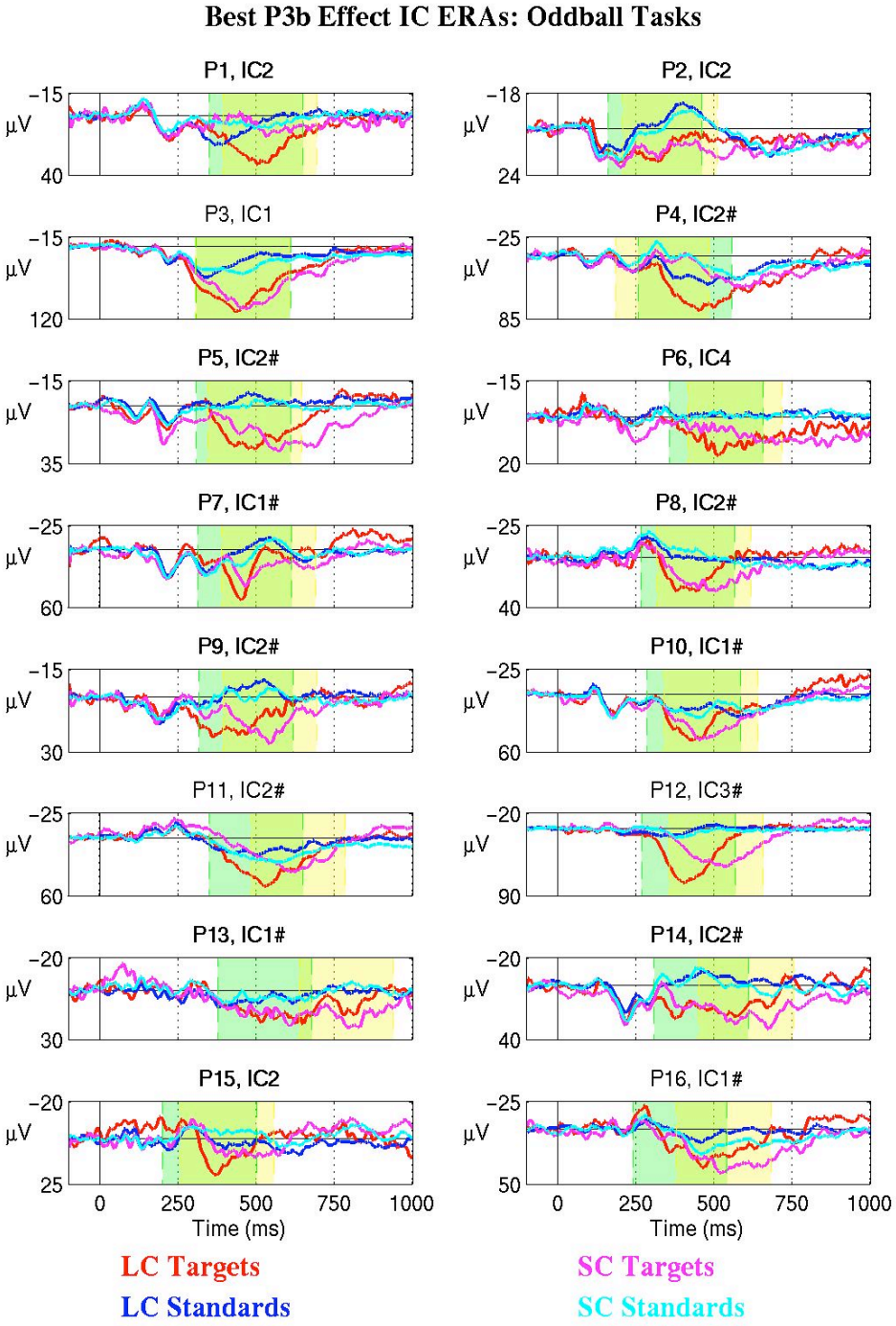

Figure VI.27 – Average oddball task activations of each participant's "best P3b" IC. Bold headings indicate that the IC is common to both the oddball and sentence acceptability tasks. # indicates that the IC is reliable. Green and yellow shading indicate a participant's LC and SC task P3b windows, respectively. Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component, LC=Letter Case, SC=Semantic Category.

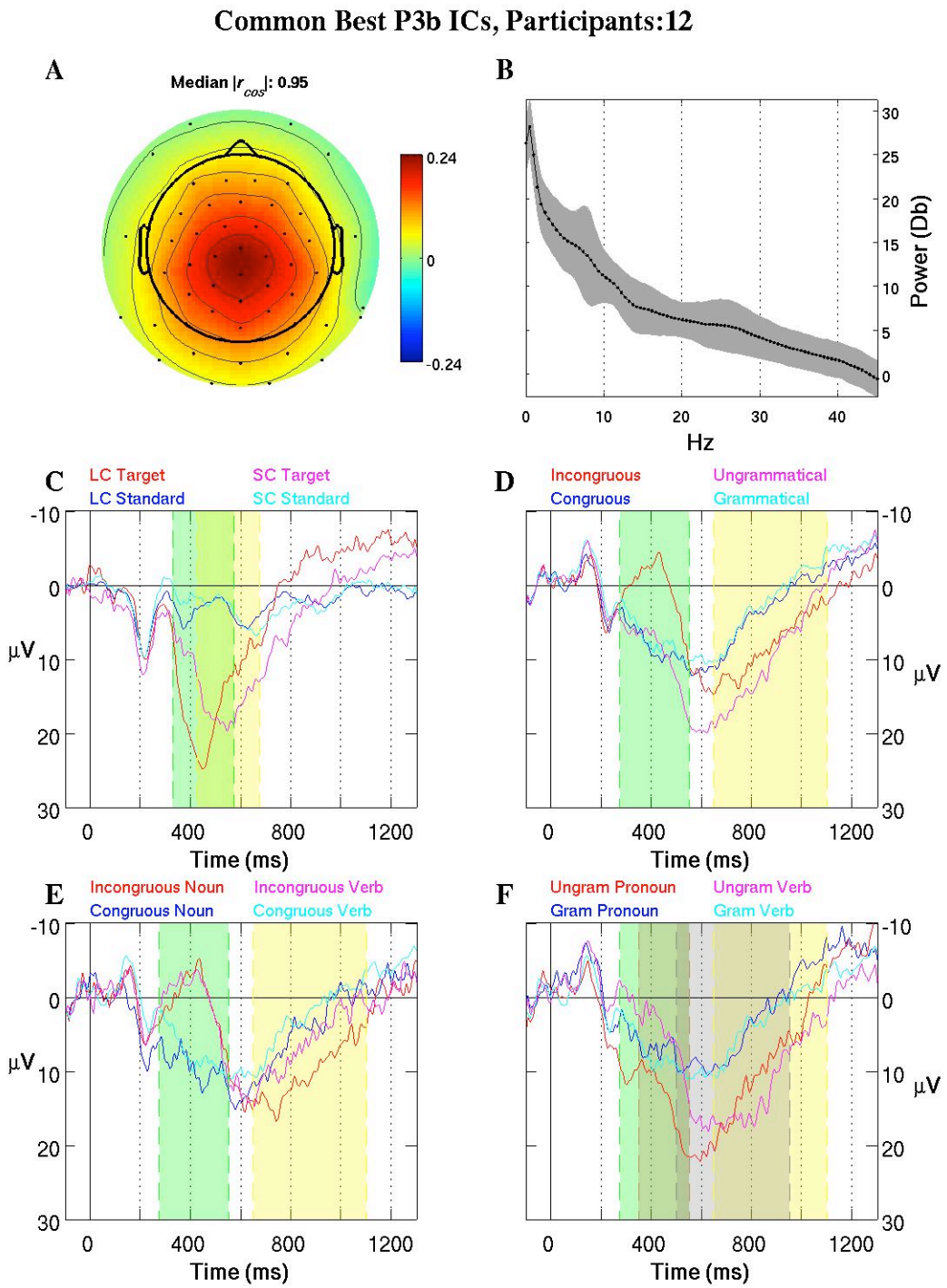

Figure VI.28 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one SD). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

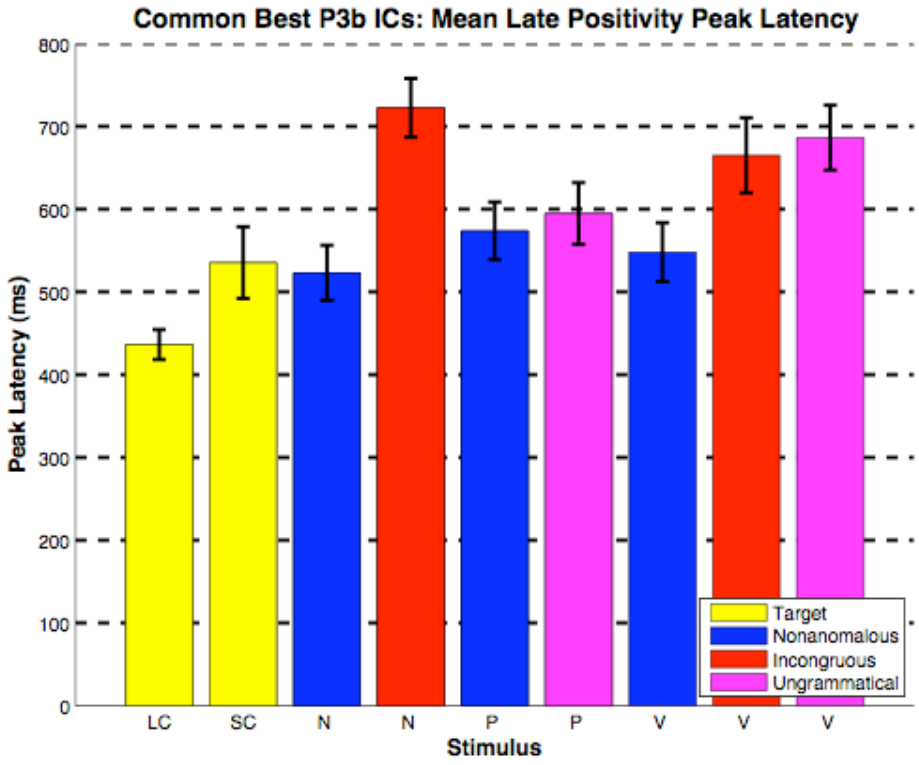

Figure VI.29 – Mean peak latency of the late positivity of the 12 common best P3b ICs. Error bars are standard error. The peak latency of the late positivity to targets was measured as the latency of the most positive point in an IC's ERA to targets in an oddball task from 0 to 1300 ms after low pass filtering the ERA at 7 Hz. The peak latency of the late positivity to sentence endings was measured as the latency of the most positive point in an IC's ERA to a particular class of sentence ending from 300 to 1000 ms after low pass filtering the ERA at 7 Hz. Abbreviations: LC=Letter Case (Oddball Task), SC=Semantic Category (Oddball Task), N=Noun, P=Pronoun, V=Verb.

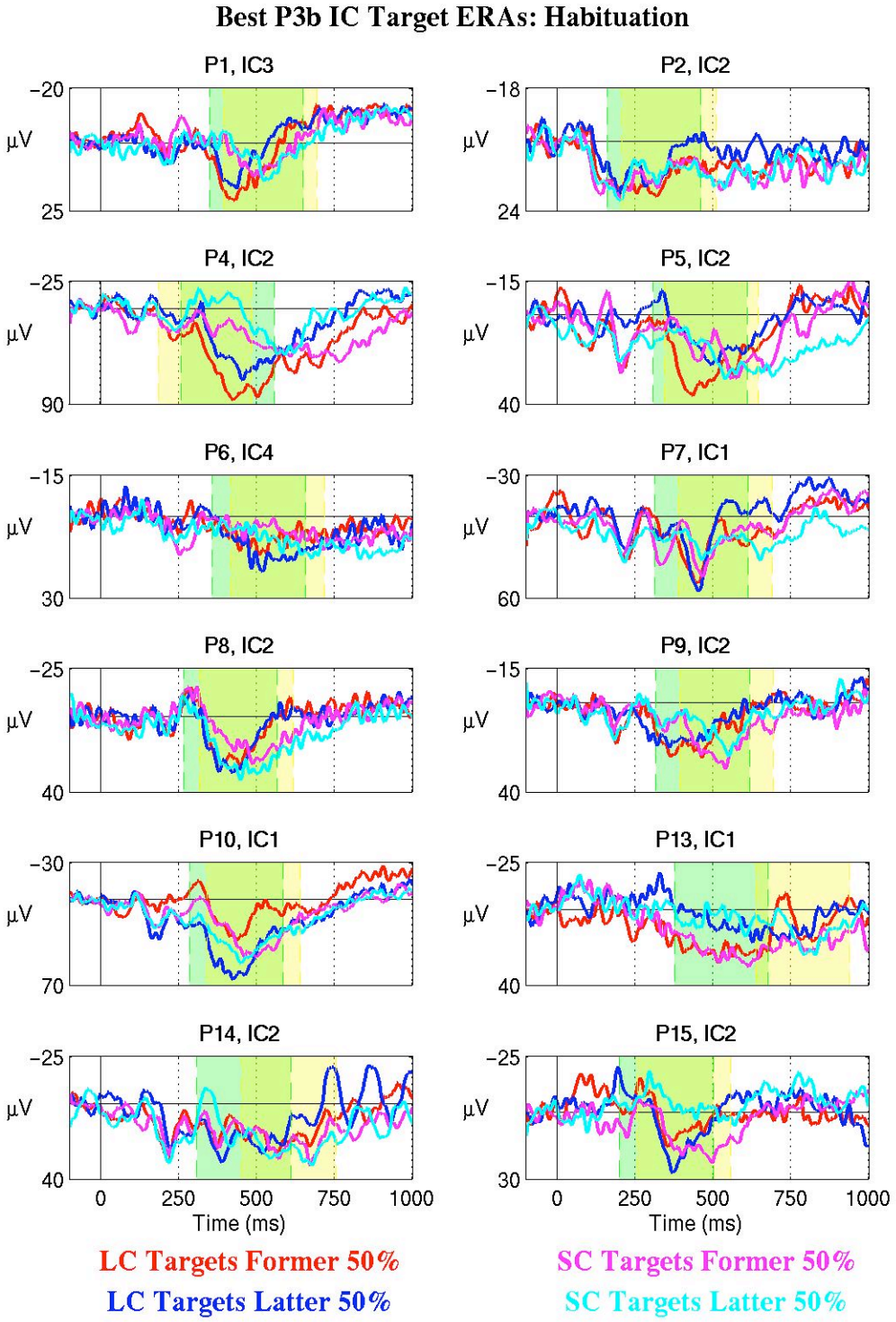

Figure VI.30 – Average responses to targets from the first and second half of both oddball tasks of the best P3b ICs that are common to the oddball and sentence acceptability tasks. Green and yellow shading indicate a participant's LC and SC task P3b windows (respectively). Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component, LC=Letter Case, SC=Semantic Category.

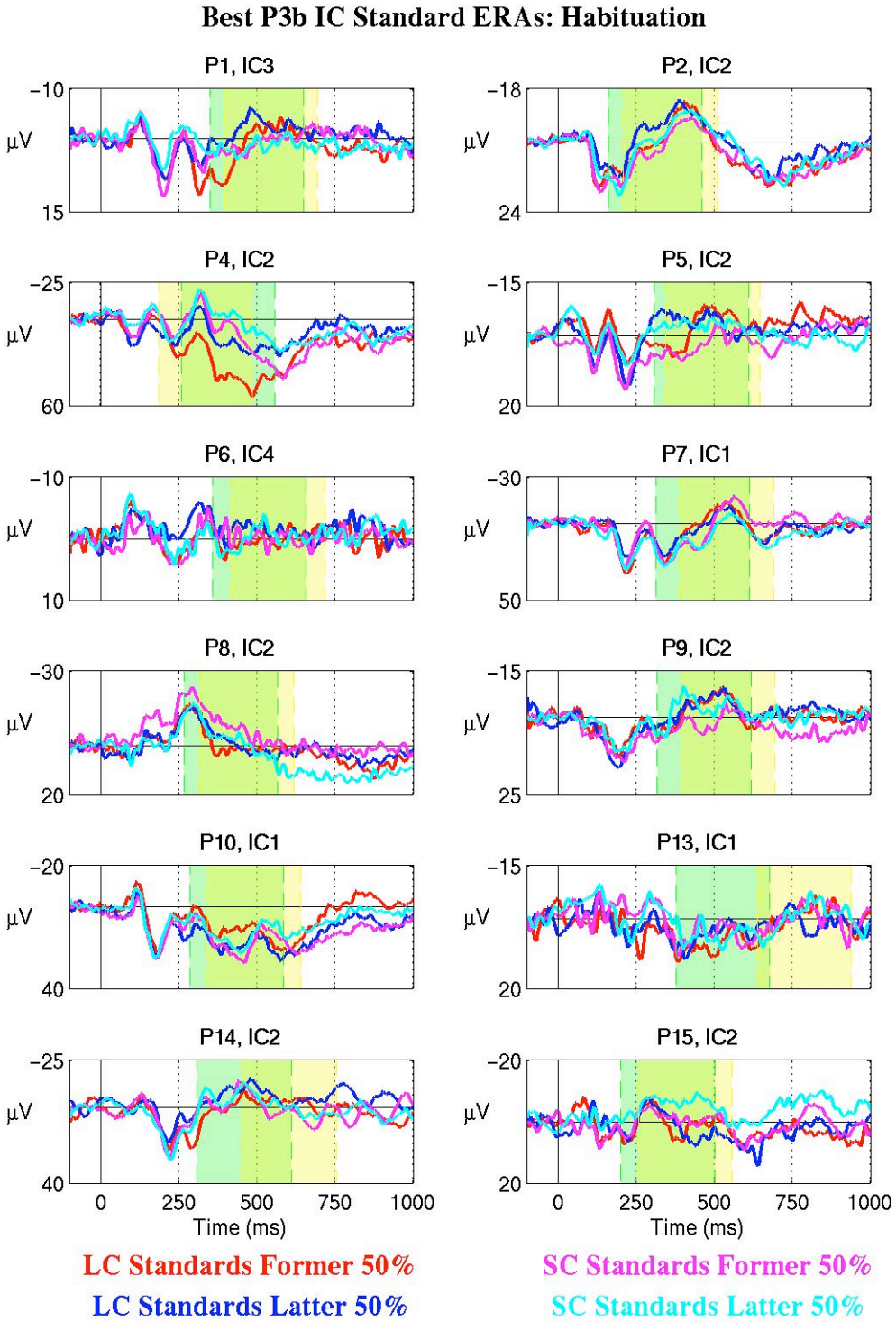

Figure VI.31 – Average responses to standards from the first and second half of both oddball tasks of the best P3b ICs that are common to the oddball and sentence acceptability tasks. Green and yellow shading indicate a participant's LC and SC task P3b windows (respectively). Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component, LC=Letter Case, SC=Semantic Category.

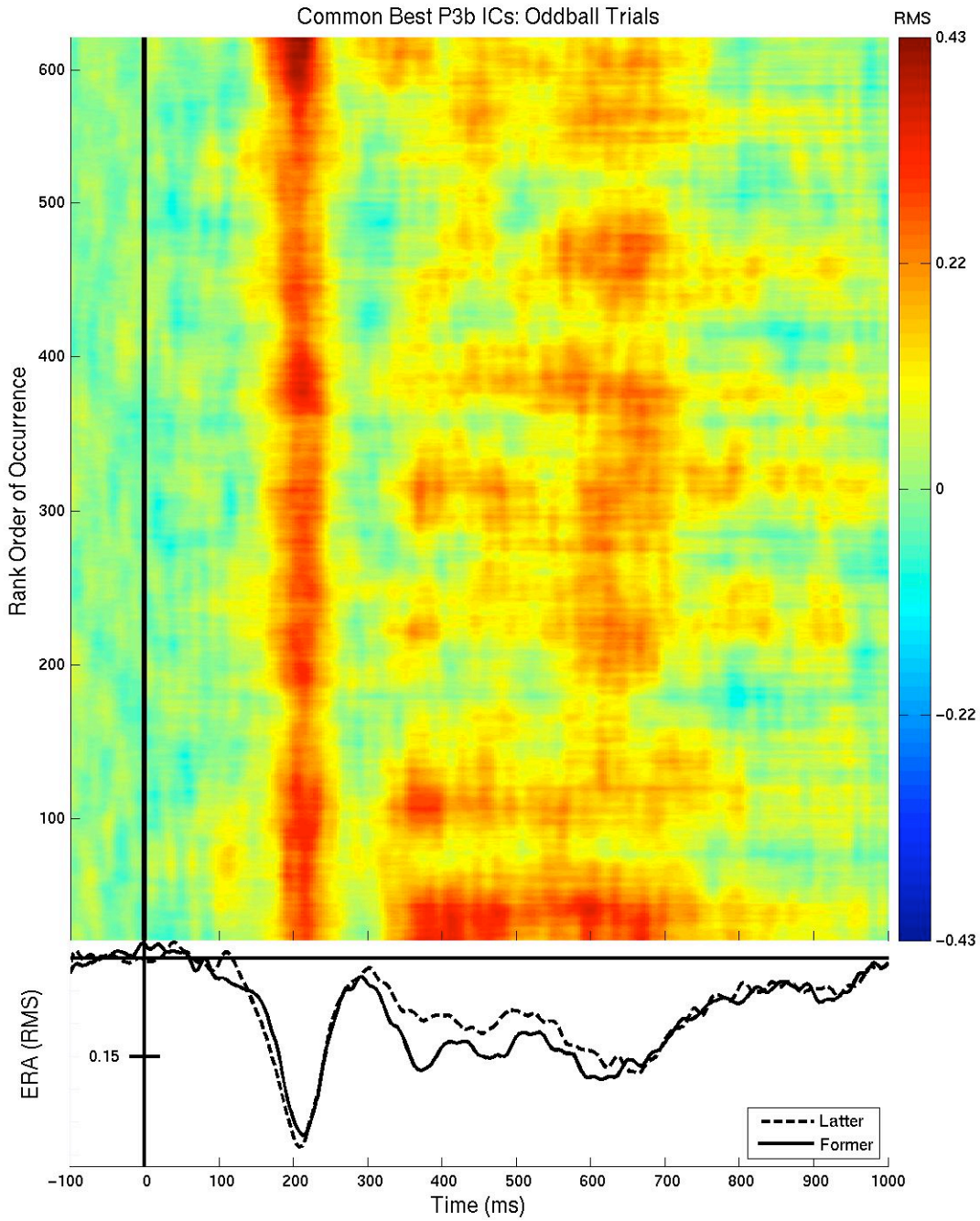

Figure VI.32 – **(Top)** 7530 single trials of the 12 common best P3b IC responses to standards and targets in both oddball tasks. Each IC's activation has been normalized to unit root mean square amplitude and their polarities are as shown in Figure VI.26. Trials are sorted by order of occurrence (1 indicates the earliest trial), stacked along the y-axis, and smoothed using a 500 trial vertical moving average. **(Bottom)** Averages (ERAs) of the former and latter 50% of the single trials.

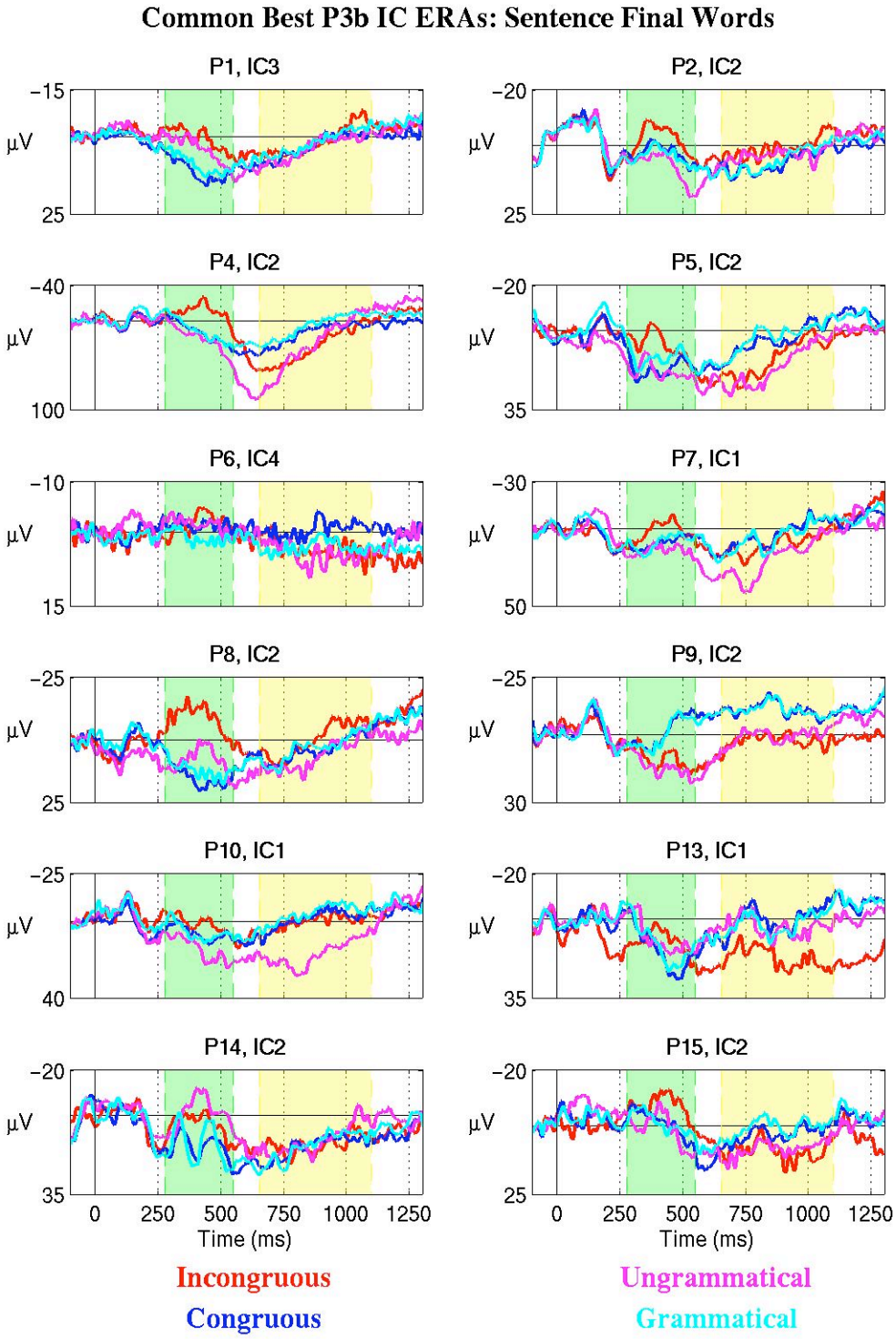

Figure VI.33 – Average activations of common best P3b ICs to sentence final words. Green and yellow shading indicate grand average noun/verb N400 and P600 windows, respectively. Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

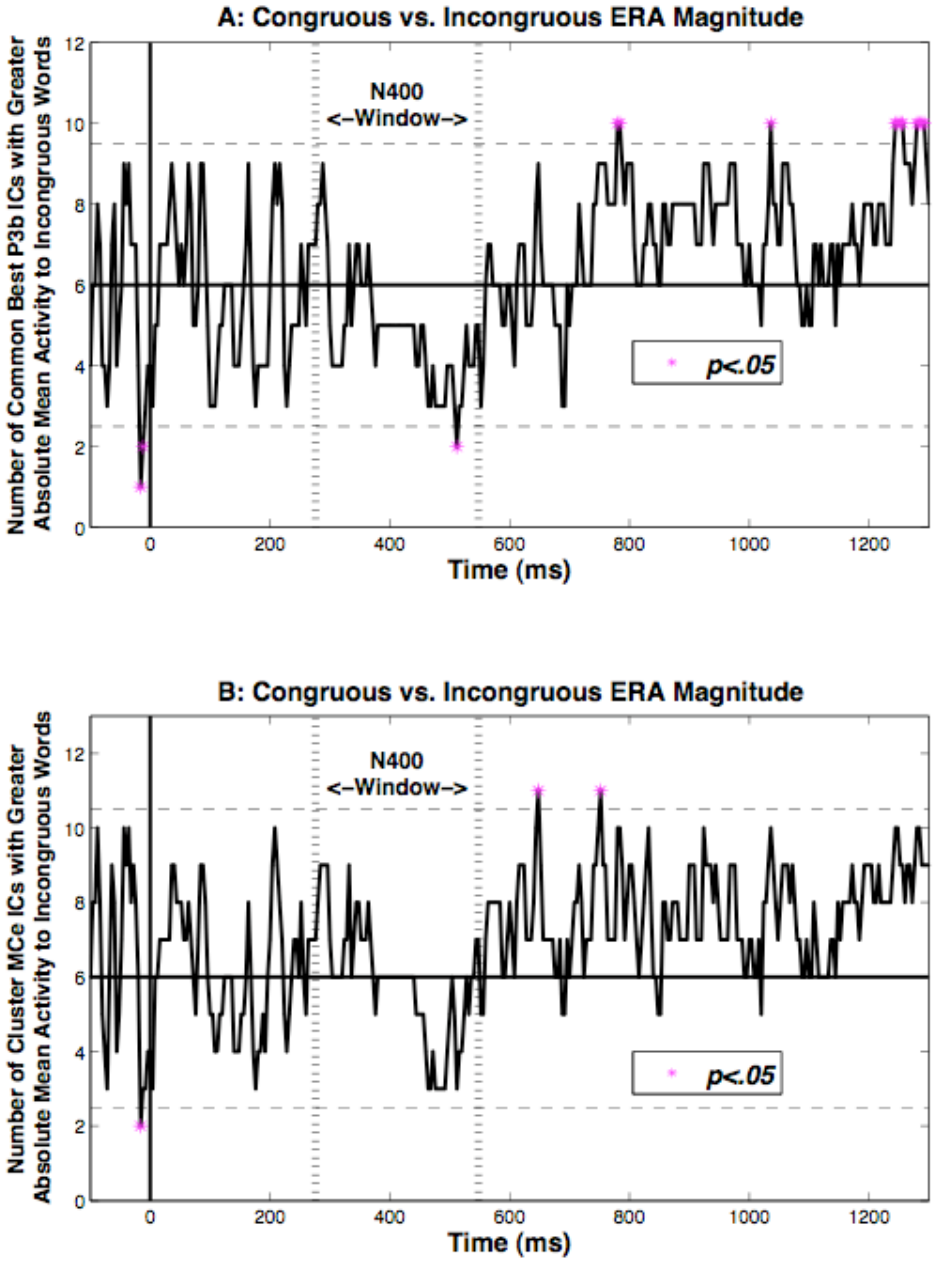

Figure VI.34 – (**A**) Number of common best P3b ICs whose ERAs to incongruous words are of greater absolute magnitude than their ERAs to nonanomalous noun and verb endings. Significantly many or few numbers of ICs are indicated with pink asterisks (two tailed sign test). Horizontal dashed lines indicate cut-off for statistical significance. (**B**) Same as A, but for the 13 ICs in the MCe cluster.

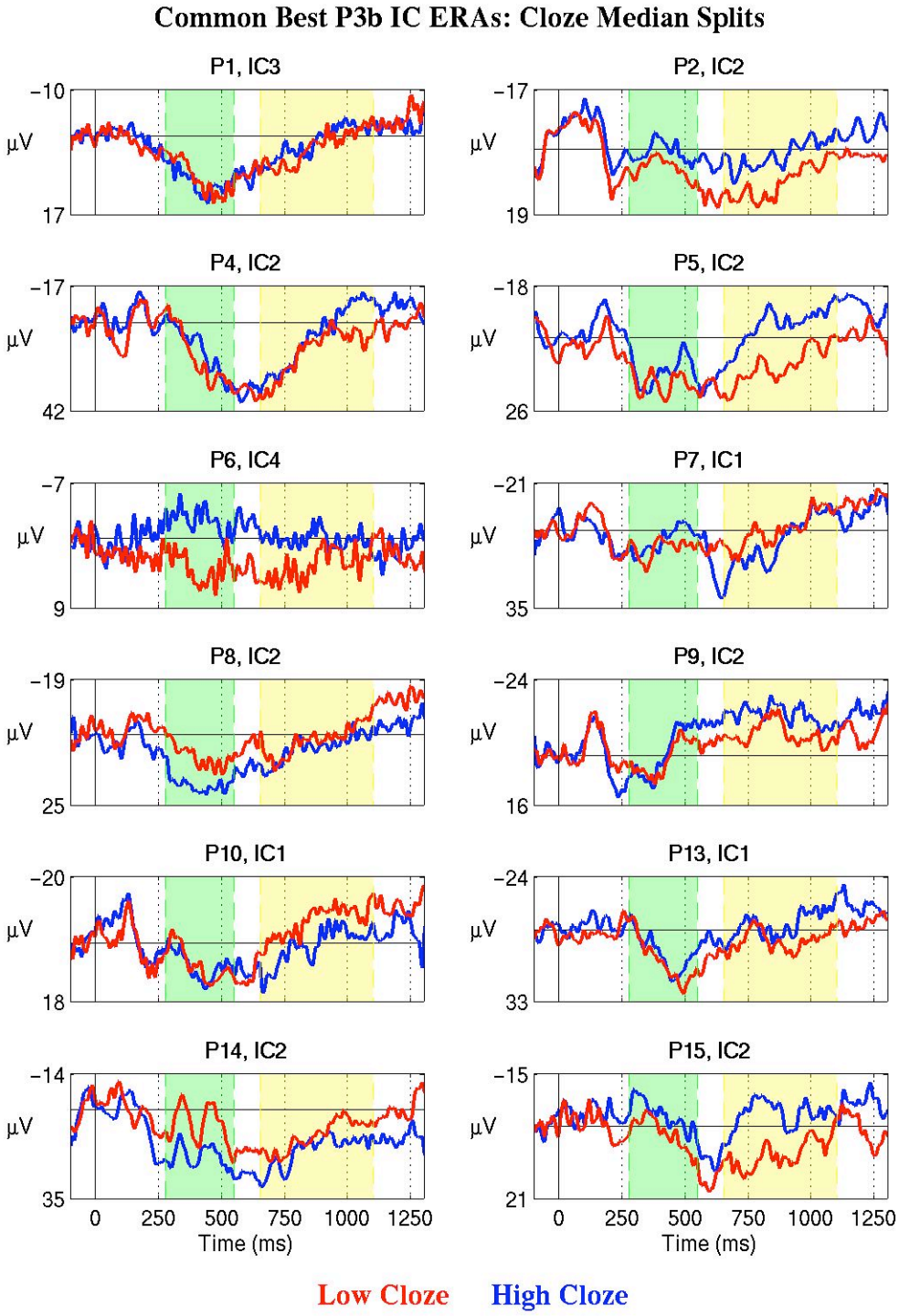

Figure VI.35 – Average responses to low and high cloze nonanomalous noun and verb endings of the best P3b ICs that are common to the oddball and sentence acceptability tasks. Green and yellow shading indicate grand average N400 and P600 windows (respectively). Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

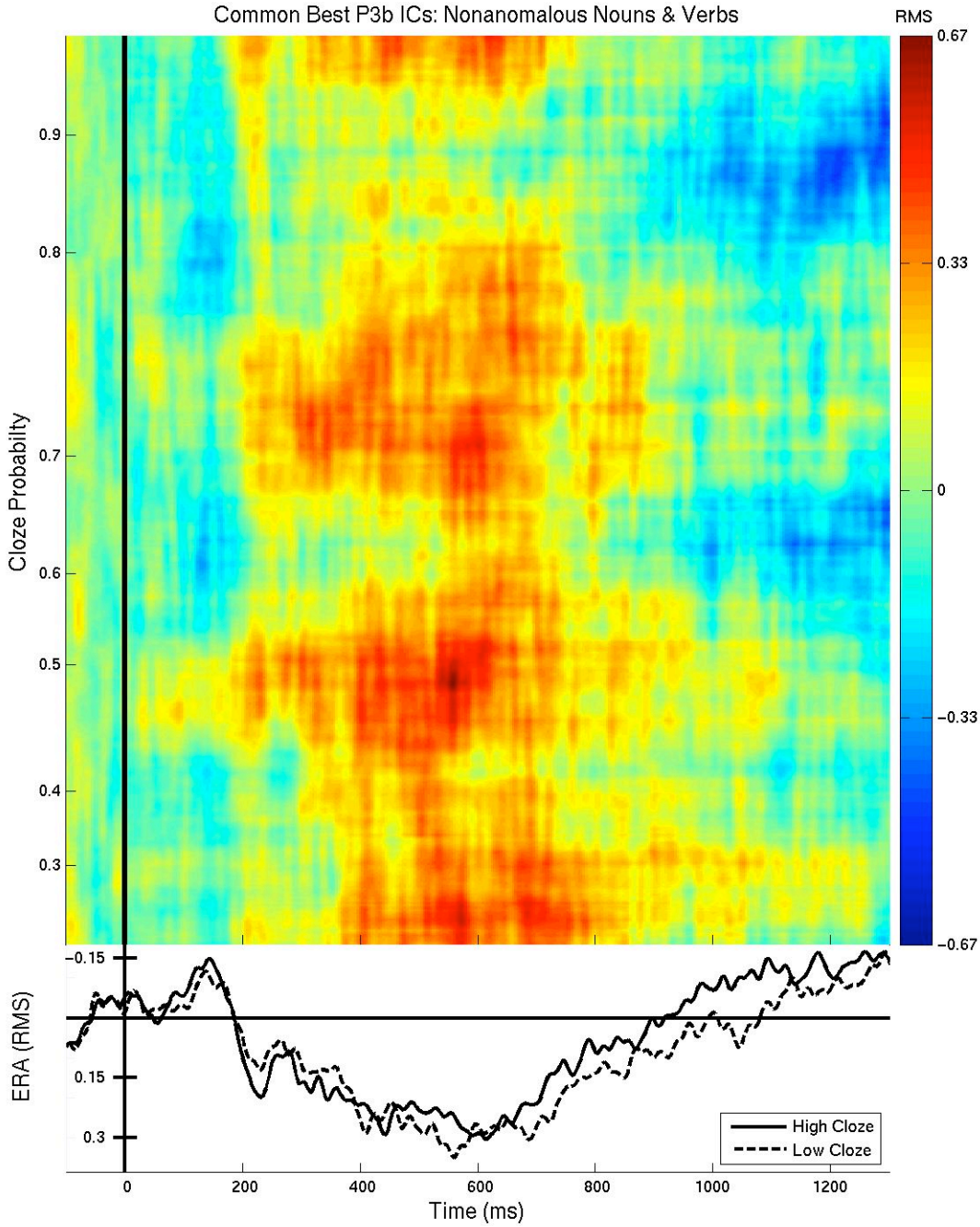

Figure VI.36 – **(Top)** 1053 single trials of the 12 common best P3b IC responses to nonanomalous noun and verb sentence endings. Each IC's activation has been normalized to unit root mean square amplitude and their polarities are as shown in Figure VI.26. Trials are sorted by cloze probability, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERAs) of the top and bottom 50% of the single trials.

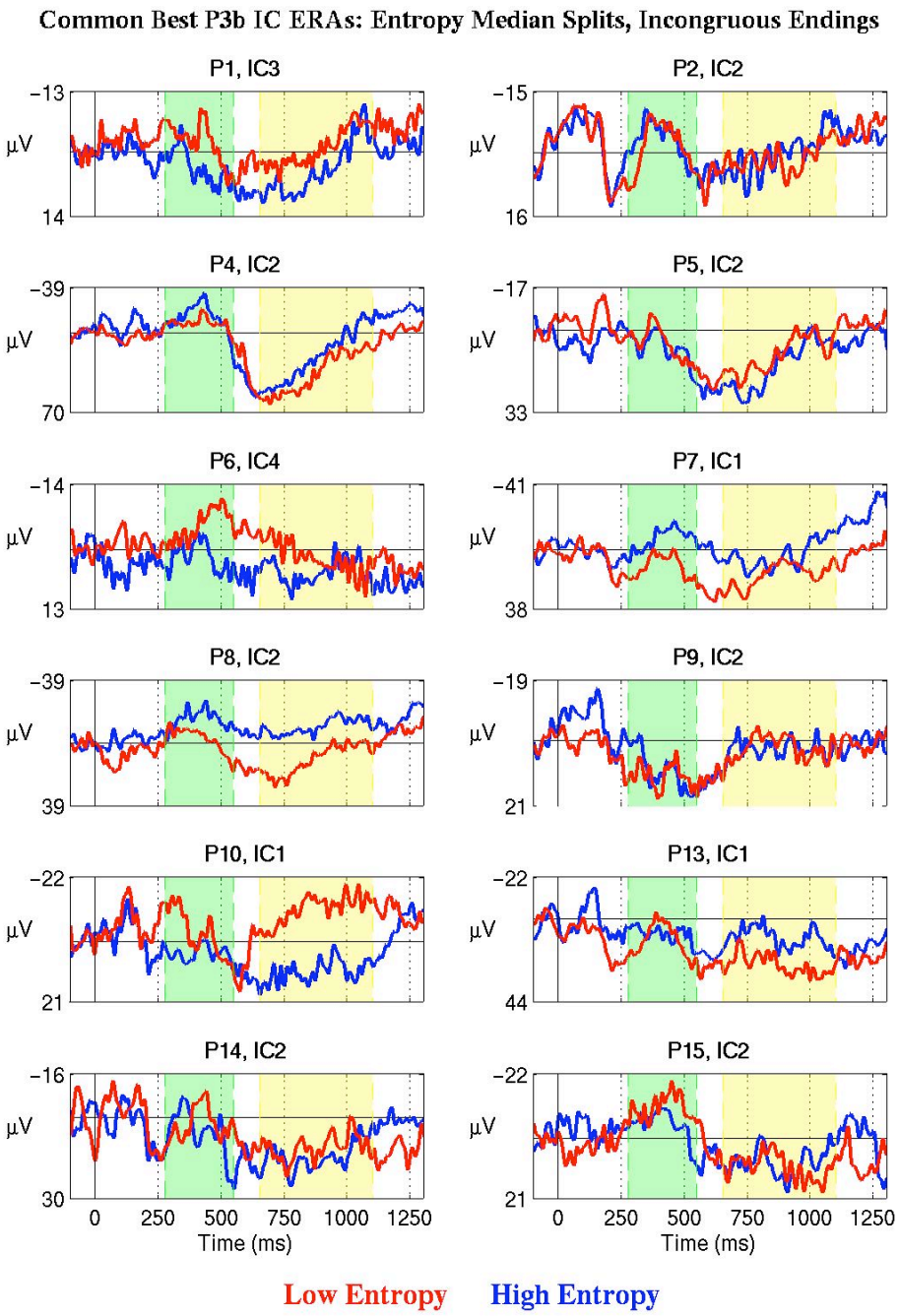

Figure VI.37 – Average responses of the best P3b ICs that are common to oddball and sentence acceptability tasks to the incongruous endings of low and high entropy sentence stems. Green and yellow shading indicate grand average N400 and P600 windows (respectively). Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

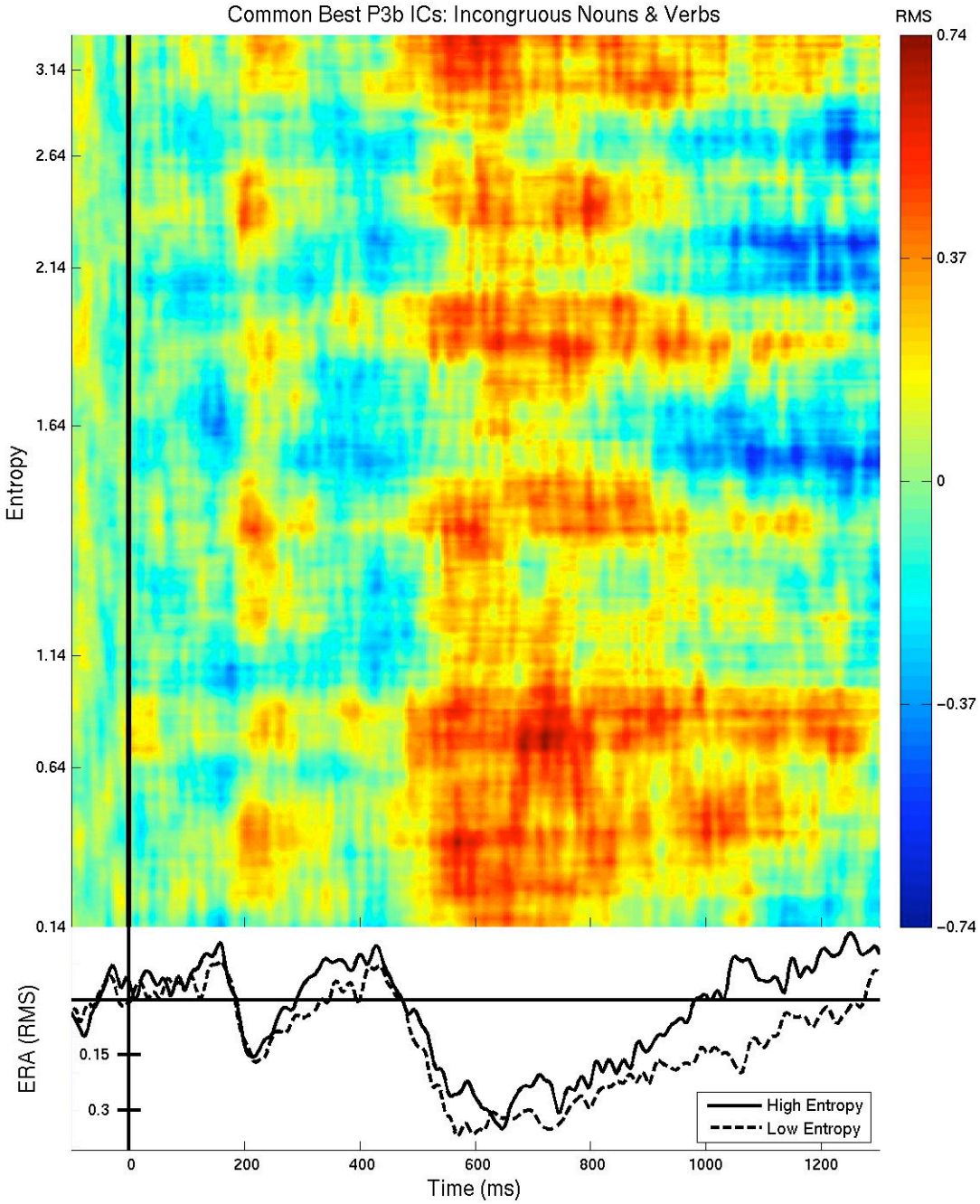

Figure VI.38 – **(Top)** 728 single trials of the 12 common best P3b IC responses to incongruous noun and verb sentence endings. Each IC's activation has been normalized to unit root mean square amplitude and their polarities are as shown in Figure VI.26. Trials are sorted by entropy (in bits), stacked along the y-axis, and smoothed using a 50 trial vertical moving average. **(Bottom)** Averages (ERAs) of the top and bottom 50% of the single trials.

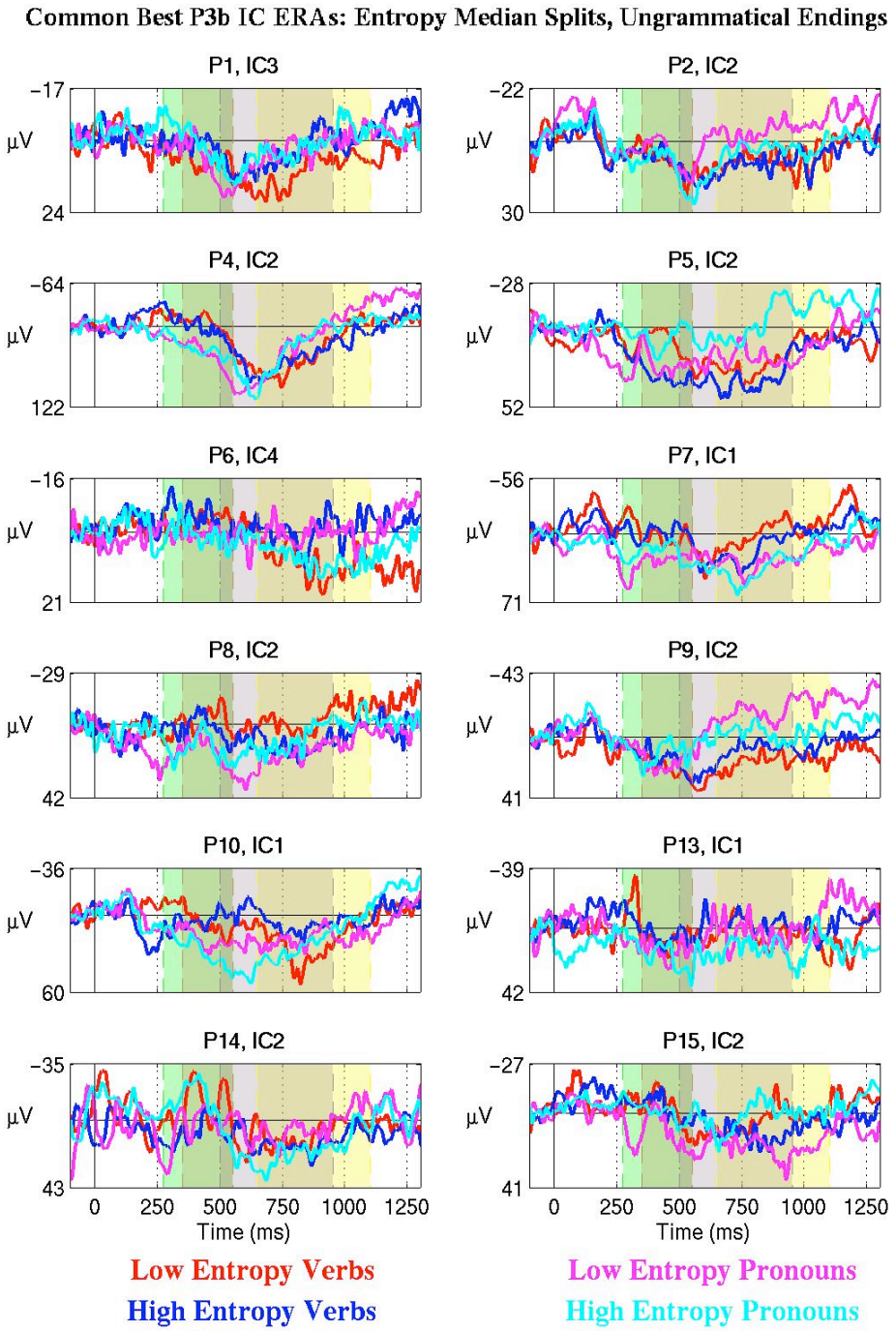

Figure VI.39 – Average responses of the best P3b ICs that are common to oddball and sentence acceptability tasks to the ungrammatical endings of low and high entropy sentence stems. Green, yellow, brown, and grey shading indicate grand average verb N400, verb P600, pronoun N400, and pronoun P600 windows (respectively). Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

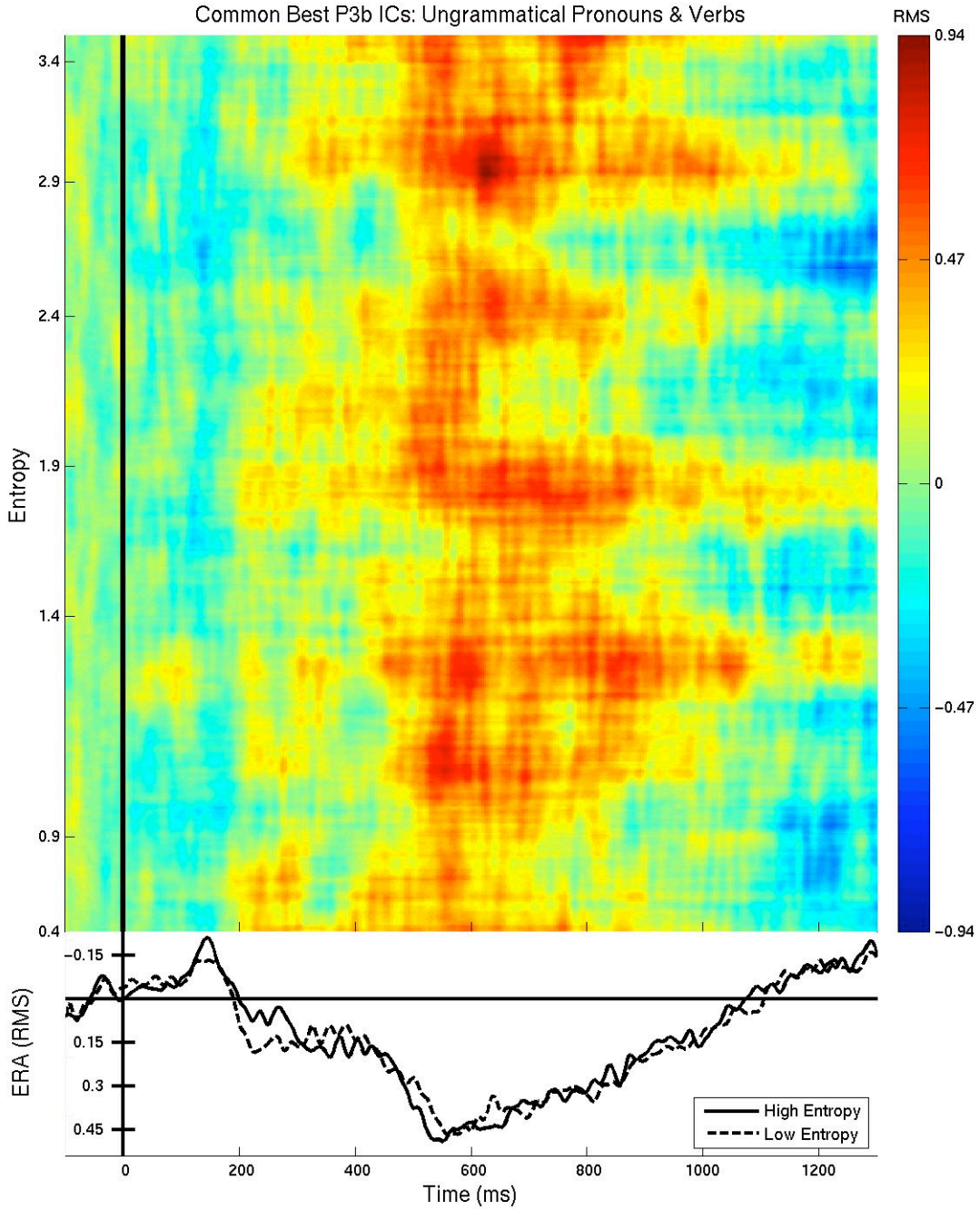

Figure VI.40 – **(Top)** 701 single trials of the 12 common best P3b IC responses to ungrammatical pronoun and verb sentence endings. Each IC's activation has been normalized to unit root mean square amplitude and their polarities are as shown in Figure VI.26. Trials are sorted by entropy (in bits), stacked along the y-axis, and smoothed using a 50 trial vertical moving average. **(Bottom)** Averages (ERAs) of the top and bottom 50% of the single trials.

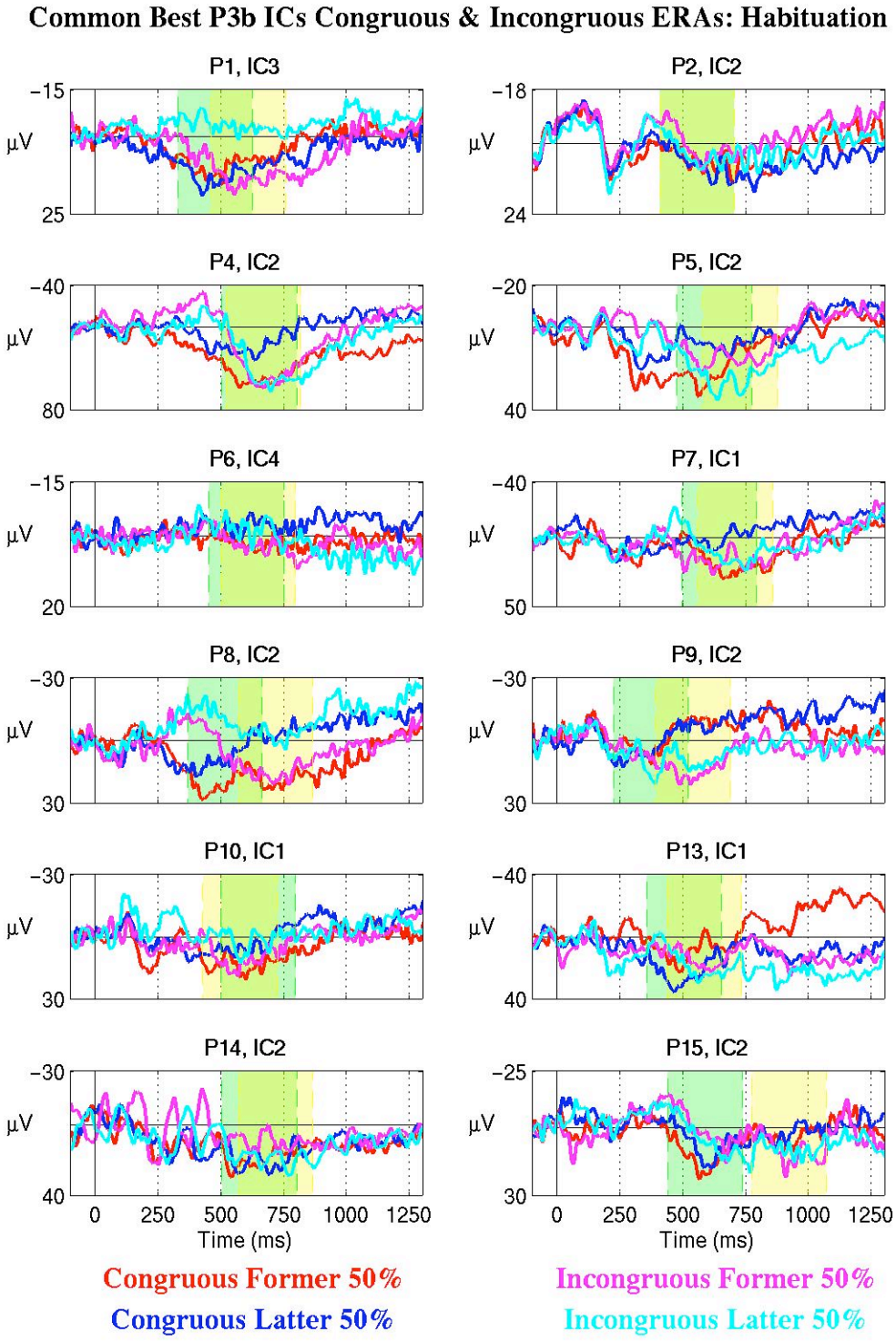

Figure VI.41 – Average responses of the best P3b ICs that are common to oddball and sentence acceptability tasks to noun and verb sentence endings from the former and latter half of the experiment. Green and yellow shading indicate individual congruous and incongruous LPC windows, respectively. Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

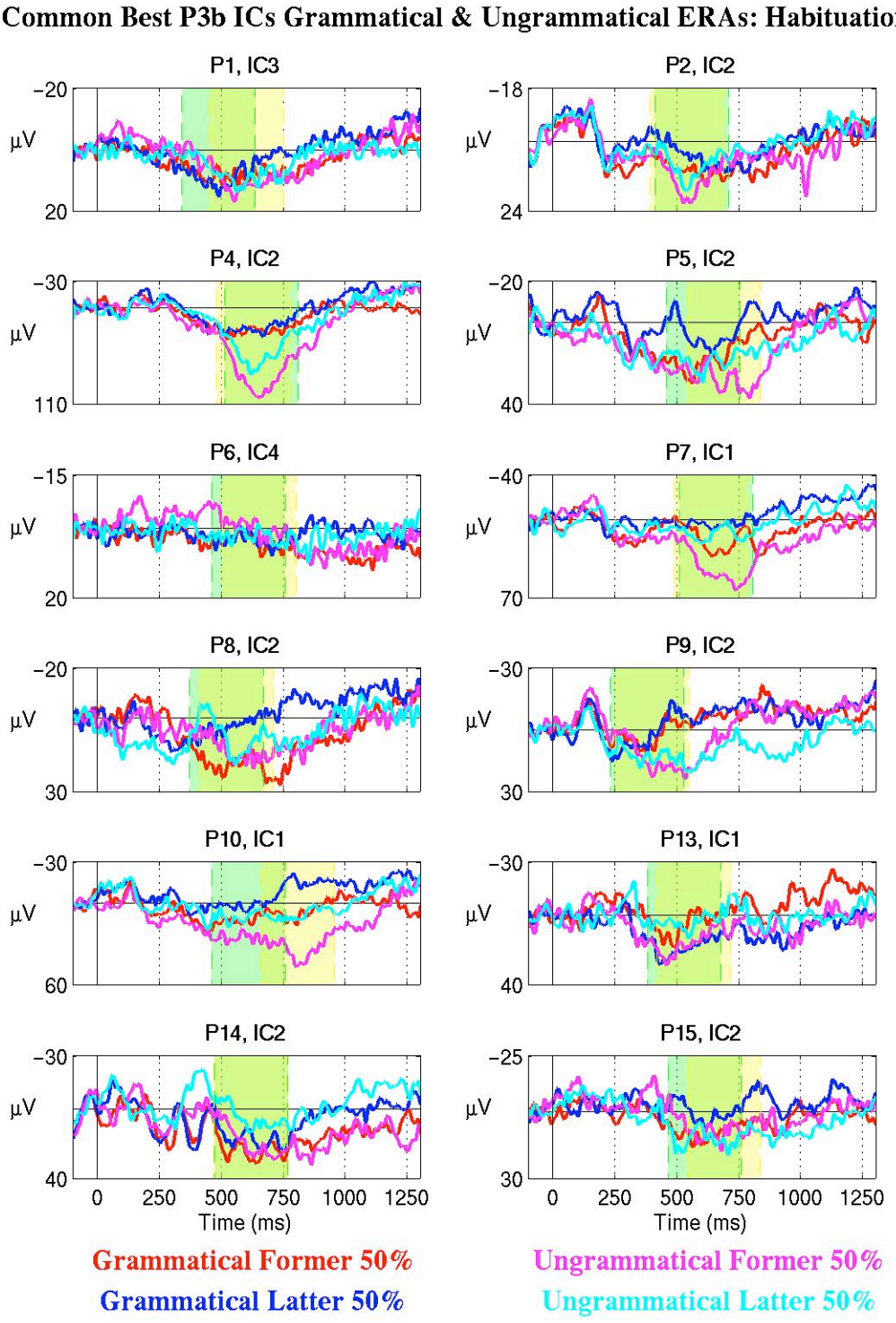

Figure VI.42 – Average responses of the best P3b ICs that are common to oddball and sentence acceptability tasks to pronoun and verb sentence endings from the former and latter half of the experiment. Green and yellow shading indicate individual grammatical and ungrammatical LPC windows, respectively. Activations are computed after normalizing IC topographies to unit length. IC polarity is as shown in Figure VI.26. Abbreviations: P=Participant, IC=Independent Component.

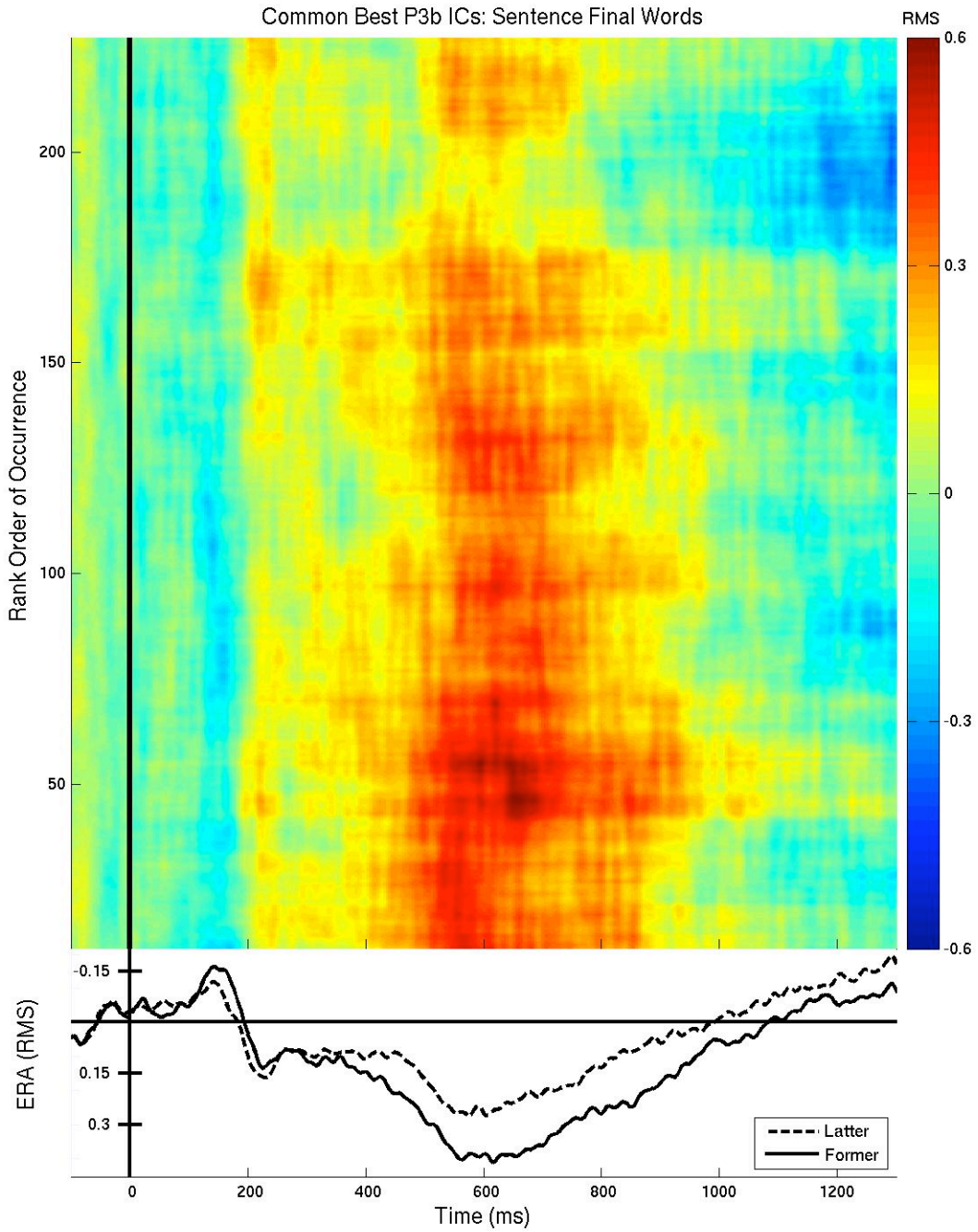

Figure VI.43 – **(Top)** 2845 single trials of the 12 common best P3b IC responses to sentence final words in the sentence acceptability task. Each IC's activation has been normalized to unit root mean square amplitude and their polarities are as shown in Figure VI.26. Trials are sorted by order of occurrence (1 indicates the earliest trial), stacked along the y-axis, and smoothed using a 250 trial vertical moving average. **(Bottom)** Averages (ERAs) of the former and latter 50% of the single trials.

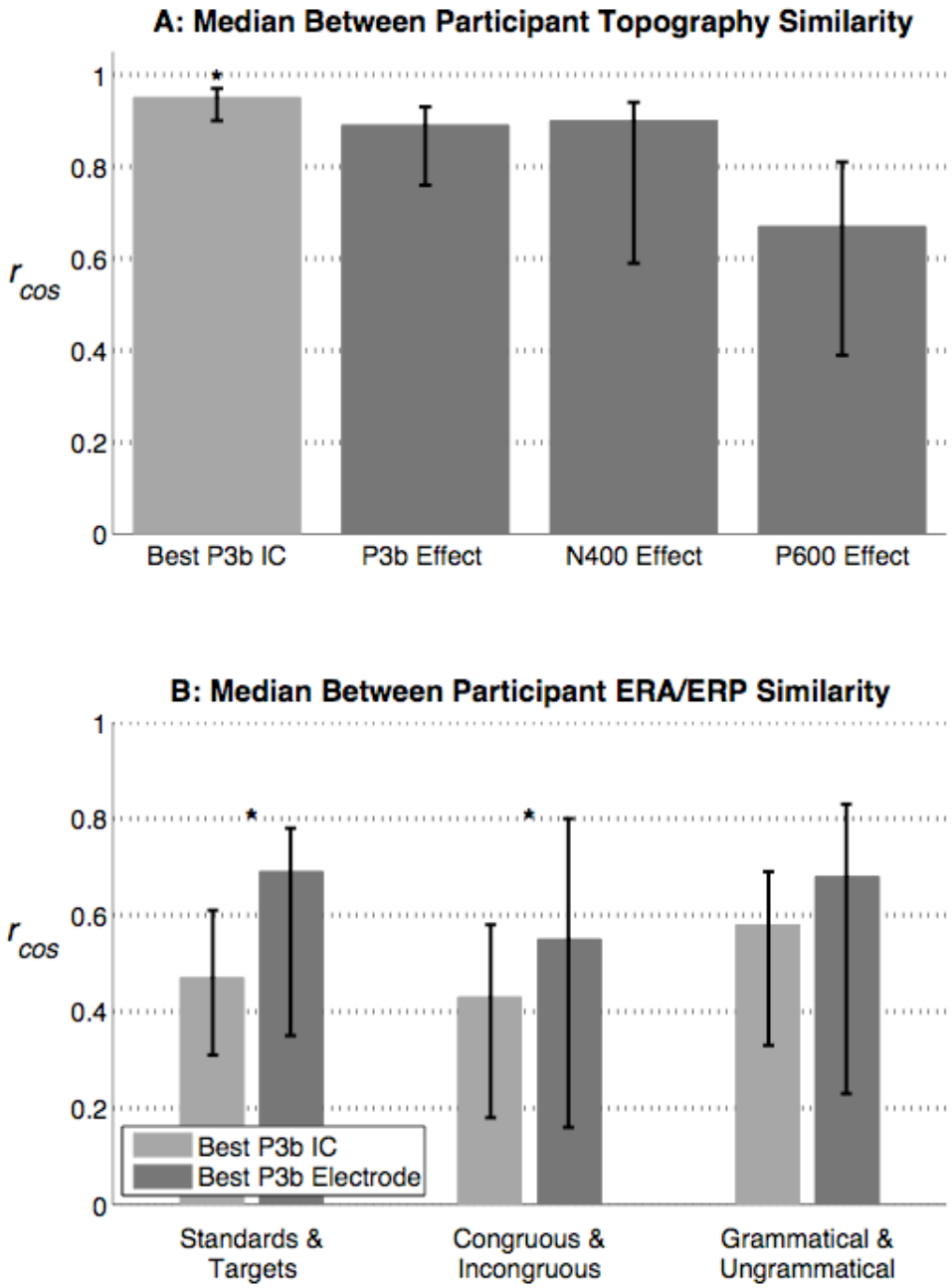

Figure VI.44 – (**A**) Median topography similarity of the best P3b IC and deviance ERP effects of the 12 participants whose best P3b ICs are common to the oddball and sentence acceptability tasks. Star indicates a significant difference between ICs and each of the three ERP effects (two-tailed sign test, binomial(66,.5)<=19, p<.0008). (**B**) Median similarity of the best P3b IC event-related activation and ERPs at the best P3b electrode of the same 12 participants. Stars indicate significant differences (two-tailed sign test, binomial(66,.5)>=44, p<.01). For both graphs, error bars indicate 25th and 75th percentiles.

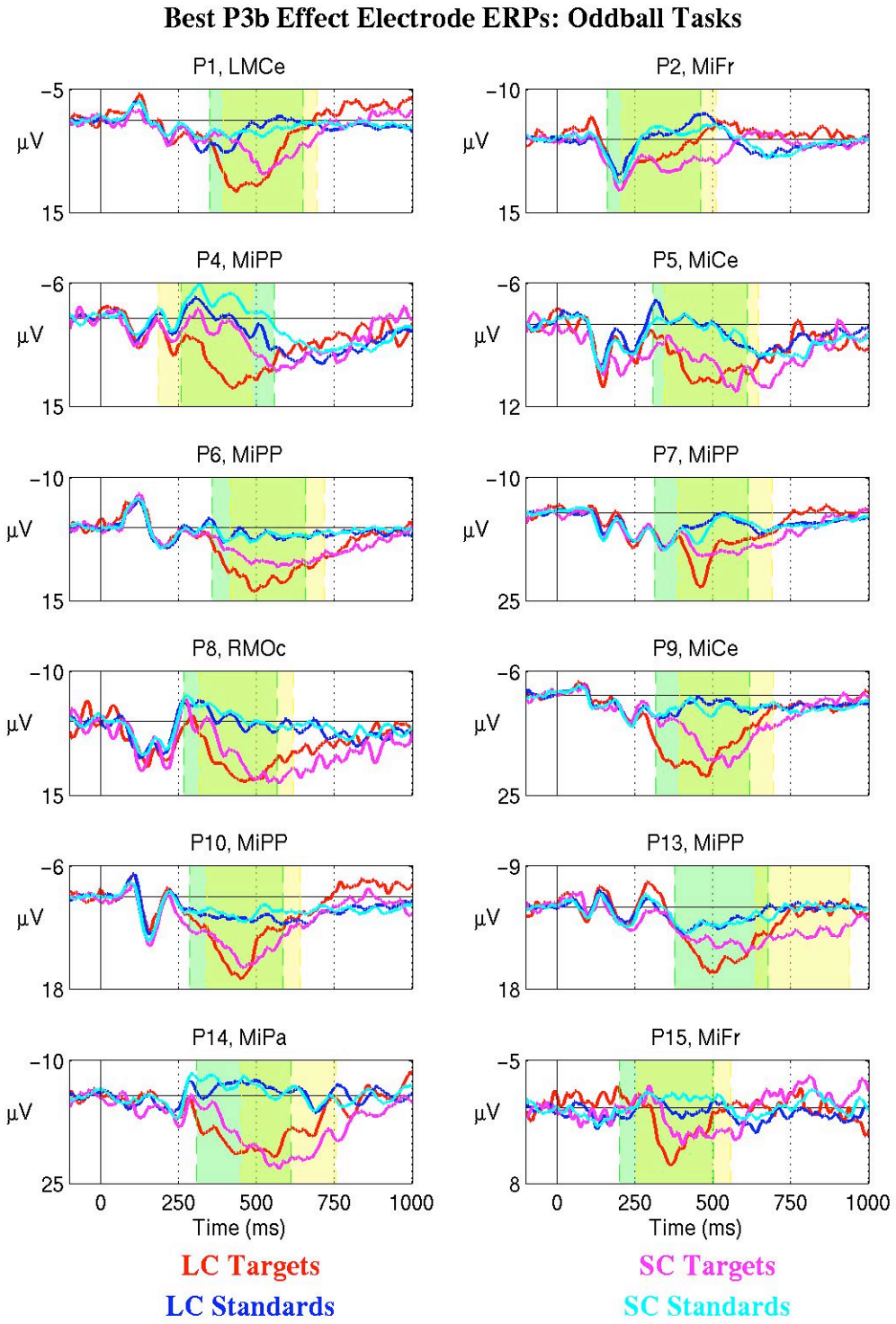

Figure VI.45 – Oddball task ERPs from each participant's "best P3b" electrode, which is the electrode that accounts for the most variance of the participant's P3b effect (averaged across both oddball tasks). Green and yellow shading indicate a participant's letter case and semantic category task P3b windows, respectively. Abbreviations: P=Participant. See Appendix 1 (Section IV.I) for electrode abbreviations.

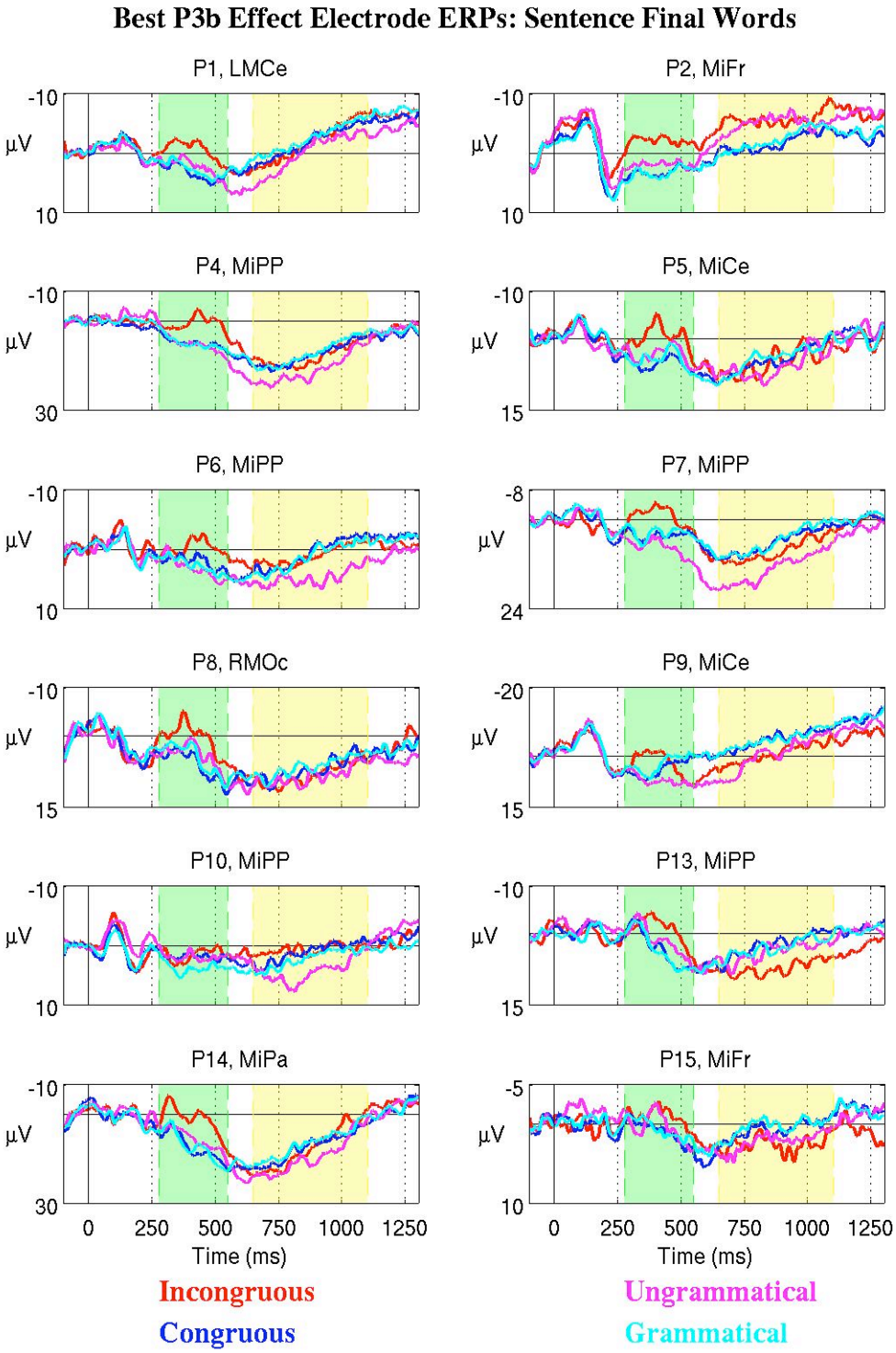

Figure VI.46 – ERPs from each participant's "best P3b" electrode (see Figure VI.45) to sentence final words in the sentence acceptability task. Green and yellow shading indicate a participant's N400 and P600 windows, respectively. Abbreviations: P=Participant. See Appendix 1 (Section IV.I) for electrode abbreviations.

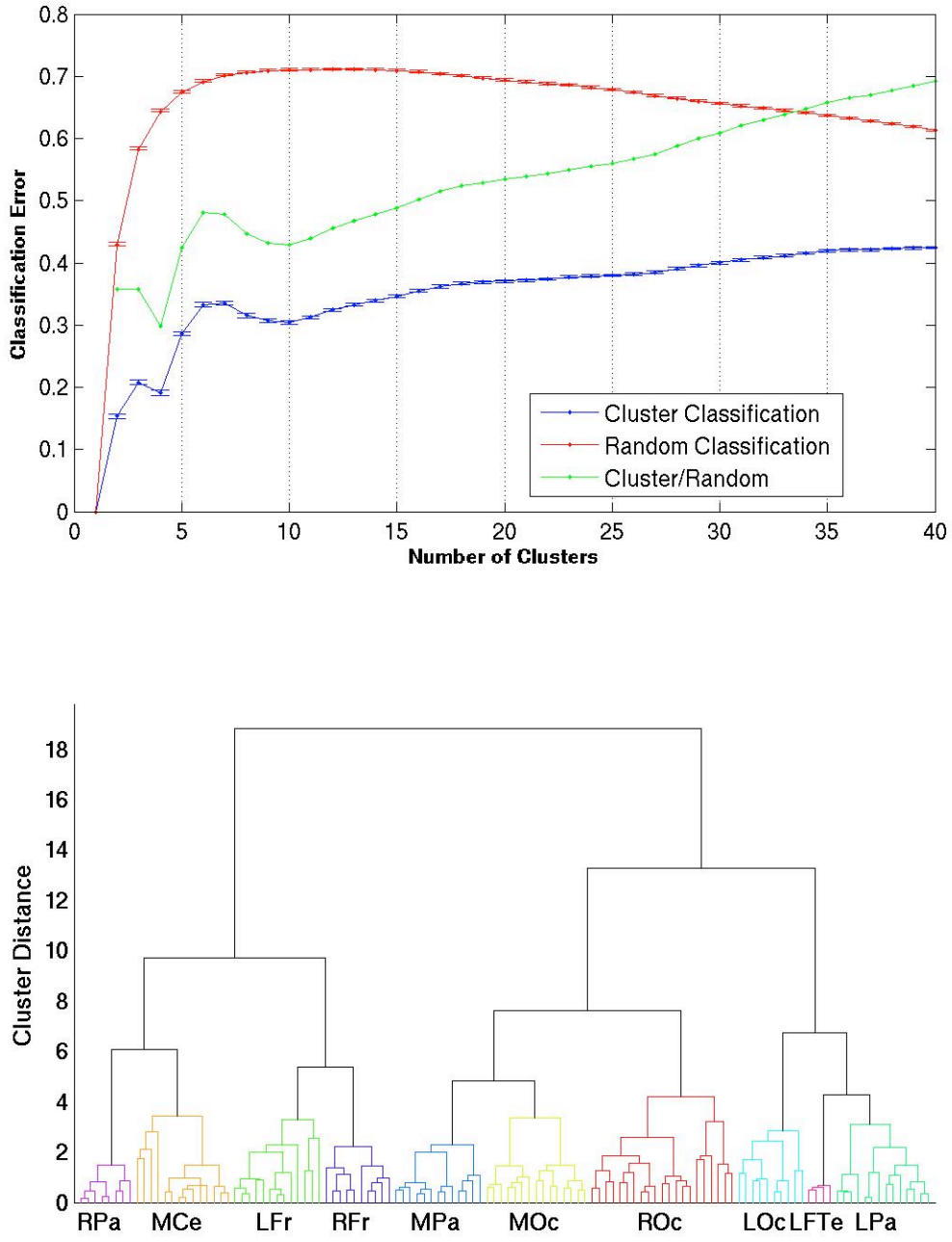

Figure VI.47 – (**Top**) Mean classification error for classifiers built from clustered halves of all ICs (error bars are standard err). (**Bottom**) Dendrogram illustrating the selected ten clusters (RPa=Right Parietal, MCe=Midline Central, LFr=Left Frontal, RFr=Right Frontal, MPa=Midline Parietal, MOc=Midline Occipital, ROc=Right Occipital, LOc=Left Occipital, LFTe=Left Fronto-Temporal, LPa=Left Parietal).

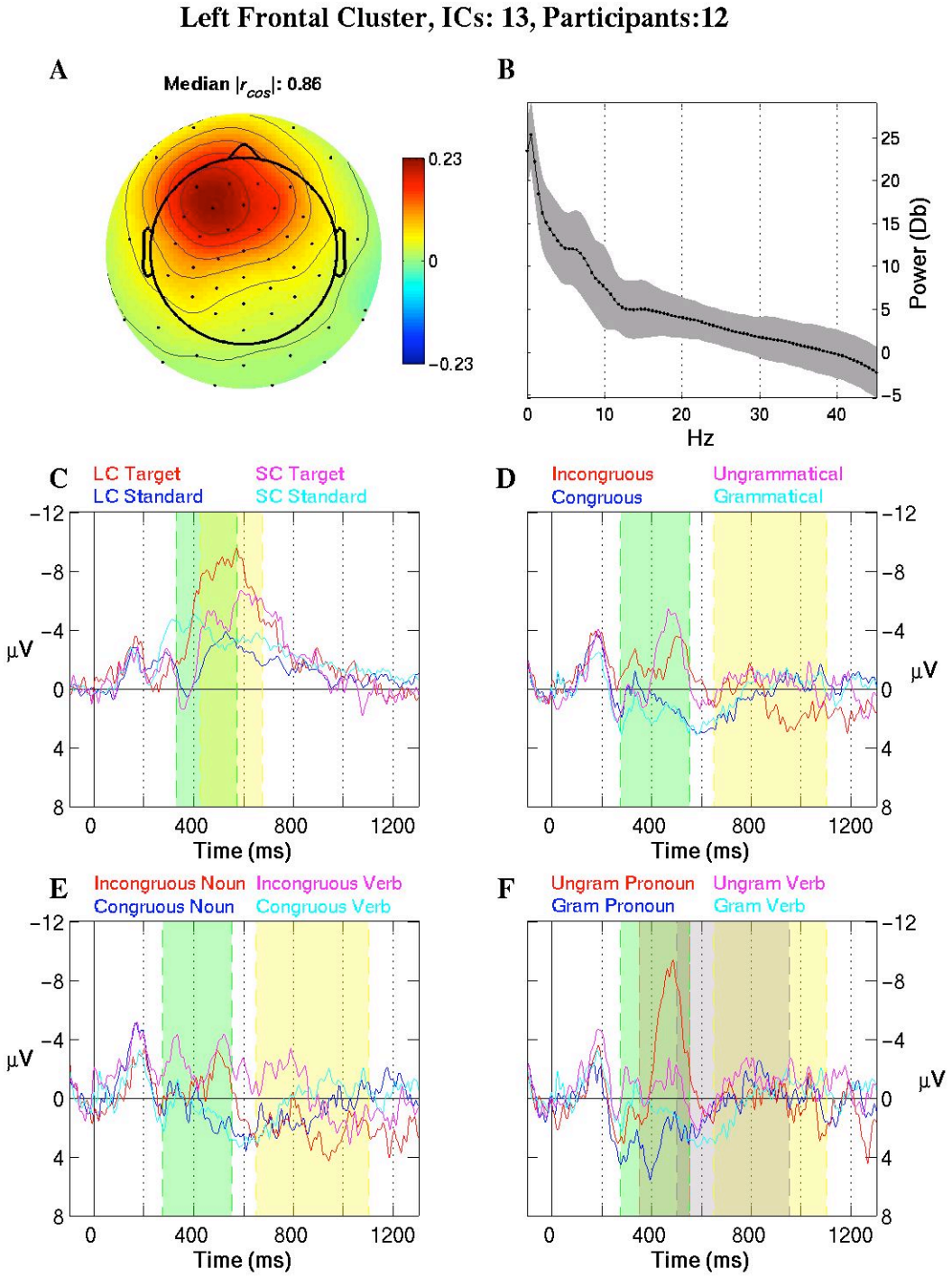

Figure VI.48 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

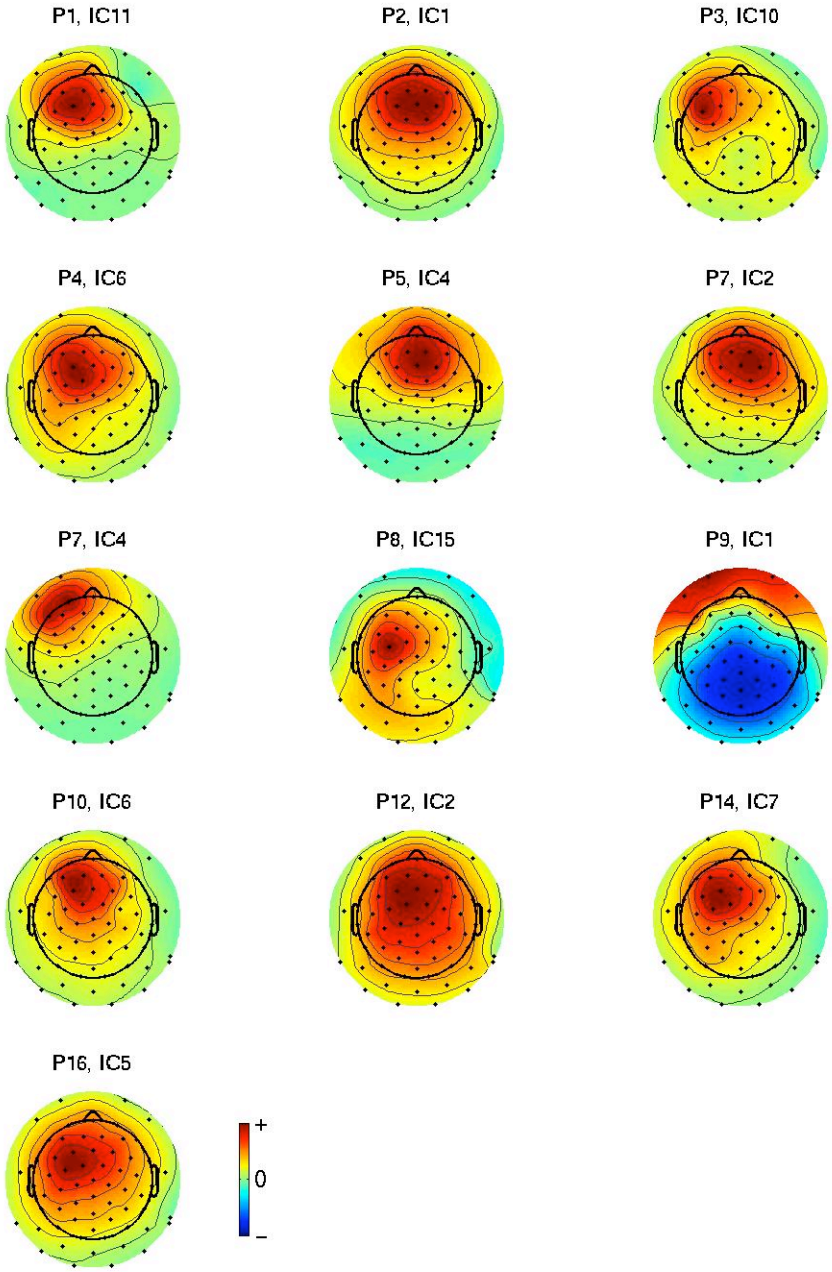

Figure VI.49 – Topographies of members of the left frontal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

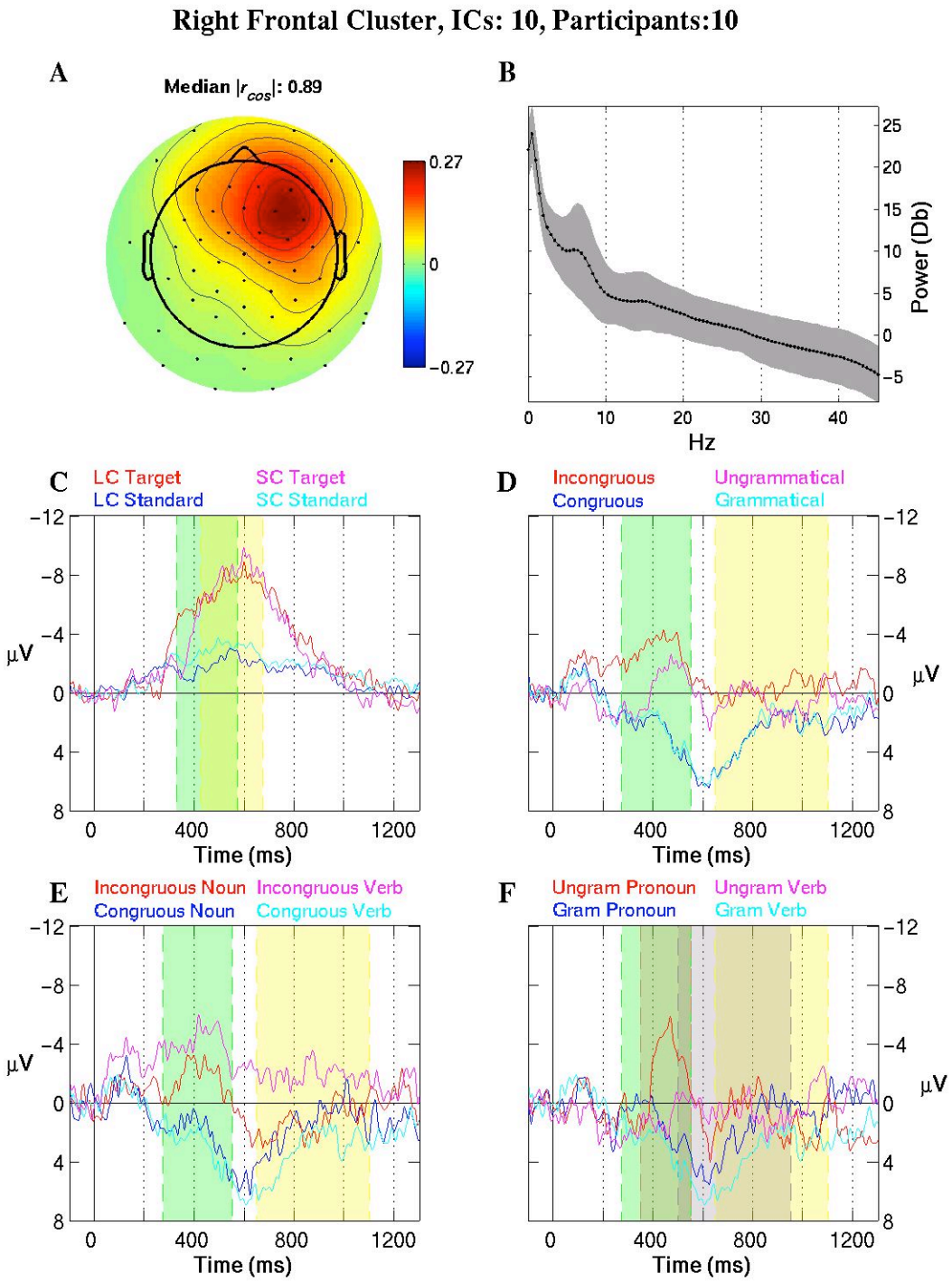

Figure VI.50 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

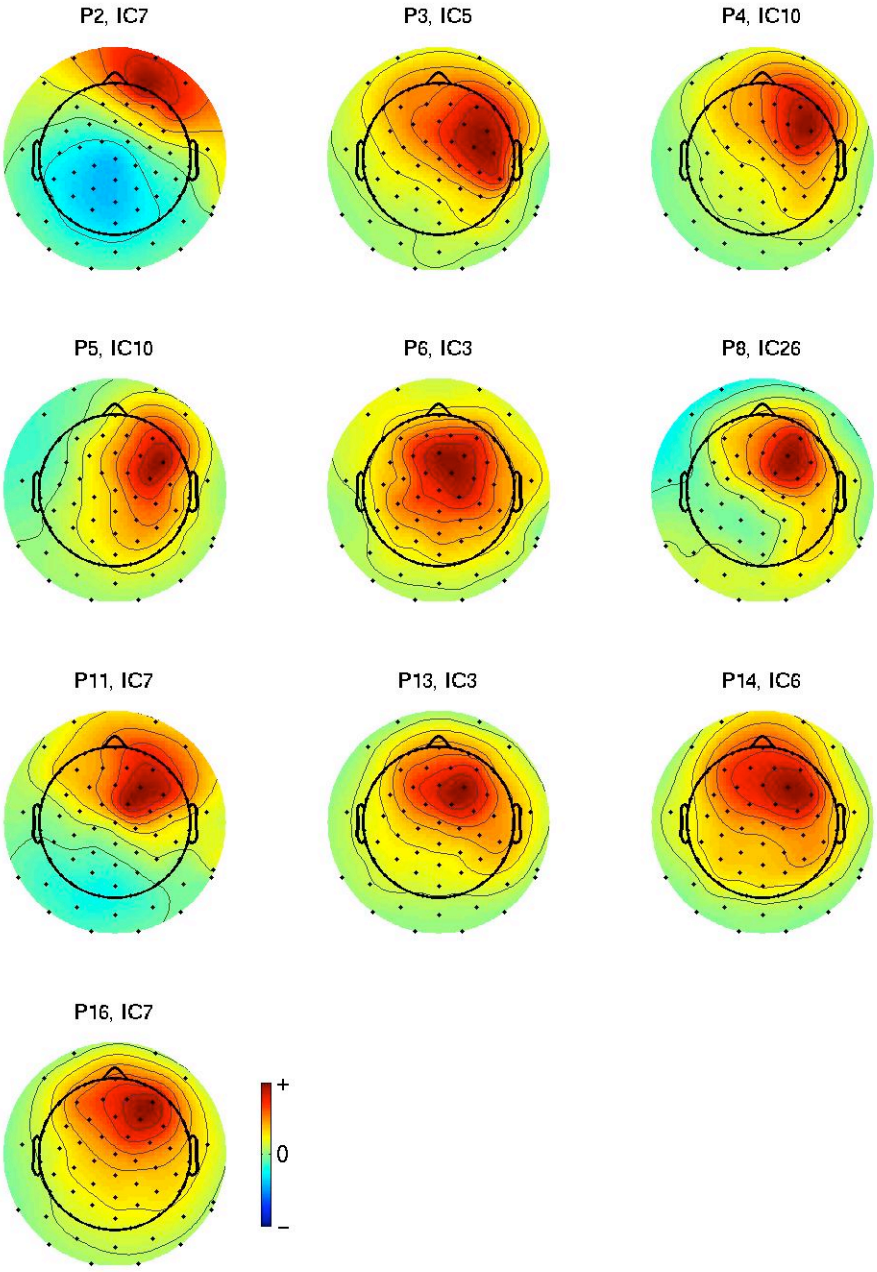

Figure VI.51 – Topographies of members of the right frontal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

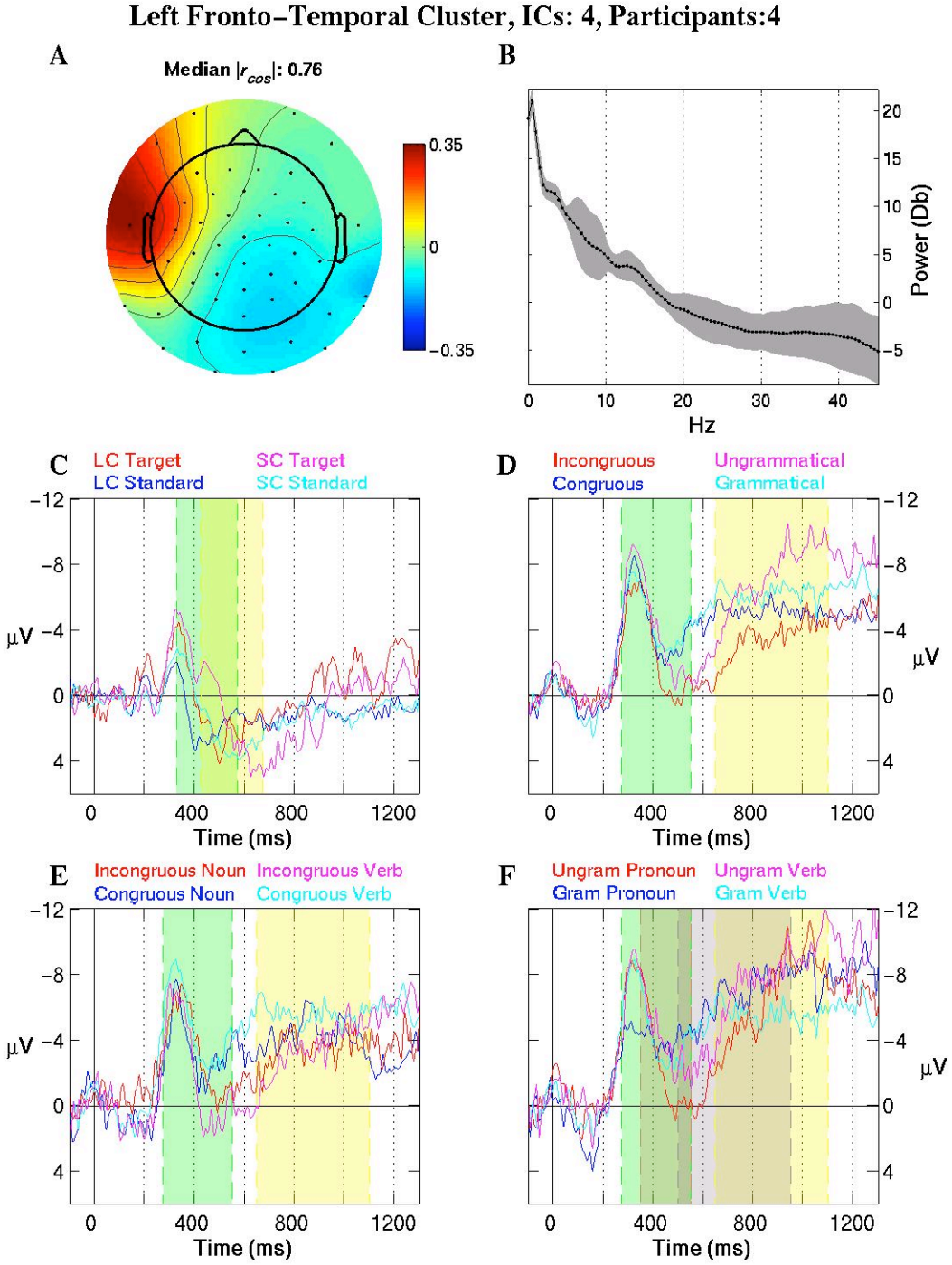

Figure VI.52 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

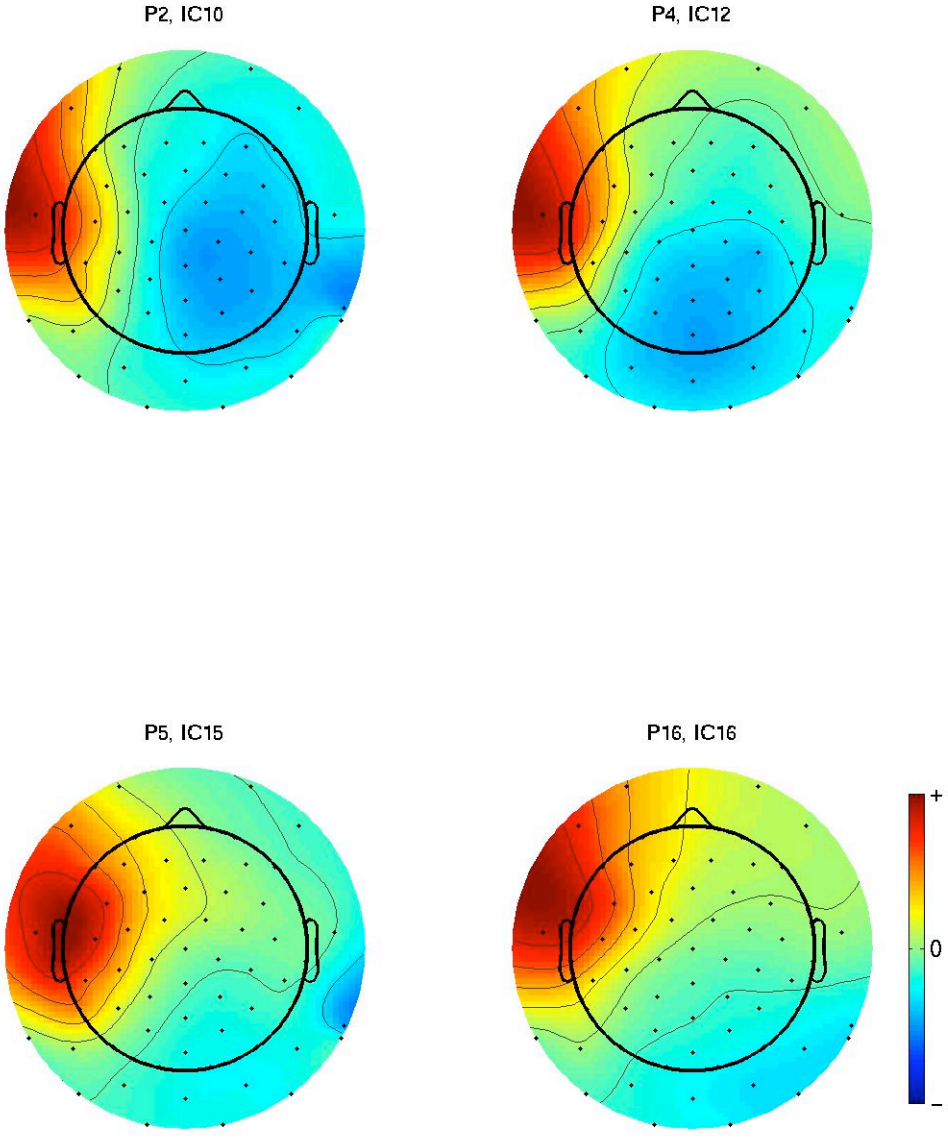

Figure VI.53 – Topographies of members of the left fronto-temporal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

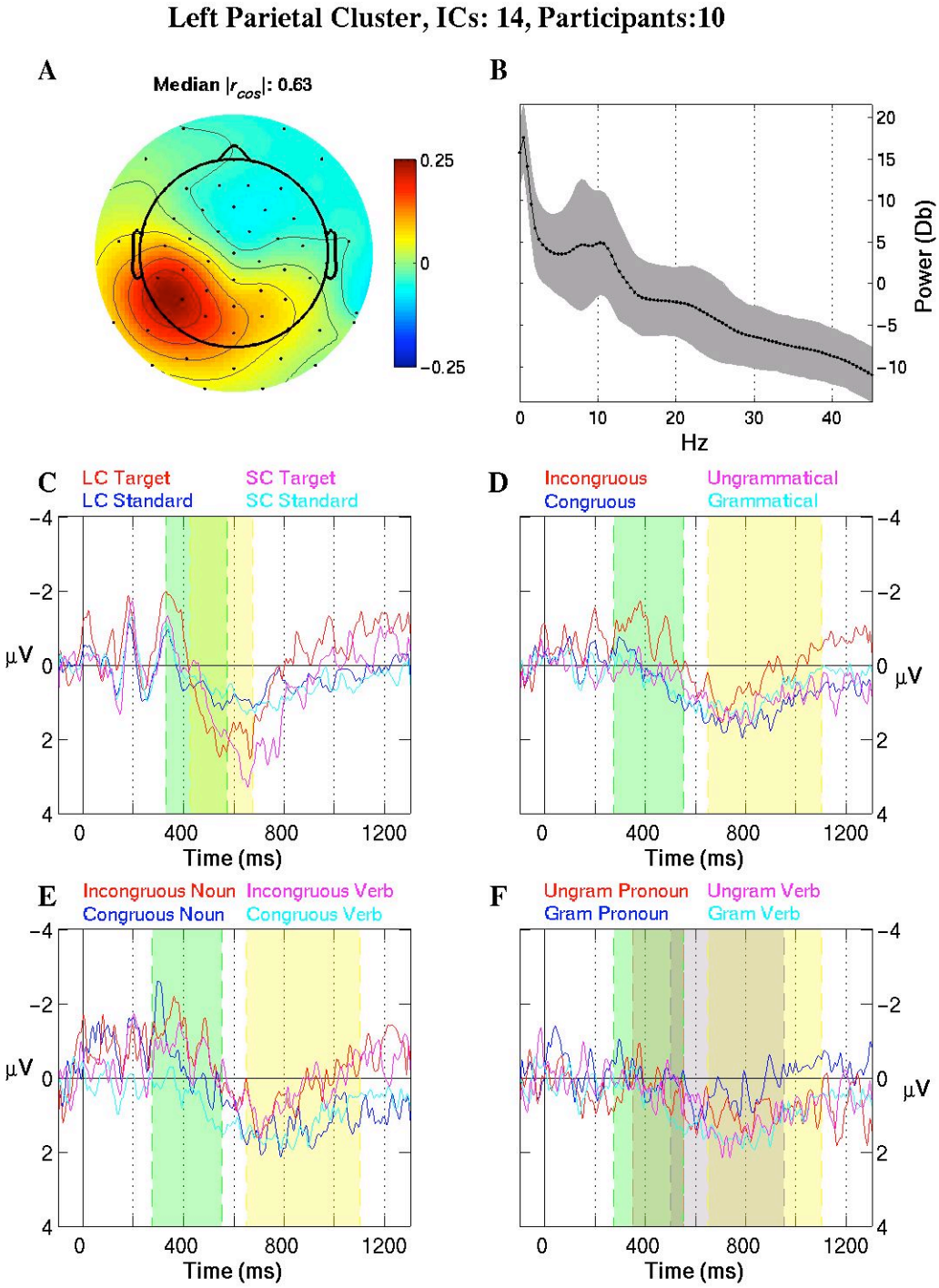

Figure VI.54 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

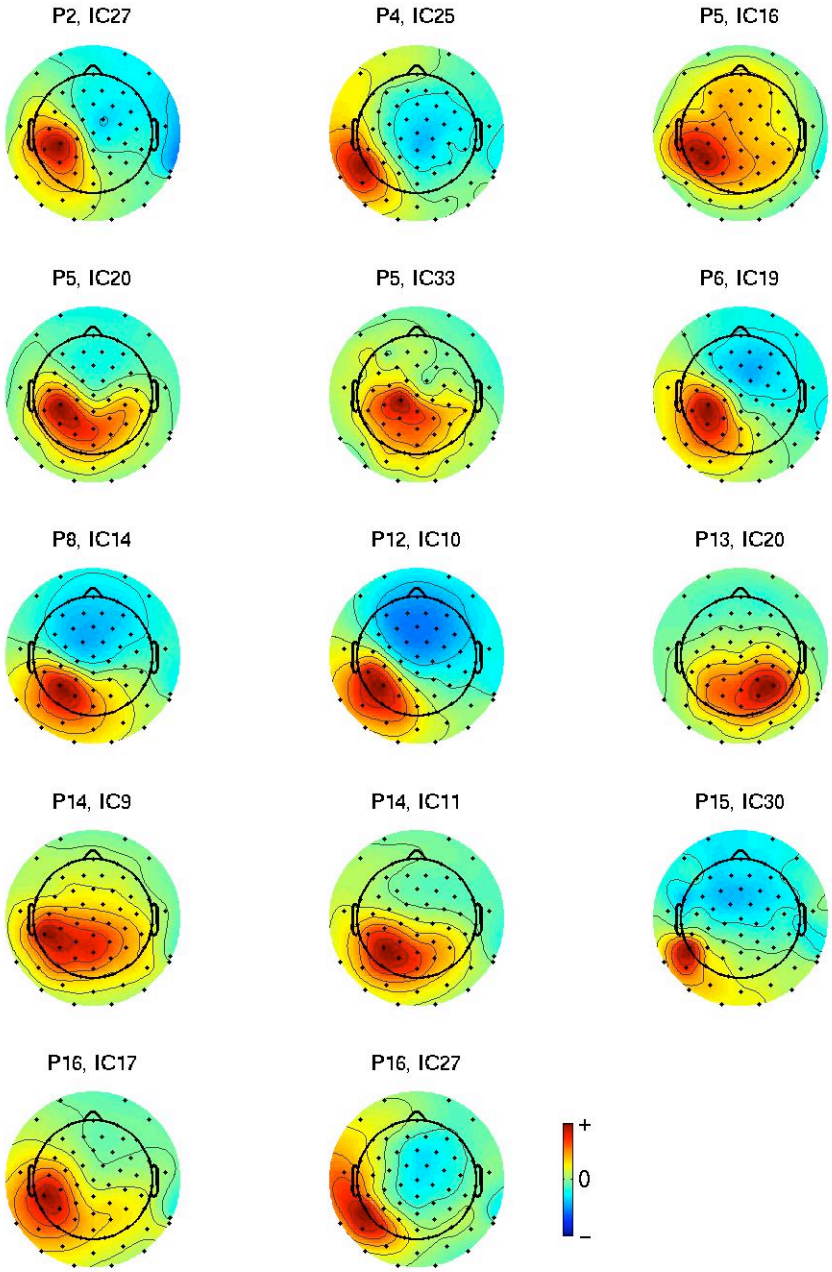

Figure VI.55 – Topographies of members of the left parietal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

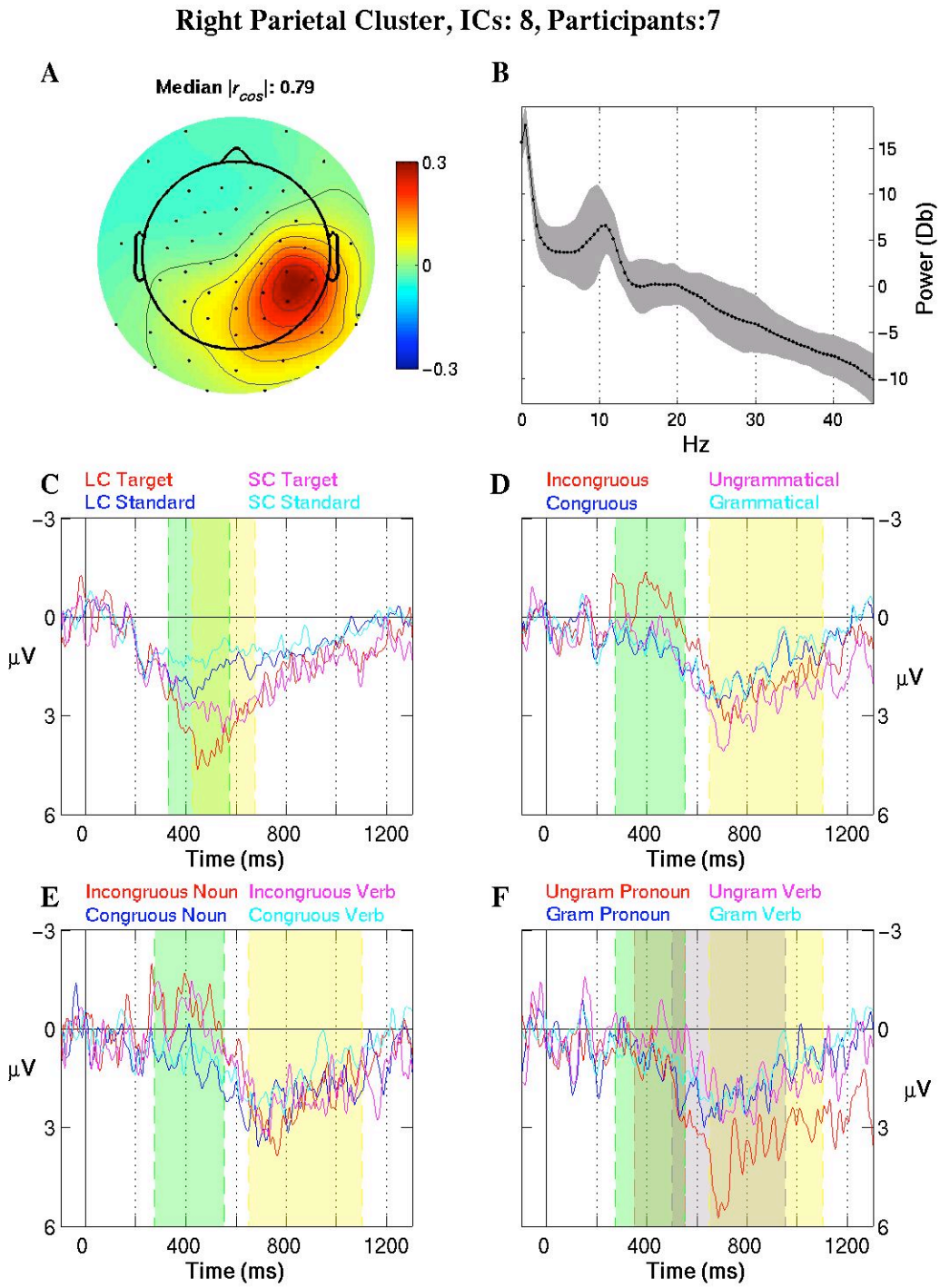

Figure VI.56 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

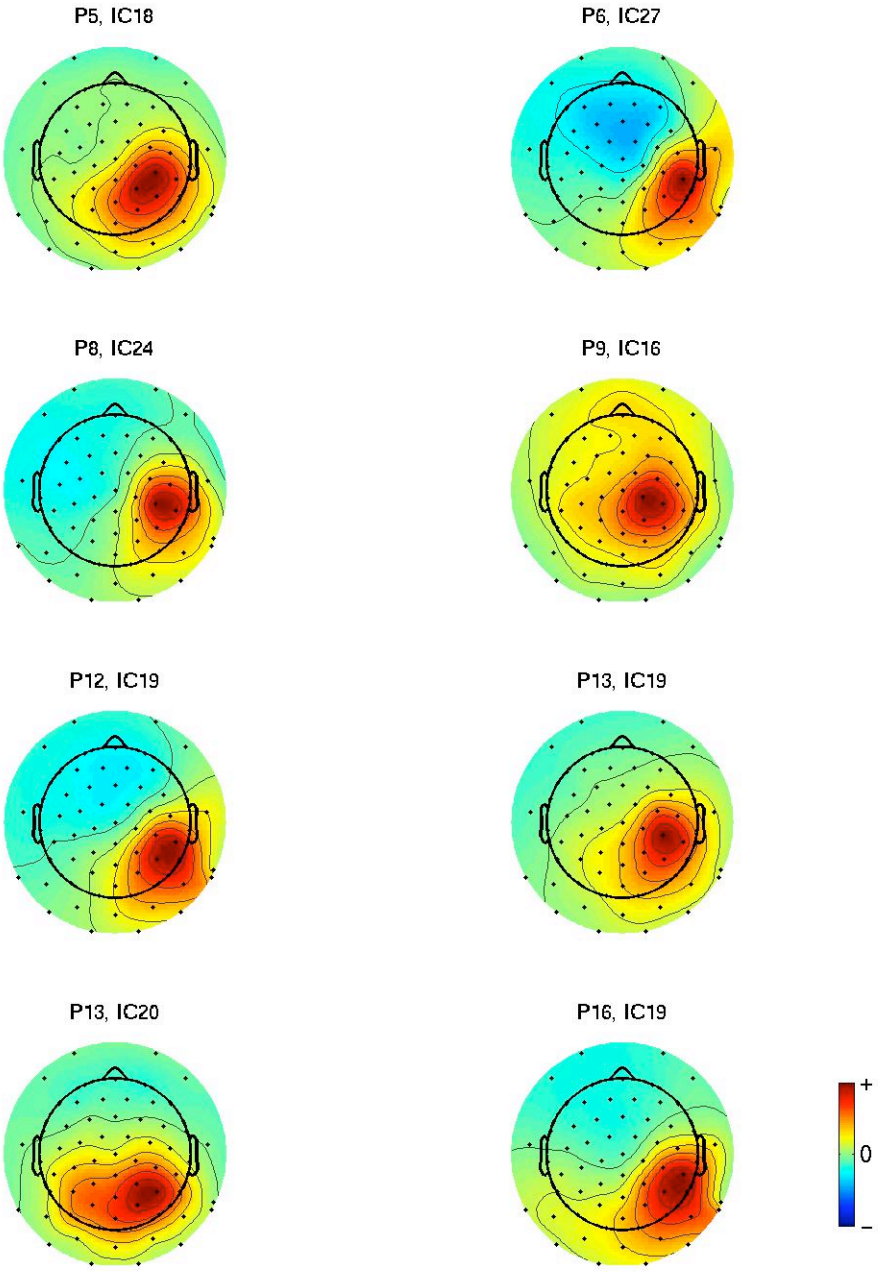

Figure VI.57 – Topographies of members of the right parietal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

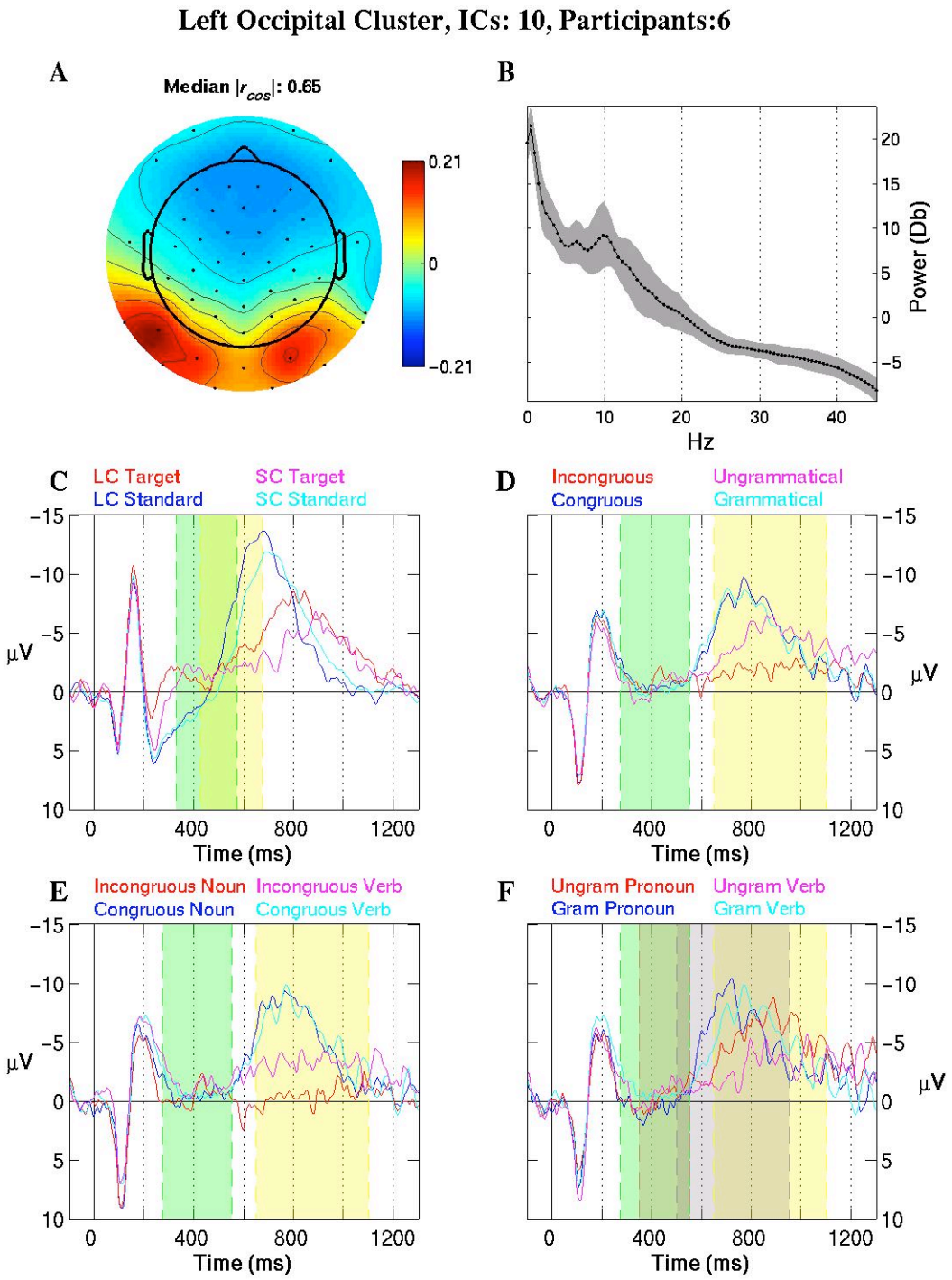

Figure VI.58 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

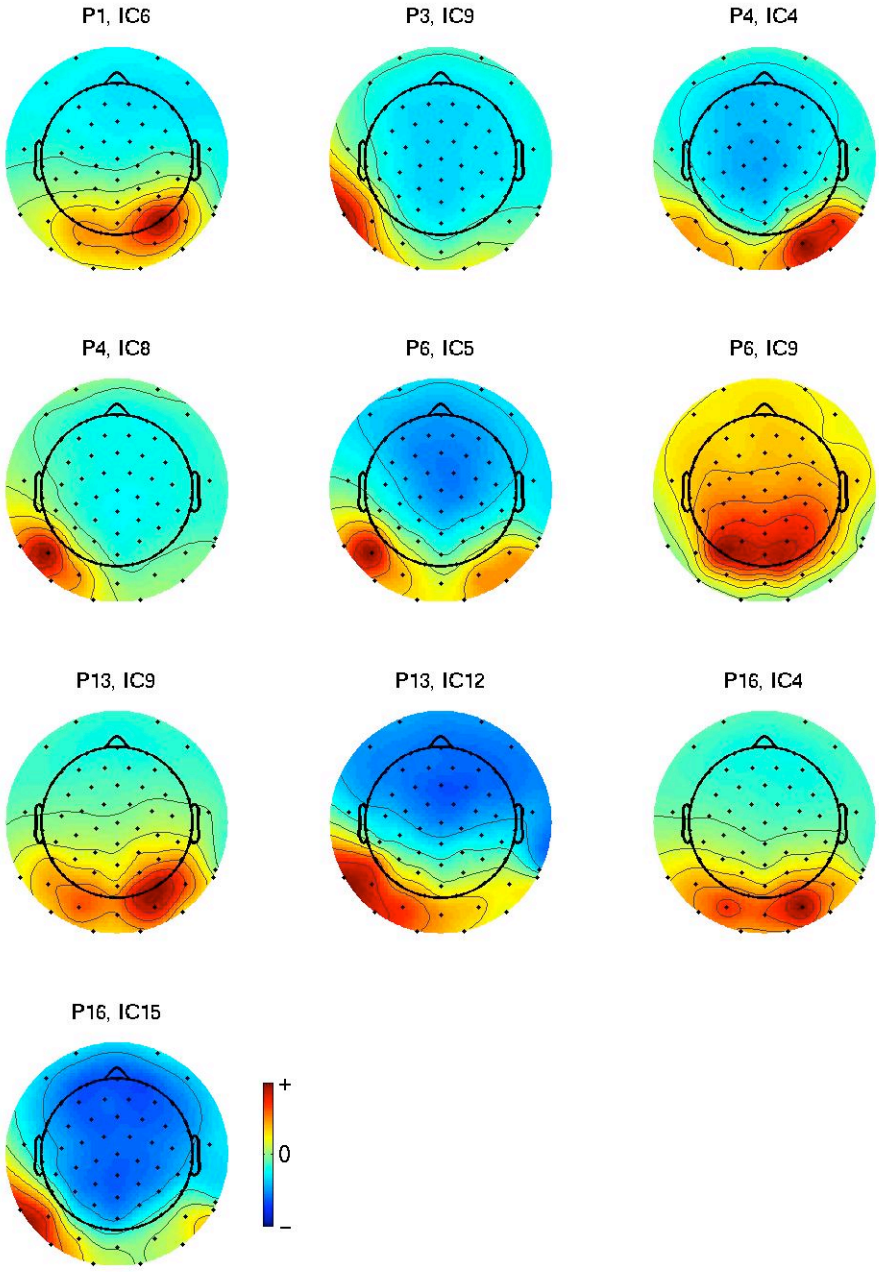

Figure VI.59 – Topographies of members of the left occipital cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

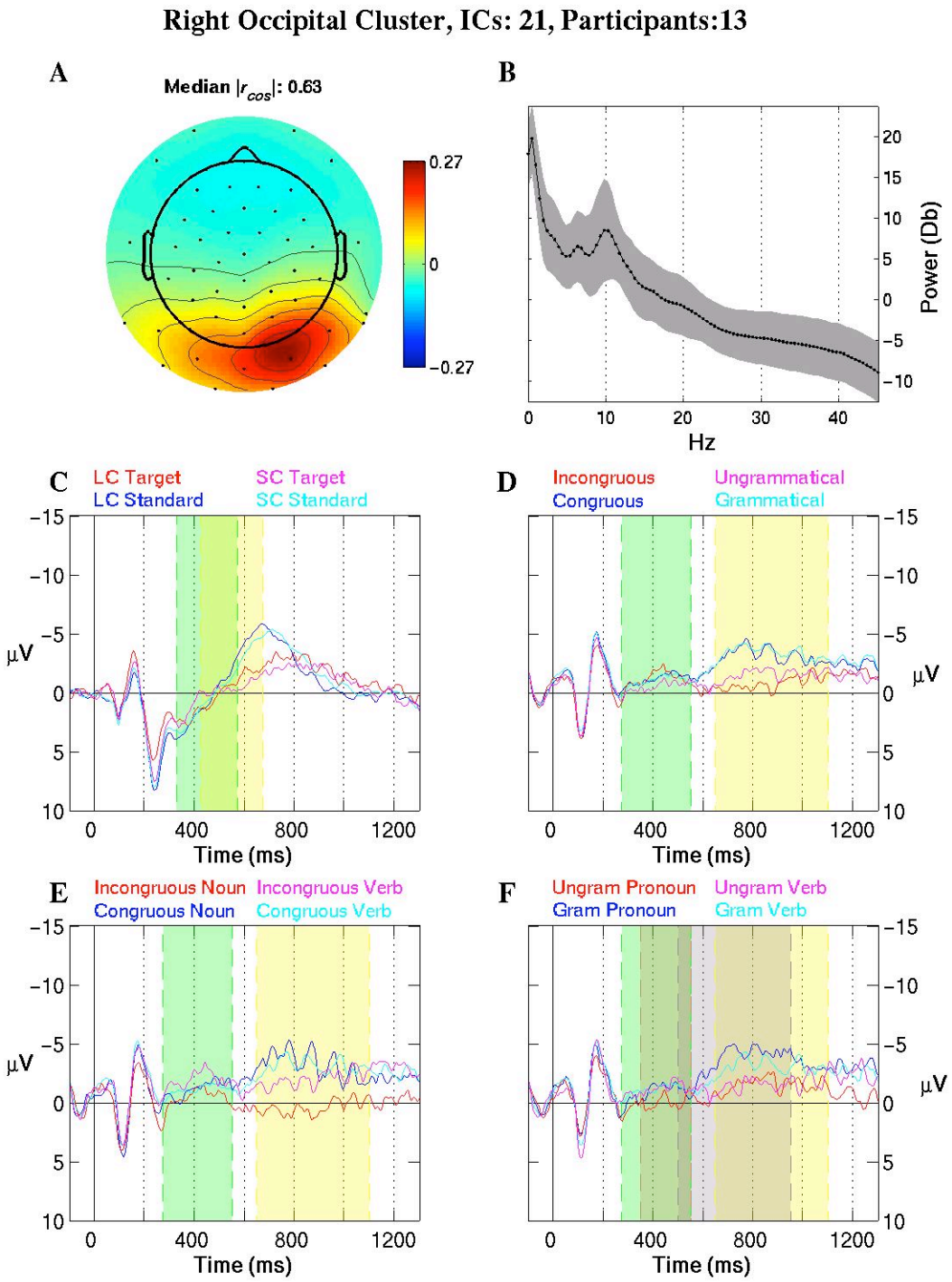

Figure VI.60 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

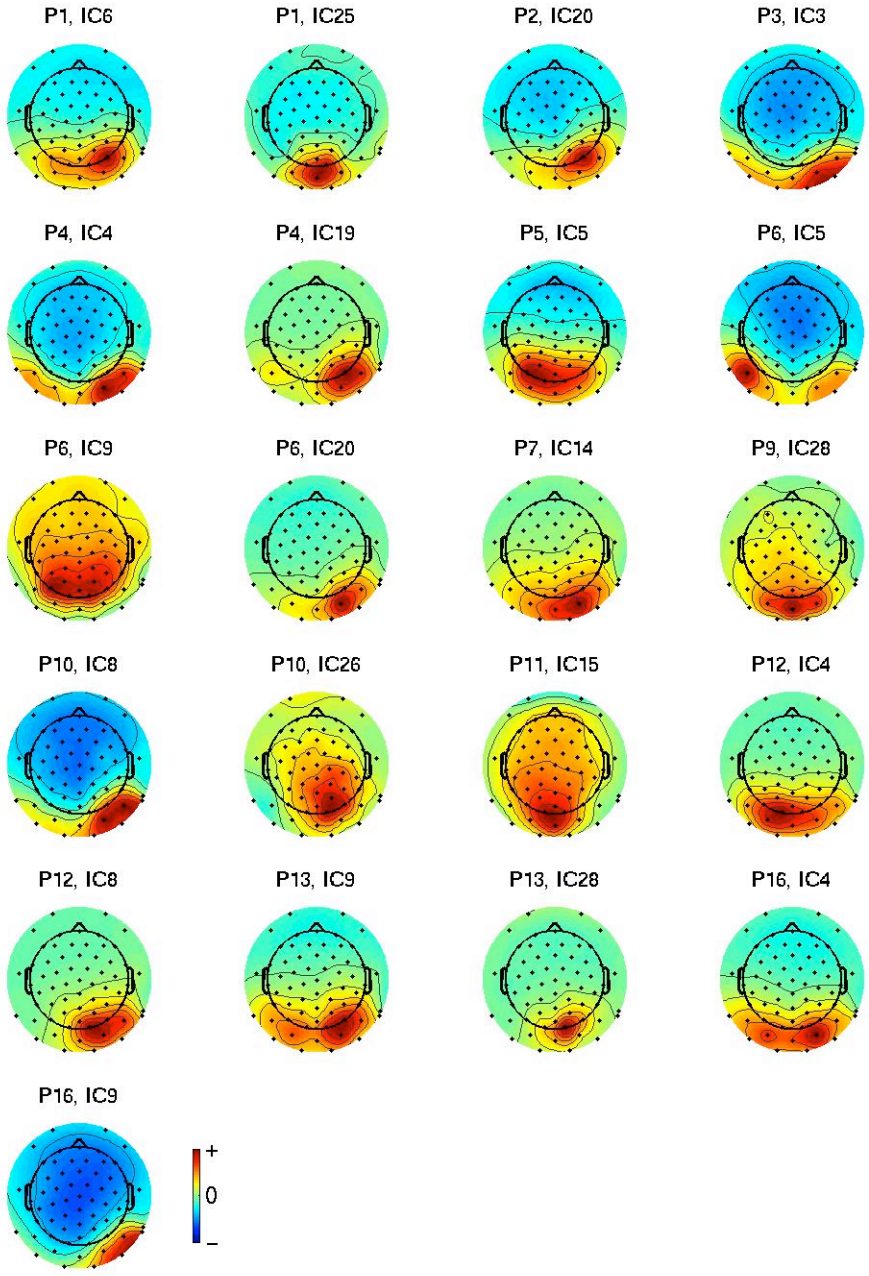

Figure VI.61 – Topographies of members of the right occipital cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

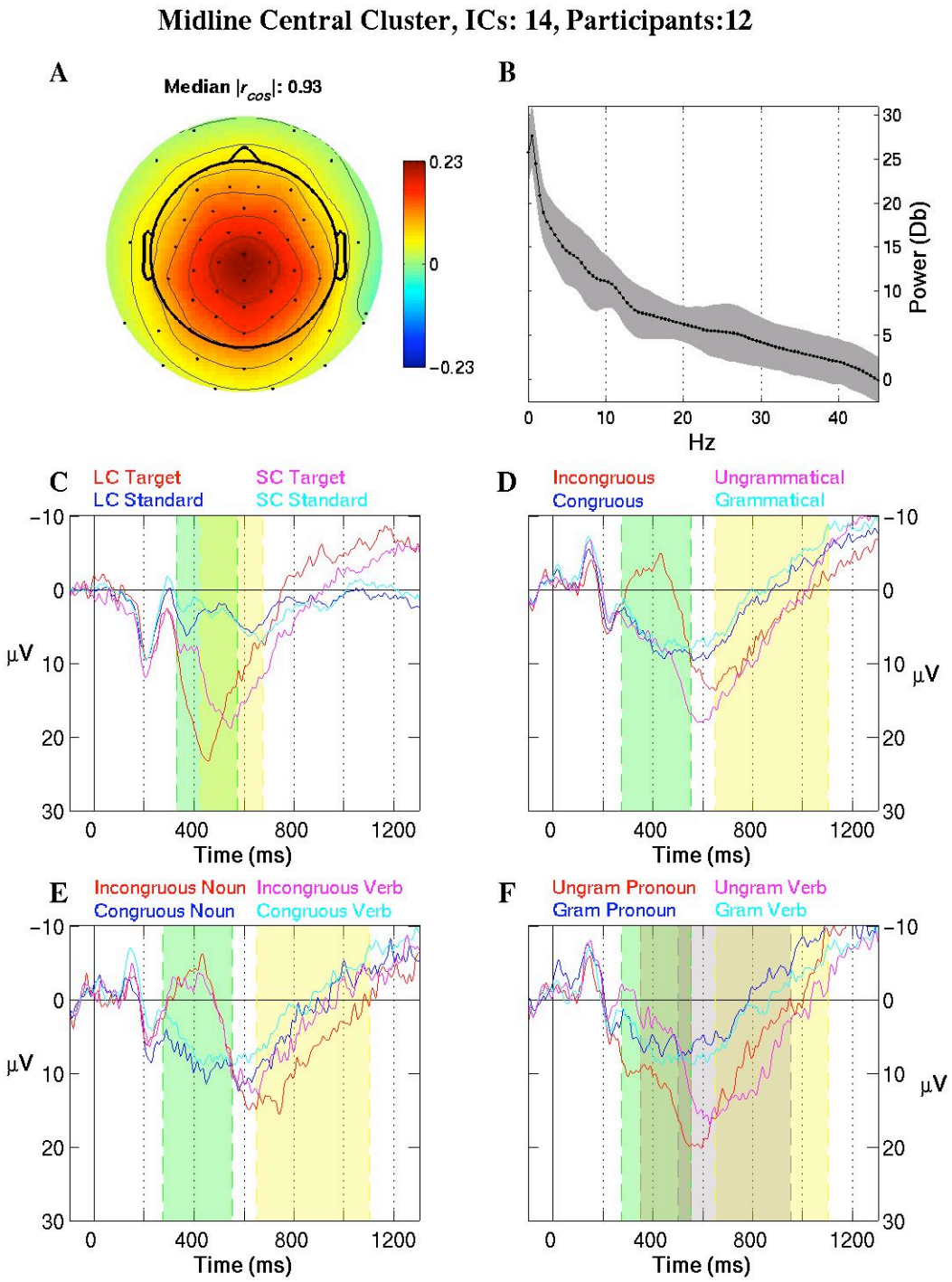

Figure VI.62 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

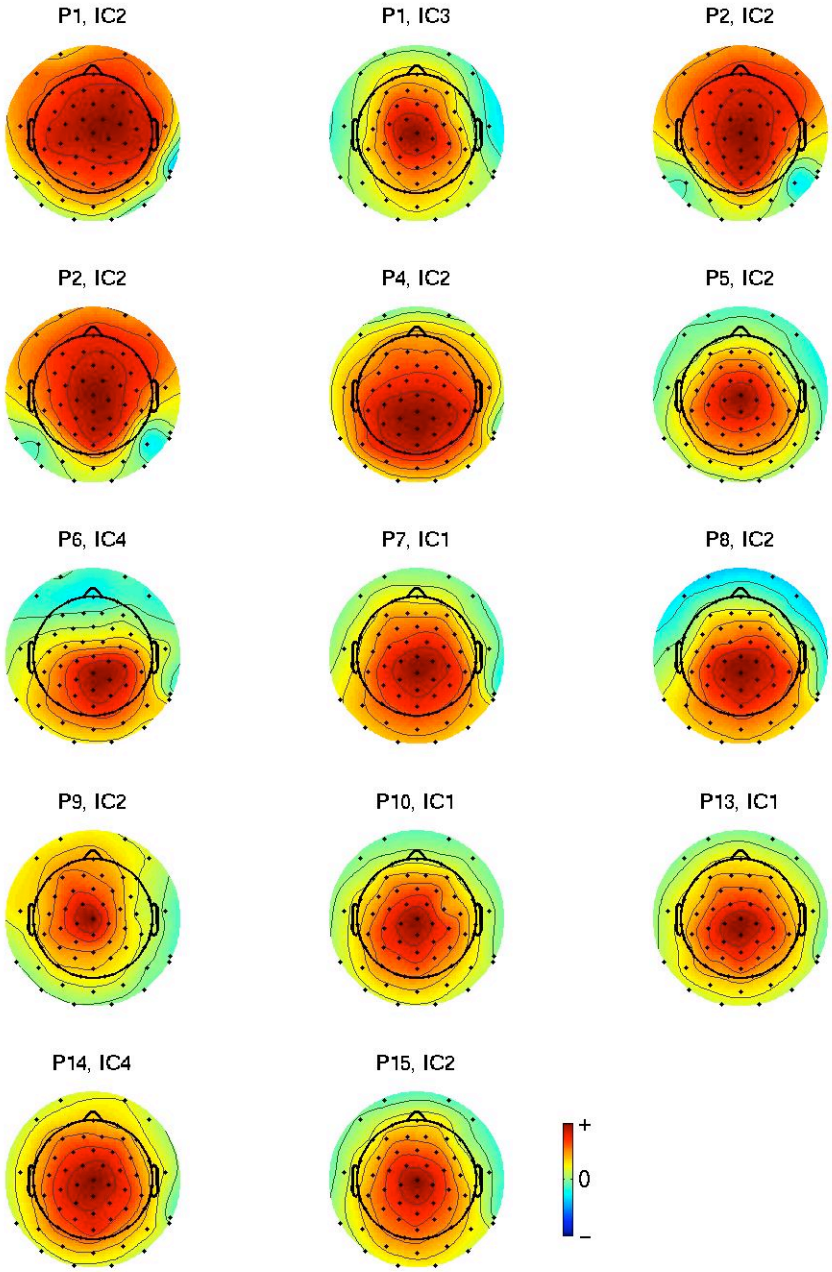

Figure VI.63 – Topographies of members of the midline central cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

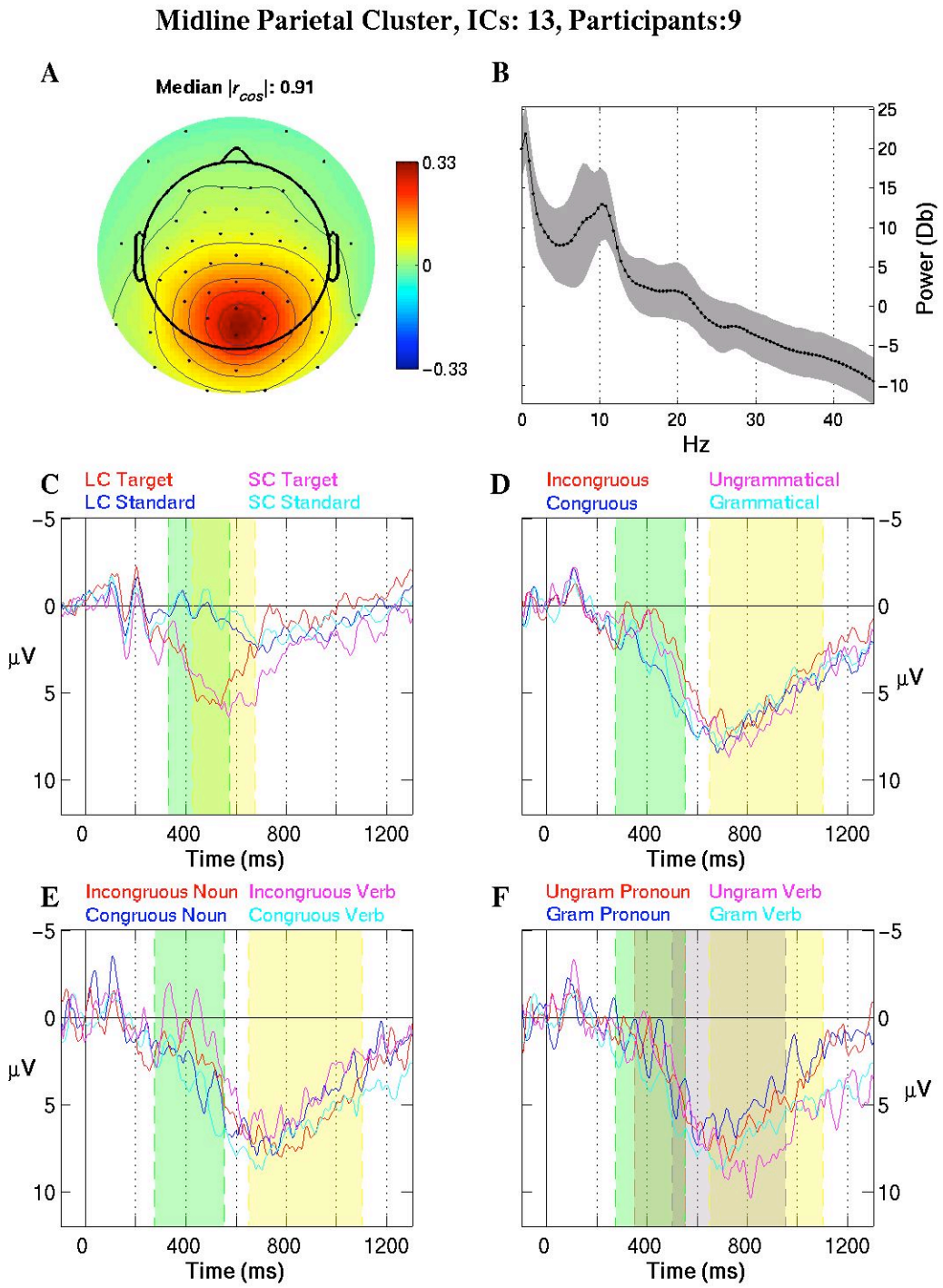

Figure VI.64 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

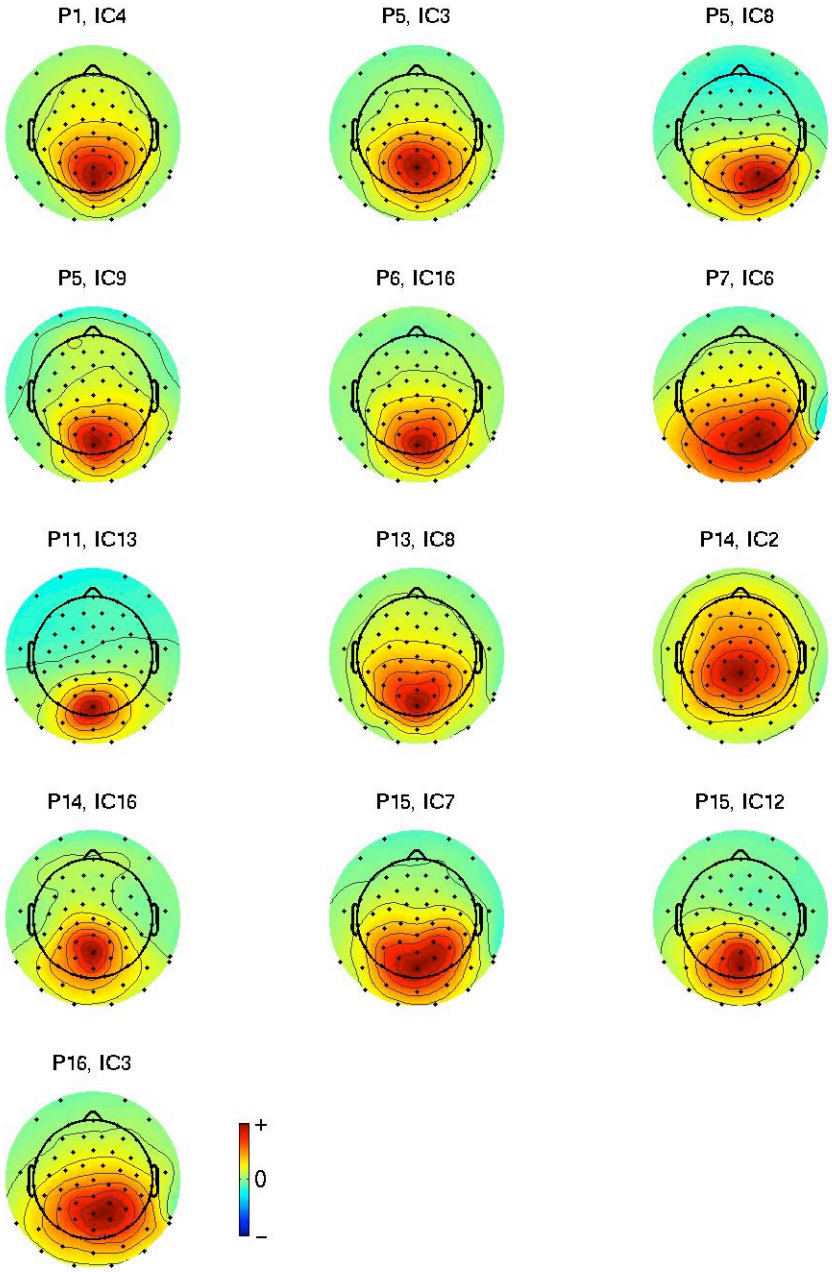

Figure VI.65 – Topographies of members of the midline parietal cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

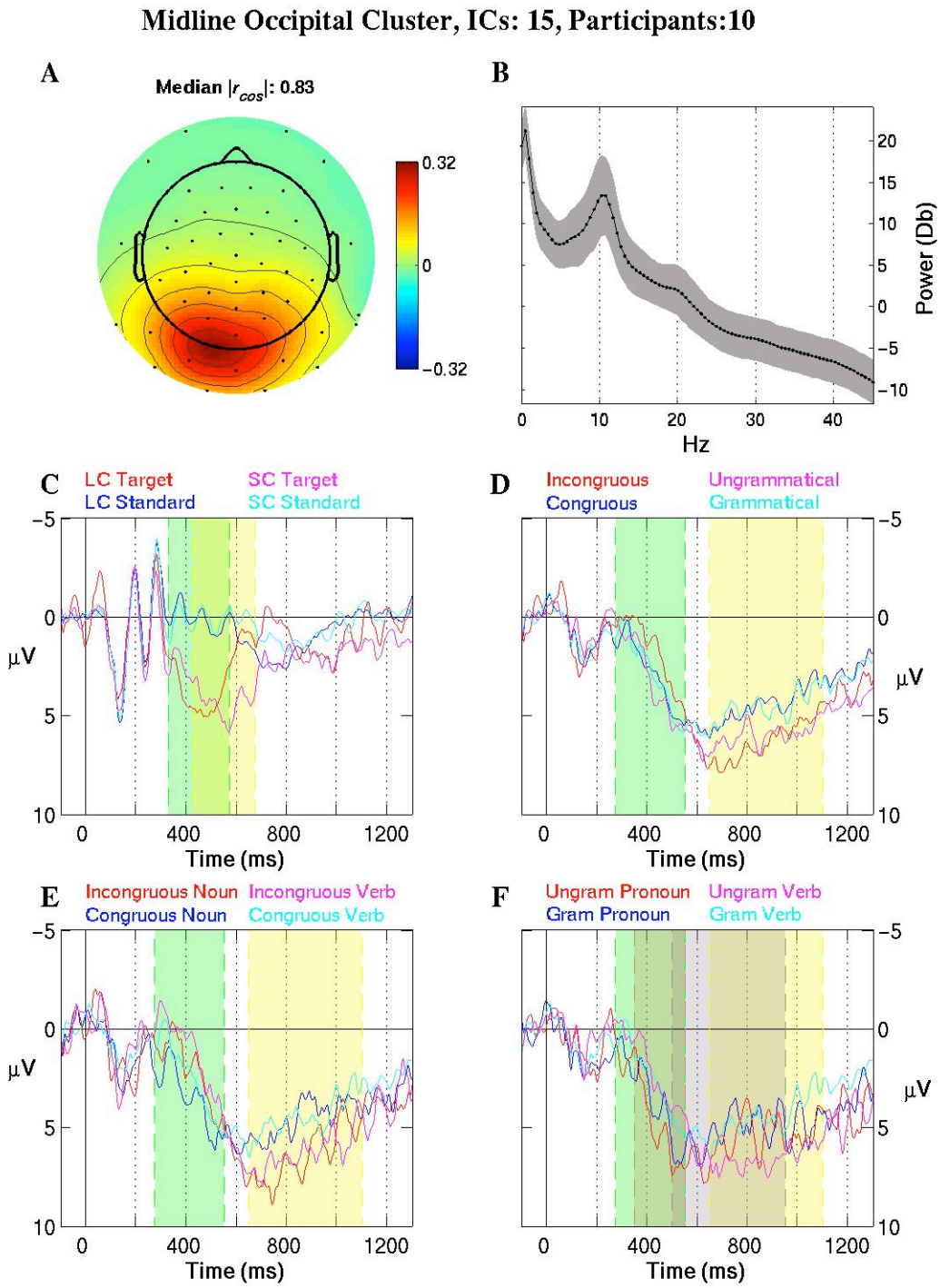

Figure VI.66 – (**A**) Mean IC topography. (**B**) Mean IC spectra (shaded region represents +/- one standard deviation). (**C**) Mean IC event-related activations (ERA) in oddball task. Green and yellow shading represent letter case and semantic category task P3b windows, respectively. (**D-F**) Mean IC ERAs from sentence acceptability task. Green, yellow, brown, and grey shading represent noun/verb N400, noun/verb P600, pronoun N400, and pronoun P600 windows, respectively.

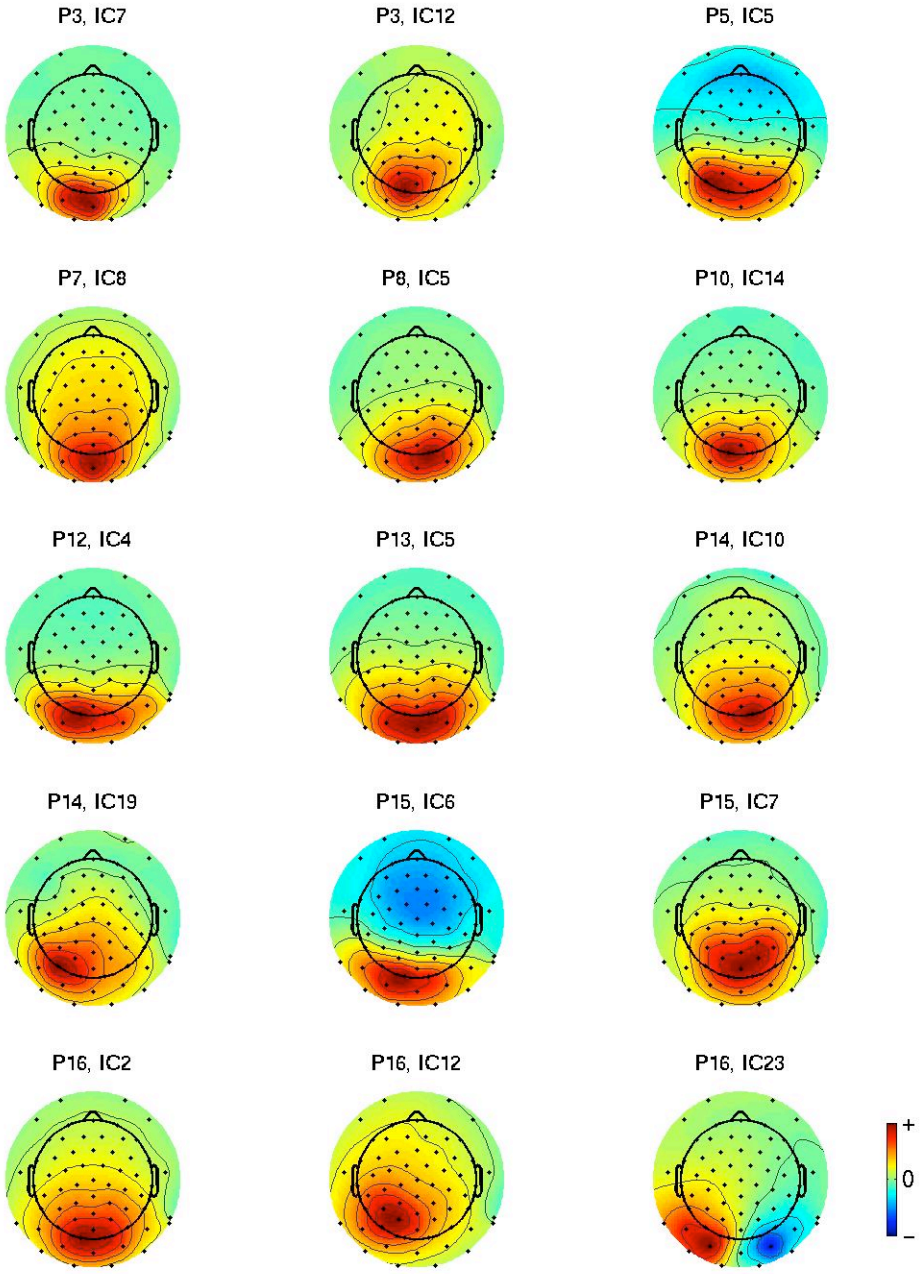

Figure VI.67 – Topographies of members of the midline occipital cluster. Unitless topography weights are represented with a color scale unique to each topography (+=maximum absolute weight, -=-maximum absolute weight). Abbreviations: P=Participant, IC=Independent Component.

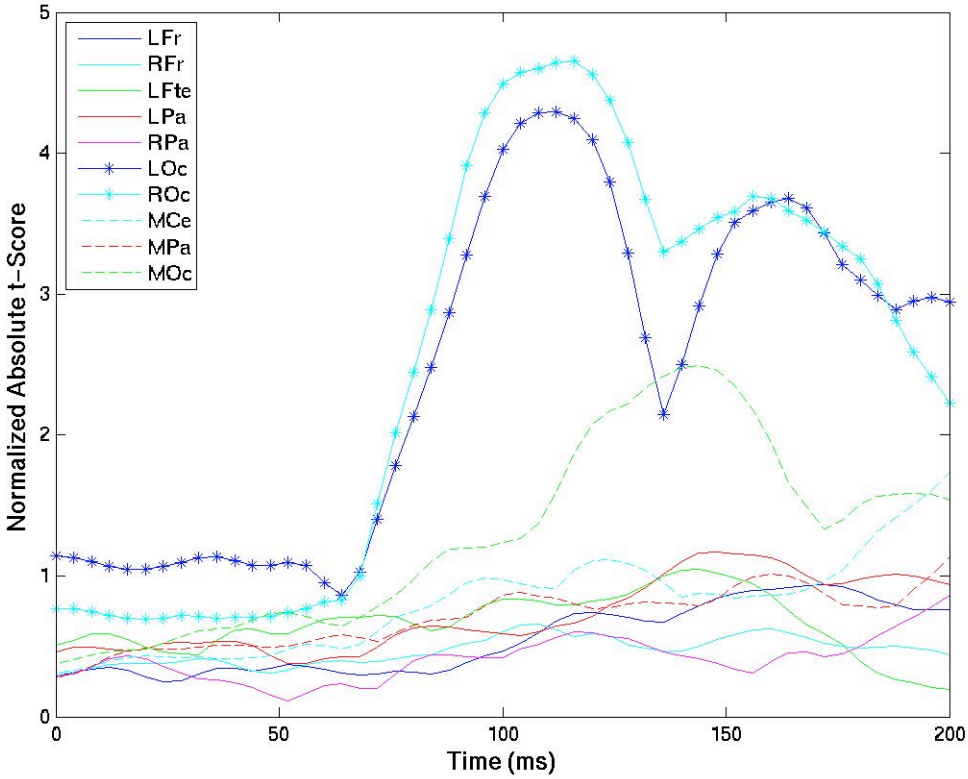

Figure VI.68 – Mean normalized absolute *t*-scores (see Section VI.B.3.a) for each IC cluster. High *t*-scores indicate a strong degree of stimulus-locked activity at that point in time. Note that the early (0-50 ms) relatively high *t*-scores for clusters LOc and ROc reflect activity elicited by a word occurring 300 ms before the beginning of almost all sentence acceptability task epochs. Cluster abbreviations: LFr=Left Frontal, RFr=Right Frontal, LFTe=Left Fronto-Temporal, LPa=Left Parietal, RPa=Right Parietal, LOc=Left Occipital, ROc=Right Occipital, MCe=Midline Central, MPa=Midline Parietal, MOc=Midline Occipital.

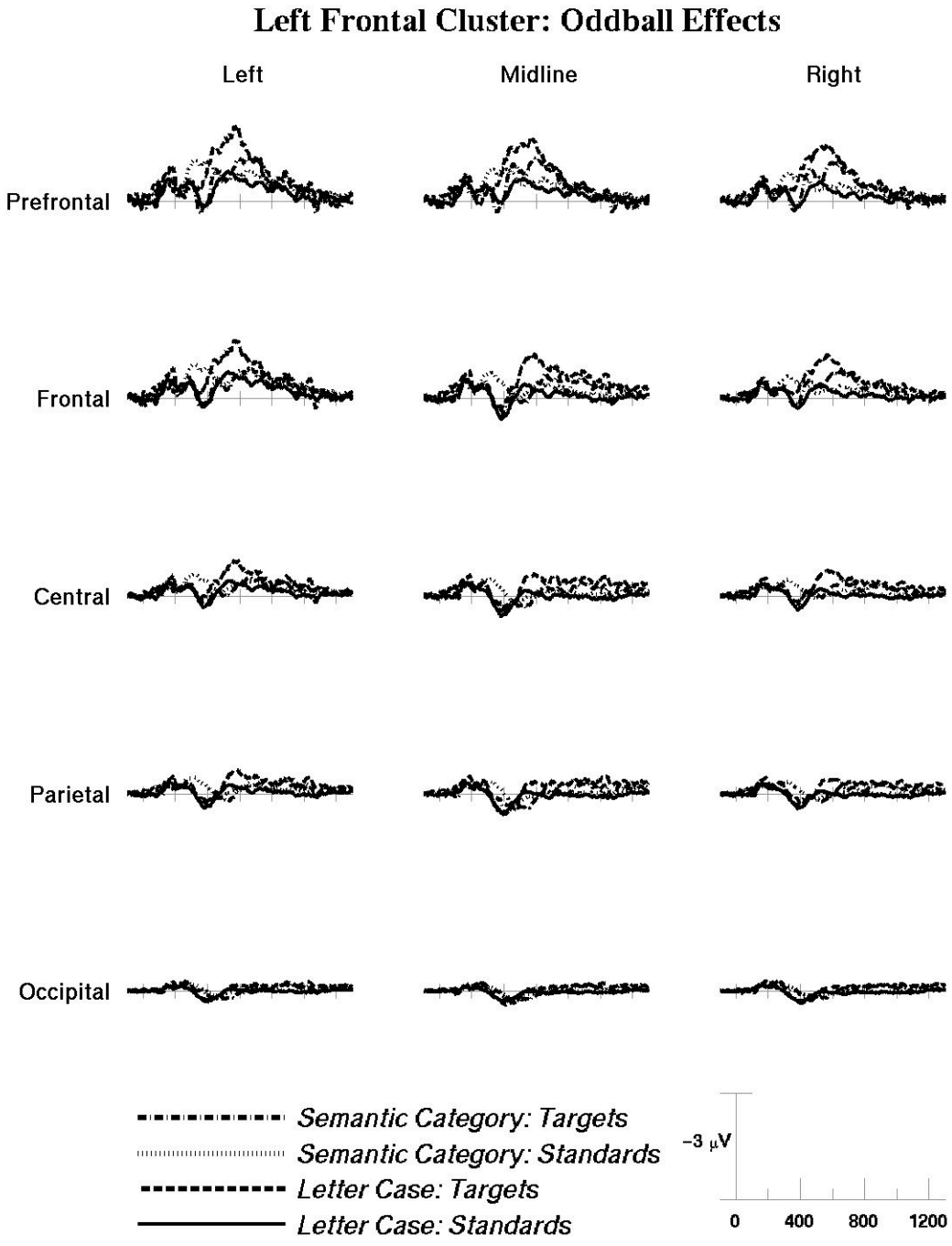

Figure VI.69 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the left frontal cluster. Averages are constructed from 12 participants.

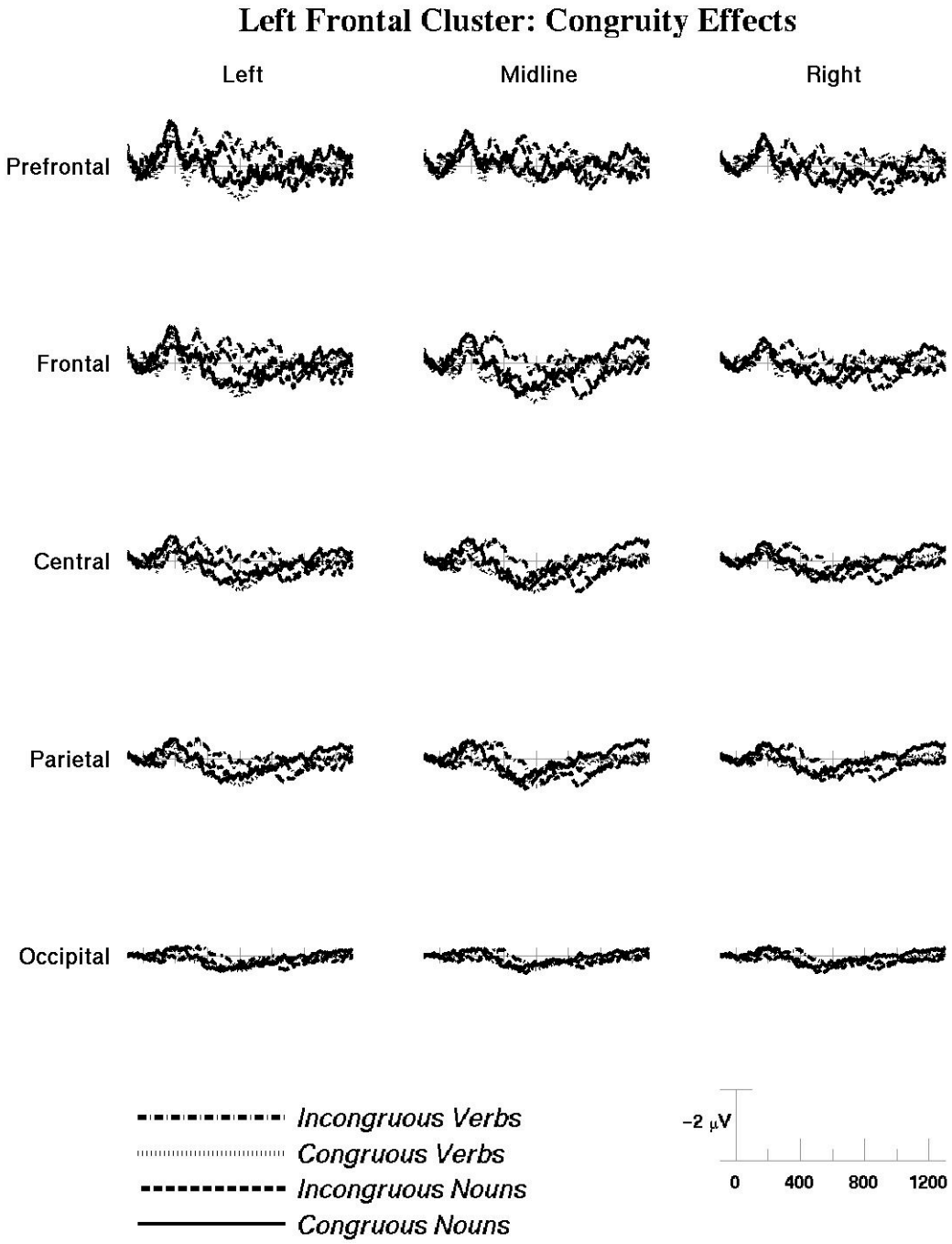

Figure VI.70 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the left frontal cluster. Averages are constructed from 12 participants.

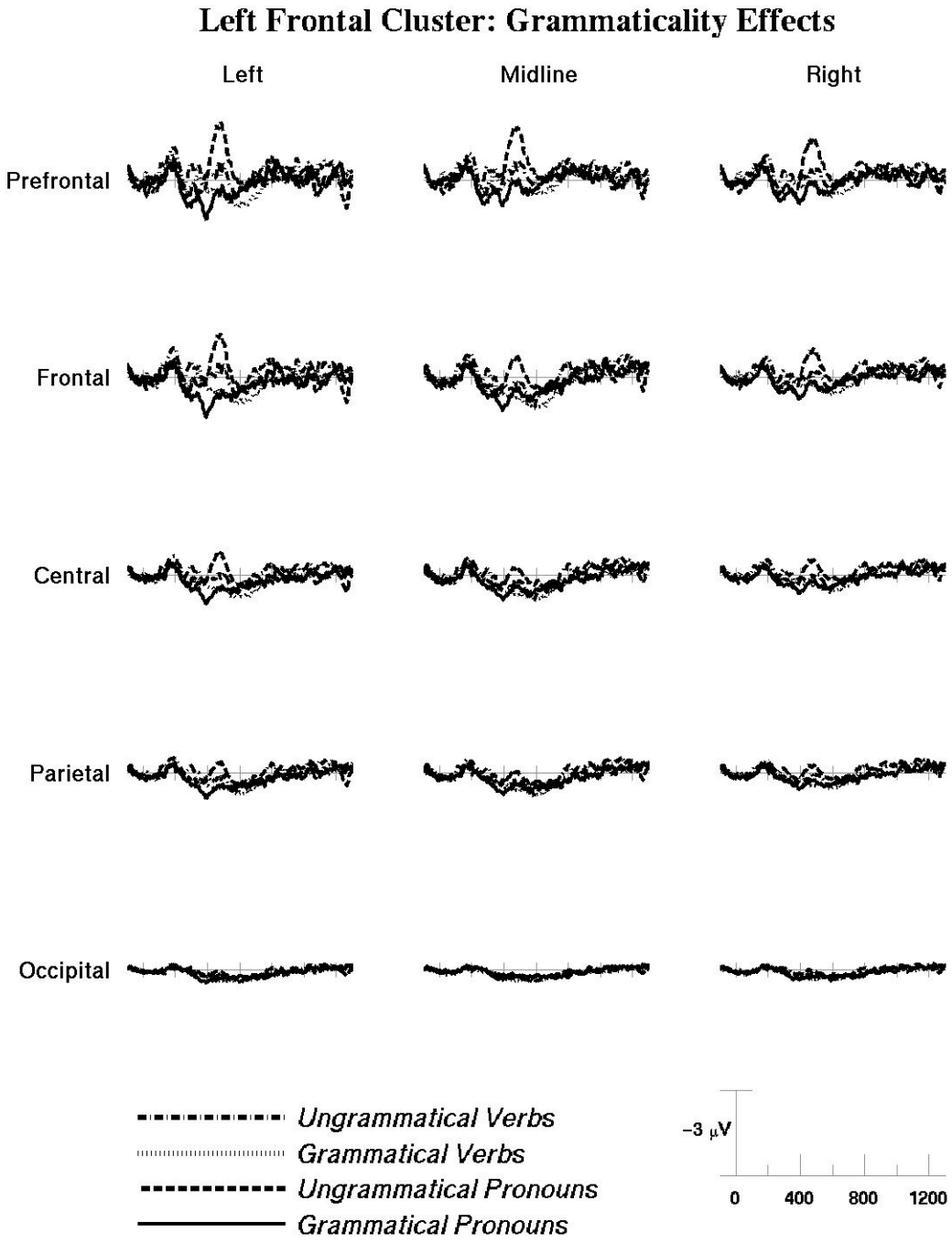

Figure VI.71 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the left frontal cluster. Averages are constructed from 12 participants.

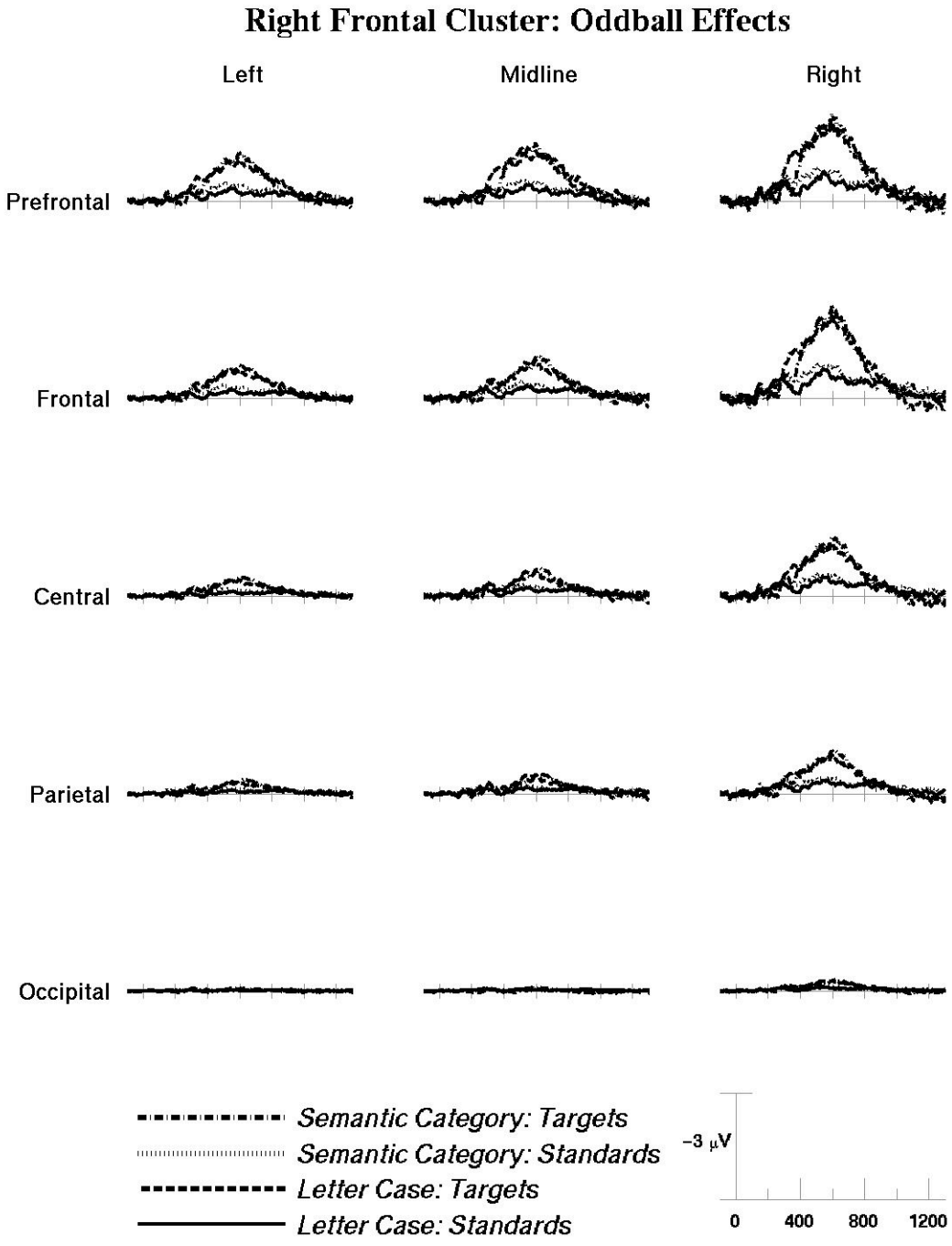

Figure VI.72 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the right frontal cluster. Averages are constructed from 12 participants.

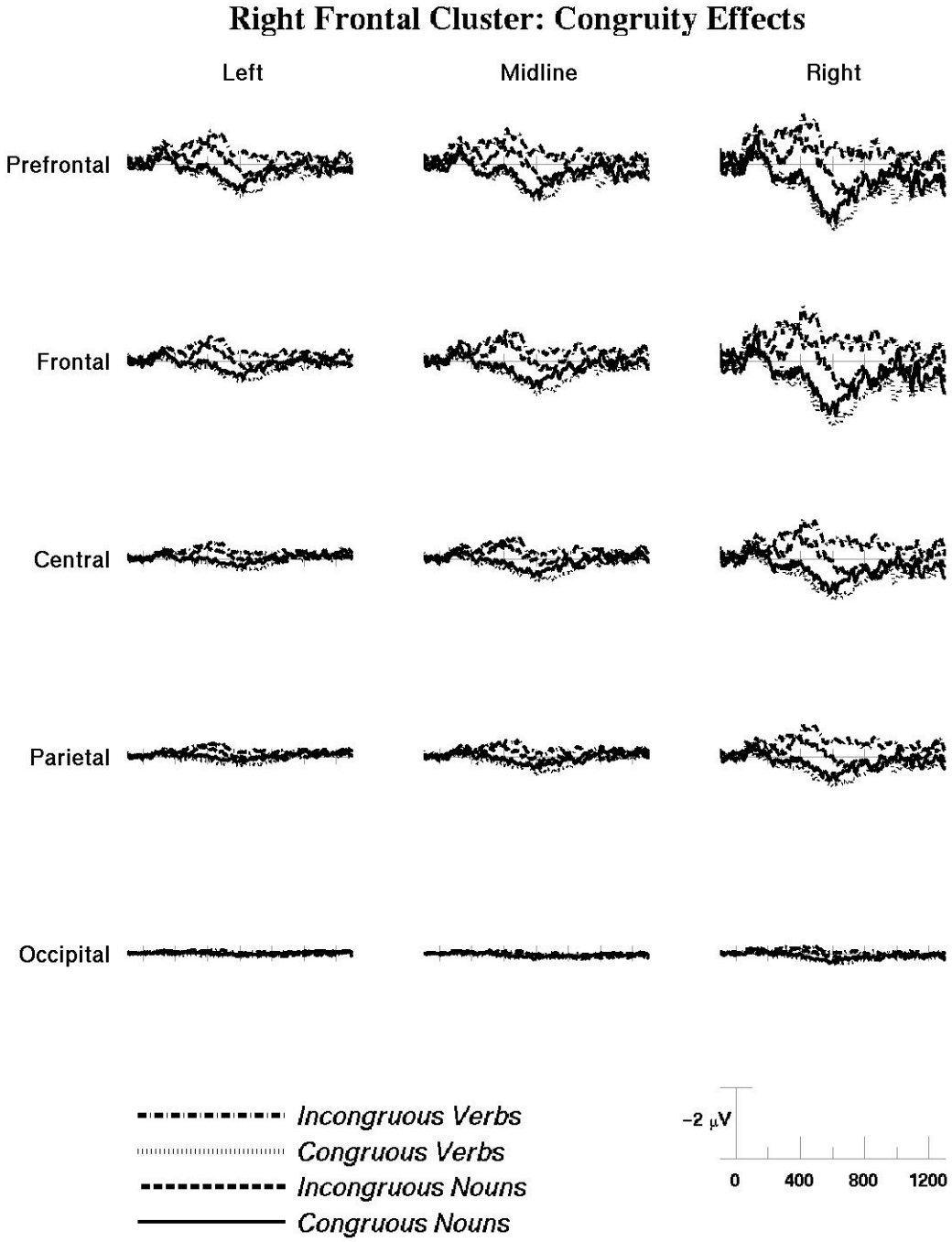

Figure VI.73 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the right frontal cluster. Averages are constructed from ten participants.

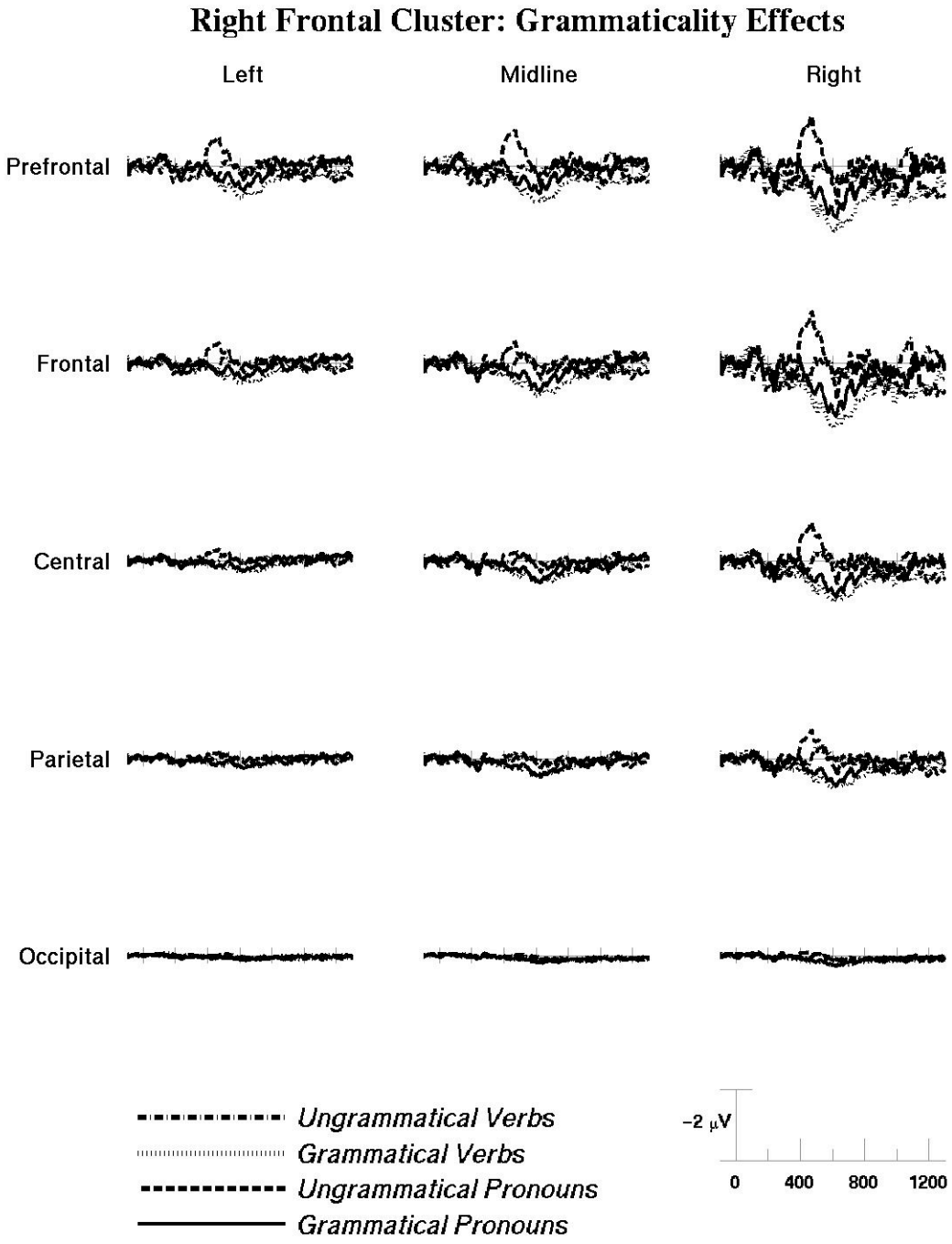

Figure VI.74 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the right frontal cluster. Averages are constructed from ten participants.

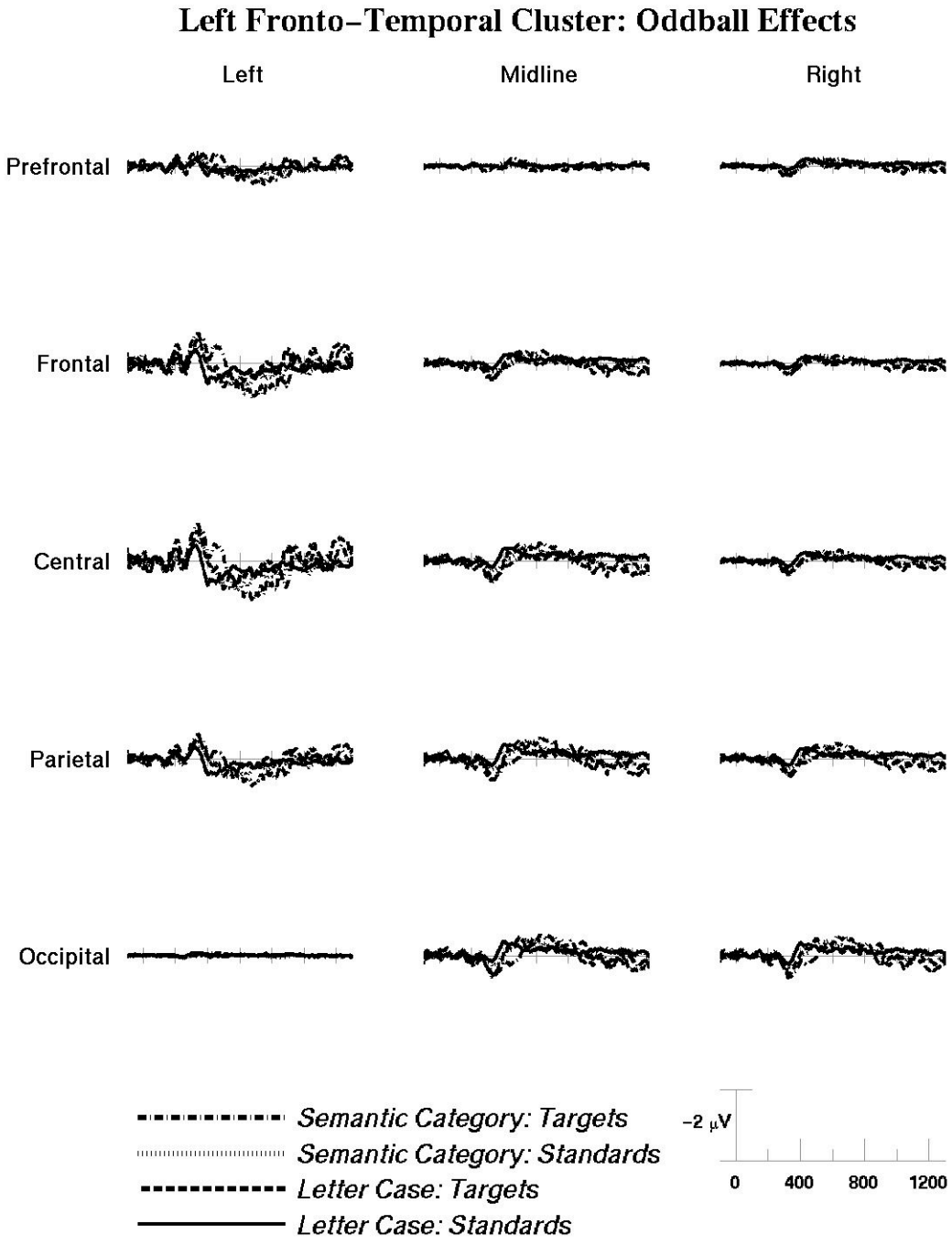

Figure VI.75 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the left fronto-temporal cluster. Averages are constructed from four participants.

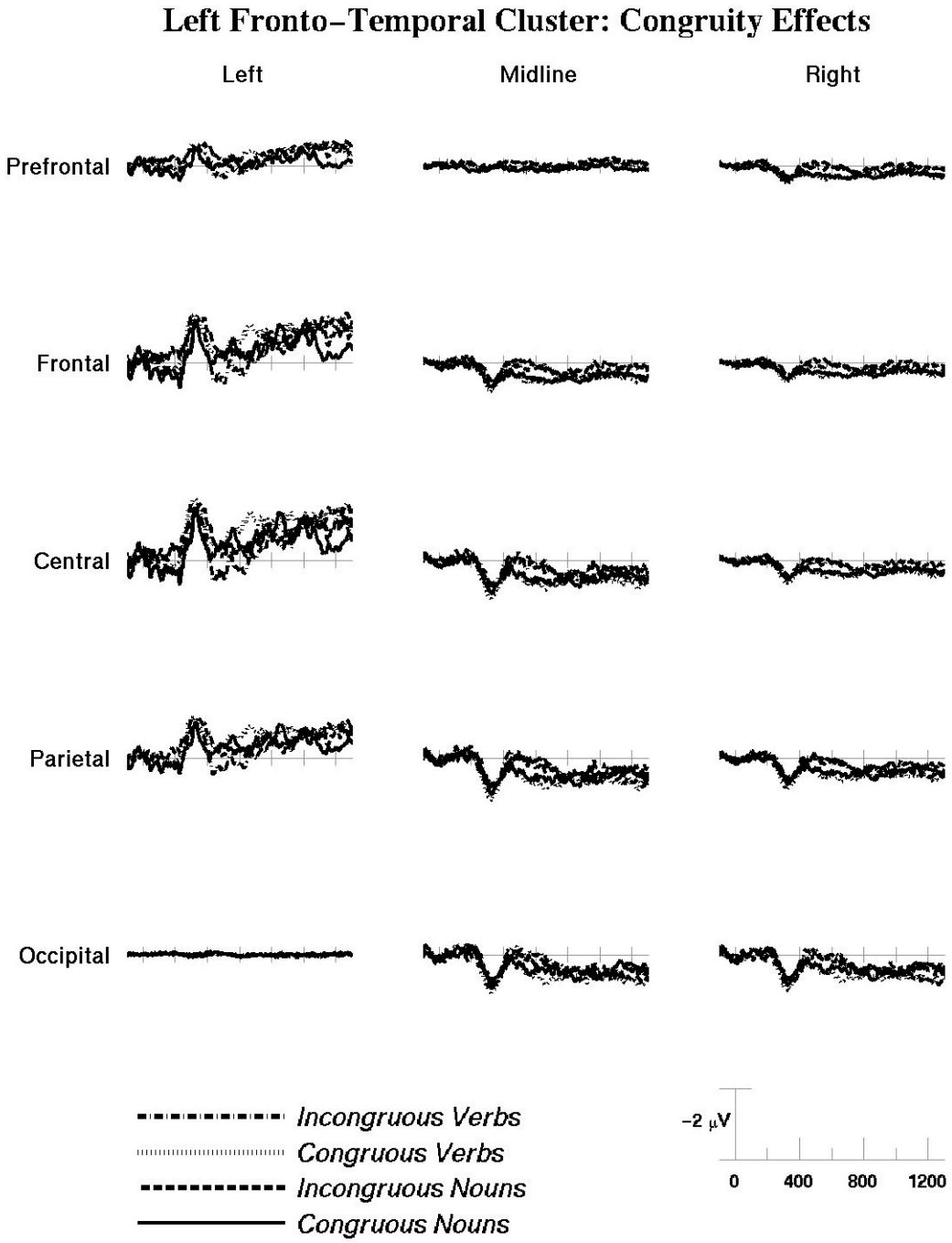

Figure VI.76 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the left fronto-temporal cluster. Averages are constructed from four participants.

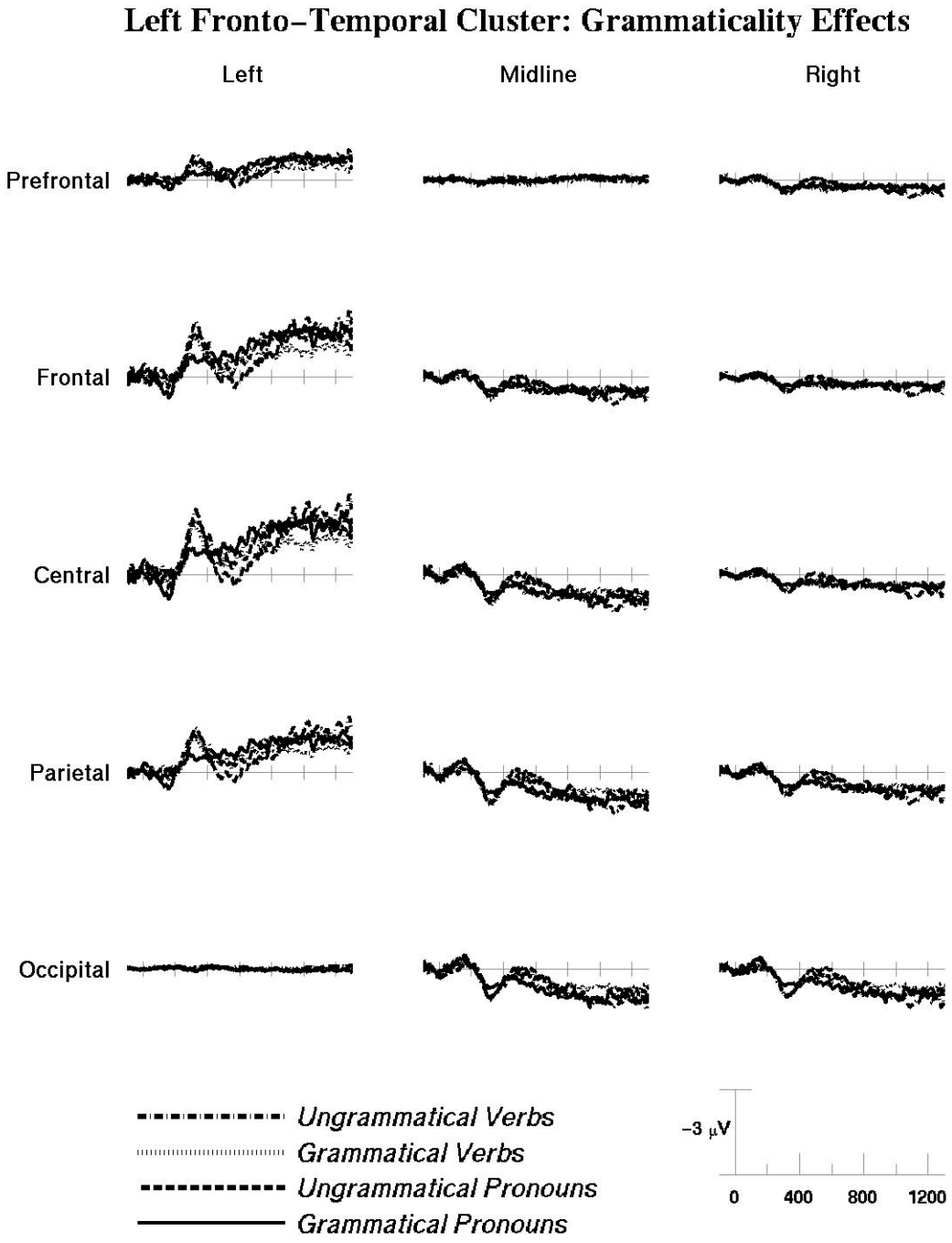

Figure VI.77 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the left fronto-temporal cluster. Averages are constructed from four participants.

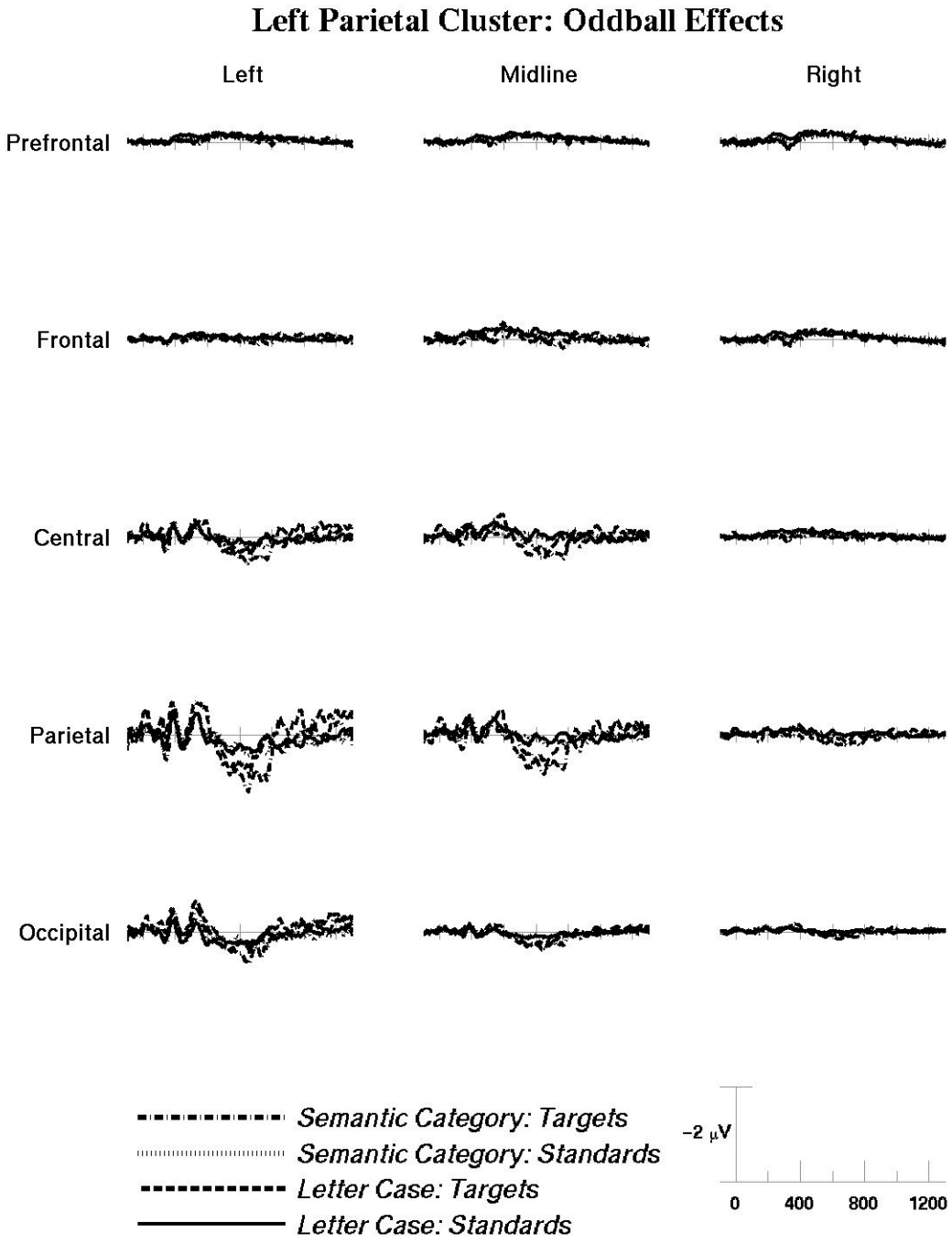

Figure VI.78 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the left parietal cluster. Averages are constructed from ten participants.

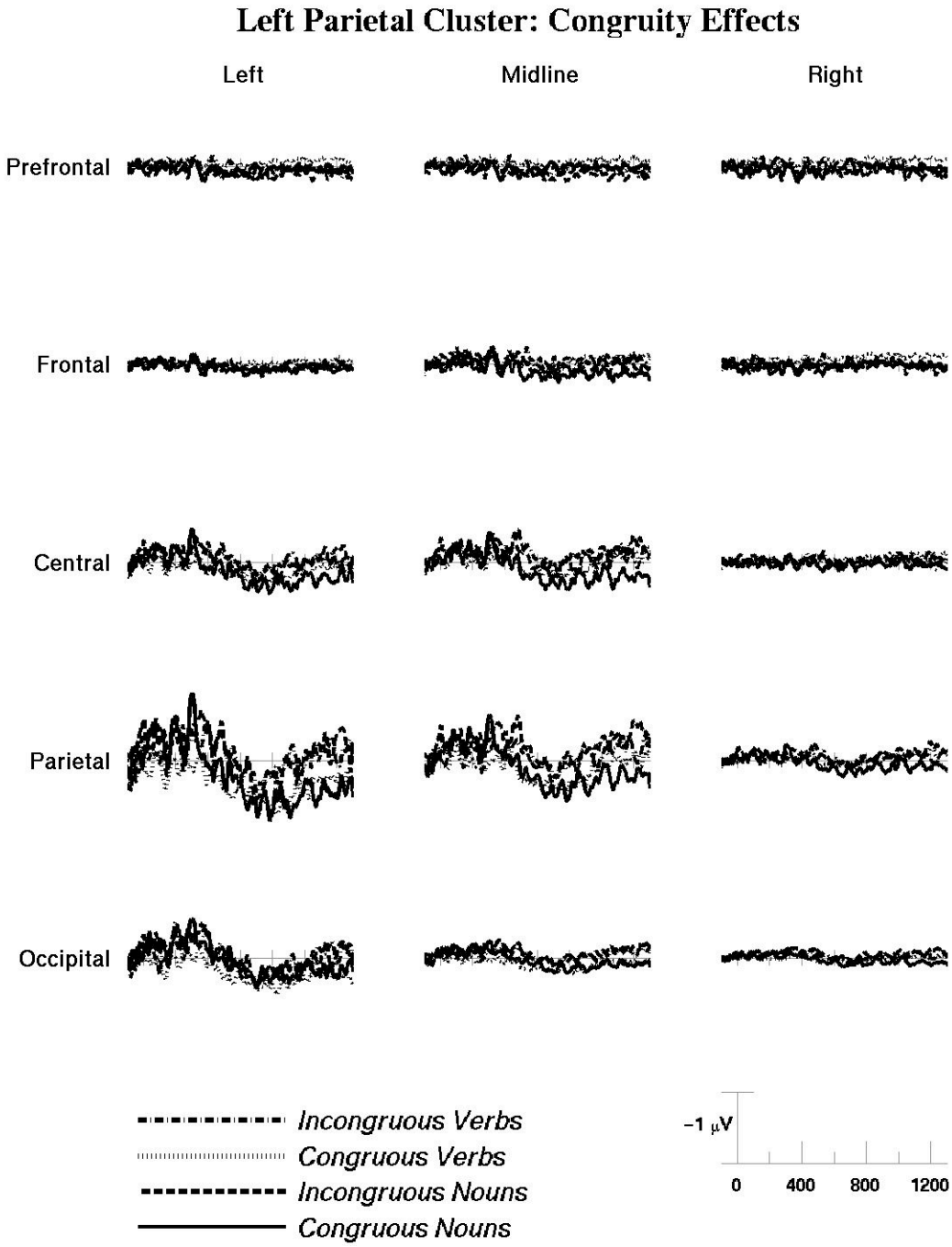

Figure VI.79 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the left parietal cluster. Averages are constructed from ten participants.

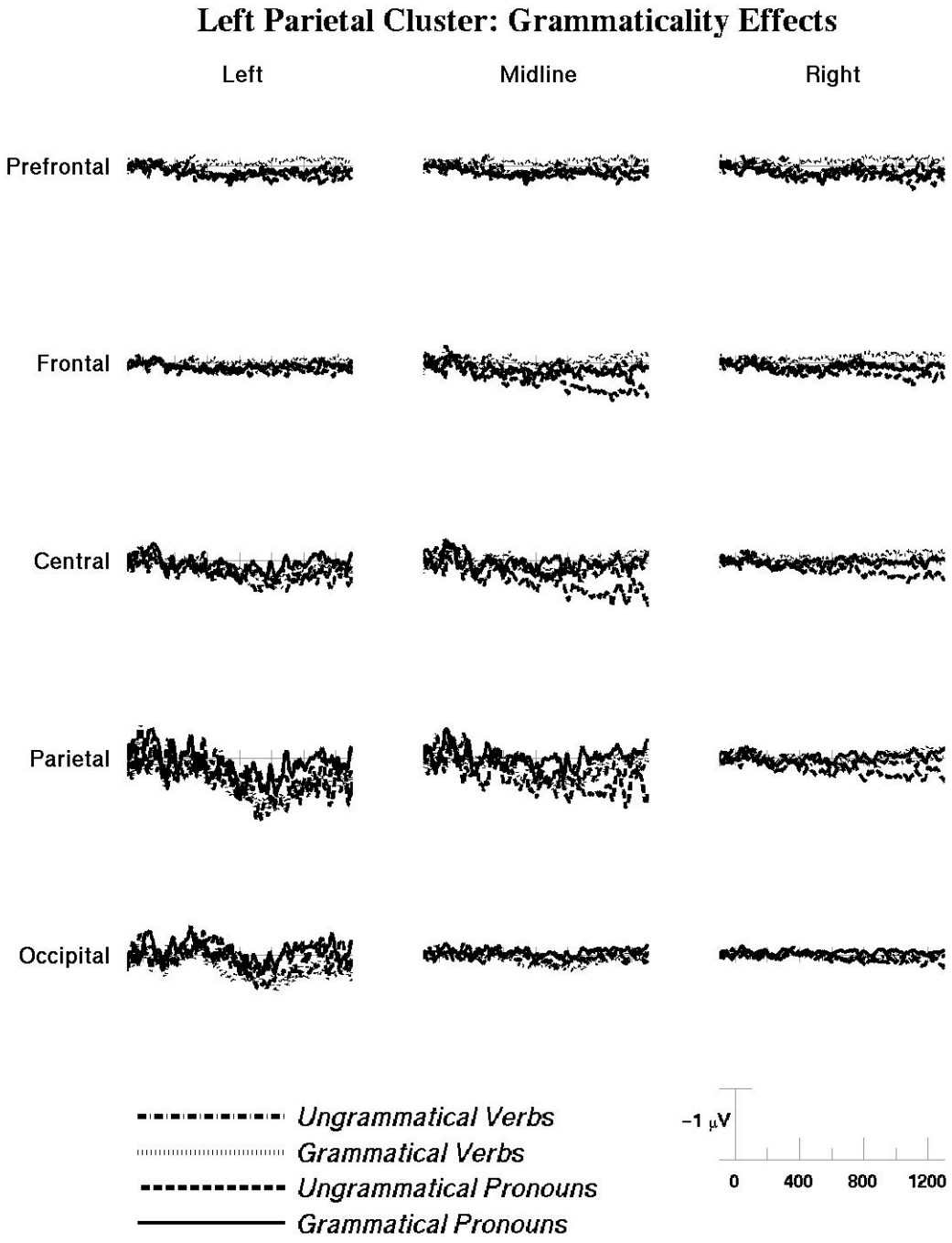

Figure VI.80 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the left parietal cluster. Averages are constructed from ten participants.

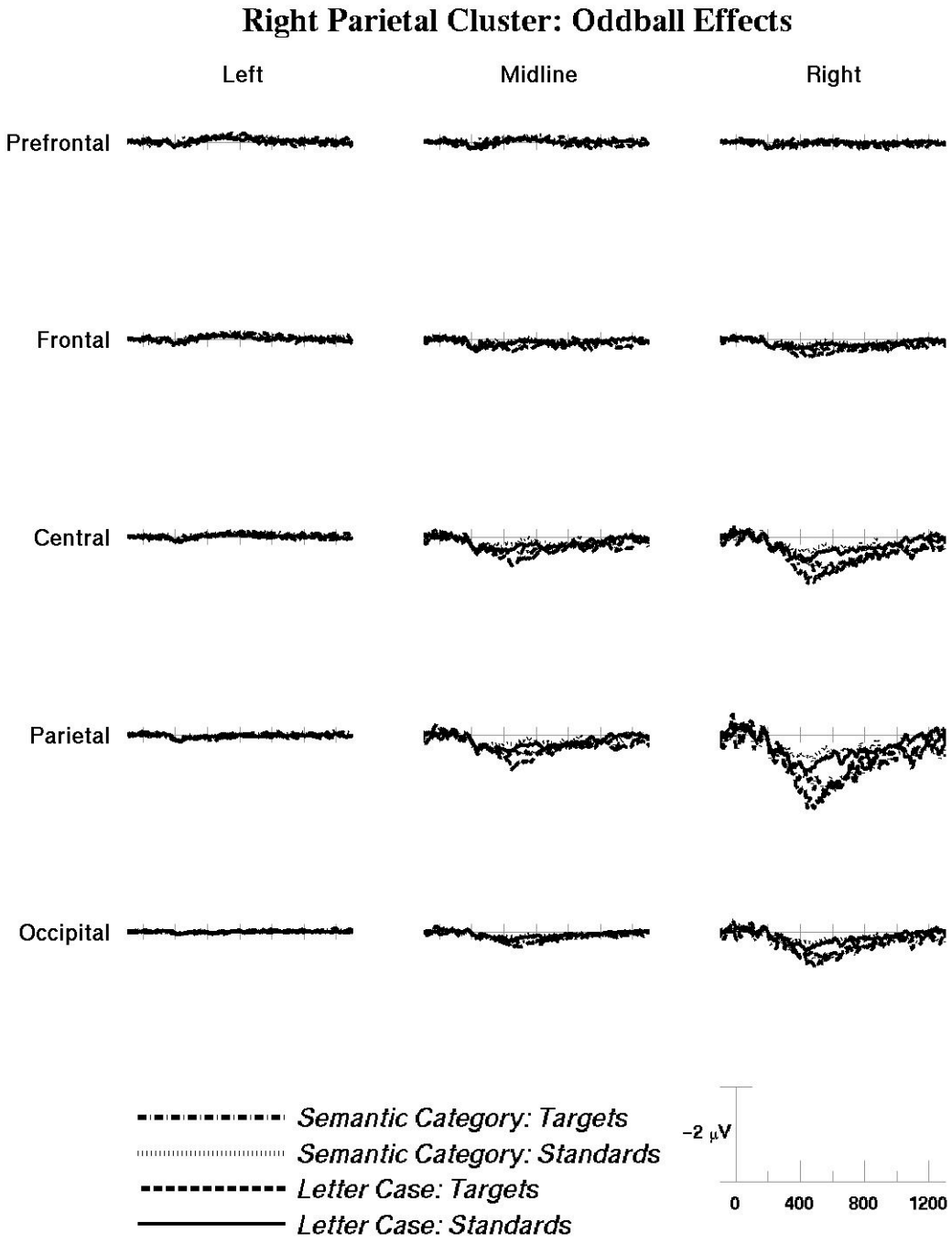

Figure VI.81 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the right parietal cluster. Averages are constructed from seven participants.

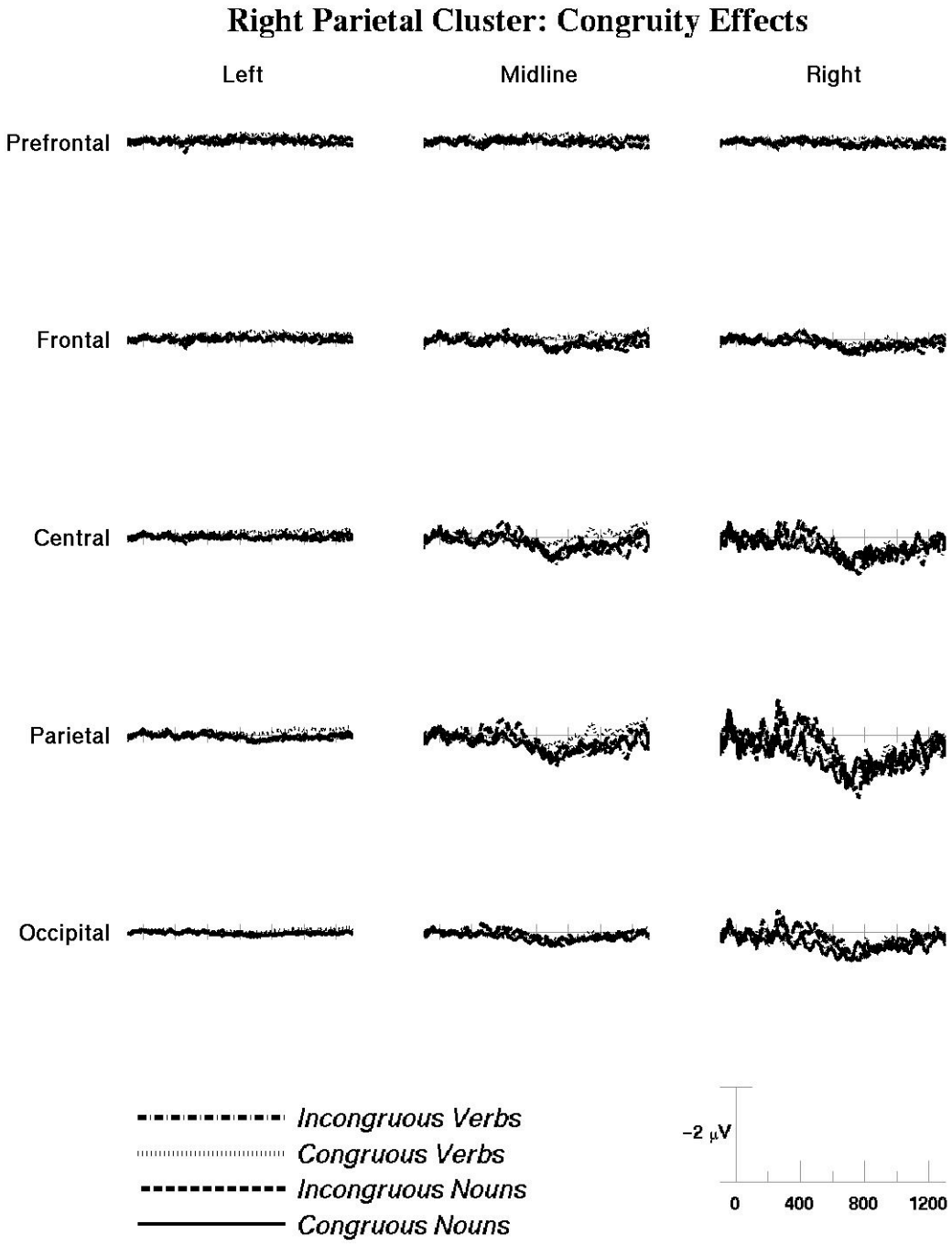

Figure VI.82 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the right parietal cluster. Averages are constructed from seven participants.

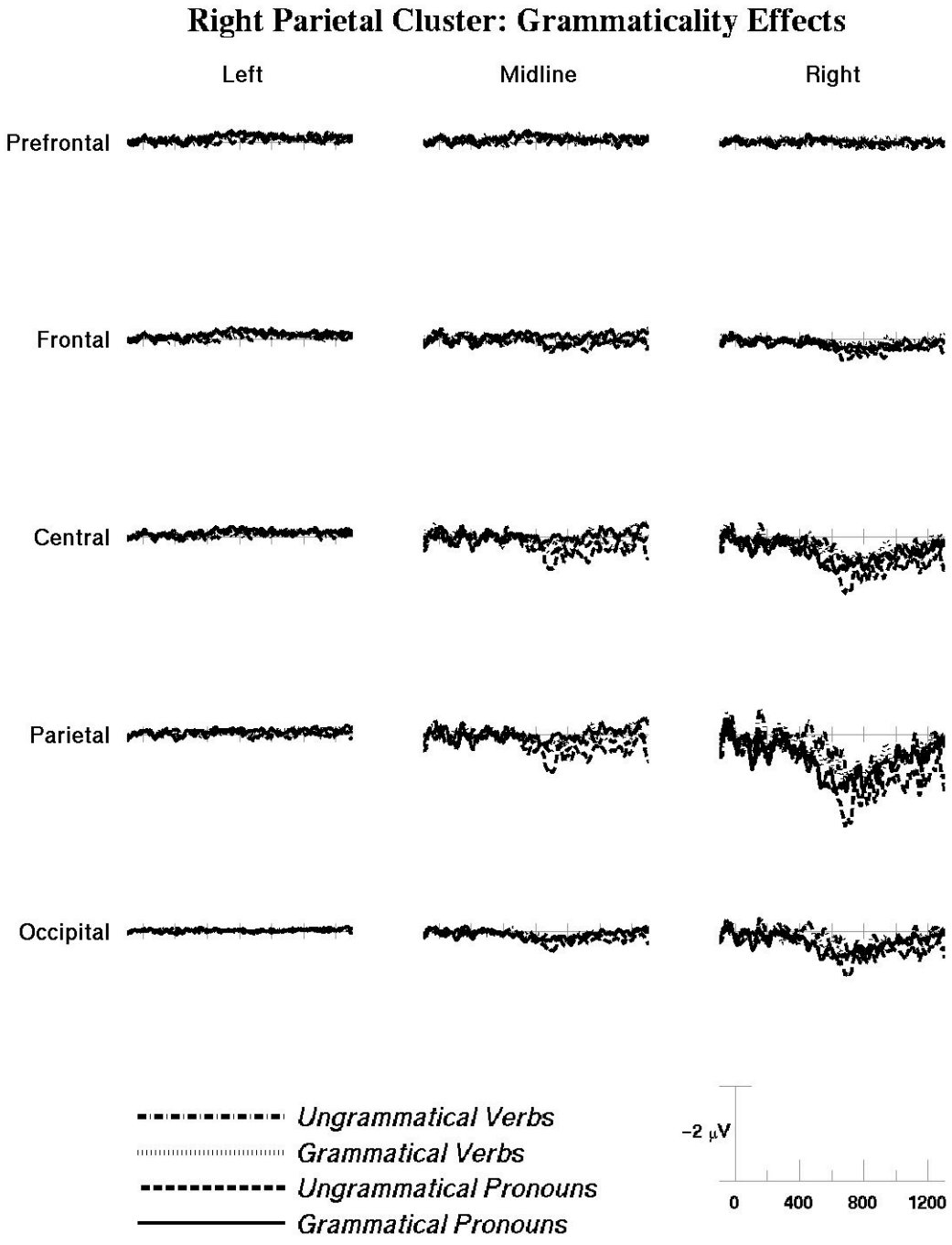

Figure VI.83 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the right parietal cluster. Averages are constructed from seven participants.

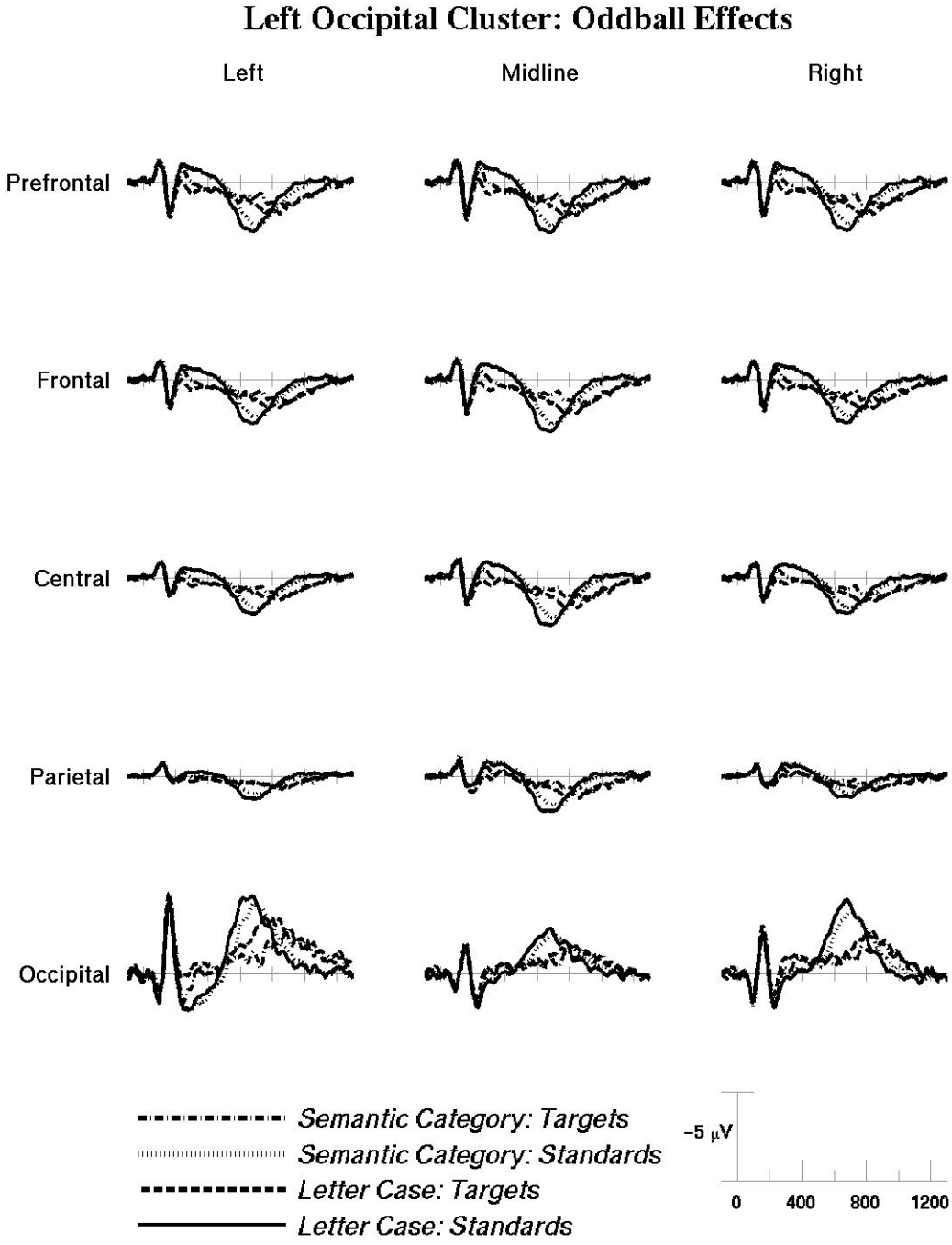

Figure VI.84 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the left occipital cluster. Averages are constructed from six participants.

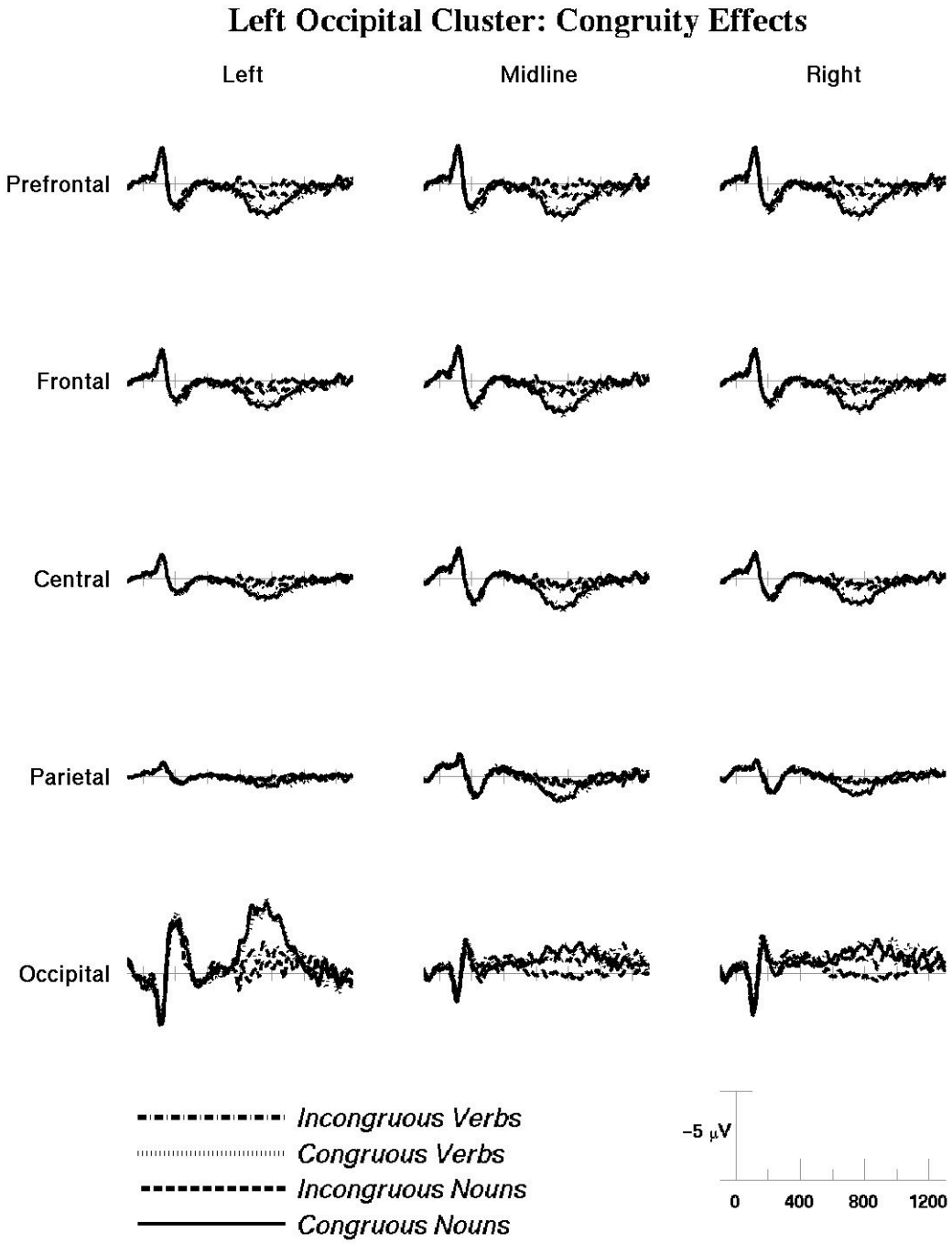

Figure VI.85 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the left occipital cluster. Averages are constructed from six participants.

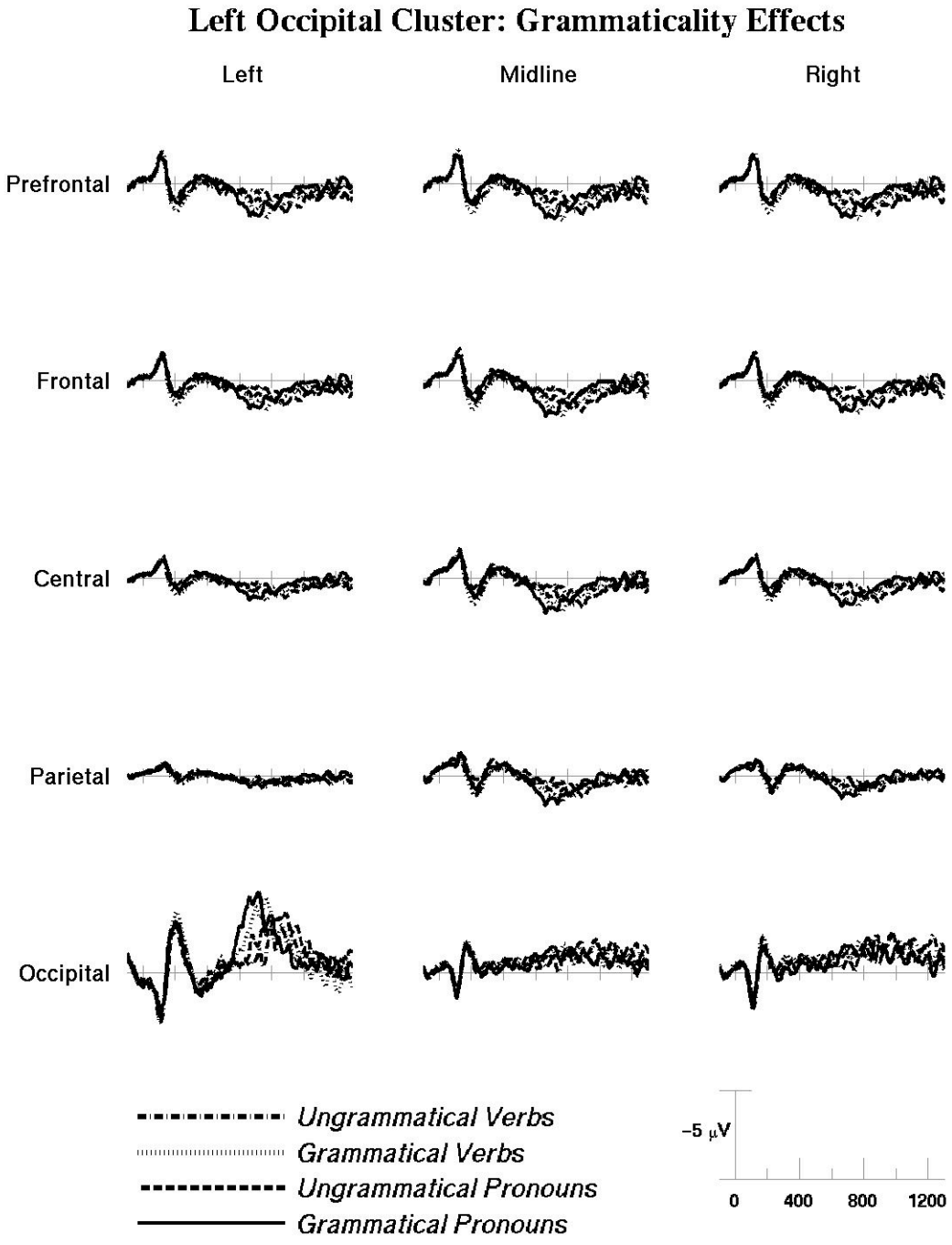

Figure VI.86 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the left occipital cluster. Averages are constructed from six participants.

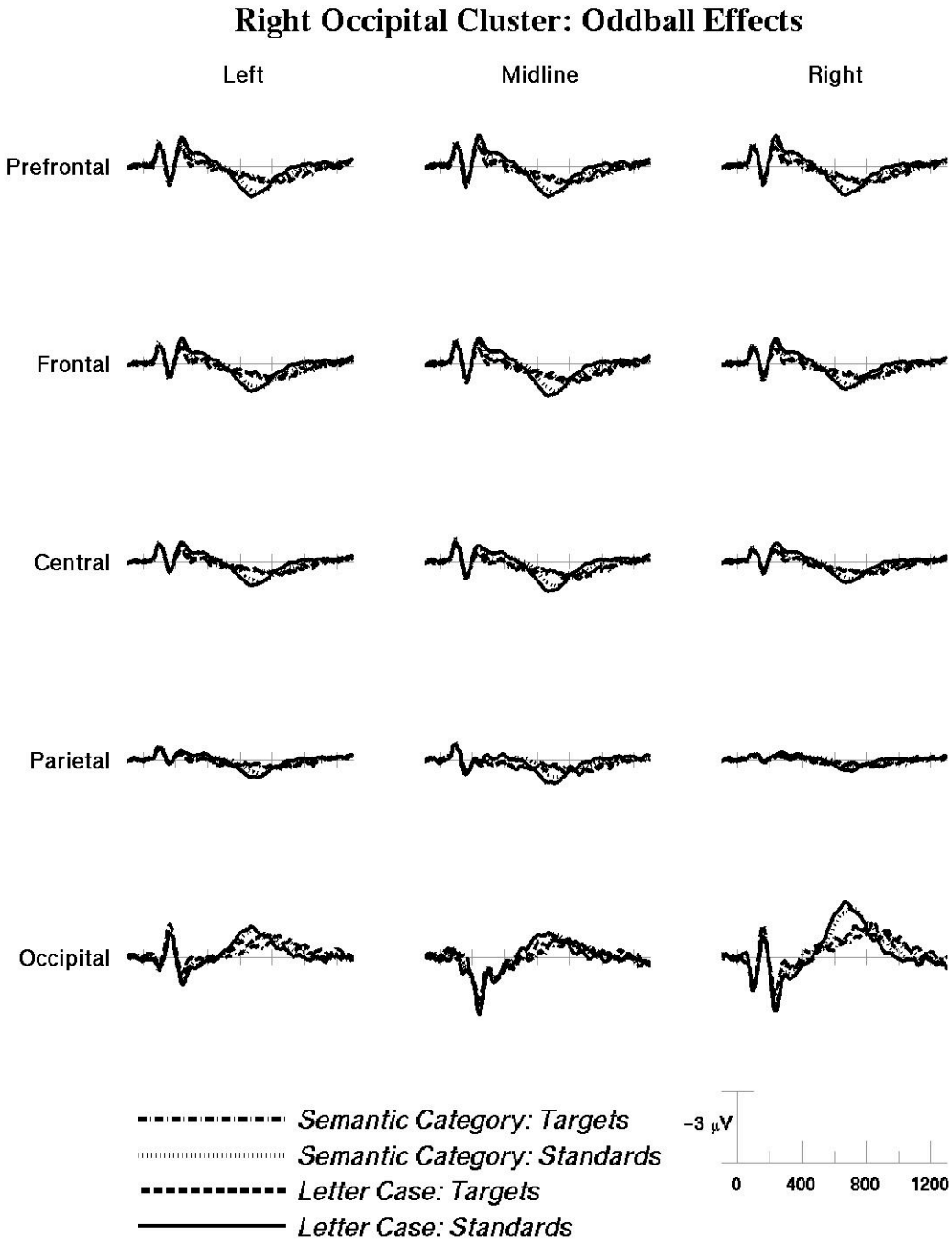

Figure VI.87 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the right occipital cluster. Averages are constructed from 13 participants.

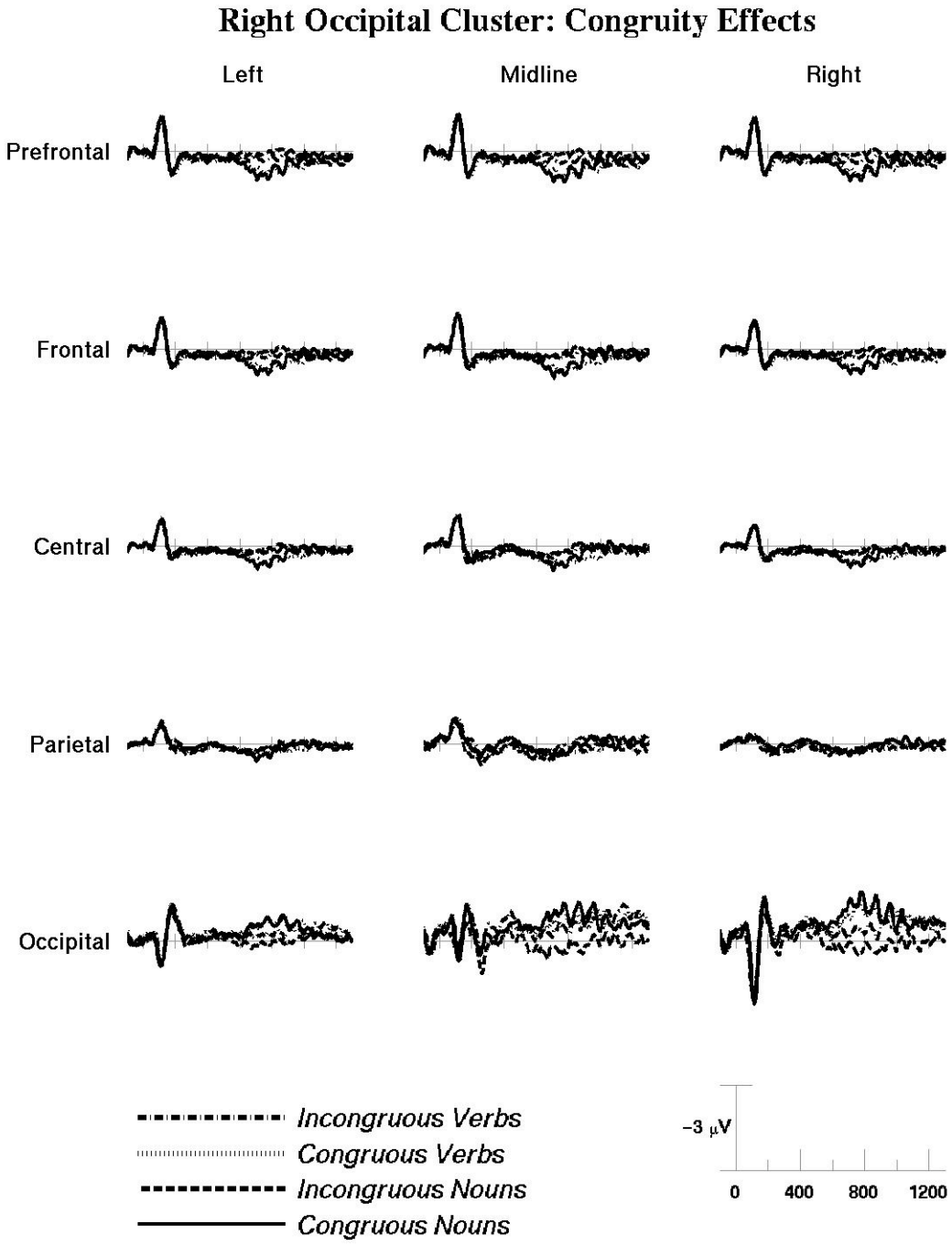

Figure VI.88 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the right occipital cluster. Averages are constructed from 13 participants.

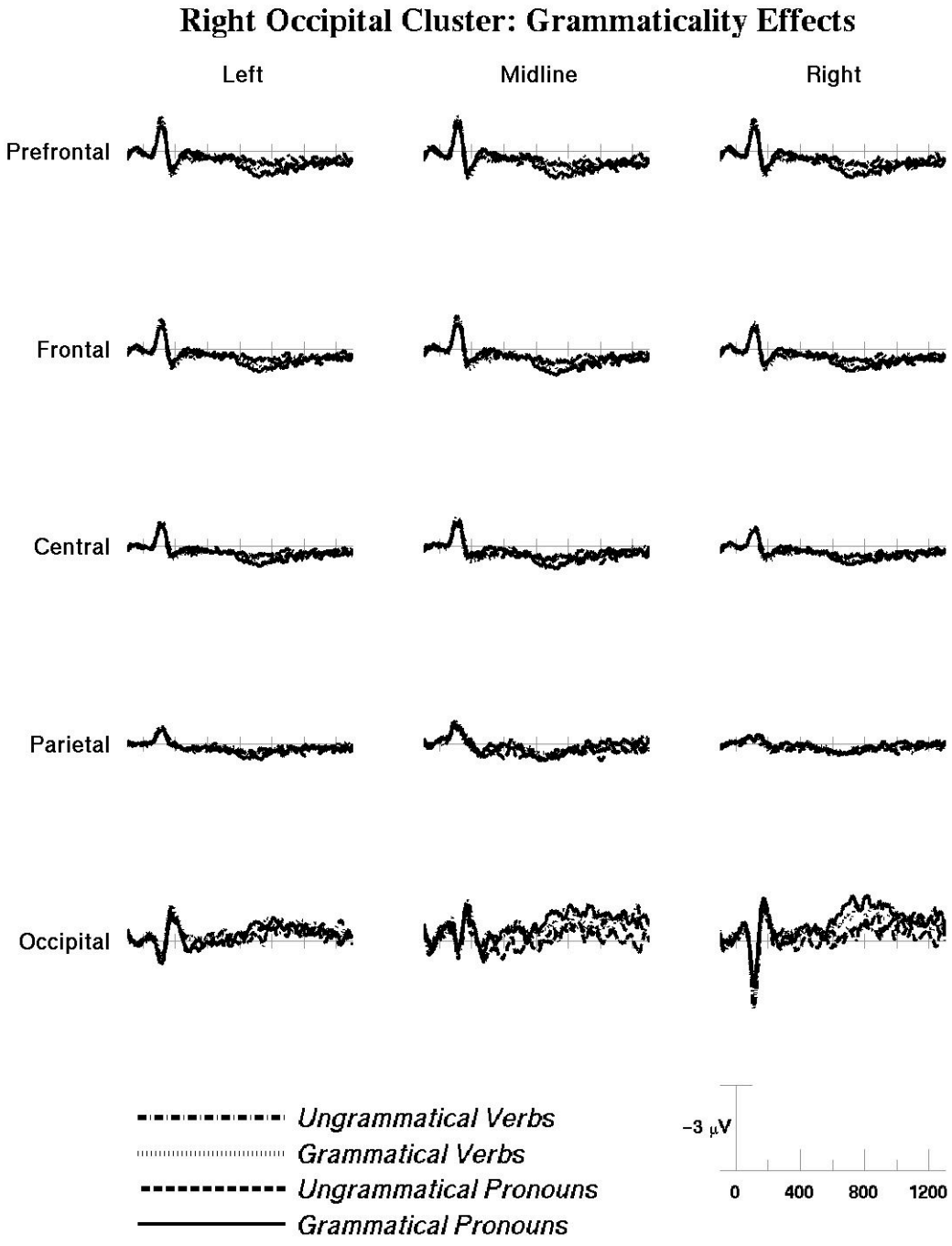

Figure VI.89 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the right occipital cluster. Averages are constructed from 13 participants.

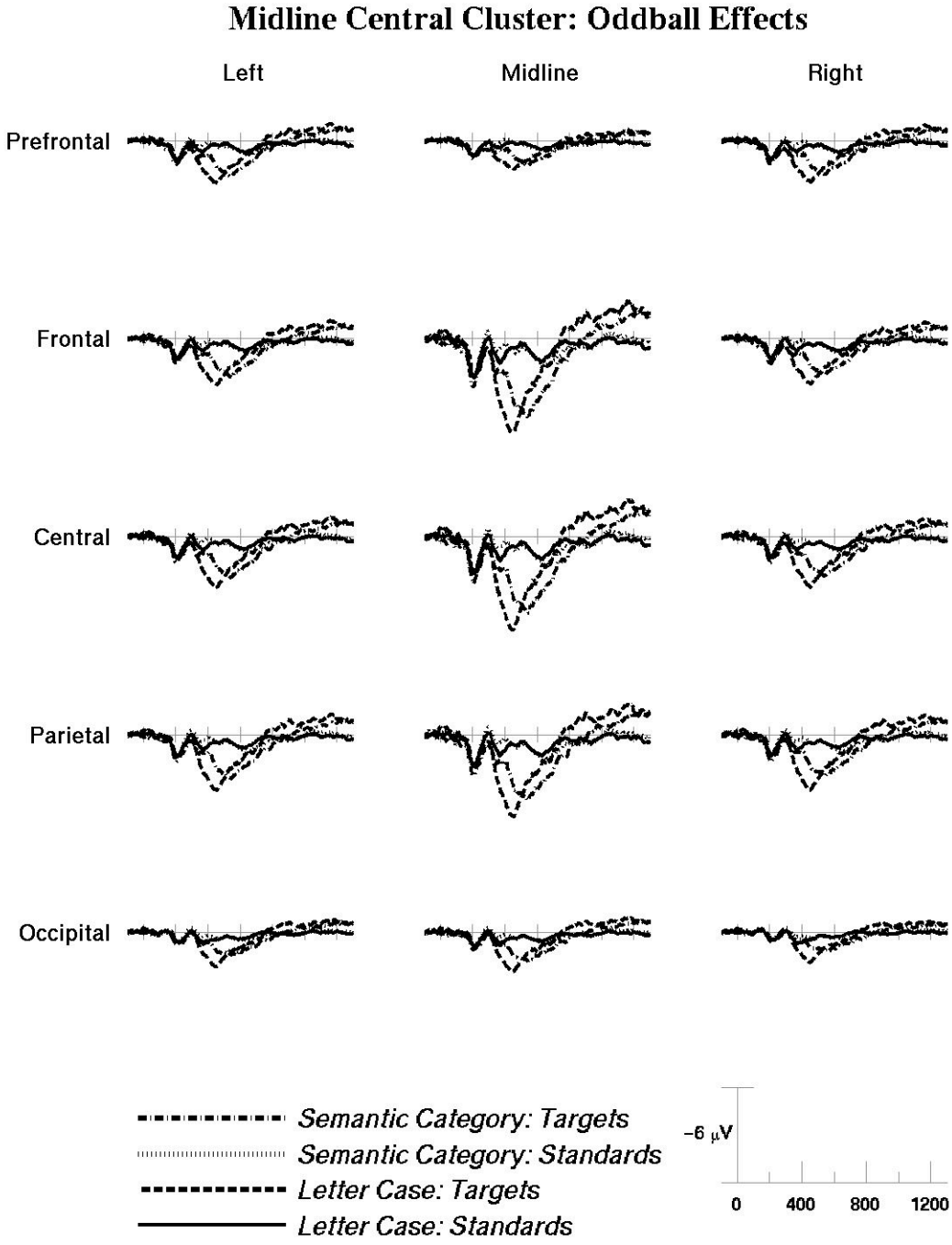

Figure VI.90 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the midline central cluster. Averages are constructed from 12 participants.

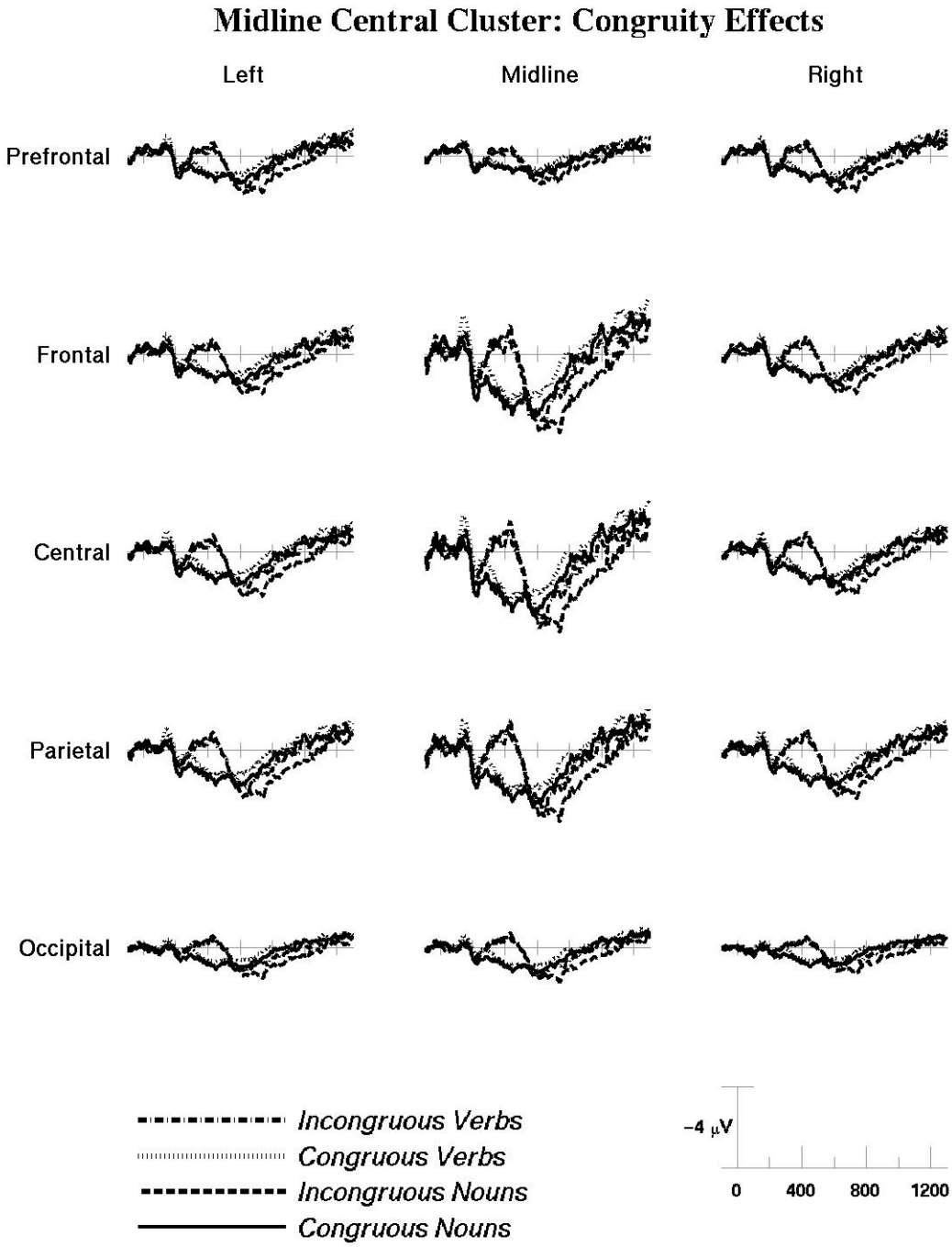

Figure VI.91 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the midline central cluster. Averages are constructed from 12 participants.

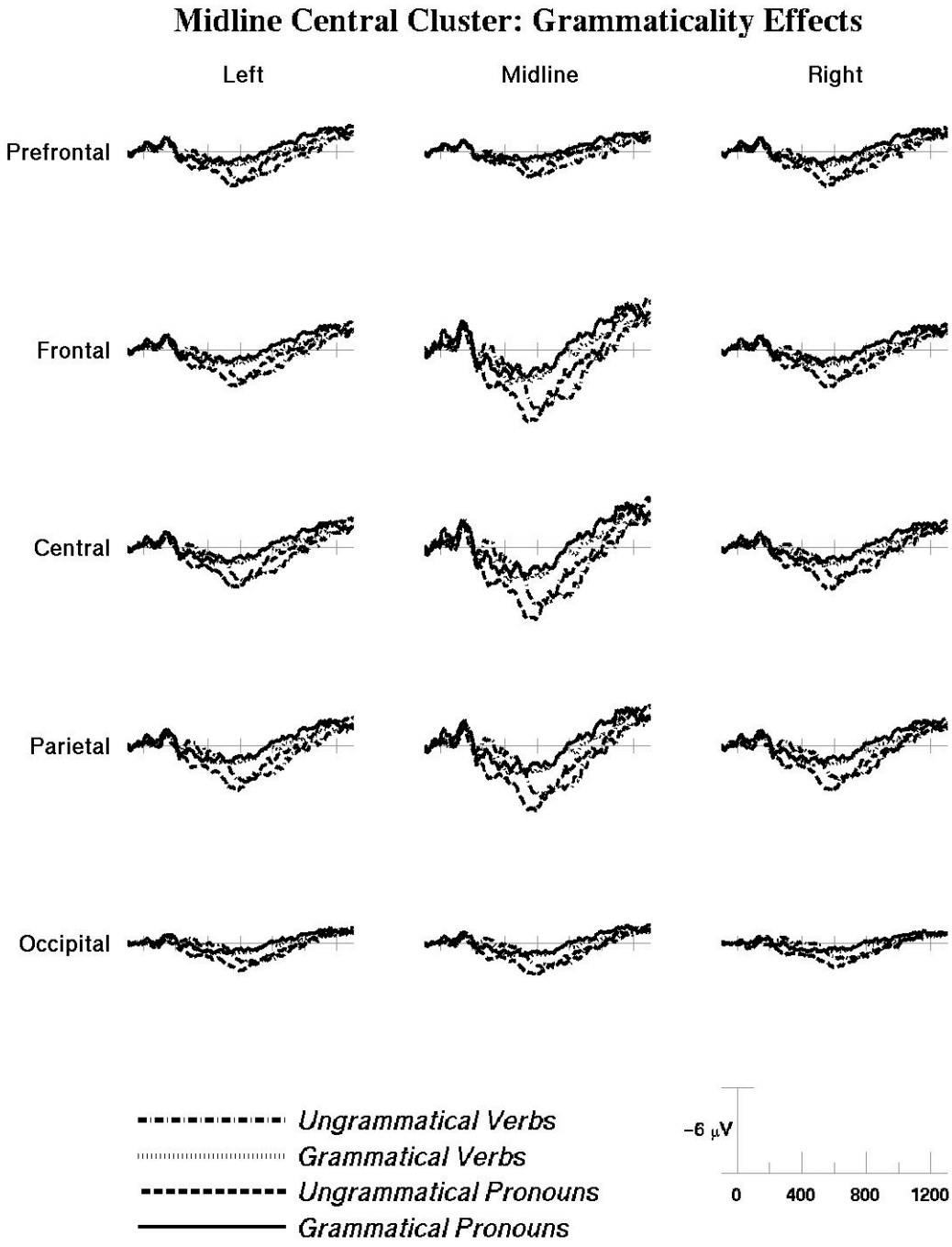

Figure VI.92 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the midline central cluster. Averages are constructed from 12 participants.

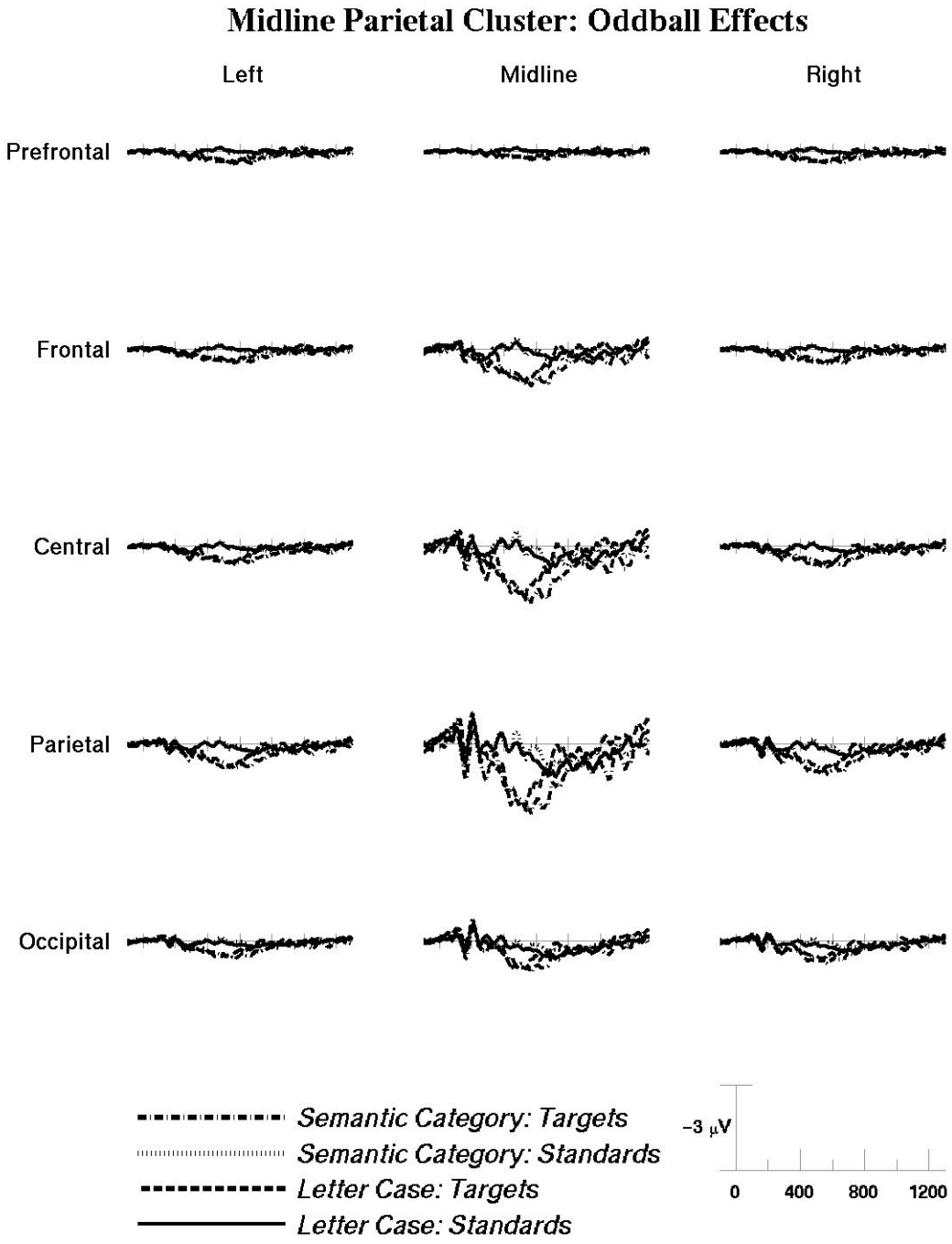

Figure VI.93 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the midline parietal cluster. Averages are constructed from nine participants.

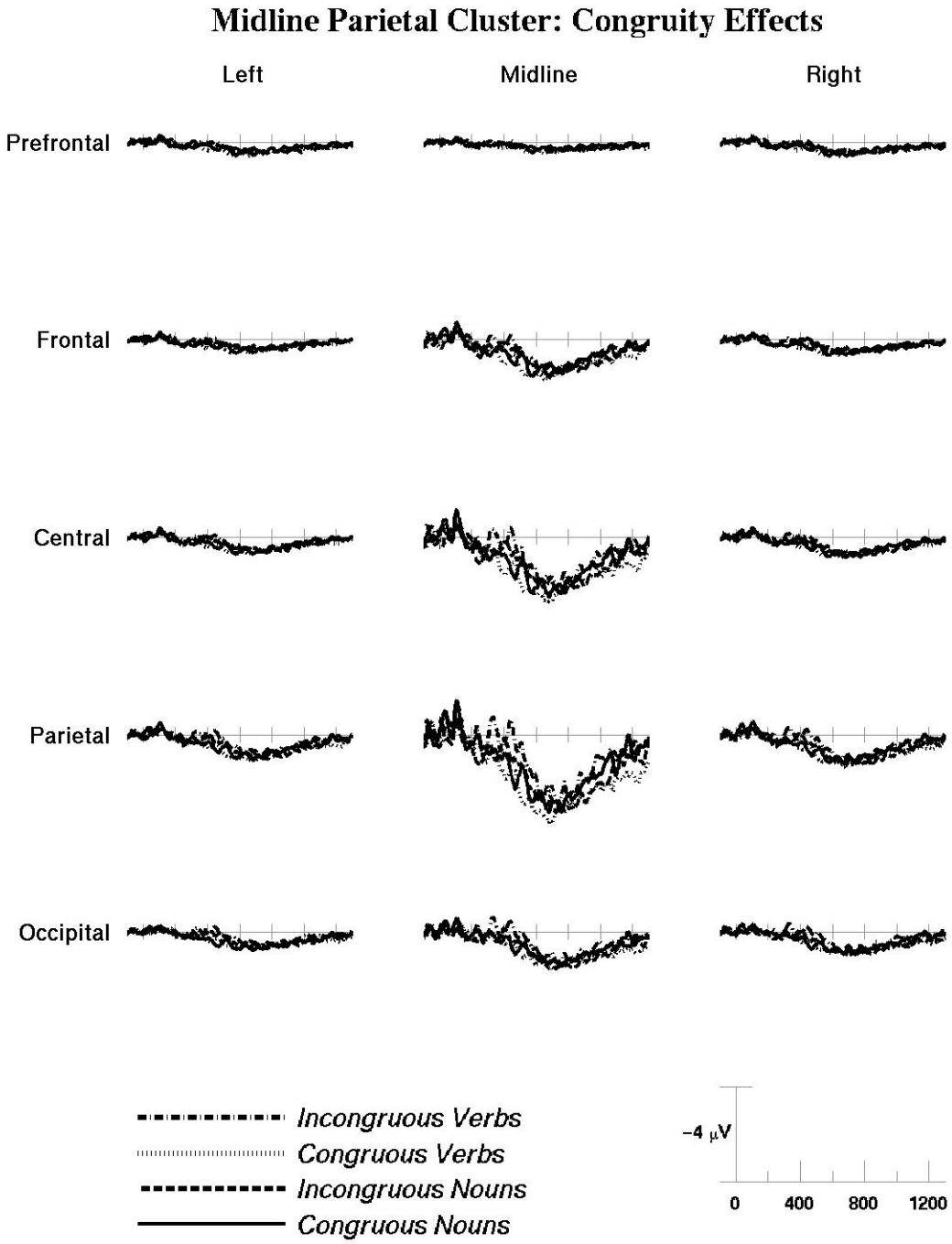

Figure VI.94 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the midline parietal cluster. Averages are constructed from nine participants.

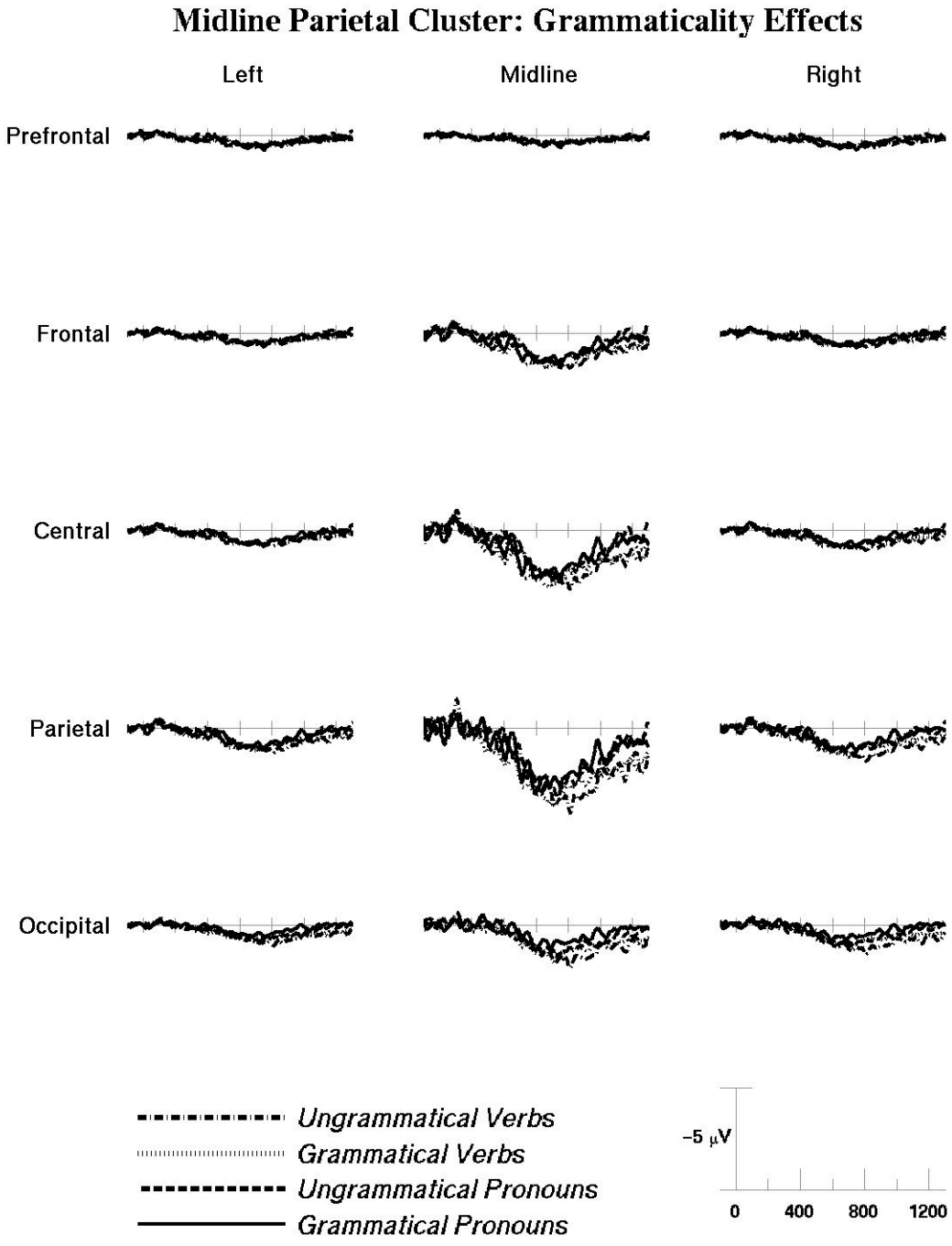

Figure VI.95 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the midline parietal cluster. Averages are constructed from nine participants.

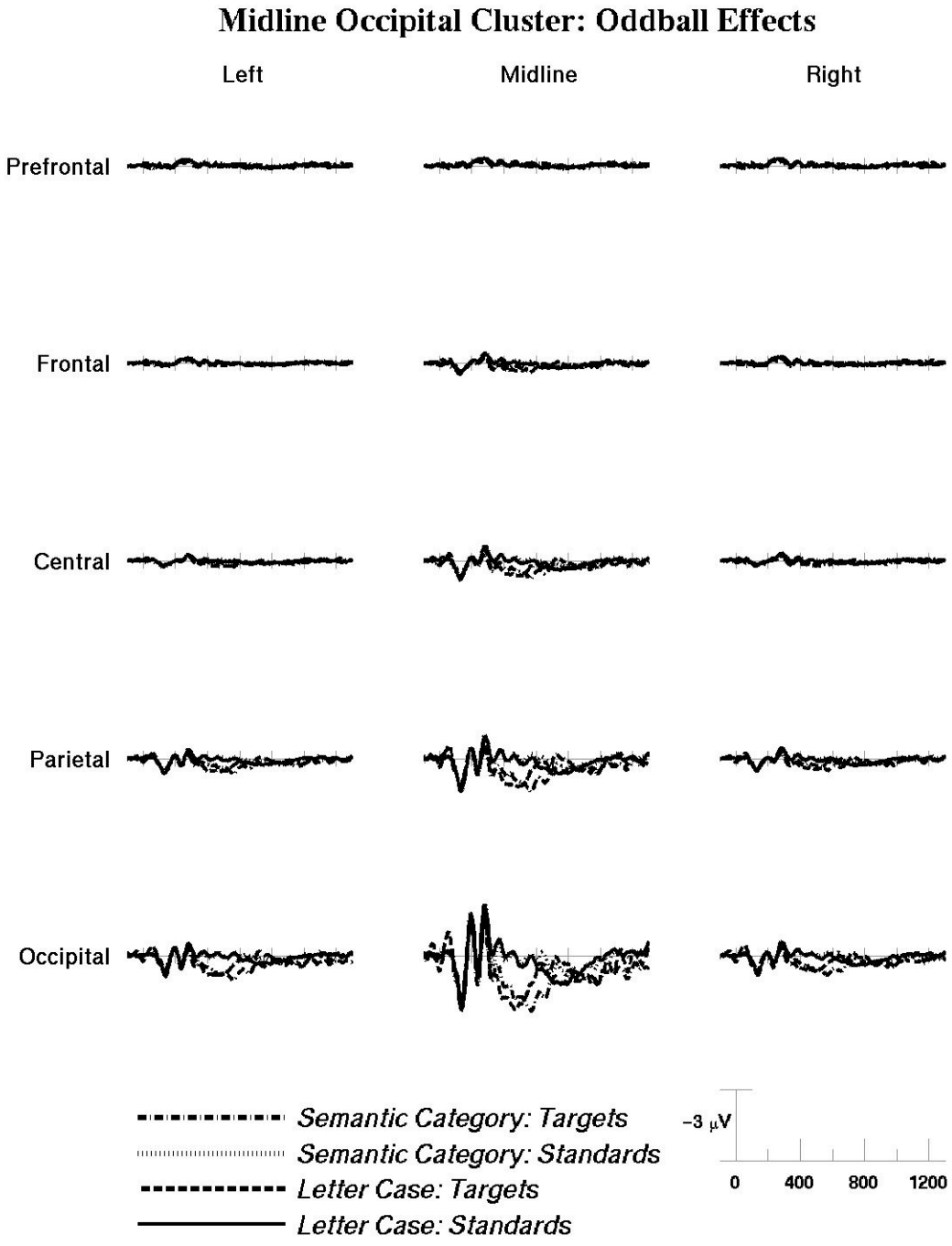

Figure VI.96 – Example grand average scalp-projected oddball task event-related activations (ERAs) of the midline occipital cluster. Averages are constructed from ten participants.

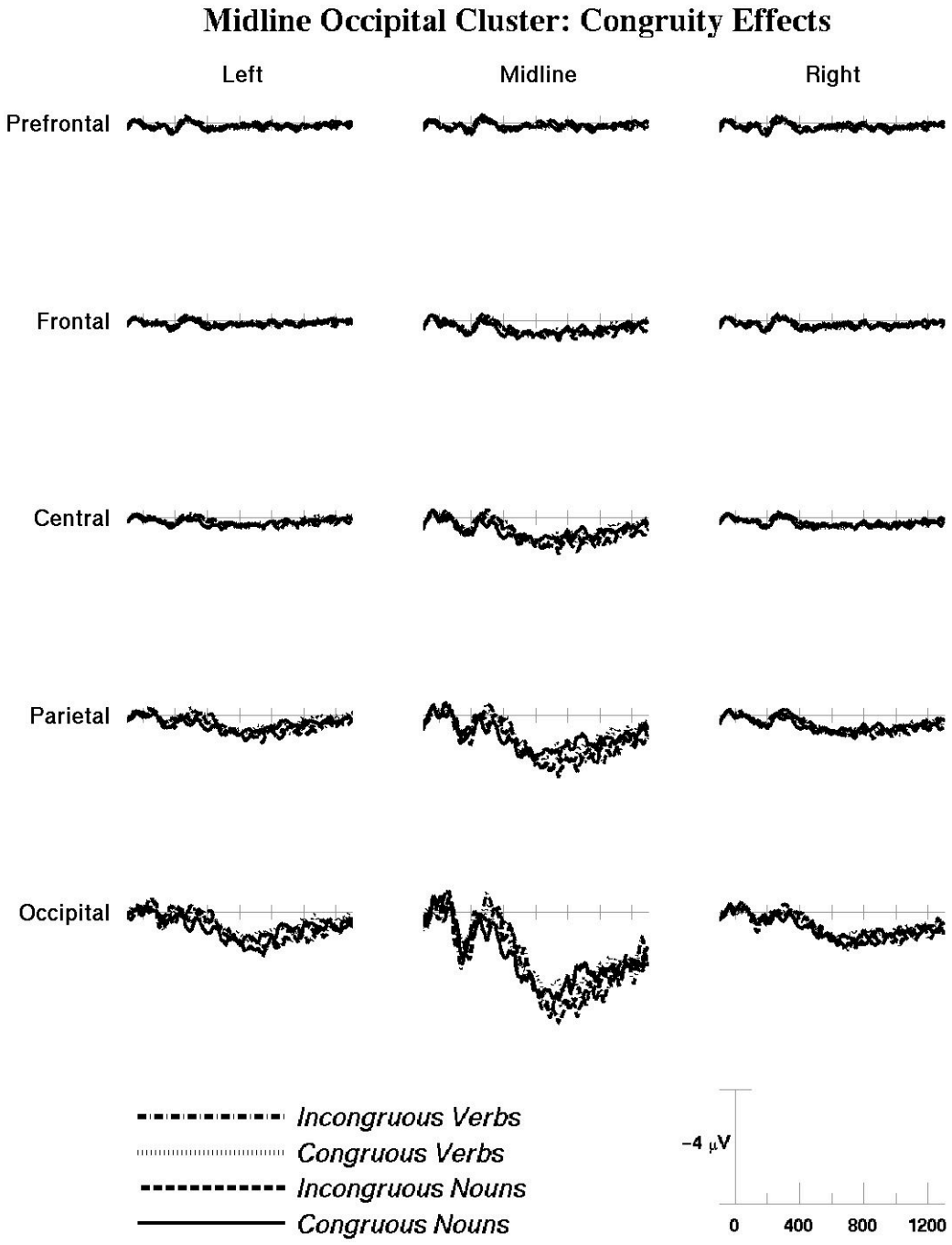

Figure VI.97 – Example grand average scalp-projected event-related activations (ERAs) to congruous and incongruous sentence endings of the midline occipital cluster. Averages are constructed from ten participants.

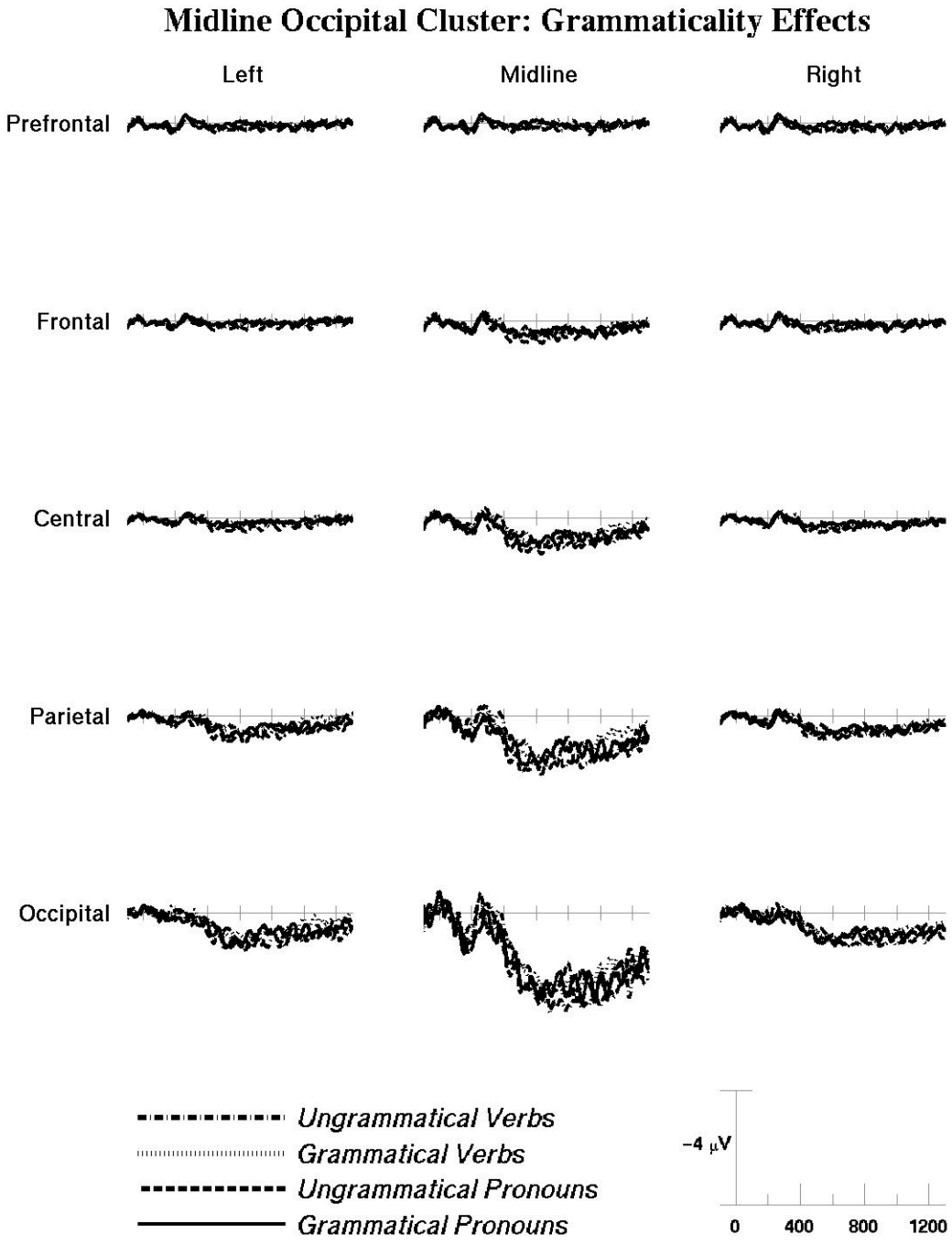

Figure VI.98 – Example grand average scalp-projected event-related activations (ERAs) to grammatical and ungrammatical sentence endings of the midline occipital cluster. Averages are constructed from ten participants.

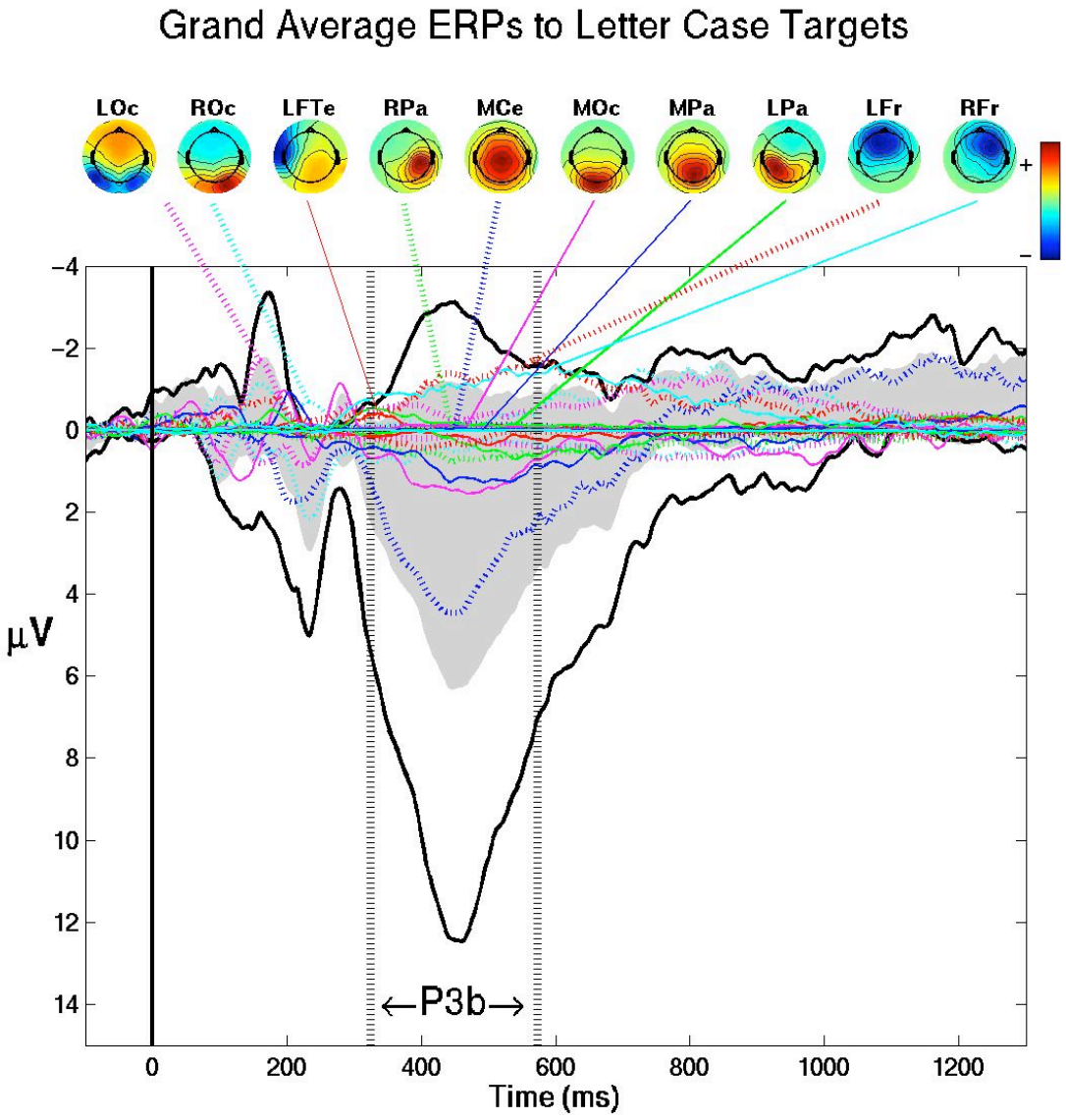

Figure VI.99 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to targets in the letter case oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

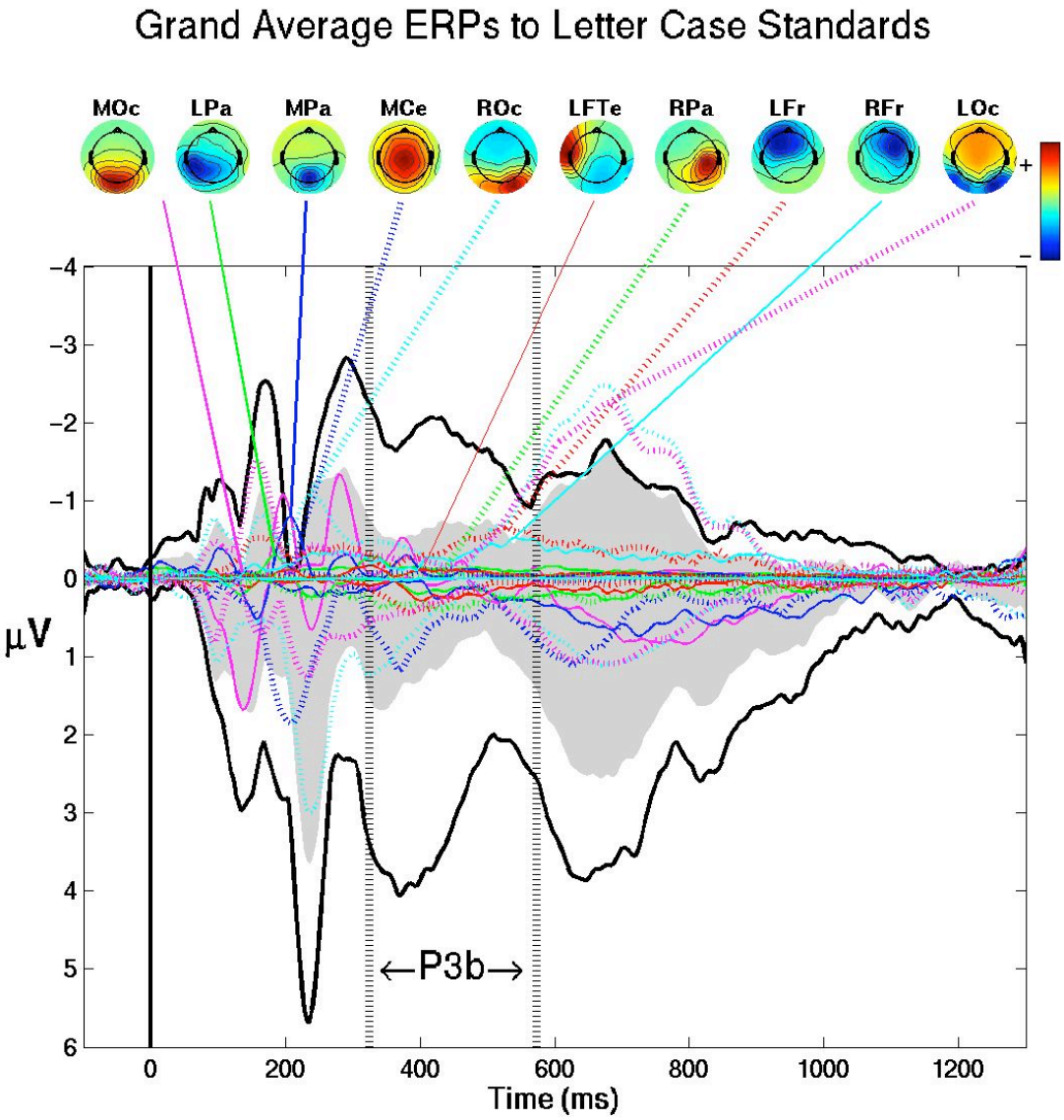

Figure VI.100 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to standards in the letter case oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

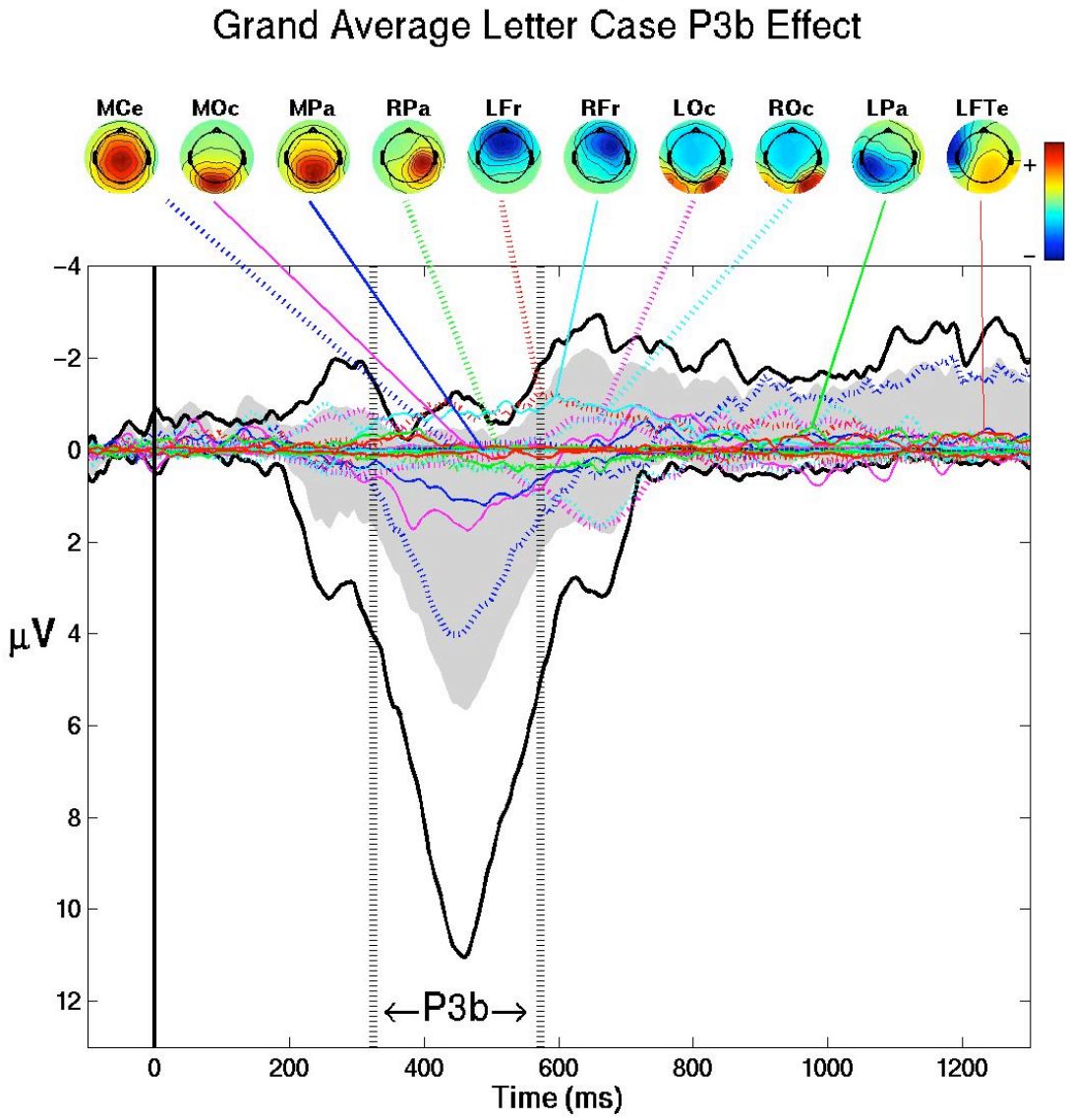

Figure VI.101 – Envtopo plots of each common IC cluster's contribution to the grand average P3b effect (ERPs to targets minus ERPs to standards) in the letter case oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

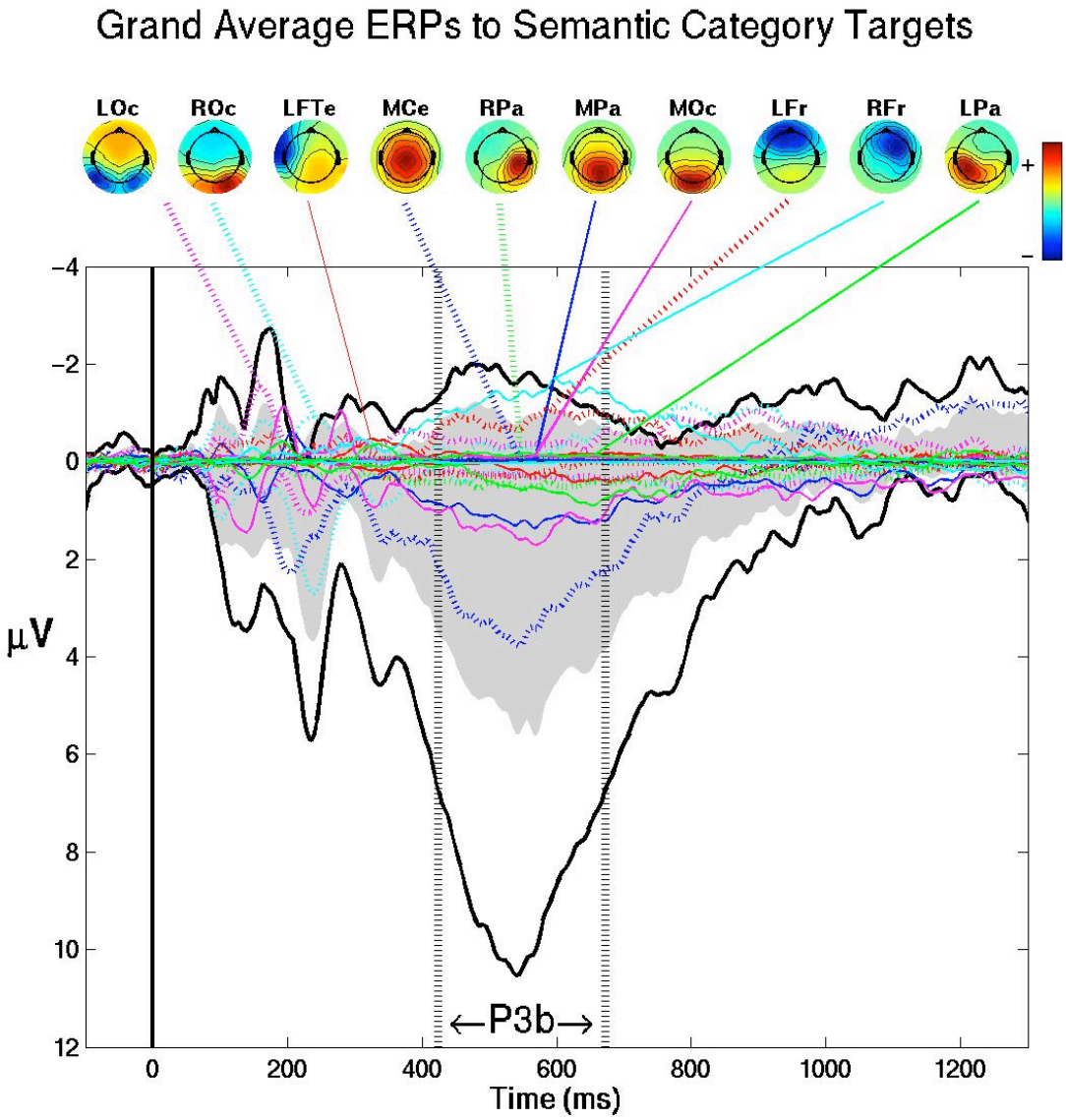

Figure VI.102 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to targets in the semantic category oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

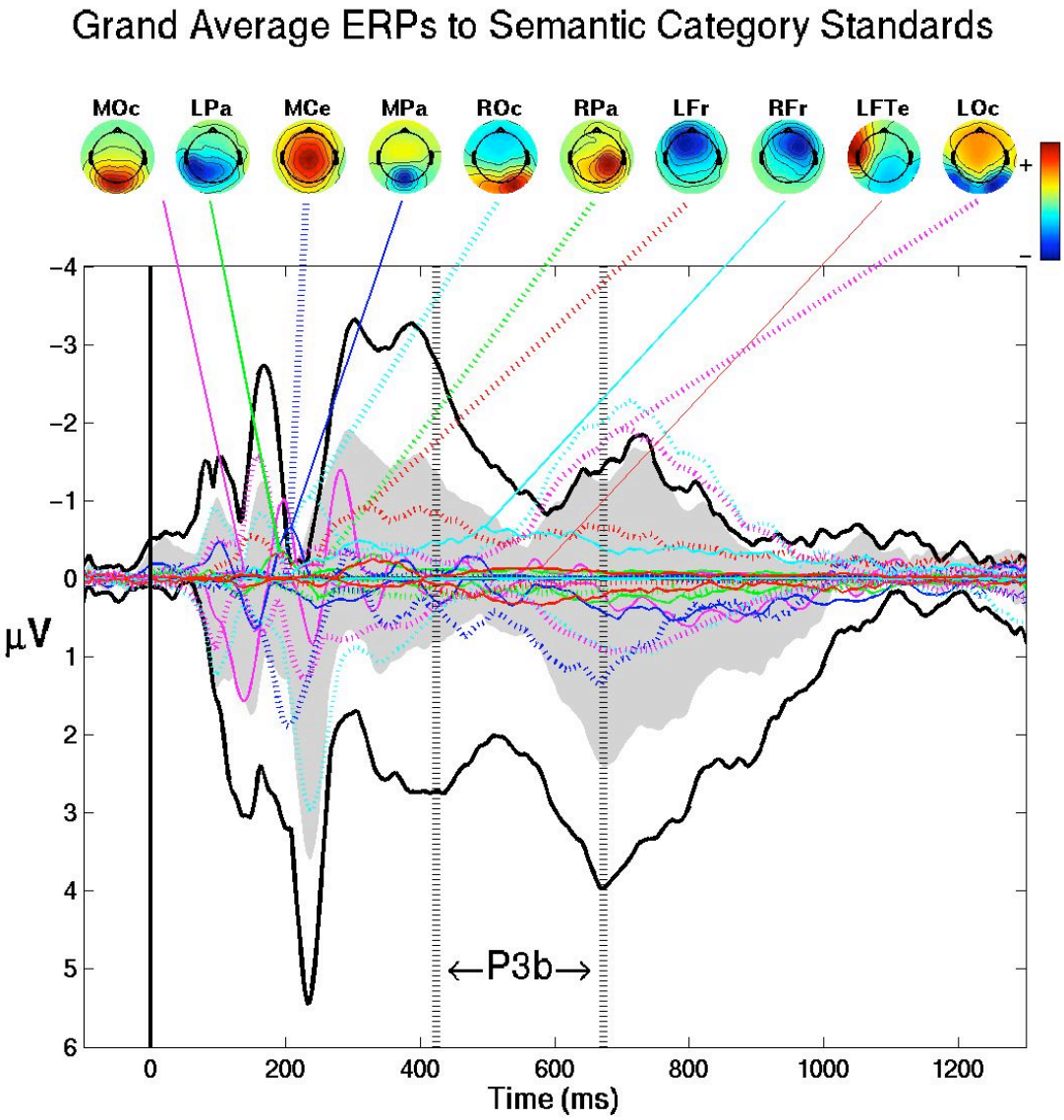

Figure VI.103 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to standards in the semantic category oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

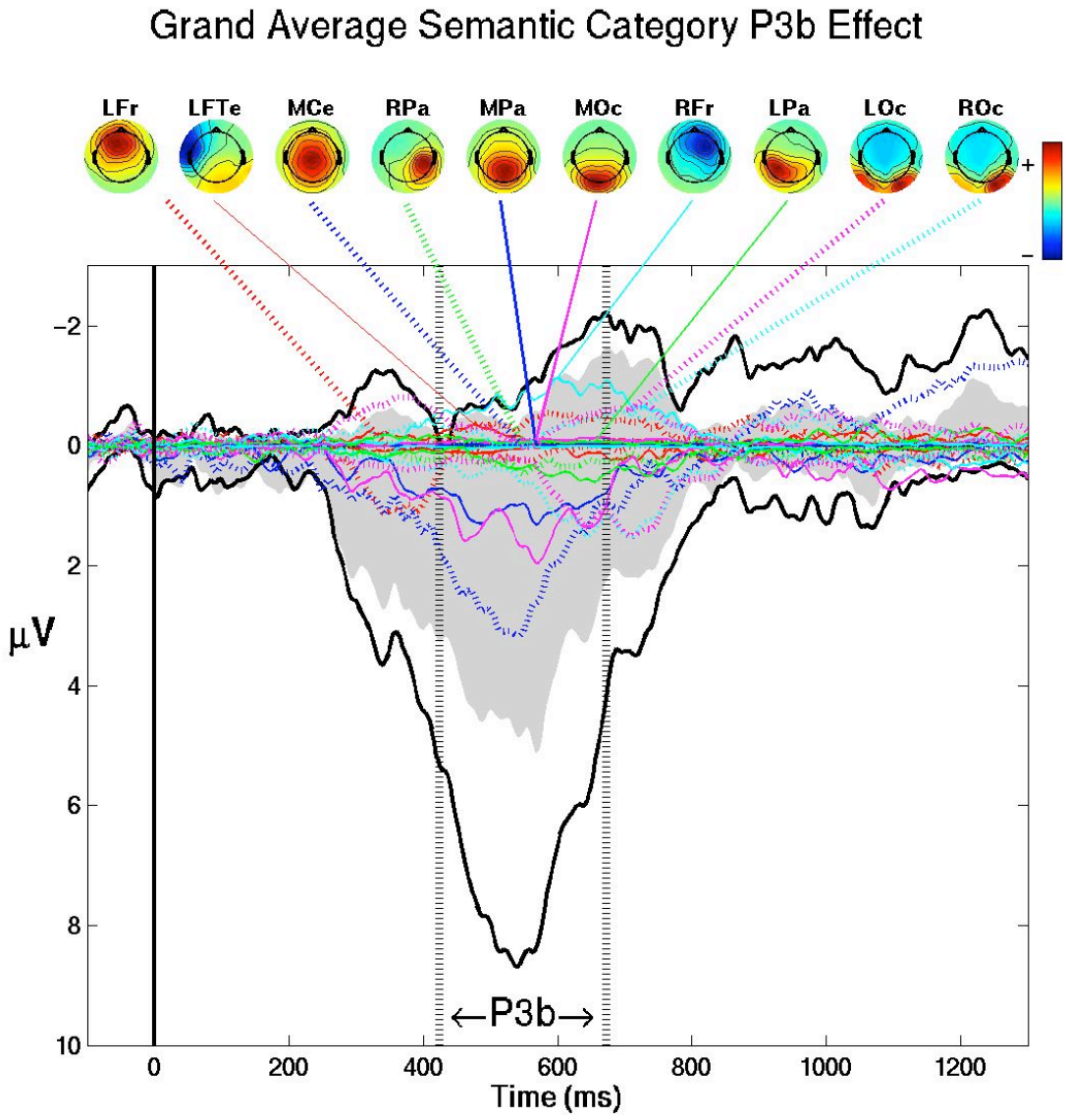

Figure VI.104 – Envtopo plots of each common IC cluster's contribution to the grand average P3b effect (ERPs to targets minus ERPs to standards) in the semantic category oddball task. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

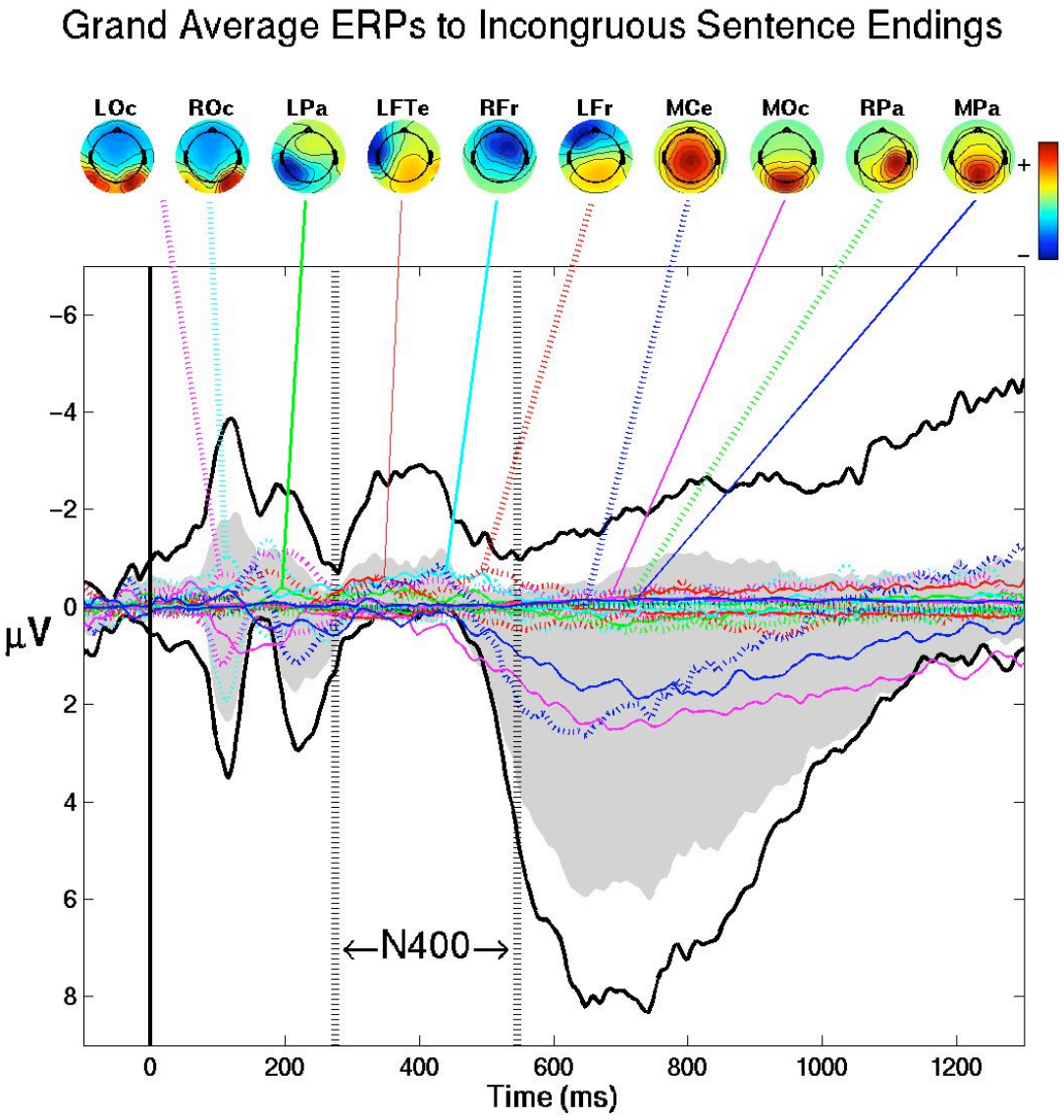

Figure VI.105 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to incongruous sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

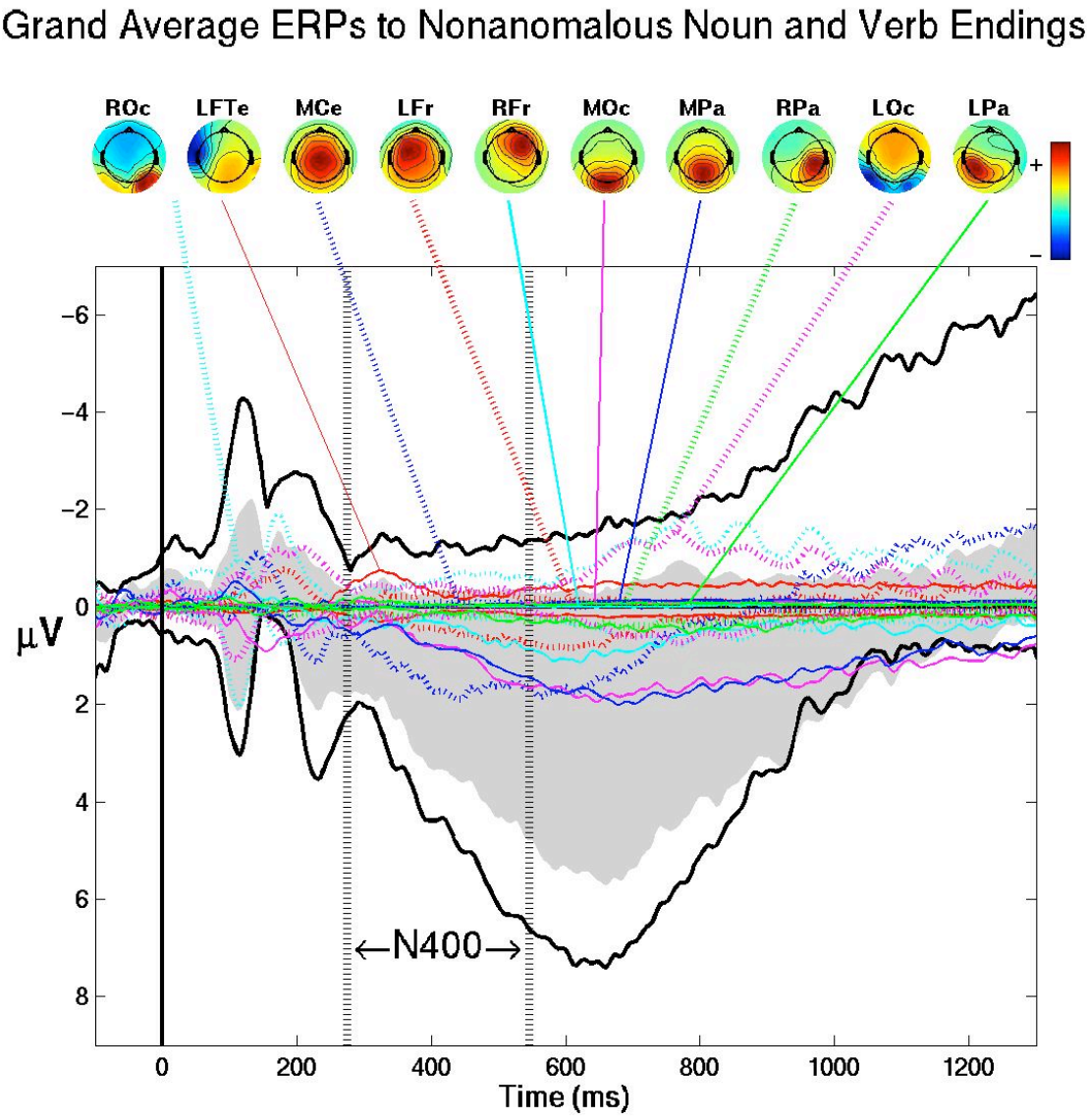

Figure VI.106 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to nonanomalous noun and verb sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

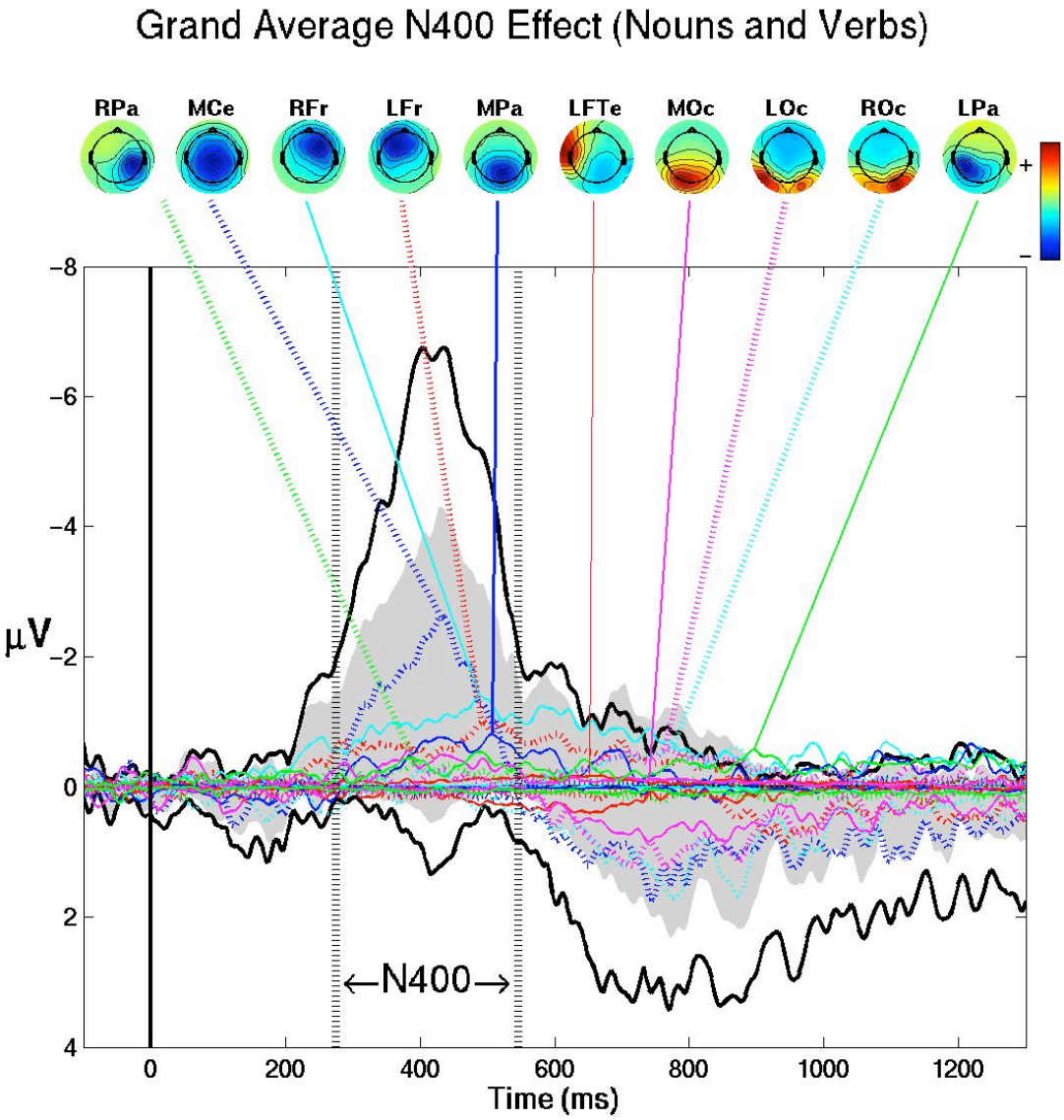

Figure VI.107 – Envtopo plots of each common IC cluster's contribution to the grand average N400 effect (ERPs to incongruous words minus ERPs to nonanomalous noun and verb sentence endings). Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the N400 window.

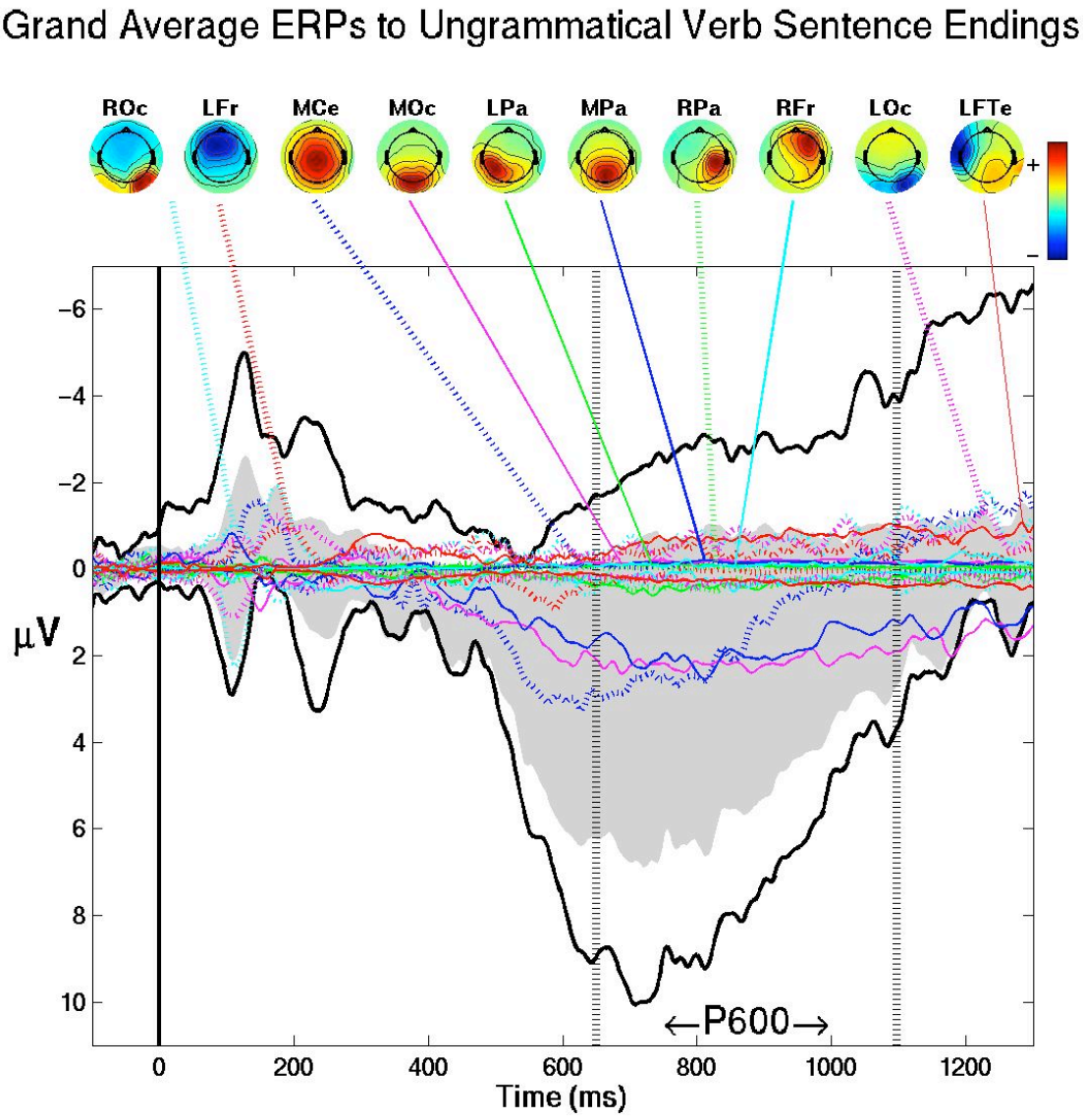

Figure VI.108 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to ungrammatical verb sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the verb P600 window.

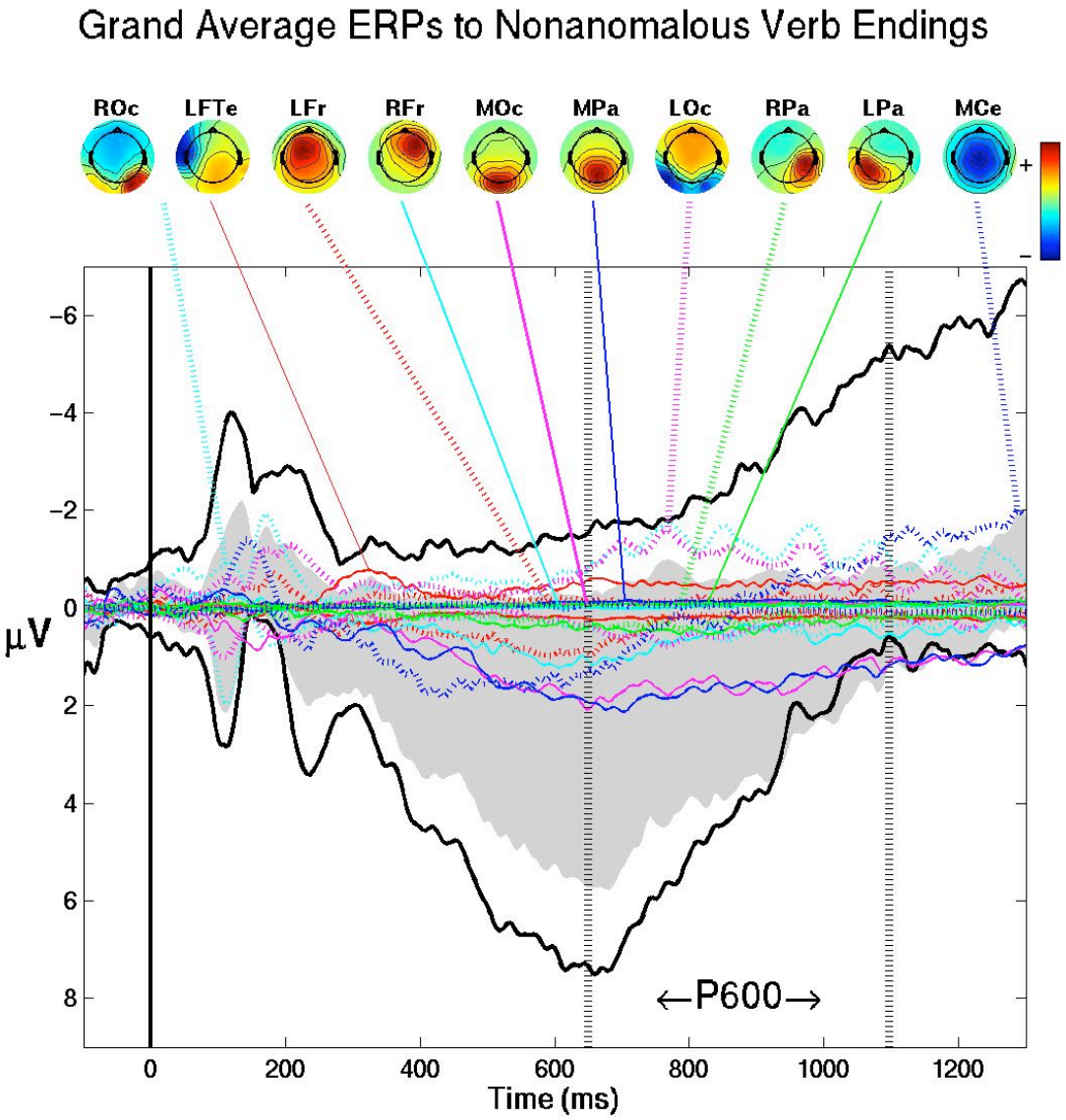

Figure VI.109 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to nonanomalous verb sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the verb P600 window.

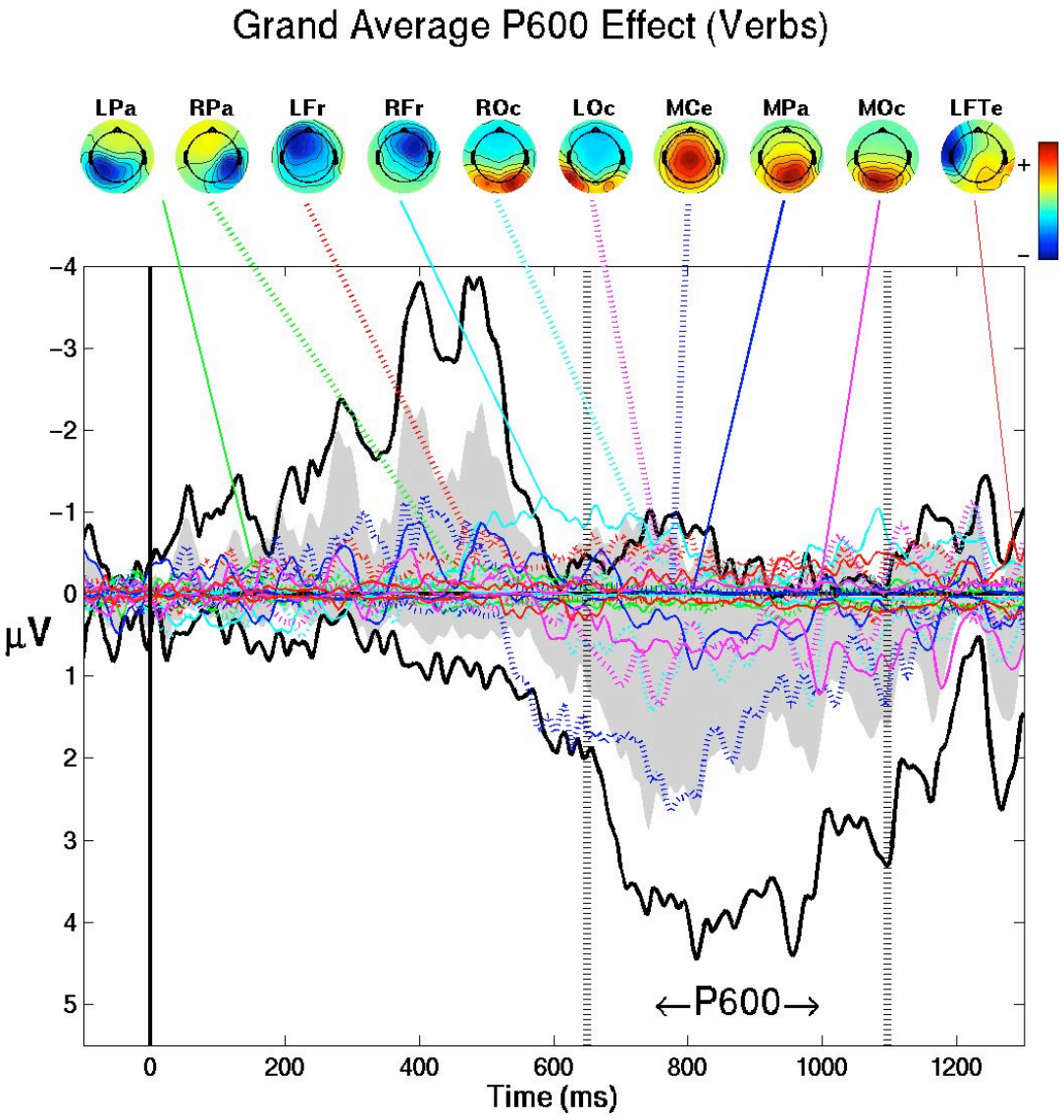

Figure VI.110 – Envtopo plots of each common IC cluster's contribution to the grand average verb P600 effect (ERPs to ungrammatical verbs minus ERPs to nonanomalous verb sentence endings). Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the verb P600 window.

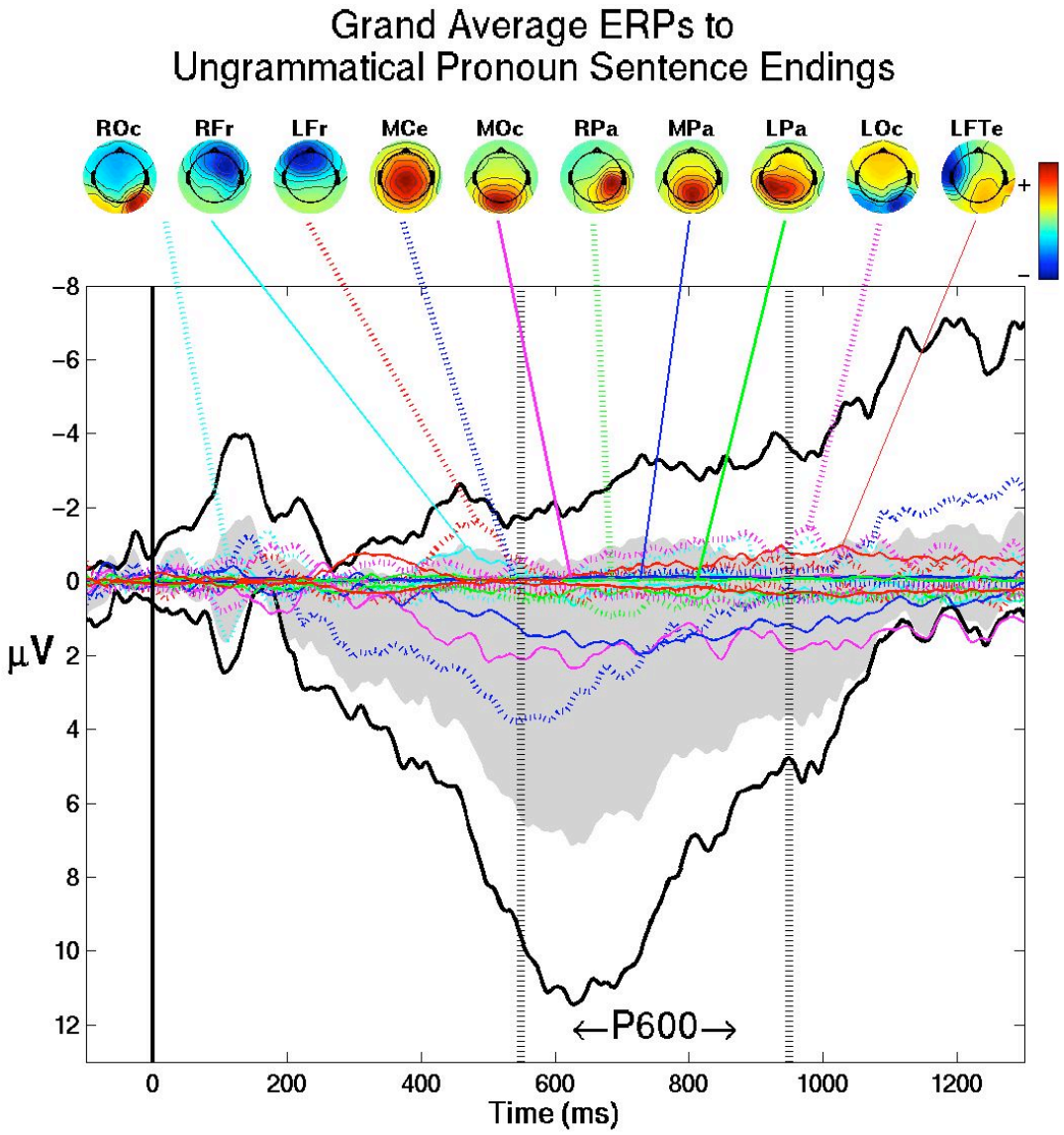

Figure VI.111 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to ungrammatical pronoun sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the pronoun P600 window.

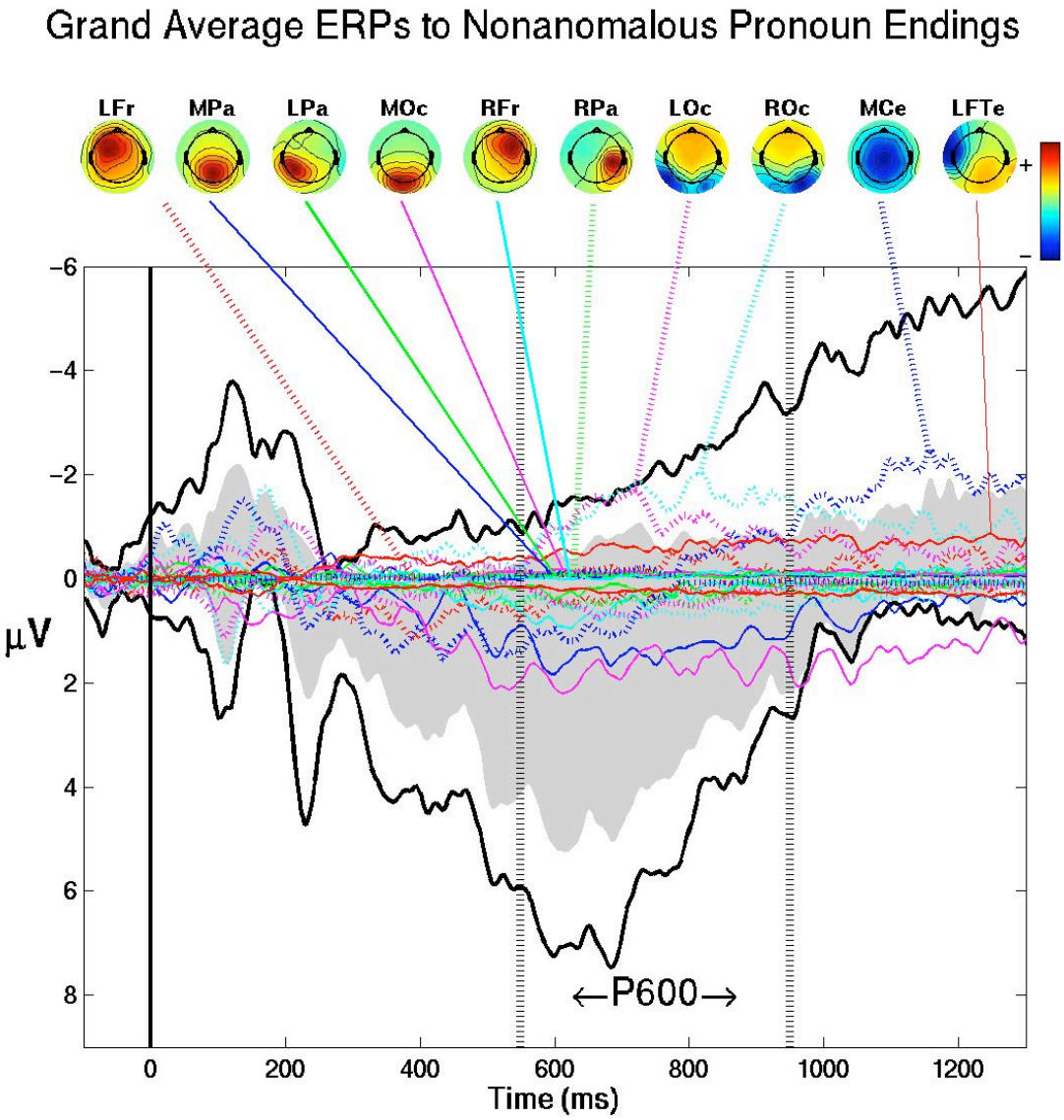

Figure VI.112 – Envtopo plots of each common IC cluster's contribution to the grand average ERPs to nonanomalous pronoun sentence endings. Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the pronoun P600 window.

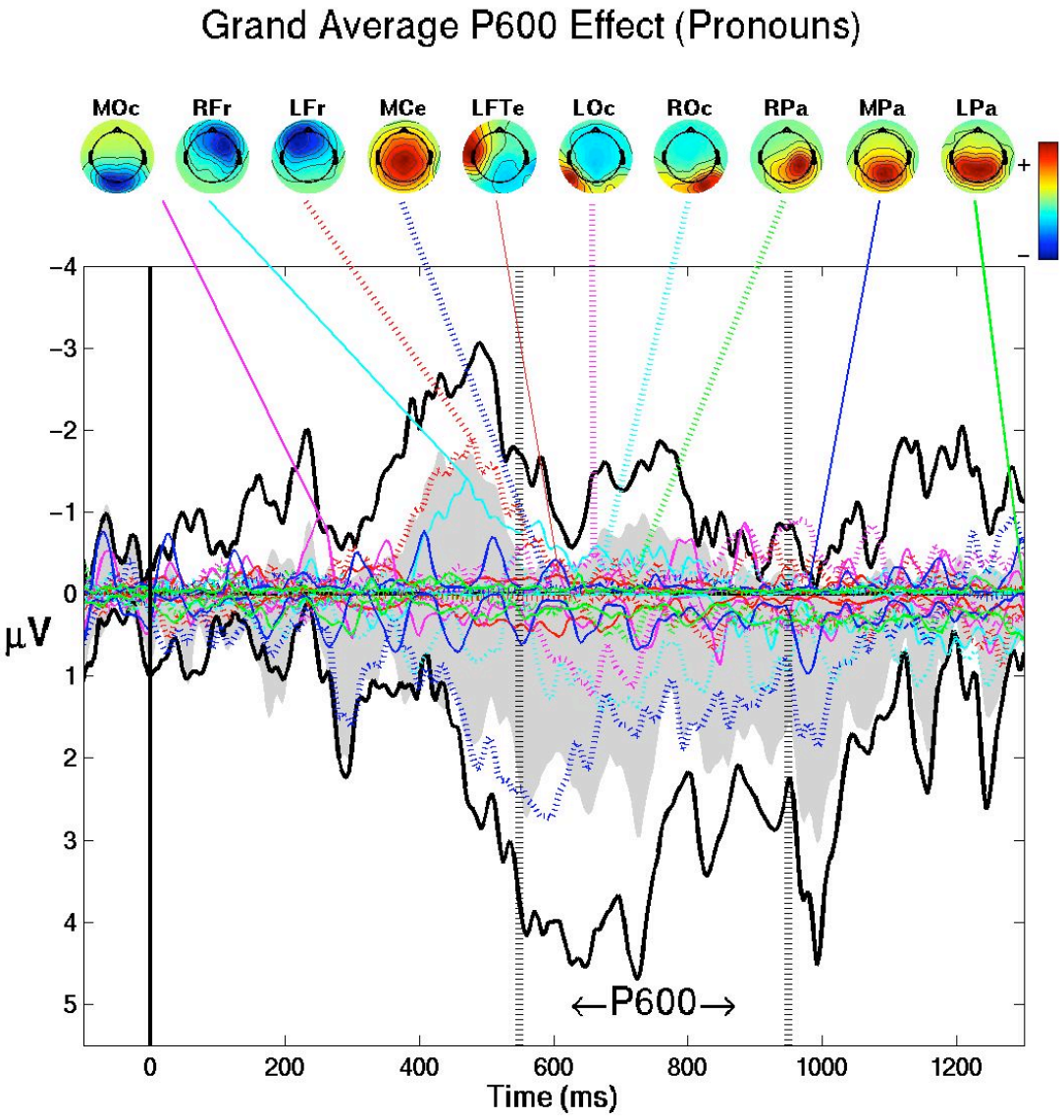

Figure VI.113 – Envtopo plots of each common IC cluster's contribution to the grand average pronoun P600 effect (ERPs to ungrammatical pronouns minus ERPs to nonanomalous pronoun sentence endings). Black waveforms trace the maximum and minimum value of the grand average ERPs at all electrodes. Colored waveforms trace the maximum and minimum value of each cluster's contribution to the grand average at all electrodes. The topography of each cluster's contribution at the time point of maximal absolute amplitude is plotted above the waveforms. Grey shading traces the total contribution of all common ICs. Vertical dashed lines indicate the pronoun P600 window.

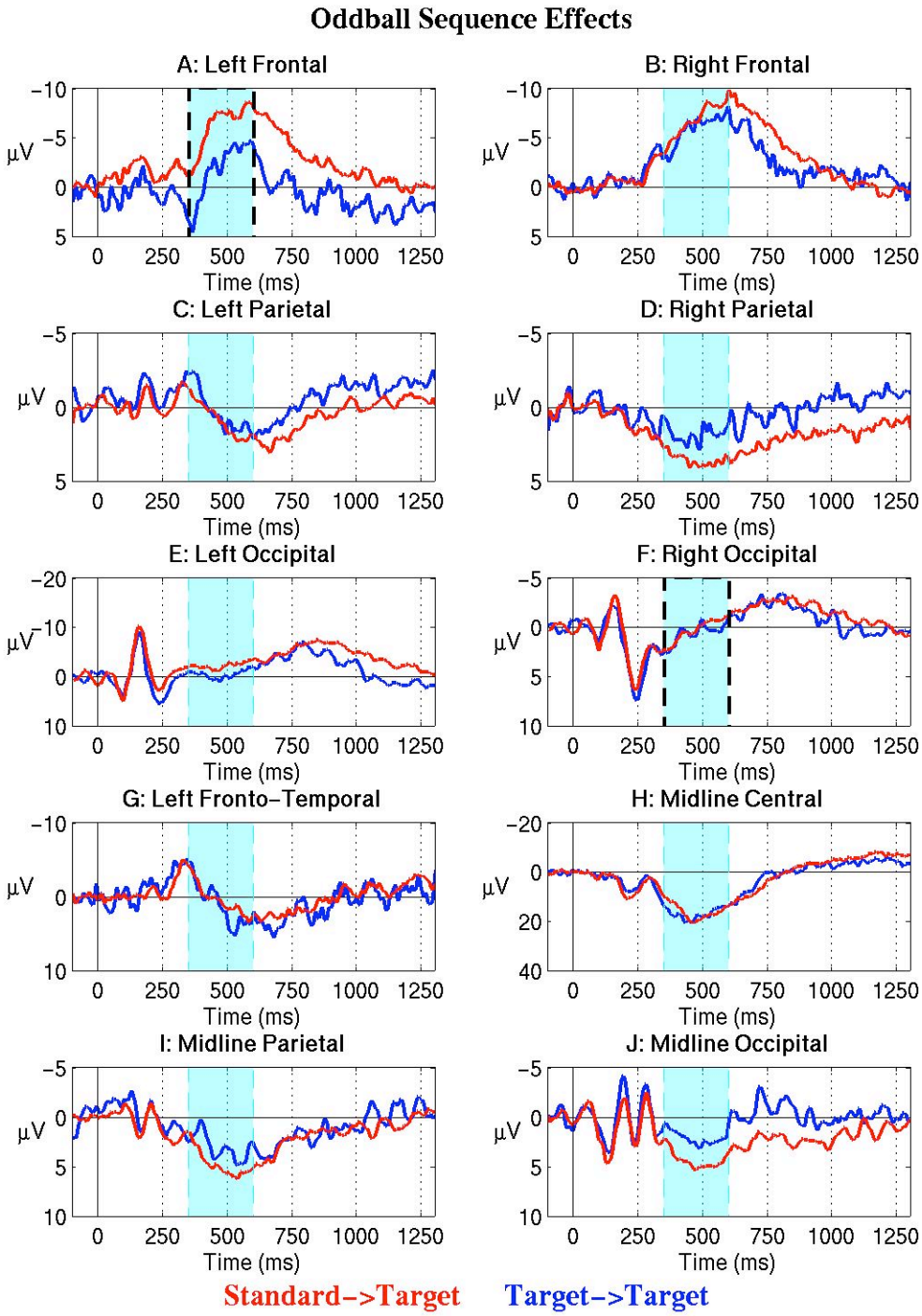

Figure VI.114 – Mean IC cluster event-related activations (ERAs) to targets in both oddball tasks. ERAs to targets preceded by a standard are plotted in red. ERAs to targets preceded by another target are plotted in blue. Light blue shading indicates window used for statistical analysis (350-600 ms). Windows bordered with black dotted lines indicate a significant difference between conditions.

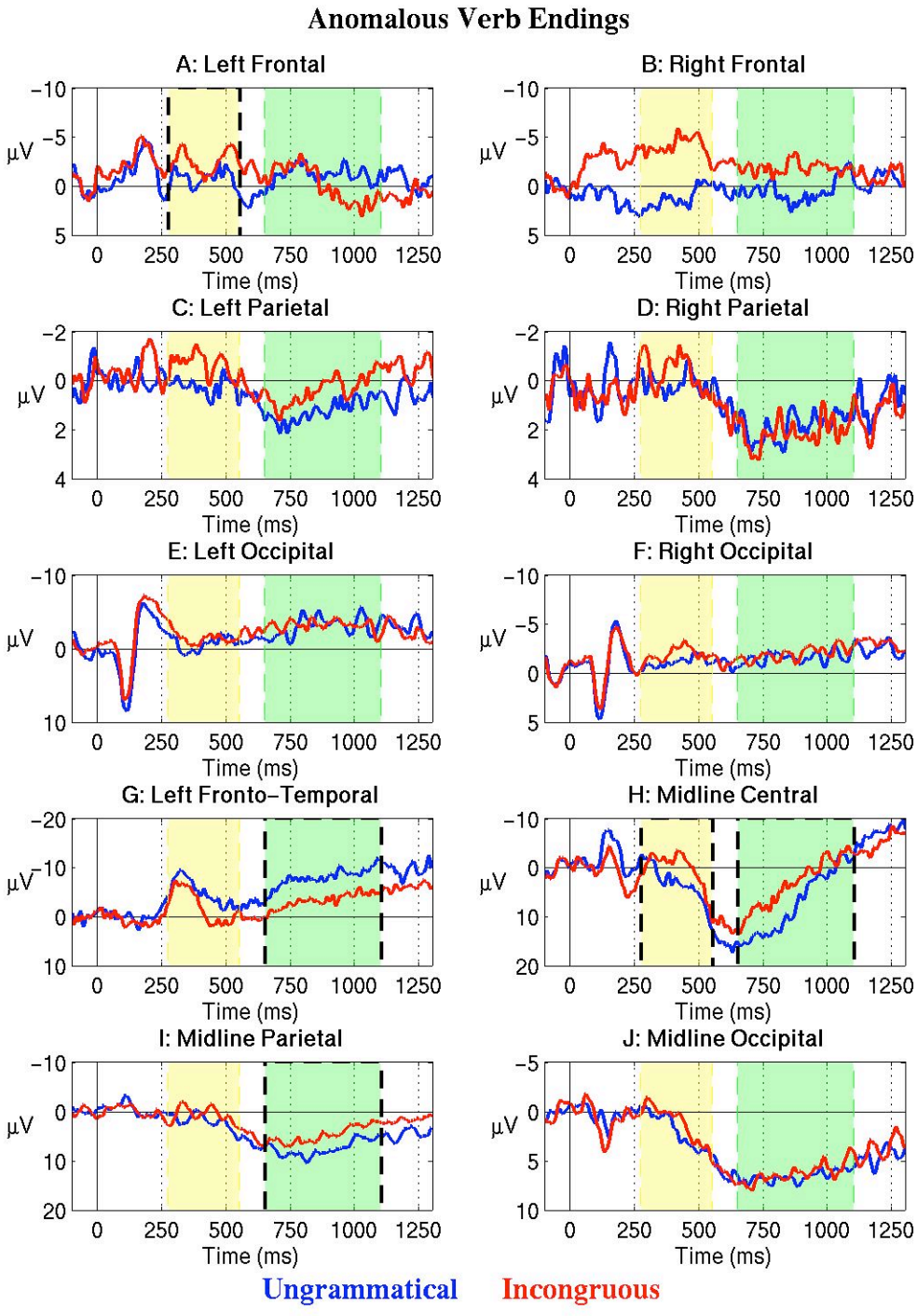

Figure VI.115 – Mean IC cluster event-related activations (ERAs) to anomalous verb endings. Yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

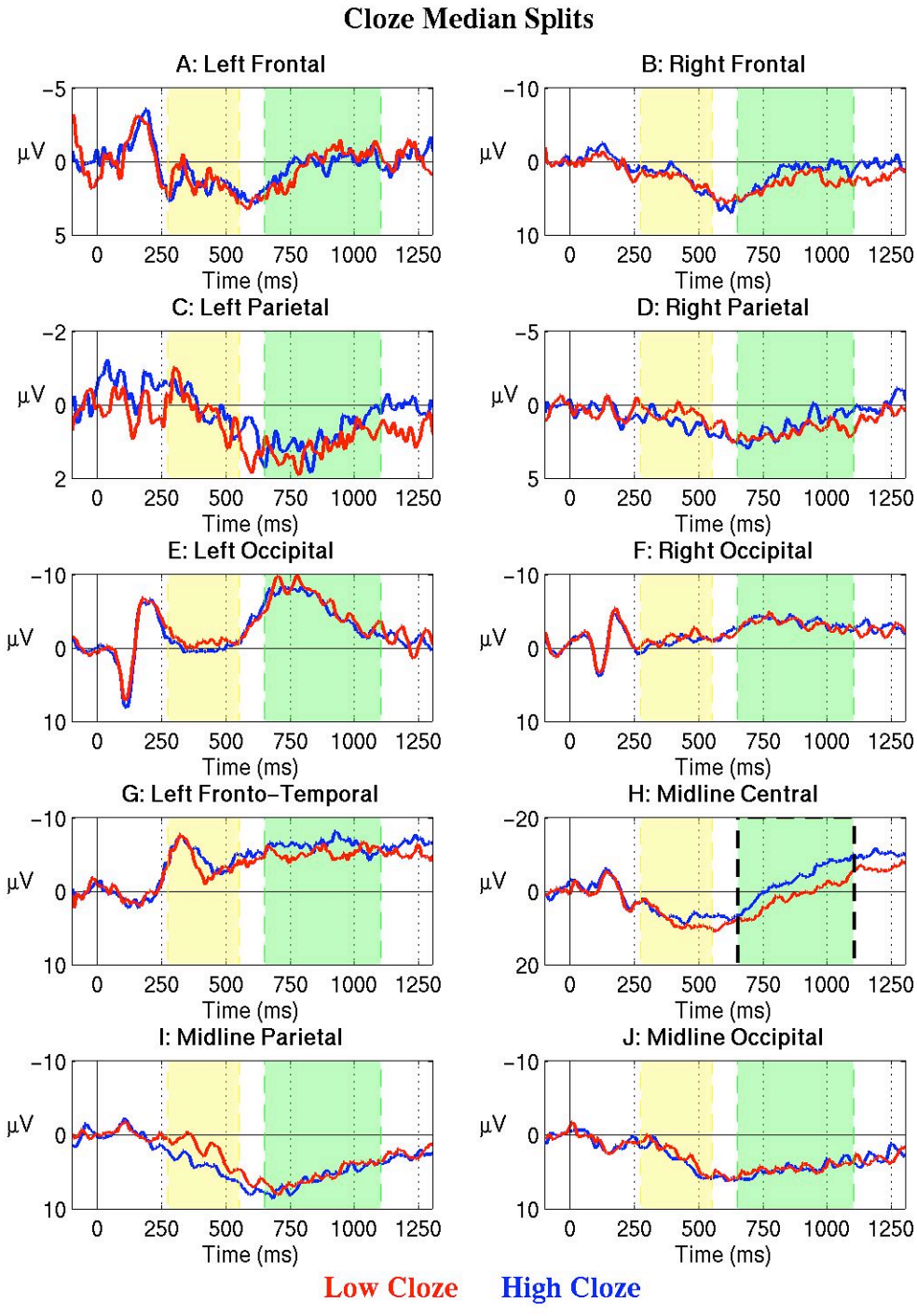

Figure VI.116 – Mean IC cluster event-related activations (ERAs) to nonanomalous sentence endings (nouns, pronouns, & verbs). ERAs to low cloze endings are plotted in red. ERAs to high cloze endings are plotted in blue. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

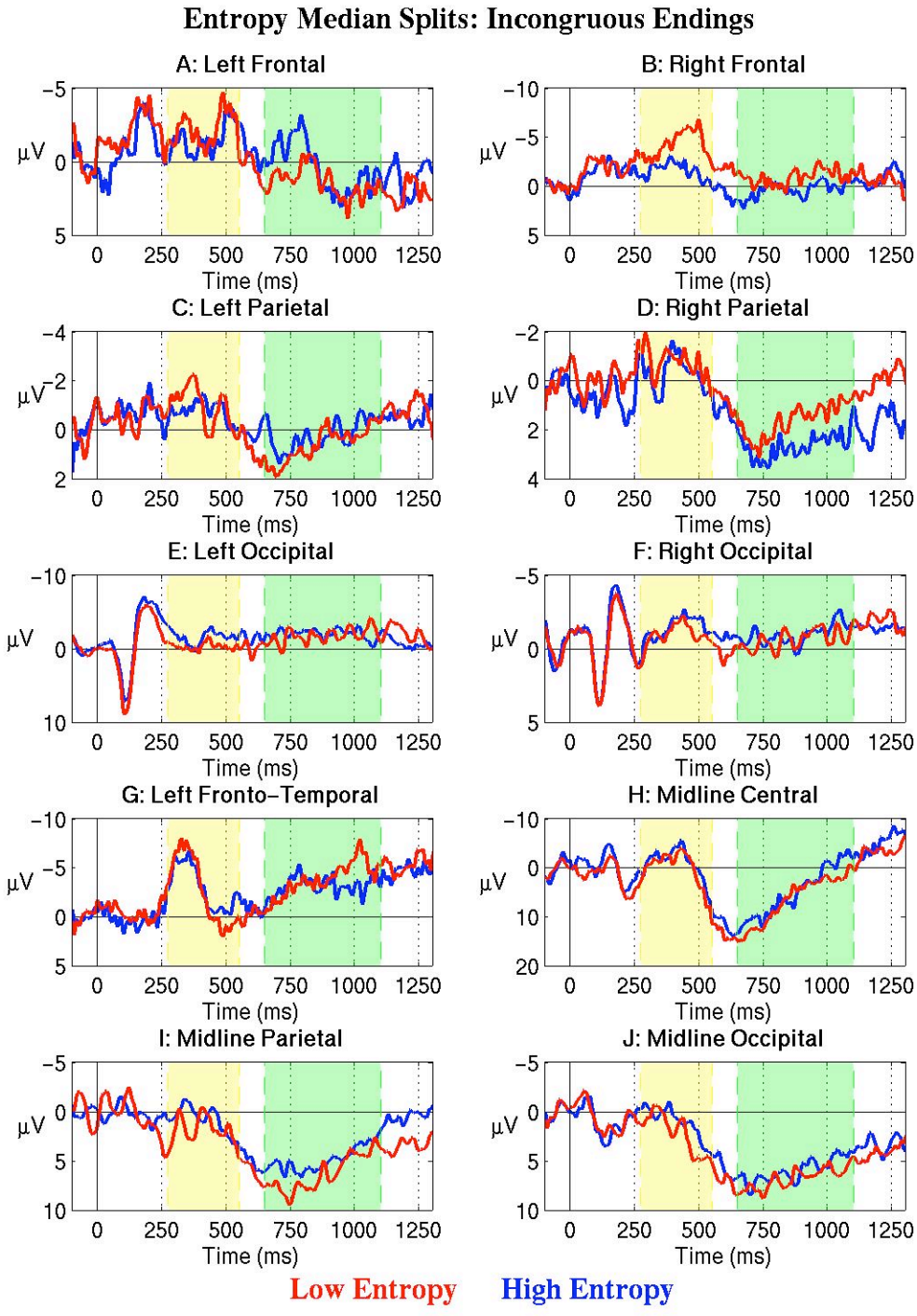

Figure VI.117 – Mean IC cluster event-related activations (ERAs) to incongruous sentence endings. ERAs to low entropy endings are plotted in red. ERAs to high entropy endings are plotted in blue. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively.

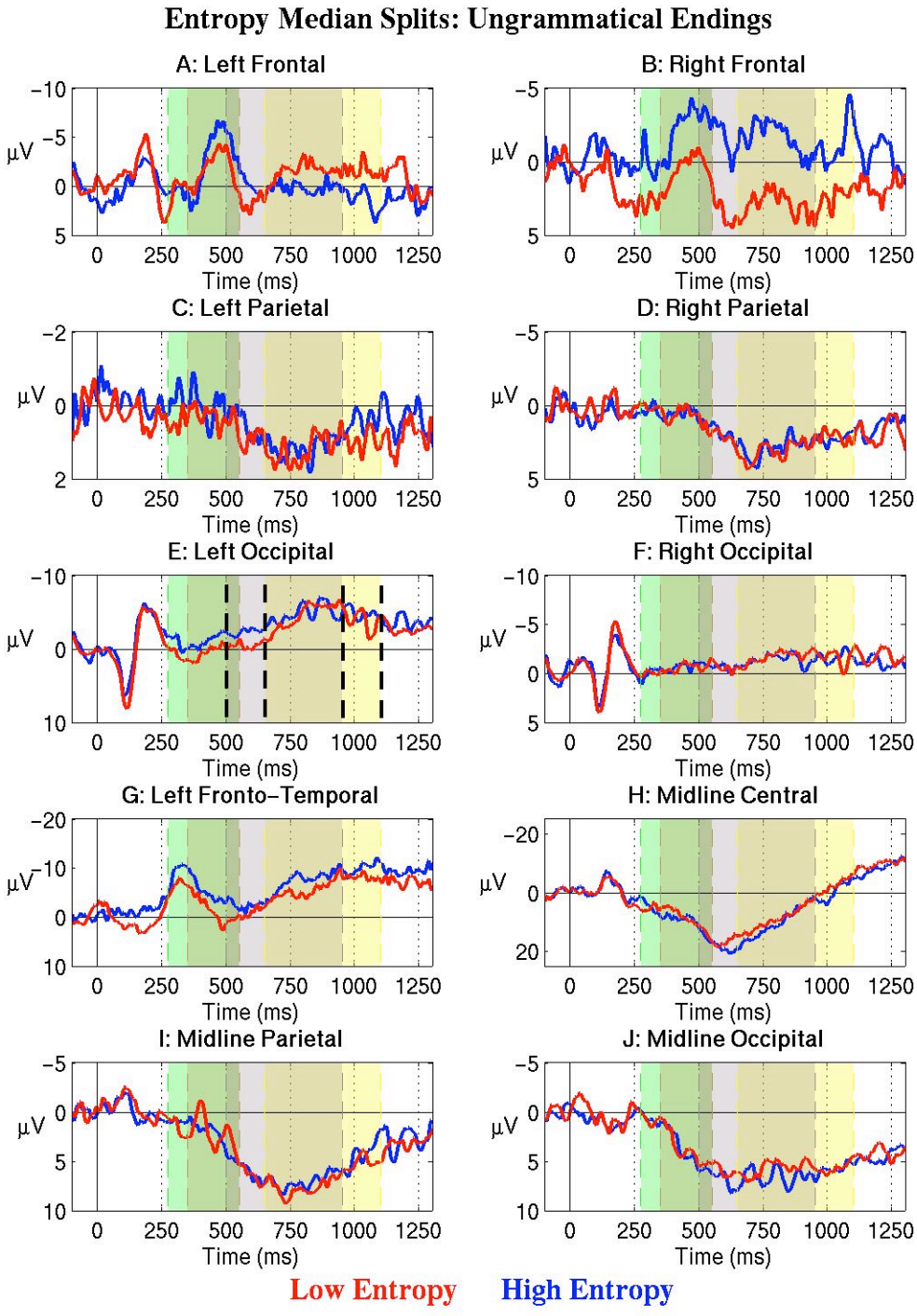

Figure VI.118 – Mean IC cluster event-related activations (ERAs) to ungrammatical sentence endings. Yellow and green shading indicate verb N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Brown and grey shading indicate pronoun N400 (350-550 ms) and P600 (500-950 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

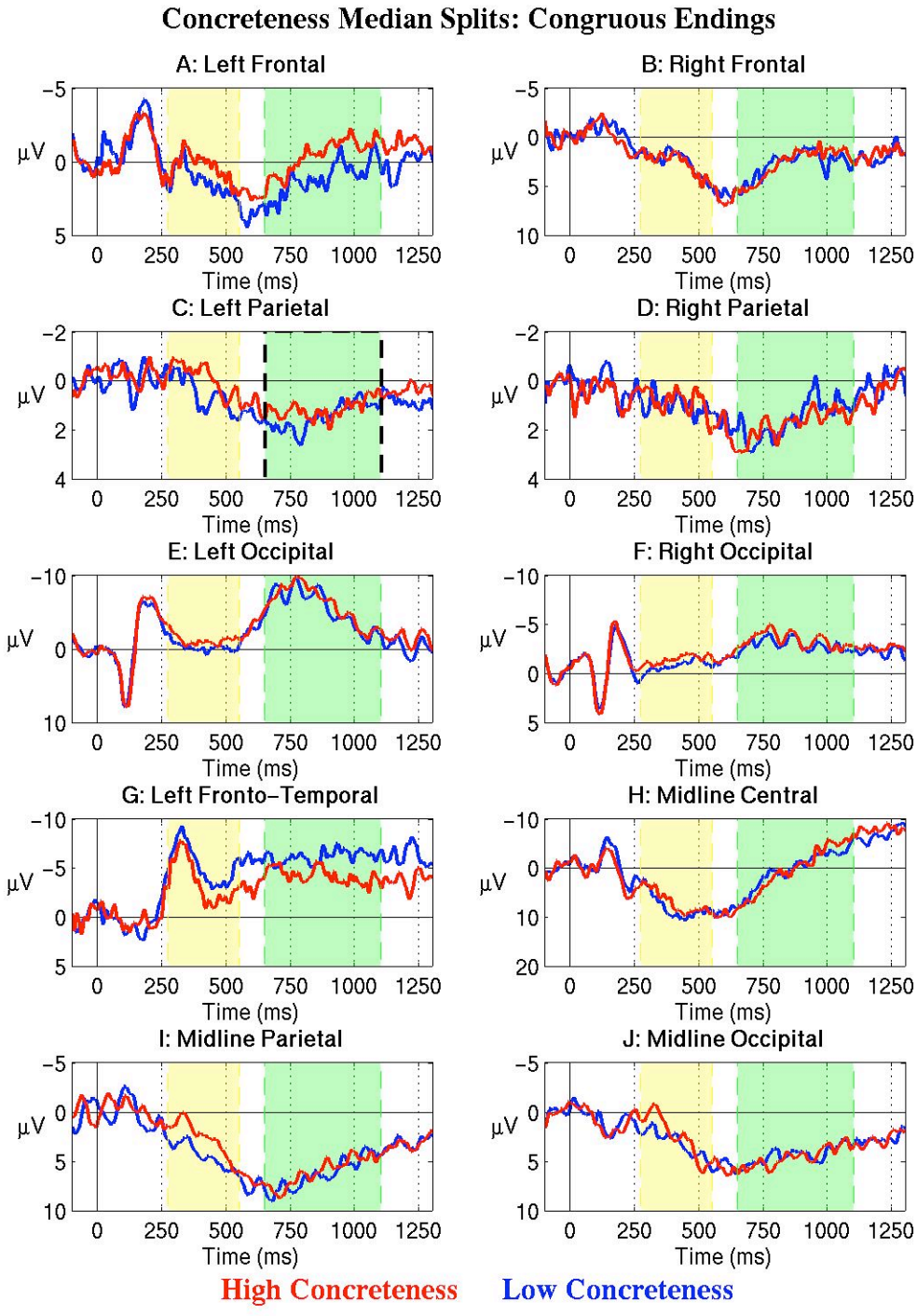

Figure VI.119 – Mean IC cluster event-related activations (ERAs) to congruous noun and verb sentence endings. ERAs to high concreteness endings are plotted in red. ERAs to low concreteness endings are plotted in blue. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

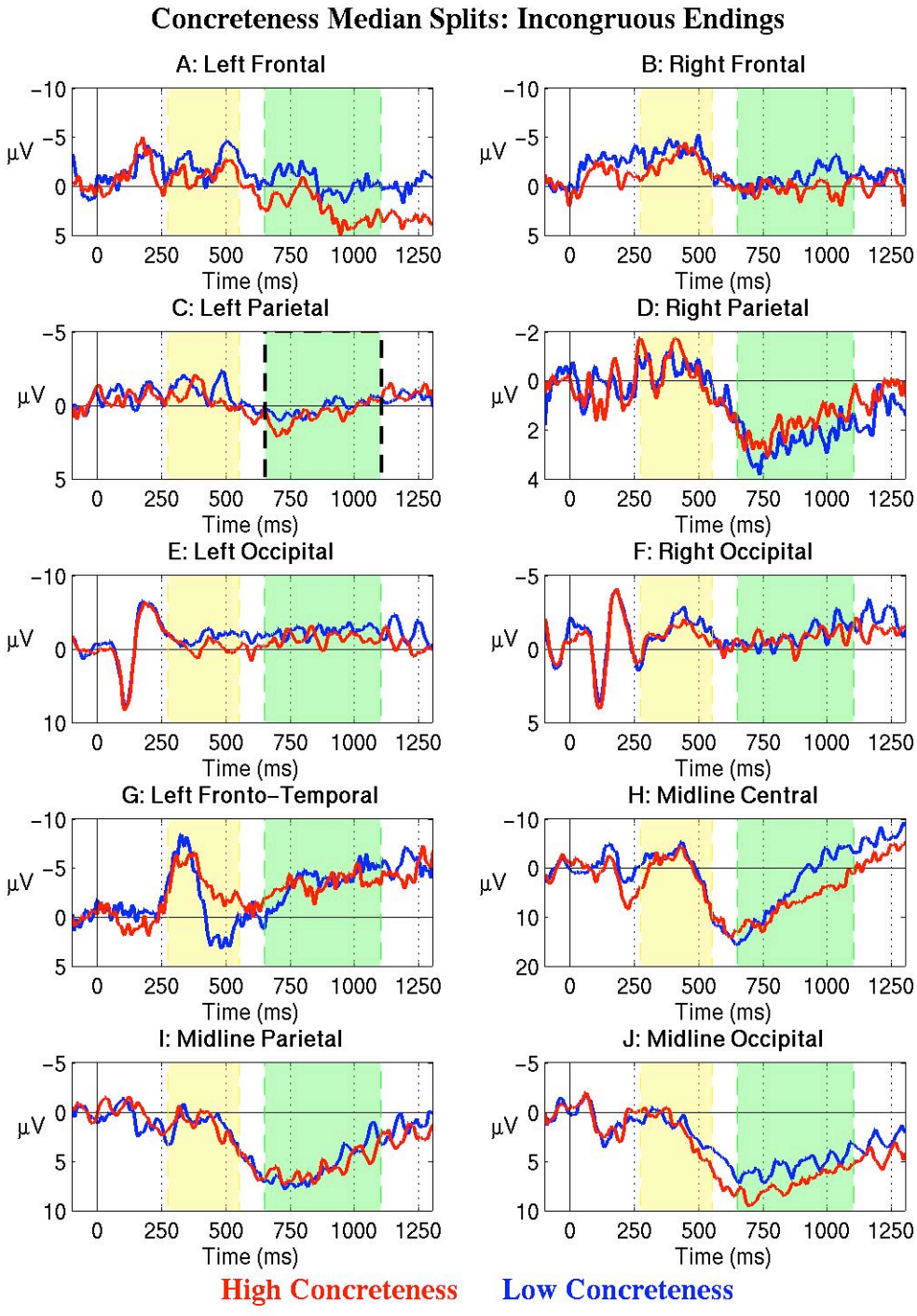

Figure VI.120 – Mean IC cluster event-related activations (ERAs) to incongruous noun and verb sentence endings. ERAs to high concreteness endings are plotted in red. ERAs to low concreteness endings are plotted in blue. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

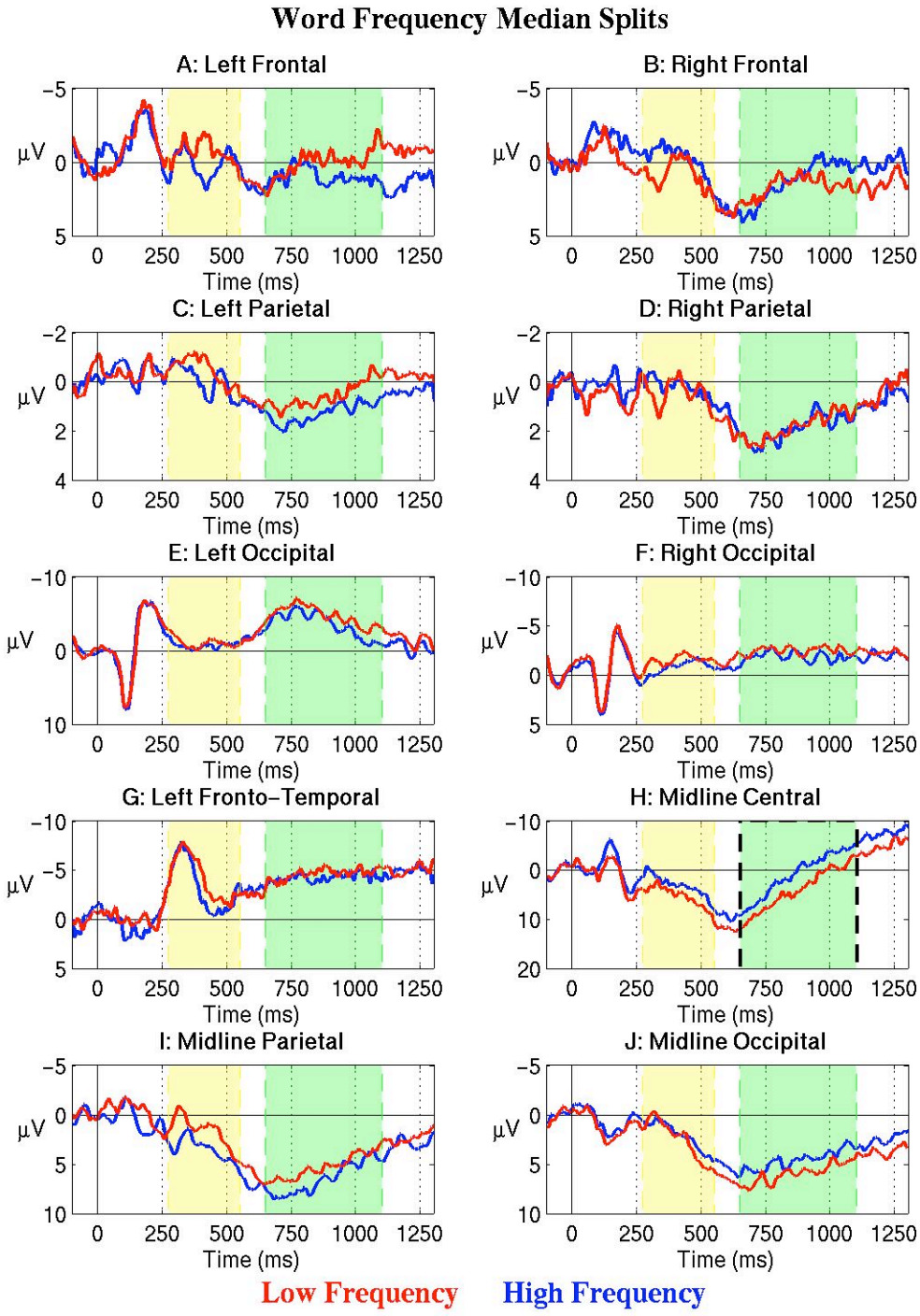

Figure VI.121 – Mean IC cluster event-related activations (ERAs) to congruous and incongruous noun and verb sentence endings after splitting trials by word frequency. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

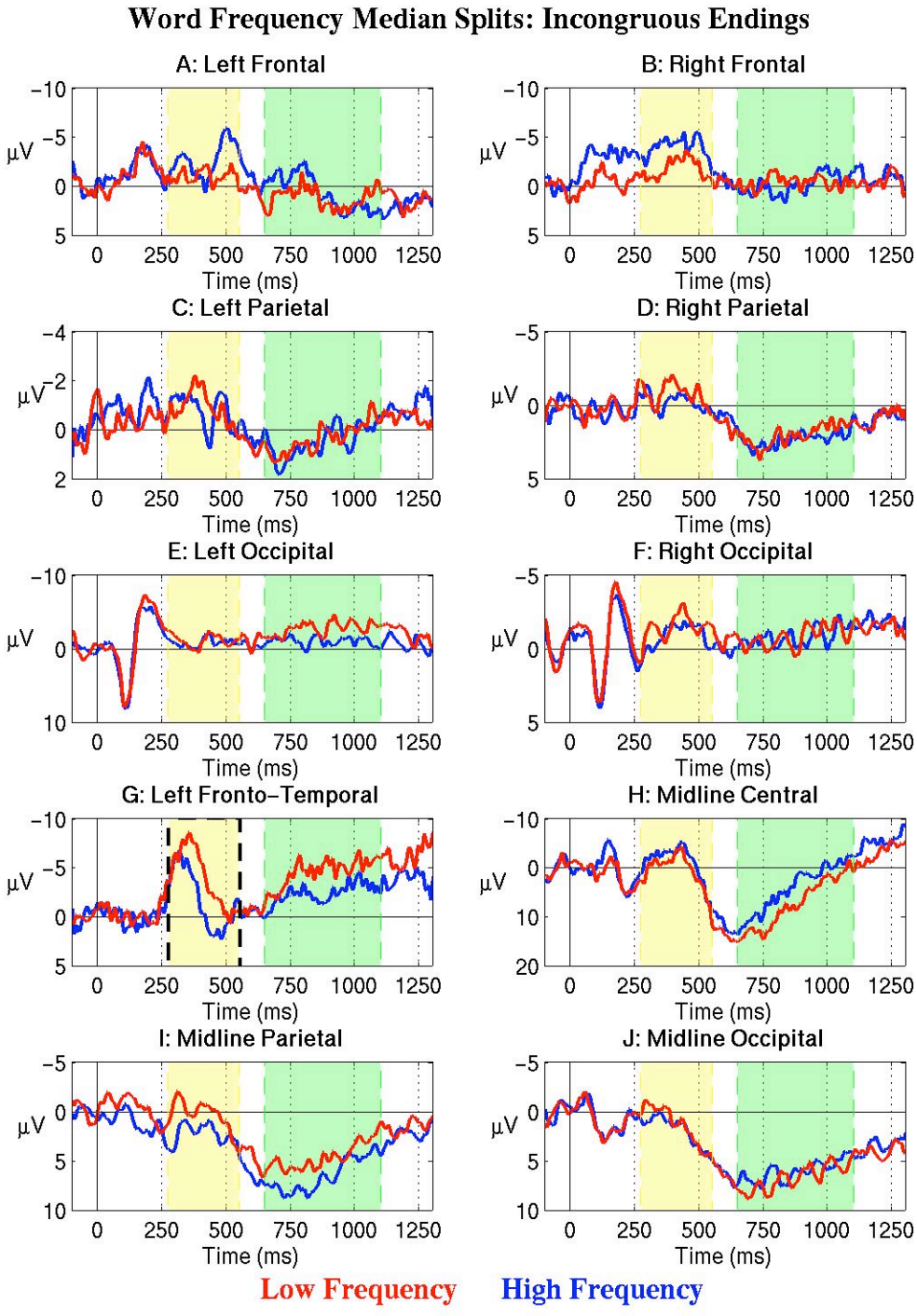

Figure VI.122 – Mean IC cluster event-related activations (ERAs) to incongruous noun and verb sentence endings after splitting trials by word frequency. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

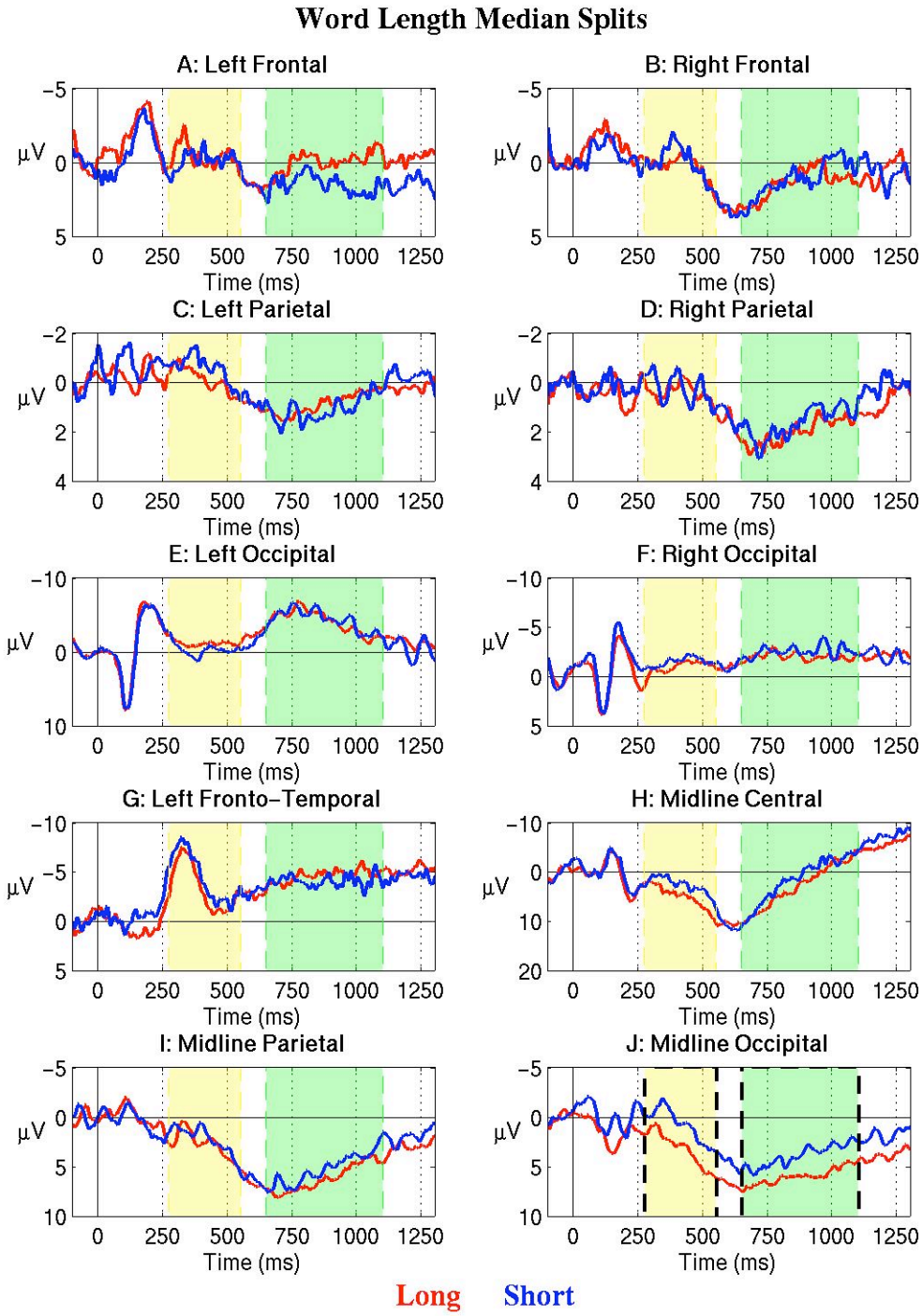

Figure VI.123 – Mean IC cluster event-related activations (ERAs) to congruous and incongruous noun and verb sentence endings after splitting trials by word length. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

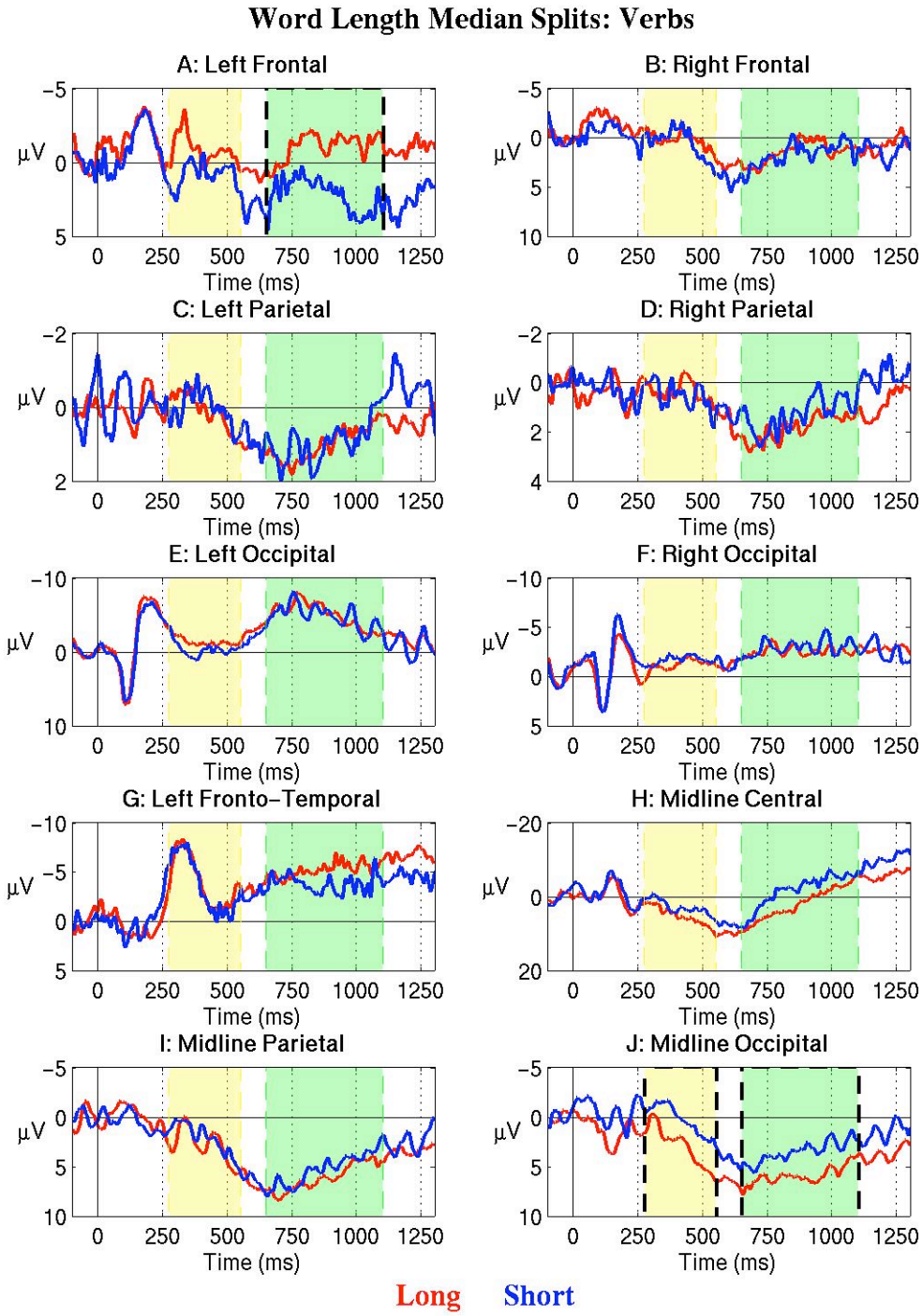

Figure VI.124 – Mean IC cluster event-related activations (ERAs) to congruous and incongruous verb sentence endings after splitting trials by word length. Light yellow and green shading indicate N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Windows bordered with black dotted lines indicate a significant difference between conditions.

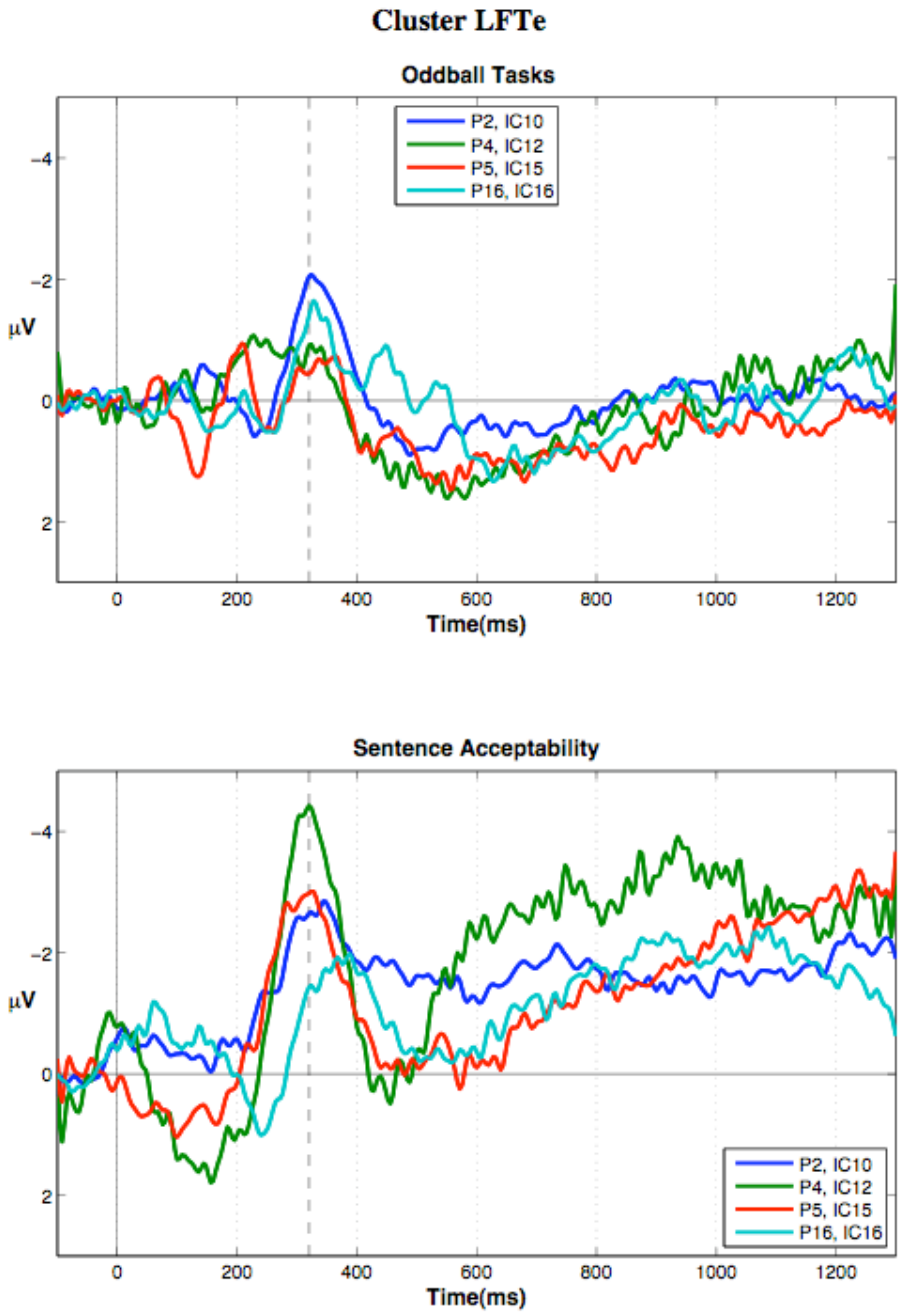

Figure VI.125 – Scalp-projected event-related activations (ERAs) of all ICs in the left fronto-temporal cluster (LFTe). (**Top**) ERAs to all standards and targets in both oddball tasks. (**Bottom**) ERAs to all sentence final words in the sentence acceptability task. Vertical dashed line marks 320 ms.

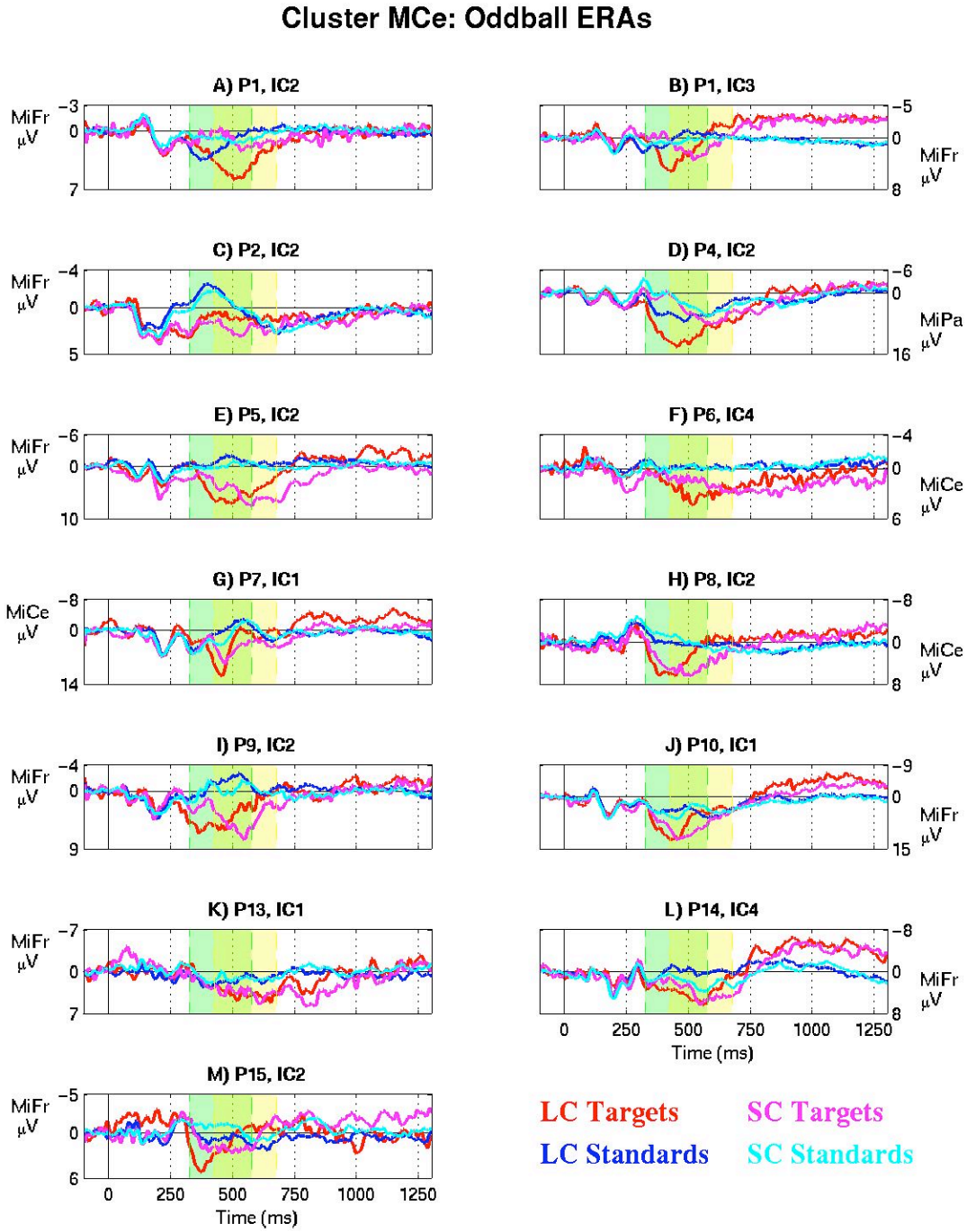

Figure VI.126 – Scalp-projected oddball task event-related activations (ERAs) of each IC in the midline central cluster (MCe). ERAs are plotted at the electrode at which the absolute IC topography is maximum (MiFr, MiCe, or MiPa). Green and yellow shading indicate LC and SC task P3b windows respectively. Abbreviations: P=Participant, IC=Independent Component, LC=Letter Case, SC=Semantic Category.

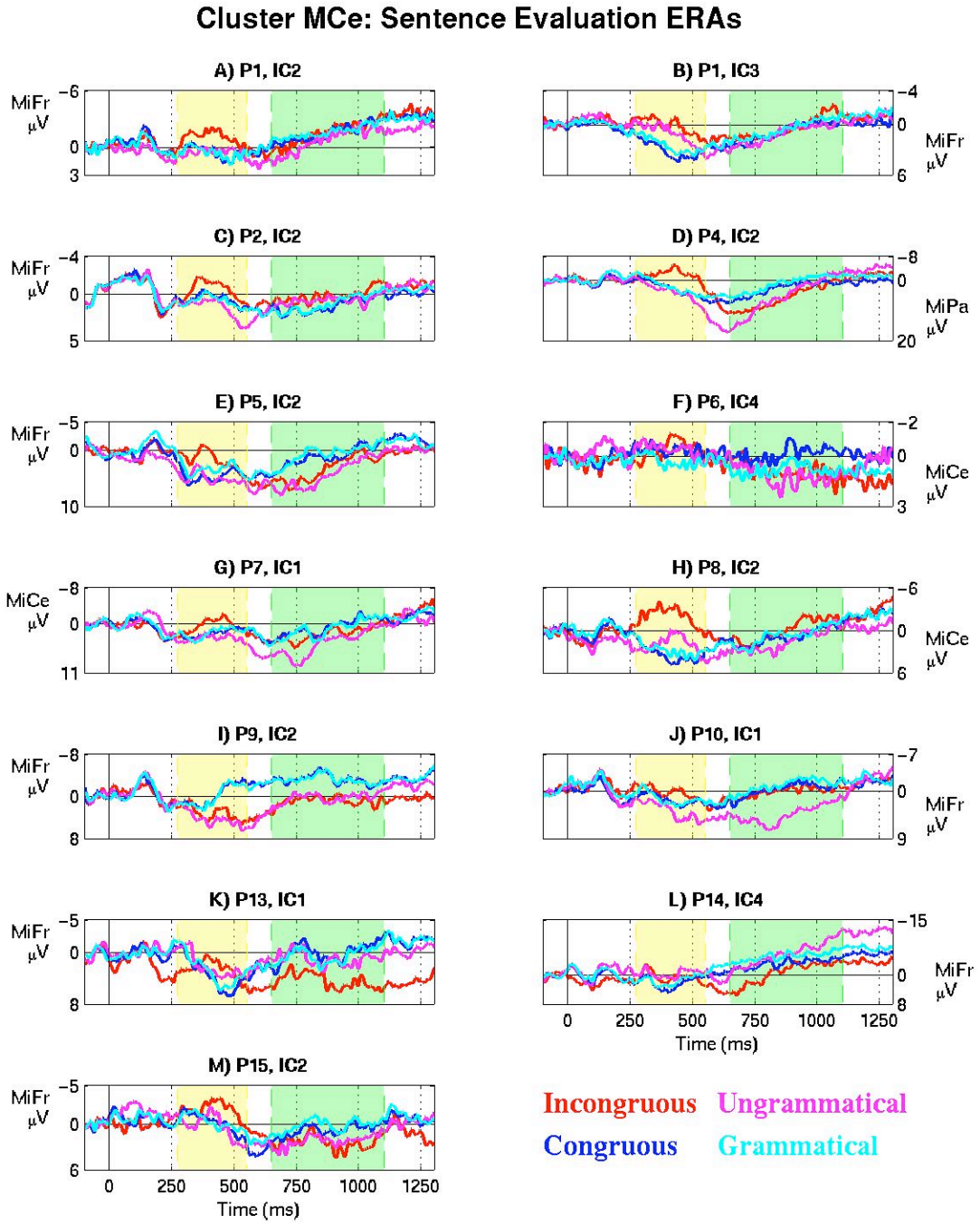

Figure VI.127 – Scalp-projected sentence acceptability task event-related activations (ERAs) of each IC in the midline central cluster (MCe). ERAs are plotted at the electrode at which the absolute IC topography is maximum (MiFr, MiCe, or MiPa). ERAs are computed across all parts of speech. Green and yellow shading indicate the noun/verb N400 (275-550 ms) and P600 (650-1100 ms) windows respectively. Abbreviations: P=Participant, IC=Independent Component.

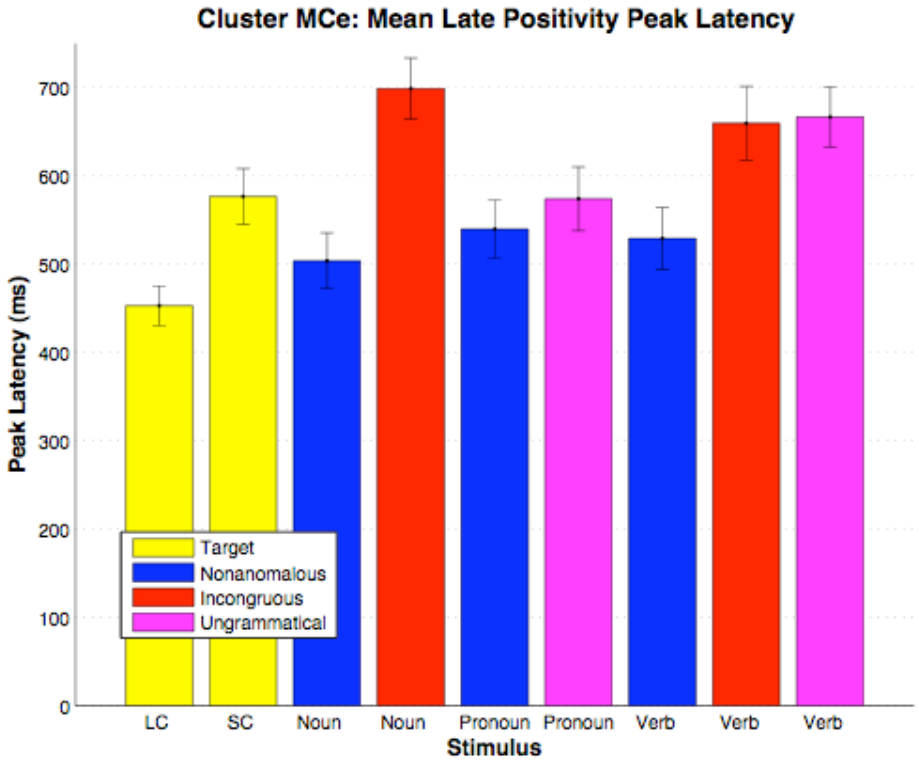

Figure VI.128 – Mean peak latency of the late positivity (LP) of ICs in the midline central cluster (error bars are standard error). Abbreviations: MCe=Midline Central, LC=Letter Case (Oddball Task), SC=Semantic Category (Oddball Task).

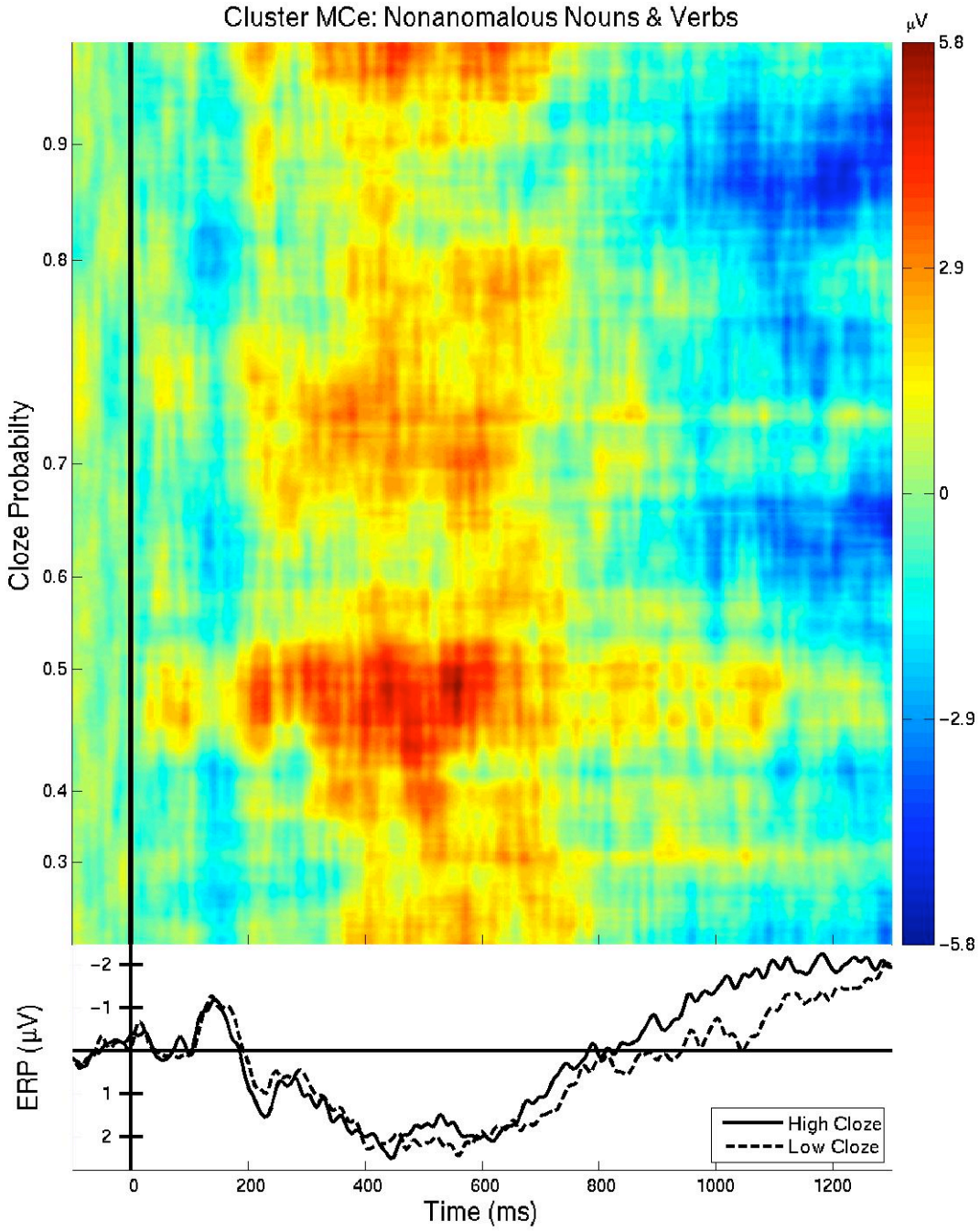

Figure VI.129 – (Top) 1157 single trials of the midline central cluster response to nonanomalous noun and verb sentence endings. Each IC's activation is set to the polarity and magnitude of its activation at its maximal electrode. Trials are sorted by cloze probability, stacked along the y-axis, and smoothed using a 100 trial vertical moving average. **(Bottom)** Averages (ERPs) of the top and bottom 50% of the single trials.

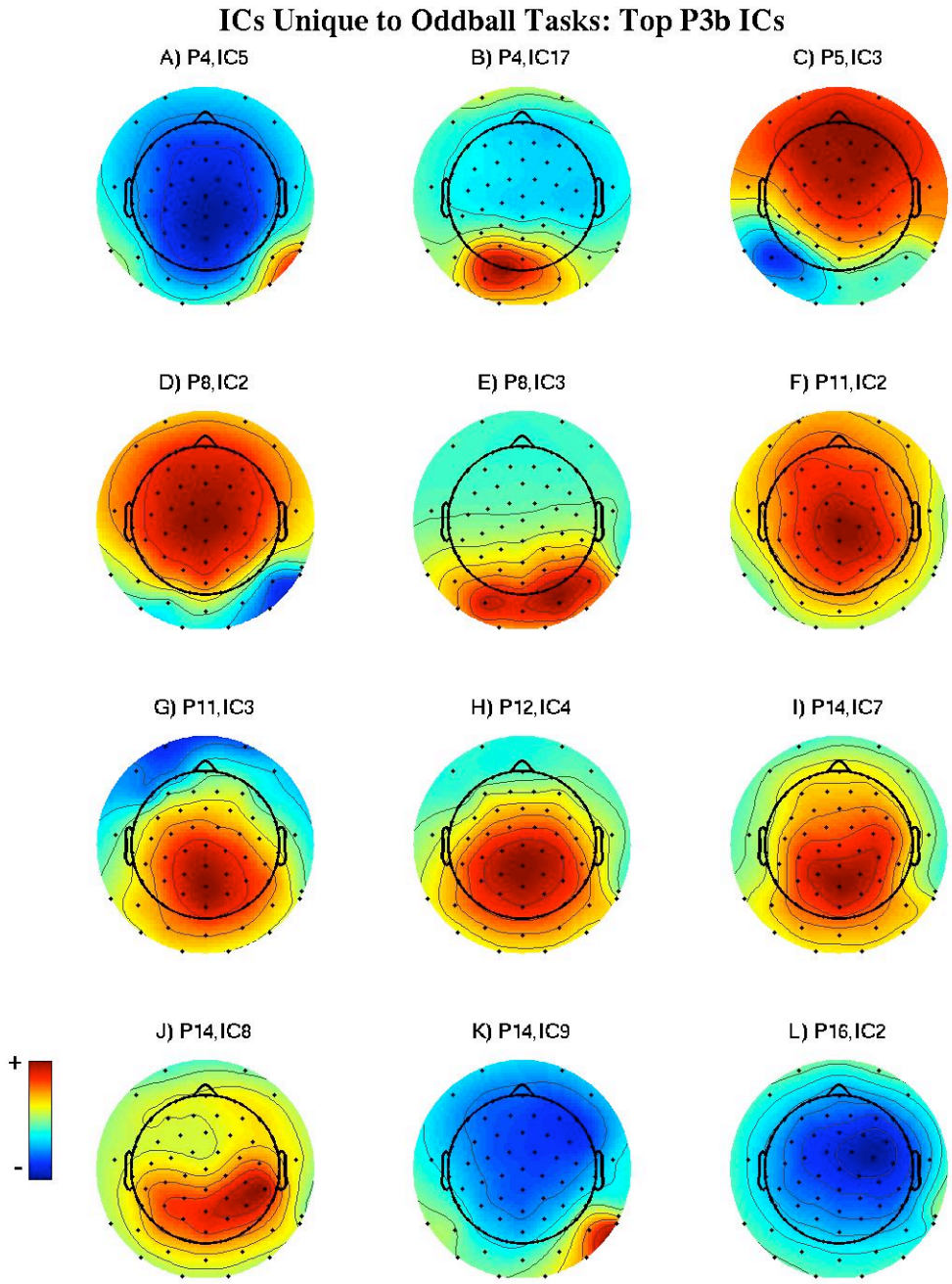

Figure VI.130 – Topographies of all ICs from the combined task data that appear to replicate in the oddball task data but **not** in the sentence acceptability task data and whose P3b effect activity is above threshold (see text). Unitless topography weights are plotted with a color scale that is unique to each topography and reflects the IC's mean contribution to the P3b effect (i.e., predominantly hot topographies mean that the IC is generally more positive to targets). Abbreviations: P=Participant, IC=Independent Component.

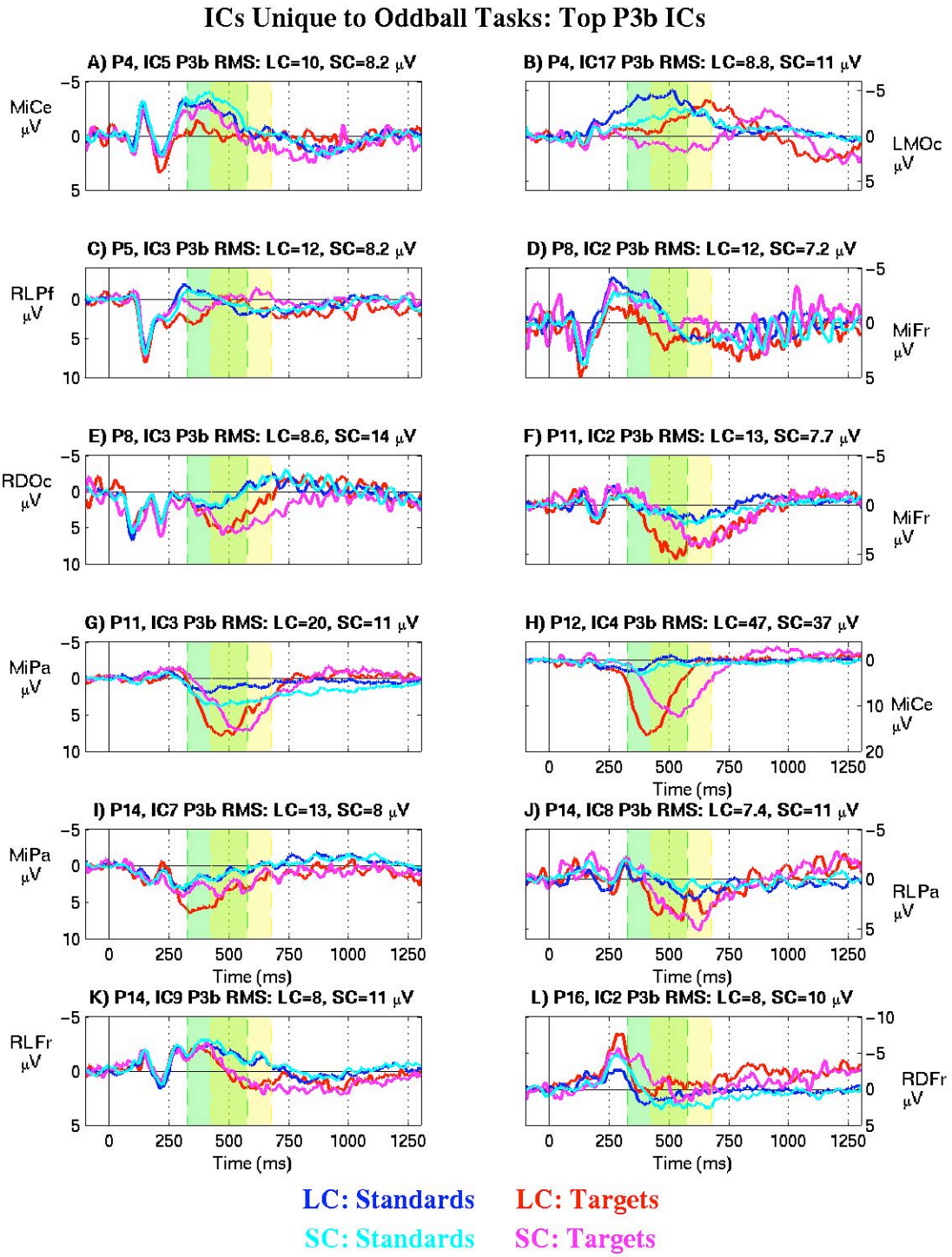

Figure VI.131 – Oddball task scalp-projected event-related activations (ERAs) of all ICs from Figure VI.106. ERAs are plotted at the electrode at which the absolute IC topography is maximum. Green and yellow shading indicate LC (325-575 ms) and SC (425-675 ms) P3b windows respectively. Abbreviations: P=Participant, IC=Independent Component, LC=Letter Case oddball task, SC=Semantic Category oddball task, RMS=Root Mean Squared activation (computed across all electrodes).

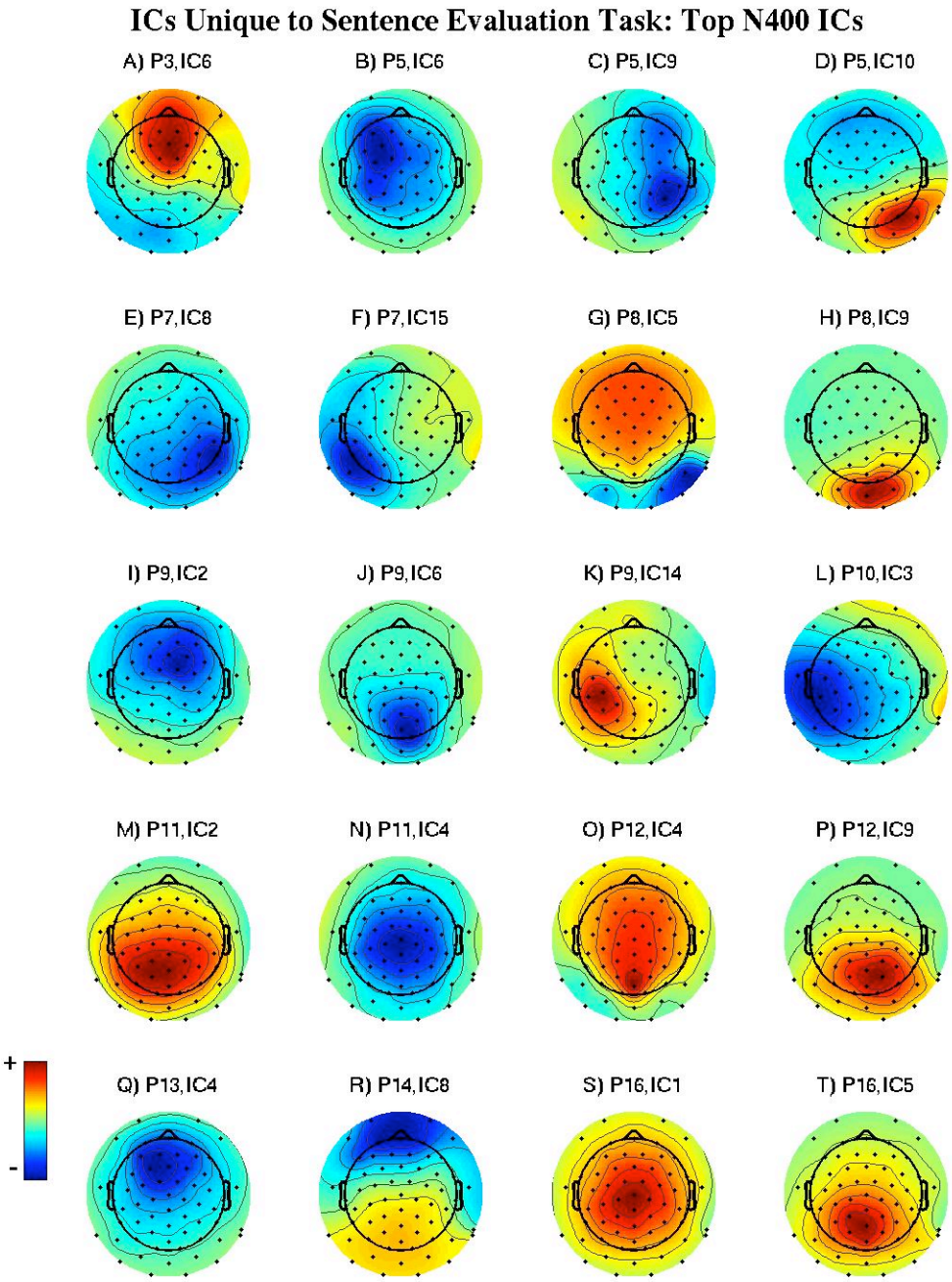

Figure VI.132 – Topographies of all ICs from the combined task data that appear to replicate in the sentence acceptability task data but **not** in the oddball task data and whose N400 effect activity is above threshold (see text). Unitless topography weights are plotted with a color scale that is unique to each topography and reflects the IC's mean contribution to the N400 effect (i.e., predominantly cool topographies mean that the IC is generally more negative to incongruous endings). Abbreviations: P=Participant, IC=Independent Component.

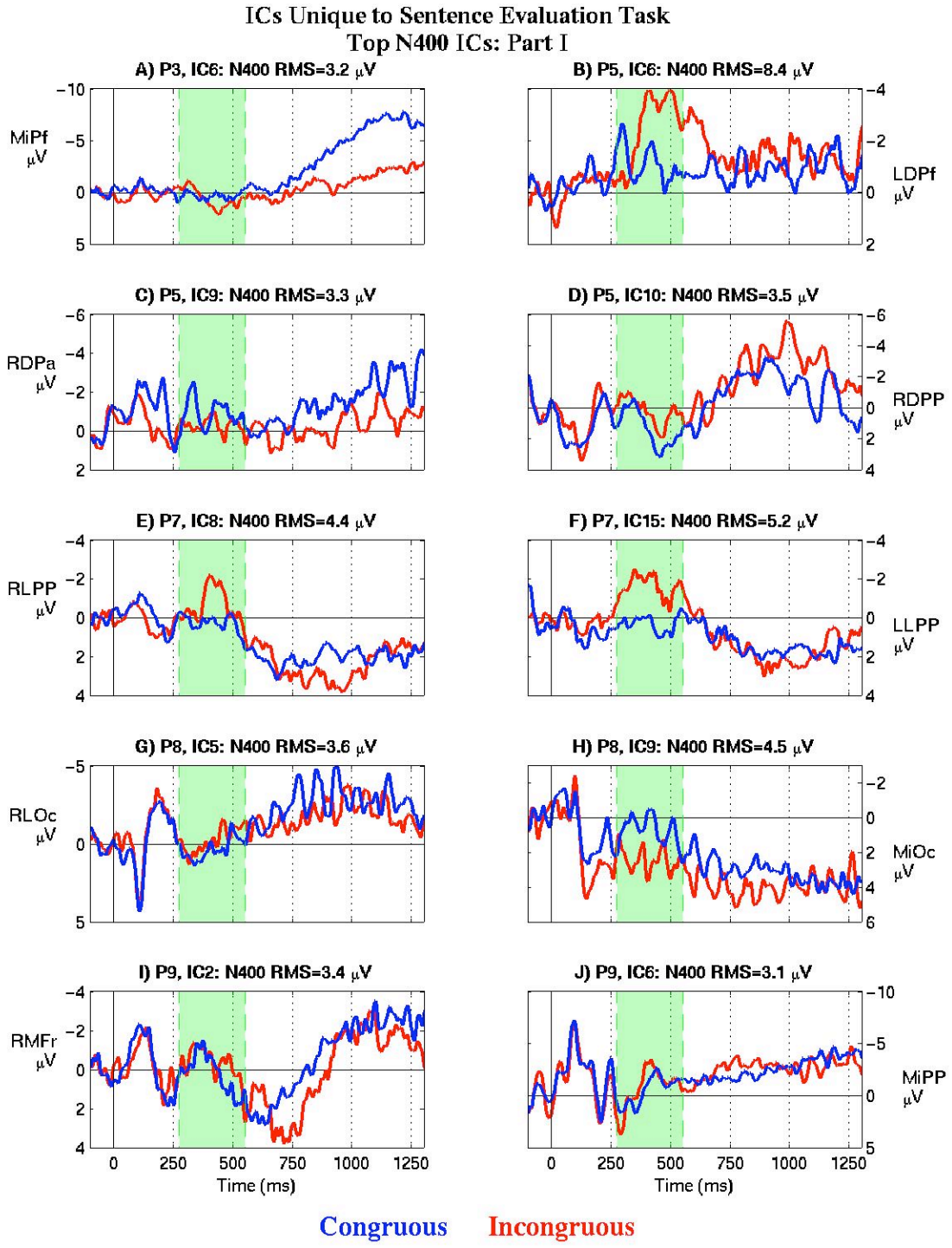

Figure VI.133 – Scalp-projected event-related activations (ERAs) of ICs A-J from Figure VI.132. ERAs are plotted at the electrode at which the absolute IC topography is maximum. Green shading indicates N400 window (275-500 ms). Abbreviations: P=Participant, IC=Independent Component, RMS=Root Mean Squared activation (computed across all electrodes).

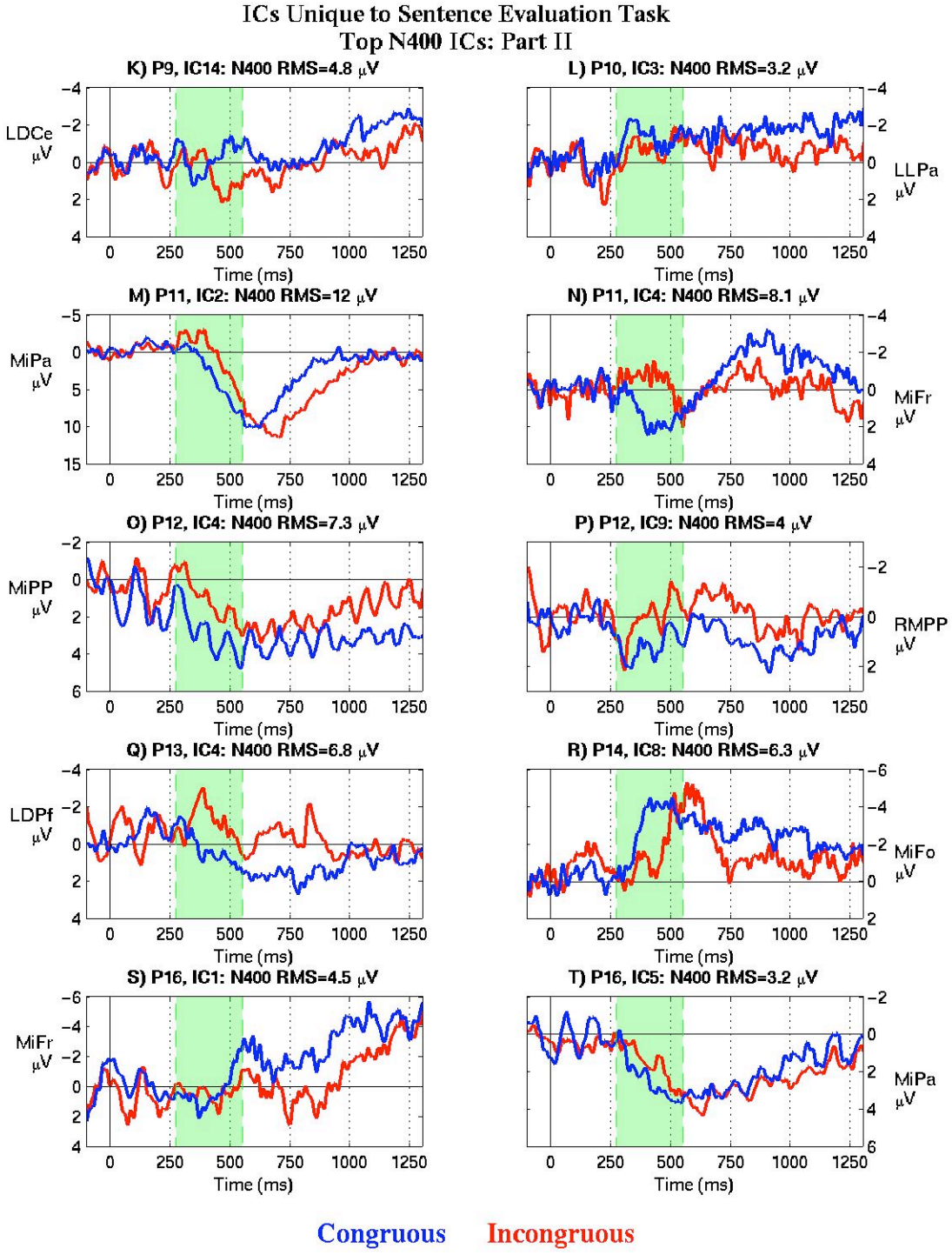

Figure VI.134 – Scalp-projected event-related activations (ERAs) of ICs K-T from Figure VI.132. ERAs are plotted at the electrode at which the absolute IC topography is maximum. Green shading indicates N400 window (275-500 ms). Abbreviations: P=Participant, IC=Independent Component, RMS=Root Mean Squared activation (computed across all electrodes).

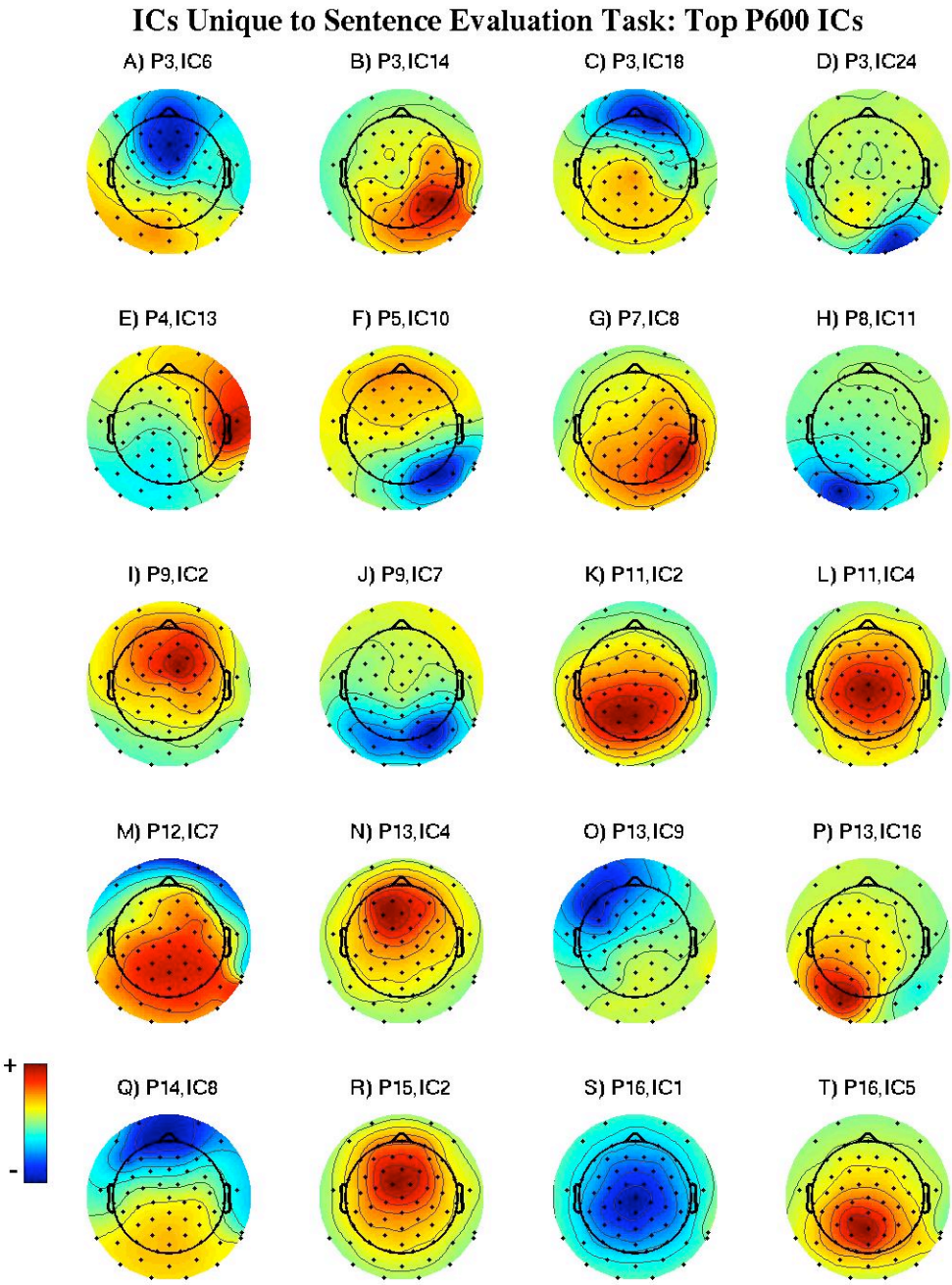

Figure VI.135 – Topographies of all ICs from the combined task data that appear to replicate in the sentence acceptability task data but **not** in the oddball task data and whose P600 effect activity is above threshold (see text). Unitless topography weights are plotted with a color scale that is unique to each topography and reflects the IC's mean contribution to the P600 effect (i.e., predominantly hot topographies mean that the IC is generally more positive to ungrammatical endings). Abbreviations: P=Participant, IC=Independent Component.

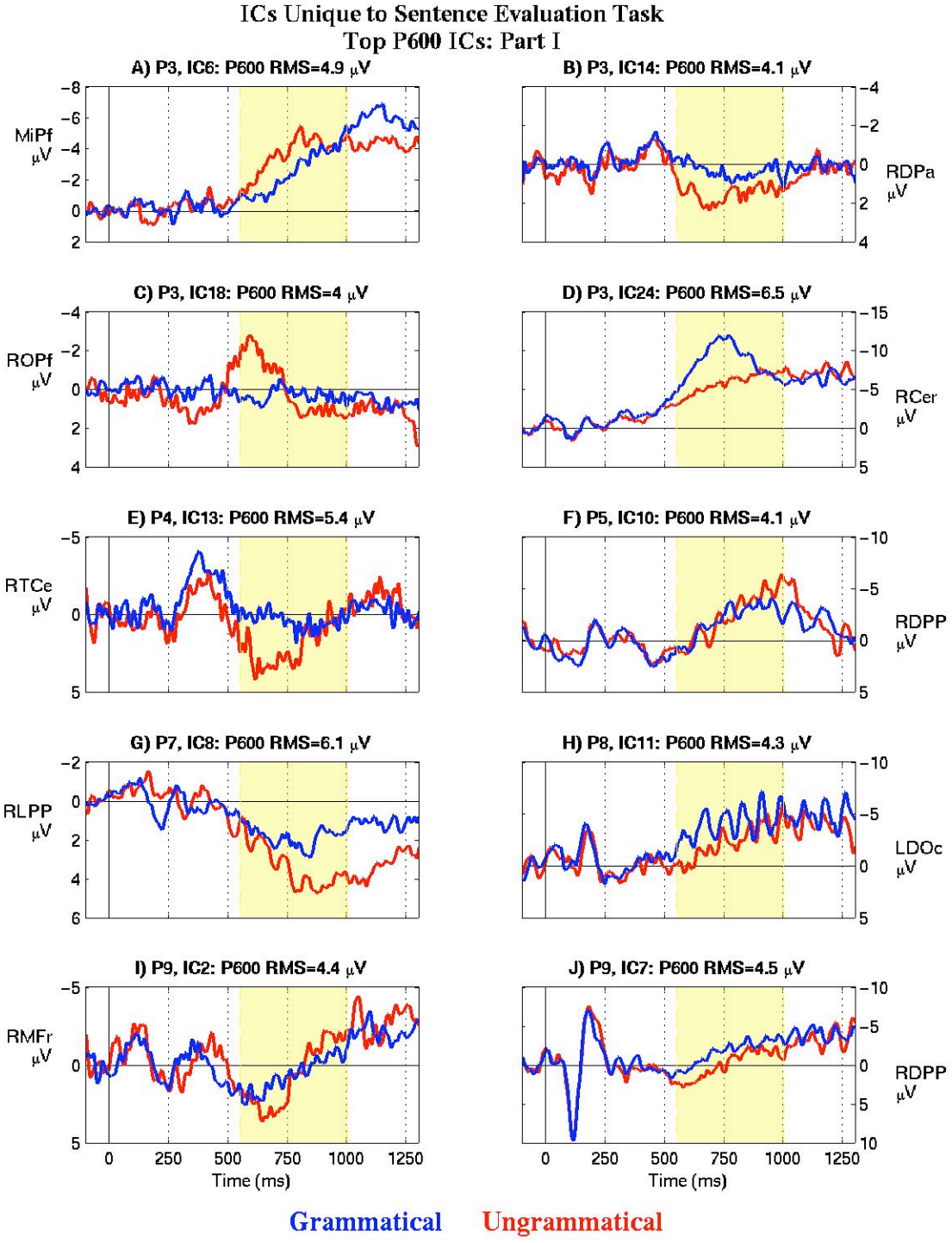

Figure VI.136 – Scalp-projected event-related activations (ERAs) of ICs A-J from Figure VI.135. ERAs are plotted at the electrode at which the absolute IC topography is maximum. Green shading indicates P600 window (550-1000 ms). Abbreviations: P=Participant, IC=Independent Component, RMS=Root Mean Squared activation (computed across all electrodes).

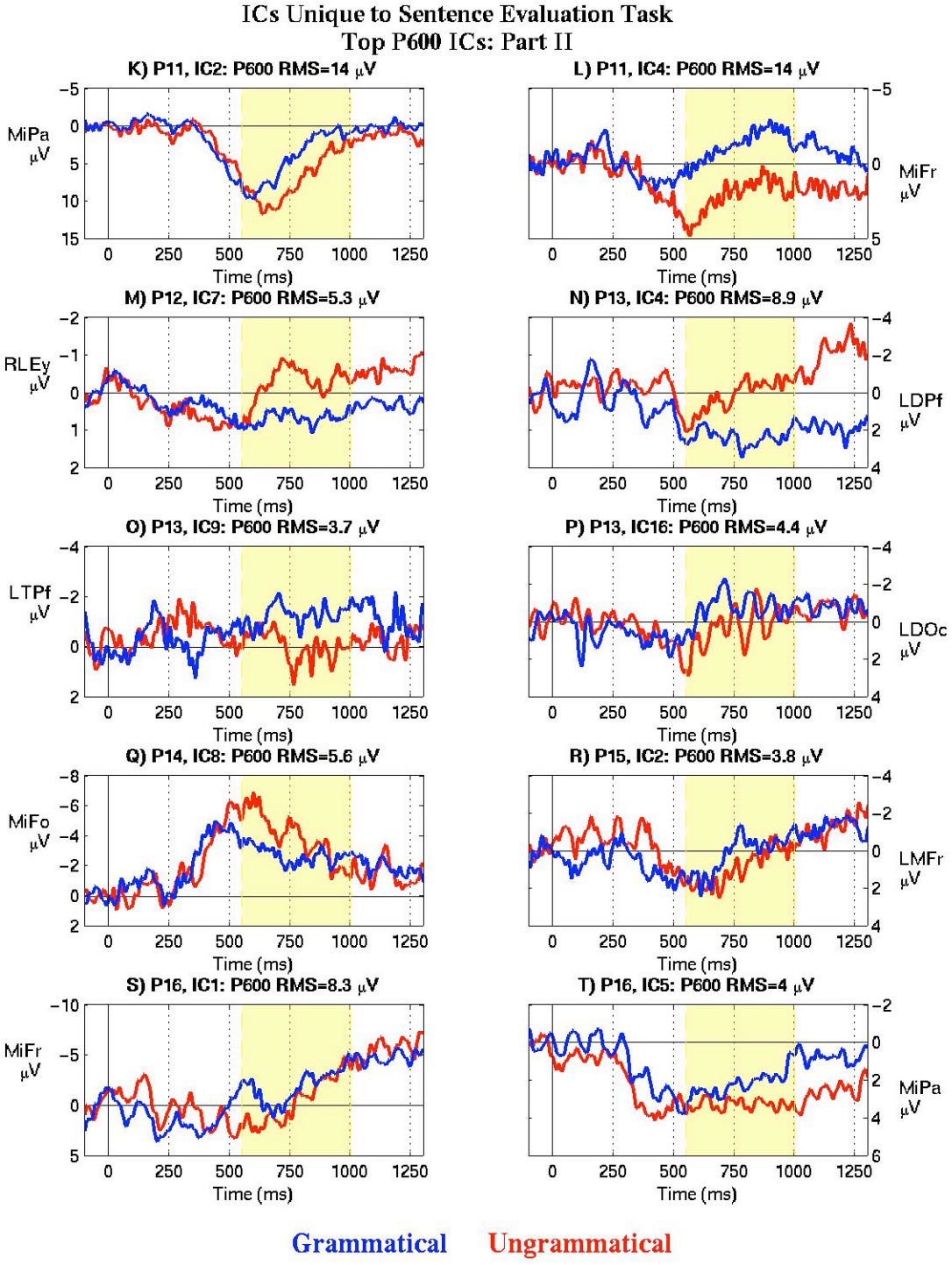

Figure VI.137 – Scalp-projected event-related activations (ERAs) of ICs K-T from Figure VI.135. ERAs are plotted at the electrode at which the absolute IC topography is maximum. Green shading indicates P600 window (550-1000 ms). Abbreviations: P=Participant, IC=Independent Component, RMS=Root Mean Squared activation (computed across all electrodes).

VI.F References

Baudena, P., Halgren, E., Heit, G., & Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. frontal cortex. *Electroencephalography and Clinical Neurophysiology, 94*(4), 251-264.

Baumgaertner, A., Weiller, C., & Buchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *NeuroImage, 16*(3 Pt 1), 736-745.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience, 11*(3), 235-260.

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., et al. (2004). Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. *The Journal of Neuroscience, 24*(42), 9353-9360.

Cardillo, E. R., Aydelott, J., Matthews, P. M., & Devlin, J. T. (2004). Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *Journal of Cognitive Neuroscience, 16*(9), 1552-1561.

Chung, G., Tucker, D. M., West, P., Potts, G. F., Liotti, M., Luu, P., et al. (1996). Emotional expectancy: Brain electrical activity associated with an emotional bias in interpreting life events. *Psychophysiology, 33*(3), 218-233.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research, 1084*(1), 89-103.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Dhond, R. P., Buckner, R. L., Dale, A. M., Marinkovic, K., & Halgren, E. (2001). Spatiotemporal maps of brain activity underlying word generation and their modification during repetition priming. *The Journal of Neuroscience, 21*(10), 3564-3571.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 112*(3), 357-374.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: Wiley.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M. L., Quiroga, R. Q., Nordby, H., et al. (2005). Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 102*(49), 17798-17803.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: The novelty oddball. *Psychophysiology, 32*(6), 579-594.

Fogelson, N., Loukas, C., Brown, J., & Brown, P. (2004). A common N400 EEG component reflecting contextual integration irrespective of symbolic form. *Clinical Neurophysiology, 115*(6), 1349-1358.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20 Suppl 1*, S8-17.

Friederici, A. D., Ruschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170-177.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral reviews, 25*(4), 355-373.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 360*(1456), 815-836.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences of the United States of America, 95*(3), 906-913.

Gough, P. B. (1983). Context, form, and interaction. In K. Rayner (Ed.), *Eye Movements in Reading: Perceptual and Language Processes* (pp. 203-211). New York: Academic Press.

Gough, P. B., Alford, J. A., & Holley-Wilcox, P. (1981). Words and contexts. In O. J. L. Tzeng, & H. Singer (Eds.), *Perception of Print: Reading Research in Experimental Psychology* (pp. 85-102). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*(6), 883-899.

Hagoort, P., Hald, L., Bastiaansen, M. C., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438-441.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17*(3), 1101-1116.

Halgren, E., & Marinkovic, K. (1995). Neurophysiological networks integrating human emotions. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (1st ed., pp. 1137-1151). Cambridge, Mass.: MIT Press.

Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology, 106*(2), 156-164.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721-742.

Horovitz, S. G., Skudlarski, P., & Gore, J. C. (2002). Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: A parametric approach to combining fMRI and ERP. *Magnetic Resonance Imaging, 20*(4), 319-325.

Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Reading anomalous sentences: An event-related fMRI study of semantic processing. *NeuroImage, 17*(2), 842-850.

Kleunder, R., & Kutas, M. (1993). Bridging the gap - evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 58*(2), 196-214.

Kluender, R., & Kutas, M. (1993). Bridging the gap - evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology, 15*(1), 3-13.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577.

Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 804-823.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Kutas, M., & Iragui, V. J. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology, 108*(5), 456-471.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Lange, T., Roth, V., Braun, M. L., & Buhmann, J. M. (2004). Stability-based validation of clustering solutions. *Neural Computation, 16*(6), 1299-1323.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation, 11*(2), 417-441.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging, 22*(10), 1517-1531.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science, 295*(5555), 690-694.

Mangels, J. A., Picton, T. W., & Craik, F. I. (2001). Attention and successful episodic encoding: An event-related potential study. *Cognitive Brain Research, 11*(1), 77-95.

Martin, J. H. (1991). The collective electrical behavior of cortical neurons: The electroencephalogram and the mechanisms of epilepsy. In E. R. Kandel, J. H. Schwartz & T. M. Jessell (Eds.), *Principles of Neural Science* (3rd ed., pp. 777-789). New York: Elsevier.

Martin-Loeches, M., Munoz, F., Casado, P., Melcon, A., & Fernandez-Frias, C. (2005). Are the anterior negativities to grammatical violations indexing working memory? *Psychophysiology, 42*(5), 508-519.

McElree, B., & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 134-157.

Moores, K. A., Clark, C. R., Hadfield, J. L., Brown, G. C., Taylor, D. J., Fitzgibbon, S. P., et al. (2003). Investigating the generators of the scalp recorded visuo-verbal P300 using cortically constrained source localization. *Human Brain Mapping, 18*(1), 53-77.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research, 30*(3), 339-364.

Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience, 12*(1), 120-133.

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin, 131*(4), 510-532.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage, 27*(2), 341-356.

Oostenveld, R., & Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human Brain Mapping, 17*(3), 179-192.

Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language-specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience, 8*, 507-526.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Pineda, J. A., Foote, S. L., & Neville, H. J. (1989). Effects of locus coeruleus lesions on auditory, long-latency, event-related potentials in monkey. *The Journal of Neuroscience, 9*(1), 81-93.

Salustri, C., & Kronberg, E. (2004). Language-related brain activity revealed by independent component analysis. *Clinical Neurophysiology, 115*(2), 385-395.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology, 38*(4), 387-401.

Stevens, A. A., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging, 18*(5), 495-502.

Stevens, M. C., Calhoun, V. D., & Kiehl, K. A. (2005). fMRI in an oddball task: Effects of target-to-target interval. *Psychophysiology, 42*(6), 636-642.

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory & Cognition, 19*(1), 95-112.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279-293.

Verleger, R., Jaskowski, P., & Wauschkuhn, B. (1994). Suspense and surprise: On the relationship between expectancies and P3. *Psychophysiology, 31*(4), 359-369.

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology, 38*(1), 41-63.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244.

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*(6), 1024-1037.

Chapter VII: Conclusions and Closing Thoughts

VII.A Introduction

The central question of this dissertation is "Is there a P3b-like subcomponent of the N400 and P600?". In this chapter, I summarize the results of the preceding chapters that address this question and suggest that there is indeed such a subcomponent. In particular, it appears that there is a P3b-like midline central subcomponent (MCe) that contributes substantially to the P3b, N400, and P600 effects. Although the existence of such a subcomponent is suggested by previous ERP, fMRI, and reaction time measures of language comprehension, there are several reasons why this subcomponent might nonetheless be spurious. Thus, it is vital to validate the results presented here with additional experiments and, potentially, alternative algorithms for extracting ERP subcomponents. To this end, I propose future to work to assess the reliability, generality, and accuracy of this MCe subcomponent and I make predictions to test the hypothesis that the activity of this subcomponent reflects an attentional response. Subsequently, I discuss what implications an MCe-like attentional subcomponent of the N400 and P600 would have for our understanding of language comprehension.

Finally, based on my experiences these past few years using ICA and studying these ERP phenomena, I present some general thoughts on the analysis of EEG and on the study of the cognitive neuroscience of language. I first discuss the utility of using ICA to define ERP subcomponents and suggest ways to best use ICA (and to some extent other techniques such as source localization) to define ERP subcomponents. I close with some reflections on the importance of individual variability for the future of EEG research and the cognitive neuroscience of language.

VII.B A Possible P3b-like Subcomponent of the N400 Effect

Although a P3b-like subcomponent of the N400 is not frequently considered (e.g., Kutas & Federmeier, 2000; Van Petten & Luka, 2006), ICA decompositions of the P3b and N400 effects suggest that there is considerable overlap between the EEG processes that produce these two phenomena. In particular, there appears to be a common midline central (MCe) subcomponent of each, which is generally characterized by a late positivity that is enhanced to deviant events and whose peak latency roughly covaries with the difficulty of stimulus evaluation. In addition, this late positivity tends to habituate and the topography of this subcomponent often exhibits a medial central focus. This ostensible MCe subcomponent was largely defined by identifying the independent components (ICs) that contributed the most to an individual's P3b effect (the class of "Best P3b ICs", Section VI.C.3) and was, in the main, replicated by automated cluster analysis (cluster MCe, Section VI.C.4). It

accounted for about 60% of the grand average P3b and N400 effects in the data presented here and possibly reflects an attentional function that is part of the brain's "orienting response" to deviant or motivationally significant stimuli (Friedman, Cycowicz, & Gaeta, 2001; Halgren & Marinkovic, 1995). This interpretation is suggested by the similarity of this subcomponent's topography to that of the P3a subcomponent of the P3b (Section II.E.4). Nevertheless, other functions (e.g., a generative model of the stimulus environment) are quite possible.

While this MCe subcomponent was most clearly characterized by a late positivity, there was some evidence of a negative potential to incongruous words in the N400 time window, which preceded this late positivity. It is currently not clear if this negative effect is reliable but it might be partially responsible for MCe's contribution to the N400 effect (i.e., its contribution appeared to be only partly due to a delayed late positivity to incongruous words relative to congruous words) and may indicate that the late positive component (LPC) to incongruous words is part of a biphasic response.

It is important to note that there are several reasons why the results of the procedure used to identify the MCe subcomponent of the P3b and N400 effects might be misleading. For example, it is possible that the ICs largely correspond to the EEG response to sentence internal words (from which the majority of the ICA training data was derived) and are not an accurate decomposition of the EEG response to sentence final words. Moreover, it could be that the procedure used to identify ICs common to the P3b and N400 data falsely deemed some ICs as common. It is also possible that

the methods used to group ICs across participants combined together ICs that were not homologous. Indeed, the inter-participant variation in the activity of the ICs that formed the MCe subcomponent suggests that these results might not be that accurate.

However, the credibility of this subcomponent is supported by the electrode level analyses of these data (Chapter IV). Specifically, these analyses found that the amplitude of the LPC to nonanomalous endings and the amplitude of the N400 effect selectively correlated, that the latency of the LPC negatively (albeit weakly) correlated with a word's cloze probability (which could help to explain why N400 amplitude covaries with a word's cloze probability), and that the topographies of the LPC to nonanomalous endings were rather similar to that of the N400 effect. Thus, there is some type of relationship between the LPC and the N400 effect, which could be explained by something like the MCe subcomponent that contributes considerably to both and appears to be partially responsible for the N400 effect/LPC amplitude correlation.

In sum, given that several studies have found that people take longer to evaluate incongruous words relative to congruous ones (Section II.G.1), the electrode data evidence of a relationship between the N400 effect and the LPC, and the ICA derived MCe subcomponent, I think it is quite likely that the N400 effect is produced in part by a relative delay in the LPC to incongruous words. The MCe subcomponent is suggestive, but needs to be corroborated by additional research (see below: Section VII.D). Regardless of the accuracy the MCe subcomponent though, I believe that the likely contribution of a relatively delayed LPC to the N400 effect informs attempts to

identify the generators of the N400 effect. Specifically, source models of the N400 effect should not just report which cortical areas appear to contribute to the effect (e.g., Halgren et al., 2002, distributed model; Haan, Streb, Bien, & Rosler, 2000), but also whether a cortical area's contribution is due to increased activity to incongruous words, increased activity to congruous words, or to activity of different polarities (e.g., Halgren et al., 2002, 1-2 dipole model). Equivalently, such work could simply report the estimated source activity to congruous and incongruous words. Furthermore, the results here are a reminder of the possible pitfalls of relating fMRI and ERP results. If a significant source of the N400 effect is due to a relative delay in activity or to activity of different polarities, that source may be relatively invisible to fMRI.

VII.C On the Relationship between the P3b and P600

ICA decompositions of the P3b and P600 effects in these data suggest that there is considerable overlap between the EEG processes that produce the two phenomena. In particular, the MCe subcomponent also contributed considerably to the grand average P600 effect (about 60% of the variance). This is not surprising, given the previously reviewed evidence of a relationship between the P3b and P600 (Section II.G.2), although as just mentioned, the ostensible MCe subcomponent needs to be corroborated by additional research (see below). Rather, what was most surprising about the ICA analysis of these phenomena was how little the two effects

have in common. Indeed, it appears that the two phenomena may be distinct to "an interesting degree" (Osterhout & Hagoort, 1999, pg 12).

Unfortunately, the analyses reported here provide little clue as to how the two effects might be distinct. It was difficult to identify types of ICs that appeared to be unique to the P3b or P600 effects due to individual variability and the limited number of such ICs. As there are several reasons why the procedure used here might be underestimating the amount of overlap between the sources of the P3b and P600 (Section VI.D), without compelling classes of ICs unique to the P3b or P600, these results remain only suggestive. Perhaps experiments contrasting only the P3b and P600 would be better at identifying what differs between them using the ICA procedure applied here.

VII.D Future Work and Testable Predictions

There is plenty of a priori reason to believe that rare targets and sentence final words elicit an attentional response that is enhanced to anomalous words and the MCe class of ICs is suggestive of such a function. Nevertheless, the results of this one experiment are far from convincing evidence that there is a subcomponent of the P3b, N400, and P600 effects that was largely captured by the MCe ICs and largely corresponds to such an attentional response. From here several steps should be taken to establish the reliability, generality, and credibility of this ostensible subcomponent and its ostensible function.

First of all, it is important to show that this class of ICs is reliable. Given the large number of extant N400 and P600 data sets, this should be rather easy to establish. Although it is unlikely that many of these data sets were recorded from individuals from whom P3b data sets are also available, the late positive component (LPC) that follows sentence final words may be sufficient to reproduce the MCe class of ICs formed here, as the LPC appears quite reliable and 9 of the 12 common best P3b ICs which formed the MCe class, were also the IC that accounted for the most of that individual's LPC variance. The fact that most existing N400 and P600 data sets were probably recorded using 30 or fewer electrodes may also make it difficult to reproduce the results reported here, which were obtained from 64 electrode data, as decreasing the number of the electrodes decreases the ability of linear spatial filters to extract EEG sources (Section III.C). Nevertheless, this possible confound should be easy to assess and other 64+ electrode measures of the N400 and P600 exist and are relatively cheap to create.

Secondly, it is important to show that the MCe class of IC also contributes to the N400 and P600 effects elicited by sentence internal words. To my knowledge, qualitative difference between the N400 and P600 effects elicited by sentence internal and final anomalies have not been established. However, given the ambiguity of the EEG inverse problem (Section III.C), it could be that ERP effects that appear to be similar (especially when averaged across participants) might actually be produced by a different set of sources. Since sentence internal words do not elicit an LPC, and the MCe ICs appears to largely correspond to the LPC, it is possible that they do not

contribute that much to sentence internal N400 and P600 effects. Again, given the number of extant N400 and P600 data sets, this should be readily testable.

In addition to demonstrating the reliability and generality of this ostensible subcomponent of the N400 and P600 effects, replicating the MCe subcomponent with source localization techniques (Dale et al., 2000; Liu, Dale, & Belliveau, 2002) would add credence to the subcomponent and provide a better sense as to its neural underpinnings. Since it is not clear what the sources of this class of ICs might be (Section VI.C.3.d), it is difficult to predict exactly how a source localization model should correspond to this class of ICs. Rather, it would be best to independently apply ICA and source localization to the same data to determine the extent to which the two provide consistent results.

Finally, to determine if this subcomponent does index an attentional response, I propose investigating the degree to which it covaries with noninvasive measures of the brain's orienting response. Specifically, I propose investigating a relationship between the latency and amplitude of the late positivity (LP) of the MCe ICs and startle eyeblink modification (SEM--for review see Filion, Dawson, & Schell, 1998). The startle eyeblink reflex occurs when a startling stimulus, such as a loud burst of white noise or a tap on the forehead, triggers a blink. Although this response is reflexive in that it is not learned and does not require neocortex (ibid.), the magnitude of the blink, measured electromyographically, is sensitive to the amount of attention directed to a stimulus that precedes the startling stimulus. For example, in a target detection paradigm involving attended tones and ignored tones of different pitches,

when the startling stimuli rapidly follows tone presentation, the blinks elicited by the startling stimuli are smaller than those elicited by the startle stimuli presented alone (Filion, Dawson, & Schell, 1994). Moreover, attended tones produce greater blink suppression than ignored tones, but only in a brief time window (sometime within 60 to 240 ms after the tone--Filion, Dawson, & Schell, 1993). This enhanced blink suppression to task relevant stimuli has been interpreted as a facet of the brain's orienting response to such stimuli (Filion et al., 1998).

I propose measuring the SEM at various time lags (e.g., every 50 ms from 100 to 400 ms poststimulus) following sentence final word onsets like those of the stimuli used here (i.e., nonanomalous, incongruous, and ungrammatical ending) and comparing the latency and magnitude of maximal SEM suppression with that of the LP of this ostensible class of ICs from separate trials (to avoid contamination of the ERPs to sentence final words by effects of the startle stimulus). I predict that the latency of the point of maximal SEM suppression will track the latency of the LP (e.g., it should be later for incongruous words) and that the magnitude of SEM suppression will correlate with the magnitude of the LP (e.g., there should be greater suppression to ungrammatical words than incongruous words).

I propose using SEM as an index of the orienting response because it is very robust (found in 90-100% of individuals with normal eyeblink reflexes--Filion et al., 1998) and exhibits temporal resolution at least on the order of centiseconds. However, to my knowledge no one has demonstrated that the latency of SEM correlates with stimulus evaluation duration or that the magnitude of SEM is sensitive to the degree to

which a stimulus elicits an orienting response (beyond gross distinctions like the aforementioned difference between attended and ignored tones). Should SEM fail to correlate with the difficulty of stimulus evaluation and the "anomalousness" of a stimulus, there are other noninvasive indices of the orienting response that could be explored. These include reaction times to stimuli in a secondary task that immediately follow primary task stimuli (in this case sentence final words) or changes in skin conductance (see Filion et al., 1994). With regards to this latter possibility, there is evidence that oddball targets and deviant task-irrelevant stimuli reliably elicit a change in skin conductance, although skin conductance appears to be only imperfectly correlated with P3b and P3a amplitude (Rosler, Hasselmann, & Sojka, 1987).

Of course, any evidence of a correlation between the late positivity of this class of ICs and indices of the orienting response would not prove that the late positivity is directly produced by the response. It would only indicate that in some circumstances it is correlated with the orienting response. The technically much more challenging project of identifying the source of the MCe ICs and disrupting them with transcranial magnetic stimulation or studying patients with damage to those areas would be necessary to demonstrate what cognitive processes these ICs are causally involved in.

VII.E Potential Implications for our Understanding of Language Learning and Comprehension

If the MCe subcomponent of the N400 and P600 effects is accurate, and it indeed is generated by an attentional response, how would it further our understanding of language learning and comprehension? With regards to the former, I suspect that MCe may not shed much light. This is based on the evidence that the P3b and LPC are not reliable predictors of memorable stimuli (for a brief review of some of the P3b literature see Kok, 2001, pg 573). For example, Fabiani and colleagues (1990) found that a word-elicited P3b-like positivity predicted the later recall of those words for rote memorization strategies but not elaborative memorization techniques. Paller et al. (1987) found that the P3b-like positivity following single words did not predict later recall although it was larger to words that were later produced during a test of implicit memory. Neville and colleagues (1986) found that the LPC to sentence endings was greater to words that were later recalled and to repeated words that were recognized. Furthermore, recent work by Borovsky et al. (2007) using a more naturalistic word learning paradigm than the previously mentioned recall, recognition, and implicit memory measures, did not find a significantly increased positivity to novel words whose meaning participants had begun to learn. Since ICA presumably provides a cleaner measure of this MCe process, it may be that MCe will prove to have a more consistent relationship with measures of language learning than conventionally measured late positivities. After all, I think there is good a priori reason to think that the brain does implement generative models in the service of language learning

(Section II.G.3) and these models are likely to produce a lot of activity in response to highly unlikely stimuli (Friston, 2005). However, if MCe primarily corresponds to an attentional function, I suspect that it may very generally facilitate responses to stimuli and not necessarily induce learning.

I think MCe potentially has more to say about the mechanics of language comprehension. At the most general level these results are relevant to one of the central questions in the science of language: To what extent (if at all) is the processing of semantic and syntactic information distinct? This question has received a great deal of attention over the decades due to the hypothesis that syntax requires a special type of computer that is capable of dealing with the recursive structure of natural language and may be relatively specific to humans, while semantics is presumably rooted in conceptual representations that are not qualitatively different from those of other species (Hauser, Chomsky, & Fitch, 2002). Moreover, some researchers have argued that the speed at which a child acquires the syntax of a language can only be explained by some type of innate syntax faculty (Chomsky, 1980). Both theoretical motivations for the dichotomy have been challenged (Elman et al., 1996; Rodriguez, 2001) and cognitive neuroscientists have turned to the brain to help answer this question, which is fundamental to theories of language processing and potentially has broader implications for the nature of neural representation (Fodor & Pylyshyn, 1988), development (Elman et al., 1996), and human origins (Hauser et al., 2002; Pinker & Bloom, 1990).

Due to their superficial differences, the N400 and P600 have sometimes been used as evidence of a neural distinction between syntactic and semantic processing. For example, Peter Hagoort has claimed that N400 and P600 demonstrate "that the brain honors the distinction between syntactic and semantic binding operations" (Hagoort, 2003, pg 1), processes that Hagoort hypothesizes involve qualitatively different representations and parts of cortex (Hagoort, 2005). Angela Friederici and Sonja Kotz (2003) have also proposed a somewhat dichotomous framework for language comprehension in which the N400 corresponds to lexical semantic processes subserved[100] by Brodmann's area 44/45 and the posterior region of the superior and middle temporal gyrus. In contrast, she argues that the P600 corresponds to syntactic revision and late integration processes subserved by the basal ganglia and the posterior portion of the superior temporal gyrus (ibid.).

In contrast, Coulson and colleagues (1998) have interpreted the N400 and P600 as evidence of quantitative differences between semantic and syntactic processing. Specifically, they note that grammatical anomalies are more categorical than semantic anomalies, which could be explained by the prevalence of non-literal language (Lakoff & Johnson, 1980). For example, while the sentence:

I love she.

[100] Friederici and Kotz did not claim that the N400 was directly generated by these areas, simply that the N400 was an index of the activity of these areas. Such an index might be quite indirect.

is quite difficult to interpret as anything but a mistaken attempt to express "I love her," a sentence like:

I take my coffee with cream and dog.

can be interpreted sensibly with some imagination (e.g., perhaps the speaker has a habit of taking her dog for a walk to a café where she orders a cup of coffee with cream). Kuperberg et al. (2003) have also suggested a largely qualitative distinction between syntactic and semantic processing. These researchers hypothesized that there is a bilateral parietal attentional network and a left hemisphere fronto-temporal semantic memory/conceptual integration network that are both engaged by grammatical and semantic anomalies to different degrees. Whereas grammatical anomalies elicit relatively more activity from the parietal network, semantic anomalies result in relatively greater activation of the left fronto-temporal network.

If the MCe subcomponent is accurate, it would confirm to some extent the general view of Coulson, Kuperberg, and colleagues as a large fraction of the N400 and P600 effects would be produced by approximately the same EEG process. Indeed, MCe may explain why it is that grammaticality affects the N400 and congruity affects the P600 (Sections II.B, II.C). However, this would far from prove that grammatical and semantic processing are qualitatively the same. First of all, MCe leaves much of the N400 and P600 effects unexplained, thus there may be subcomponents that are not

shared by the two effects and might correspond to the dichotomy Hagoort, Friederici, and Kotz propose. Furthermore, it may be that the reason why grammatical anomalies are more categorical, easier to evaluate, and elicit more parietal activity is because of syntax specific processing upstream of the MCe contribution to the N400 and P600. Given these ambiguities, the most this dissertation can currently contribute to the debate on the distinction between syntax and semantics is to suggest a more subtle understanding of the neural processes and functions that directly produce the N400 and P600 effects. While modest, such subtleties are obviously important if we are to develop accurate models of language comprehension and are an improvement over demonstrably over-simplistic interpretations of the N400 and P600 as qualitatively distinct indices of semantic and syntax processing, respectively (Friederici & Kotz, 2003; Hagoort, 2003).

The MCe subcomponent, if accurate, would also help to inform some humbler questions in the cognitive neuroscience of language. For example, it would indicate that the P600 is in large part not directly produced by the reanalysis or revision of a grammatical interpretation of a sentence (contra Friederici, 2003) as MCe's late positivity is most pronounced to oddball task targets, which do not require grammatical processing, and is quite pronounced to nonanomalous, relatively predictable endings. In addition, MCe might shed some light on the reason why concrete words (e.g., "donkey") elicit more frontal N400 effects than abstract words (e.g., "democracy") in the absence of supportive context (Section II.E.2.a). As concrete words are generally easier to process (Schwanenflugel, 1991), it may be that

the time necessary to comprehend incongruous concrete words is less than that to incongruous abstract words. If this is the case and the onset of MCe's late positivity correlates with the duration of stimulus evaluation, then it would contribute less to the N400 effect to concrete words (assuming that congruous concrete and abstract words are equally easy to process) and perhaps help to produce a more frontal N400 effect. Thus, it is plausible that an interpretation of the more frontal distribution of the N400 to concrete words as a direct index of additional "imagistic processing" (Holcomb, Kounios, Anderson, & West, 1999) is a misattribution of a difference in stimulus evaluation duration. The possibility of such a confound illustrates the value of assessing the stimulus evaluation duration of stimuli used in language ERP experiments. Perhaps such assessments should be carried out more routinely and controlled for as commonly as lexical factors such as word length and frequency.

VII.F On the utility of ICA for Defining ERP Subcomponents

Had someone asked me when I first began my graduate studies what the biggest problems in cognitive neuroscience were, I would have listed off some of our grand questions such as "What is the format of neural representation?" or "What is special about human brains that makes us capable of species specific behaviors like language?". However, in the time since then, I have come to the conclusion that the biggest problem in cognitive neuroscience is actually quite humbler, it is the very

practical problem of how to improve our ability to measure the activity of the brain[101]. This problem is readily apparent in the ERP literature, which is full of difficult-to-interpret results because of the often ambiguous relationship between ERPs from one type of stimulus/paradigm and those of another. The relationship between the P3b, N400, and P600, which is the focus of this dissertation, is just one example of this. Other examples include the relationship between the left anterior negativity (LAN) associated with syntactic processing and the N400 (Section II.F), between the similar but somewhat different "N400s" elicited by multiple stimulus modalities (e.g., text, speech, images, and odors--for review see Kutas & Federmeier, 2000), and between a P600-like effect to grammatically difficult constructions and the canonical P600 to ungrammatical words (Kaan, Harris, Gibson, & Holcomb, 2000).

Clearly there is a great need for tools like ICA that can potentially extract spatio-temporally overlapping EEG processes and improve our knowledge of what these processes are doing and where they are coming from. As reviewed in Chapter III, ICA is theoretically well motivated in many ways and has produced some compelling empirical results. Moreover, ICA's assumptions are probably more valid than those of its sibling, principal components analysis (PCA), and it has the advantage over the complementary methods of source localization in that it does not require detailed anatomical assumptions nor the cost of deriving accurate anatomical models. ICA, like any attempt to parse the EEG/ERP into components, is potentially stymied by unreliability and inaccuracy. However, unlike the conventional method of

[101] This is especially true for the cognitive neuroscience of language, which is limited to noninvasive measures and manipulations of brain activity and patient populations who may not be representative of healthy individuals.

defining ERP components (i.e., the mean ERP amplitude in a particular time window at a particular set of electrodes) and source localization, it can be difficult to know how to combine ICA results across individuals.

With regard to the reliability issue, it is quite tractable and I believe that the IC reliability test developed as part of this dissertation can improve the application of ICA by facilitating the identification of spurious ICs, by providing a sense as to how much training data are necessary for stable decompositions (e.g., Sections V.D, VI.D.1), and by providing a principled procedure for excluding individual data sets from analyses (i.e., data sets which ICA does not reliably decompose). Although other IC reliability algorithms have been proposed, this procedure improves upon them in that it considers the reliability of an IC's topography and activation and it only requires running ICA three times per participant.

Concerning the question of IC accuracy, the issue is no more and no less of a problem for ICA than for any other method for decomposing ERPs into components. Given that the number of possible EEG sources is orders of magnitude larger than appear to be necessary to account for the late cognitive ERP effects studied here (Section IV.H.3), it is apparent that the parsing of these ERP effects is massively underdetermined. Although the physiological plausibility of many ICs is quite compelling, the fact that other methods for decomposing the EEG into somewhat independent components (e.g., PCA, the principal square root of the data covariance matrix—Section III.D.5.b) can also produce such components without successfully extracting large EEG artifacts suggests that physiological plausibility is not sufficient

evidence of accuracy. As discussed in Section III.D.5.b, evidence of IC accuracy can come from other methods for measuring brain activity (e.g., fMRI, source localization), evidence of a tighter relationship with experimental variables than the electrode data, and, most convincingly, interpretations of IC function that are tested and confirmed by further experiments.

There are several impressive examples of this latter type of validation for conventionally defined ERP components. For example, a classic review of the P3b by Donchin and Coles (1988) discusses several series of experiments that were carried out to determine if the P3b is primarily an index of sensory or motor processes, if the P3b reflects a limited cognitive resource, and the degree to which the P3b is sensitive to subjective probability. As a result, the P3b has been useful for informing some questions, such as the locus of the attentional blink (Vogel, Luck, & Shapiro, 1998) and the psychological cause of the Stroop effect (Duncan-Johnson & Kopell, 1981; Rosenfeld & Skogsberg, 2006). However, my subjective impression is that similar attempts to test interpretations of ICA, PCA, and source localization ERP componentry are unfortunately rare. Consider two attempts to localize the sources of the P3b published relatively recently in high impact journals. Using MRI-derived individual head models consisting of around 38,000 possible sources, Moores and colleagues (2003) localized the P3b primarily to the intraparietal sulcus and superior parietal lobes and found evidence of occipital and fusiform contributions in some individuals. In contrast, Bledowski et al. (2004) localized the grand average P3b using a generic spherical head model that included only 14 sources, which were determined

according to fMRI results from a non-identical group of participants[102] performing the same experimental paradigm. Bledowski and colleagues found that the P3b was primarily generated by the inferior parietal lobe, posterior parietal cortex, and the inferior temporal lobe. A contribution of posterior parietal cortex is consistent across the two studies. However, nothing like Bledowski et al.'s inferior parietal and inferior temporal sources appeared in the group average results of Moores and colleagues. Moreover, Bledowski et al. did not find any evidence of Moores and colleagues' occipital contribution to the P3b. The consistency of both studies' localization results with those of fMRI and lesion studies of the P3b is similarly mixed. Most notably, while there is some fMRI evidence of superior parietal sources of the P3b (Eichele et al., 2005; Stevens, Calhoun, & Kiehl, 2005), other studies have found no such evidence (Horovitz, Skudlarski, & Gore, 2002) and patients with superior parietal lesions elicit apparently normal P3bs (Knight & Scabini, 1998). Consequently, it is presently quite difficult to know to what extent the results of either localization attempt are accurate and to what extent they are spurious due to erroneous model assumptions. To make sense of these messy findings and the similarly ambiguous results of ICA, PCA, and other source localization attempts, I think it is essential that investigators not only interpret their results, but that they make predictions based upon those interpretations that they or other researchers can (and then do) test. Given the indeterminacy of ERP componentry, consistent and coherent sensitivity of putative component activity to experimental variables across a series of experiments is the best

[102] It is possible that some of the fMRI participants also participated in the ERP study. It is not clear from their report if this was case.

evidence of their credibility. Although both Moores et al. and Bledowski et al. provided thoughtful interpretations of their results, these investigators did not make clear, testable predictions based on their findings, and, to the best of my knowledge, no one has validated their results/interpretations with further experiments.

A final complication of using ICA, which is shared by PCA but less so by source localization and conventional ERP component definitions, is the difficulty of combining ICA results across participants. Ideally, ICs (or any other linear spatial filter) are derived for each experimental participant to accommodate individual differences in cortical folding, head shape, and EEG artifacts (e.g., participants differ in the magnitude and location of the muscle artifact they produce), which can greatly affect the ability of an LSF to accurately extract an EEG source. Subsequently, homologous ICs across participants would be identified using objective functional and anatomical criteria. In practice, however, this ideal is difficult to realize. Functional criteria are difficult to use because of the great trial-to-trial and inter-individual variability in the EEG. Anatomical criteria are difficult to use because of the EEG inverse problem (Section III.C.1), though perhaps with detailed EEG source models based on an individual's anatomy this may be less ambiguous than the results of idealized electrode coordinates and spherical head models (e.g., Section VI.C.4.a, Debener, Makeig, Delorme, & Engel, 2005; Makeig et al., 2004).

This dissertation employed two alternative methods for grouping together ICs across participants. The first of these defined a class of ICs based on a prominent ERP feature (i.e., the P3b--Section VI.C.3). This is similar to the approach used by Onton

and colleagues (2005) who objectively defined a class of ICs based on their theta power (5-7 Hz activity) and an estimate of their source location. This approach may be quite successful if the class of ICs is present in a high percentage of experimental participants and if the contribution of a single IC to that effect typically stands out from the rest. For example, based on the results of this dissertation, it appears that attempting to define a "Best P600" class of IC would not be successful as it is not generally clear which IC accounts for more P600 effect variance than any other IC (Figure VI.25: D).

The other approach employed by this dissertation was automated cluster analysis. Automated cluster analysis has the advantage over the "best X IC" approach in that it produces multiple classes of ICs and all participants are not expected to contribute exactly one IC to the class. However, cluster analysis is probably best used as an initial exploratory procedure that suggests types of ICs, which an investigator can subsequently hone by developing an objective functional and anatomical definition. This stems from the fact that clustering results typically depend on a somewhat arbitrary similarity metric, they might not be that reliable, and it might not be clear how many clusters the analysis should produce. For example, in this dissertation it appeared that the chosen 10 clusters were not remarkably reliable and that anywhere from 8 to 12 clusters were about as valid a choice as 10. It is quite possible that other IC similarity metrics (e.g., Makeig et al., 2004), clustering algorithms (e.g., de Sa & Ballard, 1998; Makeig et al., 2004), or increasing the number of participants would produce more reliable and compelling results than were found

here. However, to the best of my knowledge this remains to be demonstrated, as previous applications of cluster analysis to multi-participant ICs did not assess cluster reliability or employ an objective criteria for determining the number of clusters (Fogelson, Loukas, Brown, & Brown, 2004; Makeig et al., 2004).

Yet another approach to IC grouping is subjective investigator judgment (Debener et al., 2005; Debener et al., 2005; Onton, Westerfield, Townsend, & Makeig, 2006). This technique might be useful if the class of ICs is particularly distinguishable. For example, in my own experience, lateral occipital P1-N1 ICs (the lateral occipital IC clusters in Section VI.C.4.e.4, also see Makeig et al., 2004) and "mu ICs" (Makeig et al., 2004) are often quite identifiable. However, this approach can be problematic when the class of ICs is less distinct, as it is difficult to know how much an investigator's grouping reflects post-hoc bias and it is not clear how well that grouping will replicate. Moreover, if a class of ICs is readily identifiable to an investigator, it might be straightforward to derive an objective definition of that class of IC, which would be preferable.

In sum, I believe that there is good evidence that ICA can accurately extract some EEG sources and is a potentially useful tool for parsing ERPs. However, like any other method for defining ERP components, there are complications and possible pitfalls. ICA is probably best used in conjunction with source localization algorithms based on individual participant anatomy, as such anatomical models may greatly simplify the identification of homologous ICs across participants. Furthermore, having a large number of participants might help to deal with individual variation, as

with large samples, homologous classes of ICs might become more apparent and outlier ICs in a class will have less influence on the analysis of the class. Measures of individual cognitive abilities (e.g., working memory span, vocabulary size) and behavior may also help to make sense of individual participant variation (e.g., Makeig et al.'s analysis of faster versus slower responders, 1999, 2004), and employing an IC reliability test like that presented in this dissertation will enhance the reliability and, hopefully, the fidelity of results. Finally, any proposed class of ICs should be validated by a series of experiments, as should the results of any other method for dividing an ERP into components.

VII.G On the Potential Importance of Individual Variability in Studying the Electroencephalogram and the Cognitive Neuroscience of Language

"*Every brain's a snowflake.*"

-Anders M. Dale[103]

While my work with ICA these years has been complicated by individual variability, I believe it has given me a valuable appreciation of the degree to which

[103] This quote is taken from a lecture by Anders M. Dale, "Multi-Modal Imaging," presented to the students of COGS 1: Introduction to Cognitive Science at the University of California, San Diego on November 2, 2006.

ERP phenomena can vary from head to head. These differences have important implications for our ability to understand the brain via EEG and to make sense of how brains "do" language. Without exaggeration, I think the future of EEG research and the cognitive neuroscience of language very much depends on controlling for and understanding this variation instead of simply averaging over it.

In the work presented here, individual variation was most clearly apparent in the electrode level analysis of the P3b, N400, and P600 effects (Chapter IV). While all three effects were reliable across participants and the grand average ERPs very much resembled that of previous findings, there was considerable variation in the presence of these effects and their morphologies across participants. While all 16 participants significantly exhibited some type of greater positivity to targets, only 88% of participants showed a significantly greater negativity to incongruous sentence endings and only 63% of participants showed a significantly greater positivity to ungrammatical endings. Regarding inter-participant topography variation, it was remarkably large for the P600 effect, though it was not clear how much of the variation was due to ERP estimation error (i.e., single trial variation--with more trials the P600 effect may have been more similar across participants). Inter-participant topography variation in the P3b and N400 effects could also be considerable (median r_{cos}, of ~.80), and for the P3b effects this was clearly not due to ERP estimation error.

Such variability is not typically evident in ERP reports, which often only report grand average ERPs due to the difficulty of simultaneously visualizing ERPs from multiple conditions as well as the standard error of those ERPs or the ERPs from

multiple participants. In such reports, the degree of inter-participant variation is only evident in the results of their statistical hypothesis tests. However, there are other instances of such individual variation in the literature. For example, Moores and colleagues (2003) and Haan et al. (2000) reported considerable individual variation in 124 electrode measures of P3b and N400 effect topographies respectively. In addition, Osterhout (1997) reported that only half of 14 participants exhibited P600 effects to sentence final grammatically dispreferred words while the other seven exhibited N400 effects[104]. Though, of these same 14 participants, 12 exhibited P600 effects to grammatically dispreferred sentence internal words.

Broadly speaking, inter-participant EEG variation could have two causes. First of all, it could be due to individual differences in brain, head, and skull morphometry (this was the type of individual variability Dale was referring to in his above quote), which may be functionally irrelevant but nonetheless could have great impact on a participant's EEG. For example, small changes in cortical folding may dramatically affect source orientation, which would not only change the scalp topography of the source but may interact with other sources and individual differences in skull morphometry to greatly affect which sources are effectively visible at the scalp.

The second possible cause of individual EEG variation, individual differences in brain function, is even more confounding and is an issue for all brain measures, not just EEG. There is some evidence for such causes of P3b, N400, and P600 variation. For example, an attempt to localize the P3b (Moores et al., 2003) and the N400 (Haan

[104] Note, however, that Osterhout did not report what criteria he used to determine if an individual participant exhibited a P600 or N400 effect.

et al., 2000) recorded at 124 electrodes found considerable inter-participant variation in estimated source localizations. This variation was nicely visualized in Moores' and colleagues' Figure 3, which shows considerable variation in the laterality of the estimated P3b source locations and the degree to which central, occipital, and ventral temporal areas contributed. Note, however, that some of this variation could be spurious due to erroneous source model assumptions. For example, using a more minimal source model (spherical head model and a single or bilateral pair of dipoles) of the N400m[105], Halgren and colleagues (2002) reliably localized the effect in or near the left (or bilateral) superior temporal sulcus (sulci) in all 10 of their participants, though these models were clearly inadequate (i.e., they explained only 56% of variance the N400m effect on average). Further evidence for the P3b's sensitivity to differences in individual brain function comes a report by Fabiani and colleagues (1998) who found that older adults (i.e., 65-88 years old) with frontally distributed P3b's performed worse than older adults with canonically posterior P3b's on tests of frontal lobe function (e.g., the Wisconsin Card Sorting Test). The only evidence I know of that the P600 is sensitive to functional differences across participants is an experiment by Bornkessel and colleagues (2004), which found that participants with low reading spans did not elicit P600s at grammatically ambiguous points in a sentence but that participants with high reading spans did. However, this effect was not significant in their main analysis and the P600 was abnormally distributed. Furthermore, considering the variability of the P600 across studies and participants

[105] The N400m is the magnetoencephalographic analog of the N400.

(Sections II.E.3.a, IV.H.2), between-participant comparisons may well be spurious when the number of participants is small (there were only 12 participants in the two memory span groups).

While the above evidence is admittedly thin, I suspect that as our understanding of brain function improves we will be shocked by the amount of functional variability across ostensibly normal individuals. I think this will be especially true of language function, given that the neural substrates of language are obviously quite plastic as evidenced by the multiple modalities of linguistic codes (speech, sign language, text, Braille), by the voluminous amount of linguistic information fluent brains must have learned, and by the ability of young brains with lesions that would likely cause debilitating aphasias in adults to achieve relatively normal linguistic proficiency (Stiles, Reilly, Paul, & Moses, 2005).

Individual variation has several important implications for EEG research and cognitive neuroscience in general. At minimum, it adds variation to our measures, which obscures underlying regularity. A greater complication is that inter-participant functional variation must be explained by adequate models of the neural mechanics and acquisition of cognitive faculties. Finally, the great danger individual variation poses is the possibility that the average brain may be greatly misleading, potentially collapsing multimodal distributions into a single measure that is not representative of any of the modes.

So how might we deal with individual variation beyond averaging? One possible course is to collect a wide array of information about experimental

participants so that we can identify factors that systematically covary with our measures of brain structure and function. For example, Kutas (2006) has suggested using batteries of tests analogous to those used to diagnose neural disorders. Such a battery might include measures of sensory/motor accuracy and speed, working memory and attention spans, degree of hemisphericity, personality traits, and the ability to learn and recover from errors (ibid.). To the extent that there does exist functional variability across participants, these types of data will be absolutely necessary to understand these differences.

However, while such an approach is needed, it is complicated by the fact that it is not clear what variables we should be screening individuals for in the limited time they spend in our labs. The examples cited above are just a few of a very reasonable set of possibilities suggested by Kutas and undoubtedly the list could grow yet longer. Furthermore, no amount of such data would correct for the differences in brain/head morphometry that might not cause qualitative differences in brain function but, as mentioned above, may greatly complicate EEG measures of that function (though this is less of a problem for fMRI which can define some homologous areas across participants).

In light of these complications, I think a valuable, complementary approach would be to study a few model experimental participants in extensive detail. While this approach alone clearly will not be able to explain inter-participant functional differences and is liable to produce idiosyncratic results, by removing the confound of individual variability, it may afford us an opportunity to more easily see underlying

regularity that might generalize to other brains, as well as an opportunity to better understand how we should be analyzing our data. This latter point is particularly important when one considers that understanding a lot of cognitive phenomena (especially on the single-trial level), will require the ability to measure the activity of multiple parts of the brain and their interactions (Makeig, Debener, Onton, & Delorme, 2004), which will require tools like ICA and source localization. However, as just stressed above, to assess the accuracy of any putative unmixing it is important to validate the model with testable predications and additional experiments. Being able to perform such a series of experiments on the same participant would greatly facilitate such a procedure, especially if yet additional experiments had made it possible to functionally identify cortical regions (e.g., early sensory and motor areas) in that individual, which could be used to guide and interpret putative source models.

In conclusion, while progress has been and will surely continue to be made studying the average workings of average brains, my experience with even a modest sample of individuals has led me to believe that this approach is a crude oversimplification that leaves much of our data unexplained. If the degree of individual variability is as great as I believe it is, cognitive neuroscience needs to do better. Given the advances over the past 20 odd years in noninvasively measuring the structure and activation of the brain and in the availability of computational tools for analyzing rich multivariate time series, I am confident that cognitive neuroscience can do much better.

VII.H References

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., et al. (2004). Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Neuroscience, 24*(42), 9353-9360.

Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research, 21*(1), 11-21.

Borovsky, A., Elman, J., & Kutas, M. (2007). Getting the gist is not enough: An ERP investigation of contextual word learning. *Proceedings of the 29th Annual Cognitive Science Society,* Austin Texas, 119-124.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

de Sa, V. R., & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation,* (10), 1097-1117.

Debener, S., Makeig, S., Delorme, A., & Engel, A. K. (2005). What is novel in the novelty oddball paradigm? functional significance of the novelty P3 event-related potential as revealed by independent component analysis. *Cognitive Brain Research, 22*(3), 309-321.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. *Journal of Neuroscience, 25*(50), 11730-11737.

Donchin, E., & Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 112*(3), 357-374.

Duncan-Johnson, C. C., & Kopell, B. S. (1981). The Stroop effect: Brain potentials localize the source of interference. *Science, 214*(4523), 938-940.

Eichele, T., Specht, K., Moosmann, M., Jongsma, M. L., Quiroga, R. Q., Nordby, H., et al. (2005). Assessing the spatiotemporal evolution of neuronal activation with single-trial event-related potentials and functional MRI. *Proceedings of the National Academy of Sciences of the United States of America, 102*(49), 17798-17803.

Elman, J. L., Bates, E. A., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Fabiani, M., Friedman, D., & Cheng, J. C. (1998). Individual differences in P3 scalp distribution in older adults, and their relationship to frontal lobe function. *Psychophysiology, 35*(6), 698-708.

Fabiani, M., Karis, D., & Donchin, E. (1990). Effects of mnemonic strategy manipulation in a von Restorff paradigm. *Electroencephalography and Clinical Neurophysiology, 75*(2), 22-35.

Filion, D. L., Dawson, M. E., & Schell, A. M. (1993). Modification of the acoustic startle-reflex eyeblink: A tool for investigating early and late attentional processes. *Biological Psychology, 35*(3), 185-200.

Filion, D. L., Dawson, M. E., & Schell, A. M. (1994). Probing the orienting response with startle modification and secondary reaction time. *Psychophysiology, 31*(1), 68-78.

Filion, D. L., Dawson, M. E., & Schell, A. M. (1998). The psychological significance of human startle eyeblink modification: A review. *Biological Psychology, 47*(1), 1-43.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*(1-2), 3-71.

Fogelson, N., Loukas, C., Brown, J., & Brown, P. (2004). A common N400 EEG component reflecting contextual integration irrespective of symbolic form. *Clinical Neurophysiology, 115*(6), 1349-1358.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20 Suppl 1*, S8-17.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews, 25*(4), 355-373.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 360*(1456), 815-836.

Haan, H., Streb, J., Bien, S., & Rosler, F. (2000). Individual cortical current density reconstructions of the semantic N400 effect: Using a generalized minimum norm model with different constraints (L1 and L2 norm). *Human Brain Mapping, 11*(3), 178-192.

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *NeuroImage, 20 Suppl 1*, S18-29.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*(9), 416-423.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17*(3), 1101-1116.

Halgren, E., & Marinkovic, K. (1995). Neurophysiological networks integrating human emotions. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (1st ed., pp. 1137-1151). Cambridge, Mass.: MIT Press.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 721-742.

Horovitz, S. G., Skudlarski, P., & Gore, J. C. (2002). Correlations and dissociations between BOLD signal and P300 amplitude in an auditory oddball task: A parametric approach to combining fMRI and ERP. *Magnetic Resonance Imaging, 20*(4), 319-325.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology, 15*(1), 3-13.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293.

Kutas, M. (2006). One lesson learned: Frame language processing - literal and figurative - as a human brain function. *Metaphor and Symbol, 21*(4), 285-325.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Lakoff, G., & Johnson, M. (1980). *Metaphors we Live By*. Chicago: University of Chicago Press.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*(1), 47-62.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences, 8*(5), 204-210.

Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., et al. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biol., 2*(6), e176.

Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T. J., et al. (1999). Functionally independent components of the late positive event-related potential during visual spatial attention. *Journal of Neuroscience, 19*(7), 2665-2680.

Moores, K. A., Clark, C. R., Hadfield, J. L., Brown, G. C., Taylor, D. J., Fitzgibbon, S. P., et al. (2003). Investigating the generators of the scalp recorded visuo-verbal P300 using cortically constrained source localization. *Human Brain Mapping, 18*(1), 53-77.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25*(1), 75-92.

Onton, J., Delorme, A., & Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage, 27*(2), 341-356.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews, 30*(6), 808-822.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494-522.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: Counterarguments to Coulson, King, and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*(1), 1-14.

Paller, K. A., Kutas, M., Shimamura, A. P., & Squire, L. R. (1987). Brain responses to concrete and abstract words reflect processes that correlate with later performance on a test of stem-completion priming. *Electroencephalography and Clinical Neurophysiology: Supplement, 40*, 360-365.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences 13 (4): 707-784., 13*(4), 707-784.

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation, 13*(9), 2093-2118.

Rosenfeld, J. P., & Skogsberg, K. R. (2006). P300-based Stroop study with low probability and target Stroop oddballs: The evidence still favors the response selection hypothesis. *International Journal of Psychophysiology, 60*(3), 240-250.

Rosler, F., Hasselmann, D., & Sojka, B. (1987). Central and peripheral correlates of orienting and habituation. *Electroencephalography and Clinical Neurophysiology: Supplement, 40*, 366-372.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The Psychology of Word Meanings* (pp. 223-250). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Stevens, M. C., Calhoun, V. D., & Kiehl, K. A. (2005). fMRI in an oddball task: Effects of target-to-target interval. *Psychophysiology, 42*(6), 636-642.

Stiles, J., Reilly, J., Paul, B., & Moses, P. (2005). Cognitive development following early brain injury: Evidence for neural adaptation. *Trends in Cognitive Sciences, 9*(3), 136-143.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279-293.

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1656-1674.