

UC Santa Cruz

UC Santa Cruz Electronic Theses and Dissertations

Title

From Abstract to Episodic: Representations in Speech Production

Permalink

<https://escholarship.org/uc/item/2z12k9rs>

ISBN

9798186336303

Author

Paramore, Jonathan Charles

Publication Date

2026-07-31

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

**FROM ABSTRACT TO EPISODIC:
REPRESENTATIONS IN SPEECH PRODUCTION**

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

with an emphasis in DATA SCIENCE

by

Jonathan Charles Paramore

September 2026

The Dissertation of Jonathan Charles Paramore is approved:

Professor Ryan Bennett, Chair

Professor Rachel Walker

Professor Jaye Padgett

Distinguished Professor Douglas Bonett

Donald Smith
Interim Vice Provost and Dean of Graduate Studies

Copyright © 2026
Jonathan Charles Paramore

Contents

1	Setting the Stage	1
1.1	Introduction	1
1.2	The substance of speech representations	2
1.3	The locus of linguistic knowledge	10
2	Covert URs in Pakistani Punjabi	17
2.1	Punjabi	19
2.2	Experiment 1: How are pre-N vowels realized?	23
2.2.1	Participants	23
2.2.2	Stimuli and Design	24
2.2.3	Procedure	26
2.2.4	Analysis	28
2.2.5	Results	28
2.3	Experiment 2: Do pre-N vowels trigger nasal harmony?	37
2.3.1	Stimuli, design, and analysis	38
2.3.2	Results	41
2.3.2.1	Optionality of Nasal Harmony	42
2.3.2.2	Contrastive nasal vowels and nasal harmony	45

2.3.2.3	Nasality of pre-N vowels	46
2.3.2.4	Nasal Harmony	50
2.3.2.5	Summary of results for Experiment 2	53
2.3.3	Nasalization Reflects Harmony, not Coarticulation	55
2.3.4	Harmony does not Reflect a Fixed Nasalization Window	58
2.3.5	Summary of Experiment 2	61
2.4	Discussion	62
2.4.1	Analyzing $[\tilde{V}:N]$ vowels in Punjabi with abstract URs	62
2.4.2	Punjabi nasal opacity: a challenge for surface-oriented frame-works	65
2.4.2.1	Opacity in exemplar models	66
2.4.2.2	Opacity in surface-oriented generative models	75
2.5	Conclusion	80
3	Learning Abstract Structure	88
3.1	The UR PHaSE Learner	90
3.1.1	Step (1): Phonotactic Learning	91
3.1.2	Step (2): n-disparity UR EM Learning	92
3.1.2.1	Expectation Step (E-step)	93
3.1.2.2	Maximization Step (M-step)	97
3.1.3	Step (3): Minimum Likelihood Threshold Evaluation	98
3.1.4	Step (4): UR Candidate Expansion	100
3.1.5	Iteration Until Convergence	101
3.2	The UR PHaSE Learner and Punjabi pre-N vowels	101
3.2.1	Step (1): Phonotactic Learning	102
3.2.2	Step (2): 0-disparity UR EM Learning	104

3.2.3	Step (3): Minimum Likelihood Threshold Eval. for 0-disparity URs	107
3.2.4	Step (4): UR Candidate Expansion	107
3.2.5	Step (2): 1-disparity UR EM Learning	110
3.2.6	Step (3): Minimum Likelihood Threshold Eval. for 1-disparity URs	110
3.3	The Free Ride Learning Algorithm	112
3.4	The MDL Evaluation Metric	124
3.5	MaxLex	130
3.5.1	MaxLex and Klamath	133
3.5.1.1	Phonotactic Learning in Klamath	135
3.5.1.2	Morphologically Aware Learning in Klamath	137
3.5.2	MaxLex and Punjabi /V:N/ vowels	141
3.6	Discussion & Future Research	149
4	Phonetically Complete Neutralization in Mankiyali	154
4.1	Mankiyali	155
4.2	Experiment 1: Is nasality neutralized on the vowel?	159
4.2.1	Participants	159
4.2.2	Stimuli and Design	160
4.2.3	Procedure	161
4.2.4	Analysis	162
4.2.5	Vowel Nasality is Phonologically Neutralized	164
4.2.6	Vowel Nasality is Phonetically Neutralized	169
4.2.6.1	Nasalance	169
4.2.6.2	F1 Modulation	175

4.3	Experiment 2: Are there anticipatory coarticulation differences?	182
4.3.1	Stimuli, Design, and Analysis	183
4.3.2	Results	184
4.3.2.1	Neutralization on the final vowel	186
4.3.2.2	Neutralization of anticipatory nasalization	188
4.4	Discussion	190
4.4.1	Accounting for neutralization in generative frameworks	194
4.4.1.1	The Blueprint Model of Production	196
4.4.1.2	Incremental Phonological Planning Effects	197
4.4.1.3	Cascading activation	199
4.4.2	Accounting for neutralization in eliminativist models	201
5	Simulating Complete and Incomplete Neutralization in an Exemplar-Based Learning System	204
5.1	Central Tenets of Eliminativist Usage-Based Theories	205
5.1.1	Redundant Representations	206
5.1.2	Similarity-Based Association	207
5.1.3	Frequency Effects	209
5.1.4	Emergent Structure	210
5.1.5	Articulatory Biases	211
5.1.6	Decay	212
5.2	Modeling sound change through exemplar accumulation	212
5.2.1	Model input	213
5.2.2	Exemplar creation, optimization, and storage	215
5.2.2.1	Model 1 - Selection, Assignment, Storage	215
5.2.2.2	Model 2 - Selection, Assignment, Noise, Storage	216

5.2.2.3	Model 3 - Selection, Assignment, Pre-N Bias, Noise, Storage	217
5.2.2.4	Model 4 - Selection, Entrenchment, Pre-N Bias, Noise, Storage	223
5.2.2.5	Model 5 - Selection, Entrenchment, Pre-N Bias, Noise, Storage, Decay	227
5.2.2.6	Model 6 - Selection, Entrenchment, Pre-N/Similarity Biases, Noise, Storage, Decay	229
5.2.2.7	Model 7 - Complete Neutralization via Frequency Effects	237
5.2.2.8	Model 8 - Inclusion of higher nasality does not shift the optimization endpoint	241
5.2.2.9	Model 9 - change in production bias strength . . .	245
5.3	Discussion	247
6	Evidence for Episodic Speech Representations	254
6.1	Evidence for Detailed Representations from Punjabi	255
7	Conclusion	268
	References	272

List of Tables

2.1	LMER model summarizing the effect of speech rate manipulations on vowel duration: $\text{lmer}(\text{VDuration} \sim \text{SPEECHRATE} + (1 + \text{SPEECHRATE} \text{speaker}) + (1 \text{word}))$. The inclusion of vowel quality as a fixed effect did not improve fit ($\chi^2(5) = 2.53$, $p = .77$), so it was excluded from the model.	27
2.2	LMER model summarizing the effect of CONDITION ($\text{C}\tilde{\text{V}}\text{:}$, $\text{CV}\text{:N}$), TIMESTEP (beginning, middle, final), the CONDITION*TIMESTEP interaction, and VOWELQUALITY (ɑ: , ɔ: , o: , e: , u: , i:) on mean nasalance for nineteen speakers: $\text{lmer}(\text{nasalance} \sim \text{CONDITION} * \text{TIMESTEP} + \text{VOWELQUALITY} + (1 + \text{CONDITION} \text{speaker}) + (1 \text{word}))$. . .	34
2.3	Experiment 2 tokens.	40
2.4	Results of the LMER model – $\text{lmer}(\text{nasalance} \sim \text{CONDITION} * \text{TIMESTEP} + \text{VOWELQUALITY} + (1 + \text{CONDITION} \text{speaker}) + (1 \text{word}))$ – estimating nasalance across the final vowel for 13 speakers who categorically nasalize pre-N [$\tilde{\text{V}}\text{:N}$] vowels. CONDITION was dummy coded, and effect-coded fixed effects include TIMESTEP and VOWELQUALITY.	49

2.5	Results of the LMER model estimating mean nasalance of the glide in the [VGV] sequence for 13 speakers. CONDITION was included as a dummy-coded fixed effect. The final / $\tilde{V}$:/ in the contrastive nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition serves as the intercept in the model, and the estimates for the other predictors describe the predicted deviation from the intercept: <code>lmer(nasalance~CONDITION + (1 + CONDITION speaker) + (1 word))</code>	52
2.6	Results of the LMER model estimating mean nasalance of the penultimate, pre-glide stem vowel in the [VGV] sequence for 13 speakers who categorically nasalize pre-N [$\tilde{V}$:N] vowels. CONDITION (penultimate stem vowel (oral [VGV], pre-N [$\tilde{V}\tilde{G}\tilde{V}$ N], contrastive nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV])) was included as a dummy-coded fixed effect. The intercept is the penultimate stem vowel in the oral [VGV] condition: <code>lmer(nasalance~CONDITION + (1 + CONDITION speaker) + (1 word))</code>	53
2.7	Results of the LMER model estimating mean nasalance of the penultimate vowel in the [VGV] sequence for 15 speakers. CONDITION with two levels (penultimate vowel (/ $\tilde{V}$:=CV/) and penultimate vowel (/ $\tilde{V}$:=NV/)) was included as a dummy-coded fixed effect: <code>lmer(nasalance~CONDITION + (1 + CONDITION speaker) + (1 word))</code> . 61	
2.8	Experiment 1 tokens, consisting of 63 monomorphemic words separated into three groups: oral vowels (CV ⚬), nasal vowels (C$\tilde{V}$ ⚬), and pre-N vowels (CV ⚬N).	82
2.9	Experiment 2 tokens, consisting of 24 di- and trisyllabic words separated into three groups: oral [VGV], contrastive nasal / $\tilde{V}$:/, and pre-N vowels /V:N/.	83

3.1	Punjabi surface form categories	102
3.2	Constraint weights with their initial and optimized weights in the phonotactic stage.	103
3.3	Constraint weights, UR probabilities, and SR likelihoods after 0- disparity EM learning.	106
3.4	UR candidates for ‘breath’ and ‘penalty’, expanded to include all possible 1-disparity URs generated by flipping the [NAS] feature. .	110
3.5	Constraint weights, UR probabilities, and SR likelihoods after 1- disparity EM learning.	111
3.6	Constraint weights after phonotactic learning with MaxLex	143
3.7	Constraint weights and UR probabilities with concrete URs only .	146
3.8	Constraint weights and UR probabilities with abstract URs included	148
3.9	Surface forms provided to the UR PHaSE Learner.	152
4.1	Distribution of words, expected lapel tokens (20 speakers x 2 rep- etitions x # of words), and expected air pressure mask tokens (20 speakers x 4 repetitions x # of words) used in experiment 1, sepa- rated by condition.	161
4.2	LMER model summarizing the effects of CONDITION, (3 levels), TIMESTEP (3 levels), VOWEL QUALITY (5 levels), and the interac- tion between CONDITION and TIMESTEP on nasalance for all vowel tokens in experiment 1: $\text{lmer}(\text{nasalance} \sim \text{CONDITION} * \text{TIMESTEP} +$ $\text{VOWELQUALITY} + (1 + \text{CONDITION} \text{speaker}) + (1 \text{word}))$	172
4.3	Pairwise comparison showing the effect of vowel condition on nasalance. Results were extracted using <code>emmeans()</code> , which averages over other predictors in the model: <code>emmeans(lmer-exp1, ~ condition)</code>	173

4.4	LMER model summarizing the effects of CONDITION, (4 levels) on F1/F3 ratio for all mid low back vowel tokens uttered by all 20 speakers in experiment 1: LMER(F1/F3~CONDITION + (1 + CONDITION speaker) + (1 word)).	181
4.5	Words, expected tokens (16 speakers x 38 words x 4 repetitions), and actual tokens used in experiment 2 separated by condition.	185
4.6	LMER model summarizing the effects of CONDITION (3 levels), TIMESTEP (3 levels), VOWEL QUALITY (4 levels), and the interaction between CONDITION and TIMESTEP on mean nasalance of the final stem vowel for all tokens uttered by 16 speakers in experiment 2: LMER(nasalance~CONDITION*TIMESTEP + VOWEL QUALITY + (1 + CONDITION speaker) + (1 word)).	187
4.7	Pairwise comparison showing the effect of vowel condition on nasalance. Results were extracted using emmeans(), which averages over other predictors in the model: emmeans(lmer-exp2, ~ condition).	188
4.8	LMER model summarizing the effects of CONDITION (2 levels), TIMESTEP (6 levels), SEGMENT QUALITY (7 levels), and the interaction between CONDITION and TIMESTEP on mean nasalance of the V:G sequence for all tokens uttered by 16 speakers in experiment 2: LMER(nasalance~CONDITION*TIMESTEP + SEGMENT QUALITY + (1 + CONDITION speaker) + (1 word)).	189
4.9	Pairwise comparisons show the effect of vowel condition on nasalance. Results were extracted from the fitted LMER using emmeans(), averaging over the other predictors in the model.	190

4.10	Mankiyali experiment 1 tokens, consisting of 62 monosyllabic words separated into five groups: oral (CV:), nasal (C $\tilde$ V:), monomorphemic pre-nasal (CV:N), genitive oral (CV:-N), and genitive nasal (C $\tilde$ V:-N).	203
5.1	The model input consists of four categories with two lexical items each, for a total of eight exemplar clouds.	214
6.1	Linear mixed-effects model results for f0. Phase (baseline vs. post-production) was dummy coded, timestep, voice imitated during Phase 3 (high-prestige, low-prestige, or None) and vowel quality were effect-coded, and the interaction between TIMESTEP and PHASE was included as well: <code>lmer(f0~PHASE*TIMESTEP + VOWELQUALITY + IMITATIONVOICE + (1 word))</code>	266

List of Figures

1.1	Lexical connections for ‘price’, ‘prices’, ‘prize’, and ‘plays’ (cf., Bybee 2001, p.22-23).	13
2.1	The Punjabi vowel inventory (Shackle 2003)	20
2.2	Nasalance traces of the word /bi:n/ ‘musical instrument’ uttered by the same speaker in quick succession that differ only in the speech rate. The traces show the normalized nasalance amplitude across the length of the word.	29
2.3	Scatterplot showing the .09 Pearson Correlation value between speech rate (by-speaker z-scored vowel duration) and mean nasalance for all CV:N vowels produced by nineteen speakers in Experiment 1. .	30
2.4	GAM curves separated by condition showing mean vowel nasalance across 11 normalized timesteps for all monosyllabic tokens from Experiment 1 for nineteen speakers.	31

2.5	Dot and whisker plot showing mean nasalance of $\text{C}\tilde{\text{V}}\text{:}$ vowels from Experiment 1. The yellow dots and tails represent the mean $\text{C}\tilde{\text{V}}\text{:}$ vowel nasalance for each speaker and $\pm$ two standard deviations around this mean nasalance value respectively. Speakers are ordered from left to right by magnitude of standard deviation. The blue point shows the mean nasalance across all speakers, and its tails represent the proposed <i>RoPE</i> of $\pm.061$	36
2.6	Nasalance traces for two different productions of $/\text{u}:\text{v}\tilde{\text{e}}\text{:}\text{d}\text{ə}/$ ‘in that way’ by the same speaker in quick succession that differ only in the application of nasal harmony. The y-axis corresponds to nasalance, calculated in Praat by taking the normalized proportion of nasal air pressure to the total amount of air pressure in the vocal tract.	43
2.7	Density plot showing the distribution of mean nasalance for penultimate pre-glide stem vowels for speaker AW0. contrastive nasal $[\tilde{\text{V}}\tilde{\text{G}}\tilde{\text{V}}]$ includes penultimate pre-glide stem vowels from both the $[\tilde{\text{V}}\tilde{\text{G}}\tilde{\text{V}}=\text{CV}]$ and $[\tilde{\text{V}}\tilde{\text{G}}\tilde{\text{V}}=\text{NV}]$ conditions. The plot shows a bimodal distribution for penultimate stem vowels preceding contrastive nasal $/\tilde{\text{V}}\text{:}/$, with two distinct peaks around .25 and .52. The vertical line marks the cutoff point for speaker AW0, in which tokens with a mean nasalance to the left of the line were removed from the analysis. The multimodal distribution in the oral condition reflects inherent nasalance differences across vowel qualities (Bell-Berti, 1993).	45

2.8	GAM curves for the oral [VG̃V] and contrastive nasal [ṼG̃Ṽ=CV] conditions in Experiment 2. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for all 15 speakers in the experiment. Vertical dashed lines represent segment boundaries.	46
2.9	GAM curves for the oral [VG̃V], contrastive nasal [ṼG̃Ṽ=CV], and pre-N [VG̃ṼN] conditions in Experiment 2. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for the 13 speakers in the experiment who categorically nasalized the stem-final vowel in the pre-N [VG̃ṼN] condition. Dashed vertical lines represent segment boundaries.	47
2.10	Scatterplot showing the estimated -.05 Pearson Correlation value between penultimate vowel duration (using by-speaker z-scored values) and mean nasalance in the contrastive nasal [ṼG̃Ṽ=CV] and [ṼG̃Ṽ=NV] conditions.	57
2.11	GAM curves comparing the [ṼG̃Ṽ=NV] condition to the oral [VG̃V] and [ṼG̃Ṽ=CV] conditions. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for each condition for all fifteen speakers.	60
2.12	GAM curves comparing the tautomorphemic pre-N [VG̃ṼN] condition from Experiment 2 to forms with an underlying oral vowel followed by a nasal consonant suffix: morphemic boundary /Ṽ:-N/ (n = 104). Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for each condition for the thirteen speakers from Experiment 2 that categorically nasalize pre-N vowels.	65

2.13	Spectrogram and waveform showing a sample annotation for /pɑ:n/ for speaker AR0. The top waveform shows amplitude for the oral channel, and the bottom waveform shows amplitude for the nasal channel.	84
2.14	Waveforms showing the intrusive [ŋ] in /bã:k/ for speaker AK2. The top and bottom waveforms correspond to the oral and nasal channels respectively.	85
2.15	GAM curves for the oral [VG̃V], contrastive nasal [ṼG̃Ṽ=CV], and pre-N [VG̃ṼN] conditions, separated by speaker. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for the thirteen speakers that exhibit categorical nasality on final stem vowels in pre-N [VG̃ṼN].	86
2.16	GAM curves for the oral [VG̃V], contrastive nasal [ṼG̃Ṽ=CV], and pre-N [VG̃ṼN] conditions. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for the two speakers that exhibit coarticulatory nasality on pre-N /V:N/ vowels.	87
3.1	Lattice for a structured search space of candidate URs for a single morpheme realized contextually as [i] and [ĩ]. Visualized disparities only include the features [HIGH], [RD], and [FRONT].	96
4.1	Maps showing the location of Mankiyali-speaking villages in Pakistan. The yellow region is Khyber Pakhtunkhwa Province. Mankiyali is spoken in two small villages near Oghi Tehsil in Mansehra District, Danna and Dameka.	157

4.2	The Mankiyali vowel inventory (Paramore, 2021; Paramore and Aurangzeb, 2025)	158
4.3	Scatterplot showing the relationship between vowel duration (ms) and mean vowel nasalance for the oral /CV:/ (n = 306; yellow dots and regression line) and neutralized oral /CV:-N/ conditions (n = 312; blue dots and regression line). Lemmas included in the plot are /ba:(-ŋ)/, /kjo:(-ŋ)/, /po:(-ŋ)/, and /sq:(-ŋ)/. For both conditions, tokens greater than two standard deviations from a speaker’s mean vowel duration were discarded, resulting in the removal of 3 tokens from the /CV:/ condition set and 12 tokens from the /CV:-N/ condition set.	167
4.4	GAM curves separated by word and condition showing mean vowel nasalance across 11 normalized timesteps from experiment 1 for the nasal /C $\tilde{V}$:/, oral /CV:/, and neutralized oral /CV:-N/ conditions.	168
4.5	GAM curves separated by condition showing mean vowel nasalance across 11 normalized timesteps for all tokens in experiment 1, grouped by vowel height and condition. Underlyingly oral vowels are represented by the solid lines, underlyingly nasal vowels by the dashed lines, and vowels followed by a tautomorphemic nasal coda with a dotted line. The pink lines depict nasalance trajectories for neutralized nasal /C $\tilde{V}$:-N/ vowels, and the light blue line shows trajectories for neutralized oral /CV:-N/ vowels with a nasal coda.	170

4.6	Dot and whisker plot showing mean nasalance of /C $\tilde{V}$:/ vowels from experiment 1. Yellow dots indicate each speaker's mean /C $\tilde{V}$:/ vowel nasalance, with whiskers showing $\pm$ two standard deviations (95% of values) around that mean. Speakers are ordered from left to right by magnitude of standard deviation. The blue dot marks the mean nasalance across all speakers, and its whiskers represent the proposed <i>RoPE</i> of $\pm.078$ for underlying nasal /C $\tilde{V}$:/ vowels, equivalent to the 2 SD range of the speaker with the narrowest nasalance distribution (SR0).	174
4.7	F1/F3 — F2/F3 vowel space showing the mean F1/F3 and F2/F3 ratio values calculated from the midpoint of all vowels, averaged across speakers, and separated by condition: oral /CV:/ (dark blue), nasal /C $\tilde{V}$:/ (yellow), neutralized oral /CV:-N/ (light blue), neutralized nasal /C $\tilde{V}$:-N/ (pink), and tautomorphic /CV:N/ (green).	177
4.8	Generalized Additive Model (GAM) curves of mean F1/F3 ratio across the middle 10 timesteps for tokens with the low back mid vowel /o:/ or /õ:/ in Experiment 1. The blue line represents oral /Co:/ vowels and the yellow line represents nasal /Cõ:/ vowels. Shaded region shows the 95% confidence interval around the mean.	179
4.9	Generalized Additive Model (GAM) curves of mean F1/F3 ratio across 10 timesteps for tokens with the low back mid vowel /o:/ or /õ:/ in Experiment 1.	180

4.10	GAM curves separated by condition showing mean vowel nasalance at 33 normalized timesteps across the /V:GV:/ sequence for all tokens in experiment 2 produced by 16 speakers. Vertical dashed lines represent segment boundaries.	186
5.1	Hypothetical exemplar cloud for the English word ‘pan’.	207
5.2	Model 1: selection, assignment, storage. The plot shows nasality values of individual exemplars over 1,000 iterations. Non-neutralizing ORAL exemplars are in yellow, non-neutralizing NASAL exemplars in grey, neutralizing ORAL-N exemplars in blue, and neutralizing NASAL-N exemplars in red. Values in the key show mean nasality of each lexical item’s cloud at the end of the simulation.	216
5.3	Model 2: selection, assignment, noise, storage.	217
5.4	Model 3: selection, assignment, pre-N bias, noise, storage.	220
5.5	Model 3: selection, assignment, pre-N bias, noise, storage. The plot shows the average of the final mean nasality of the four word categories in the model over 100 simulations. Tails represent 95% confidence intervals around the means.	221
5.6	Model 3 but with 50,000 exemplars rather than 1,000.	222
5.7	Model 4: selection, entrenchment, pre-N bias, noise, storage. . . .	225
5.8	Model 4: selection, entrenchment, pre-N bias, noise, storage. . . .	226
5.9	Model 4, run over 10,000 iterations.	226
5.10	Model 5: selection, entrenchment, pre-N bias, noise, storage, decay. .	228
5.11	Model 5: selection, entrenchment, pre-N bias, noise, storage, decay. .	229
5.12	Model 6 without a scalar weight increasing the pull of the pre-N nasality bias.	231

5.13	Model 6 objective function: penalty landscape for ORAL-N forms as determined by the two bias components that make up the objective function. The solid green curve shows the pre-N bias when scaled by 5, which favors nasality similar to the NASAL category; the dotted pink curve shows the similarity bias, which favors values closer to the non-neutralizing ORAL mean. The thick navy curve plots their combined objective function, with its minimum indicating the optimal nasality target for ORAL-N tokens. Vertical dashed lines mark the typical ORAL mean (yellow) at 0.2, the typical NASAL mean (gray) at 0.6, and the resulting optimal ORAL-N value (blue). The figure illustrates how competing pressures jointly determine an intermediate nasality target for ORAL-N forms.	234
5.14	Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	235
5.15	Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	236
5.16	Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay. Here the similarity bias is symmetrical, such that neutralizing forms are able to influence non-neutralizing forms. . .	237
5.17	Model 7: selection (low ORAL freq.), entrenchment, pre-N/similarity biases, noise, storage, decay.	239
5.18	Model 7: selection (low ORAL freq.), entrenchment, pre-N/similarity biases, noise, storage, decay.	240
5.19	Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	243

5.20	Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	244
5.21	Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay. From iterations 200–400, all exemplars generated for lexical items in the ORAL-N category were initialized with .85 nasality.	245
5.22	Model 9: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	246
5.23	Model 9: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.	247
6.1	Phase 1 baseline productions interface for the word [pĩ:g] ‘swing’.	257
6.2	Images presented for the two Punjabi voices. The left image corresponds to the ‘low-prestige’ brick-kiln laborer and the image on the right to the ‘high-prestige’ Punjabi scholar.	258
6.3	Average f0 trajectories across 66 Punjabi words produced in isolation. For each vowel token, f0 was measured at 10 equally spaced time points. Generalized additive model (GAM) curves were then fit to the f0 trajectories and averaged across tokens, grouped by voice.	259
6.4	Description and imagery for the low-prestige voice that participants were exposed to at the beginning of each ‘low-prestige’ block in Phase 2.	260
6.5	Example interface from a high-prestige voice block for Phase 2 of the experiment. The high-prestige voice uttered the word [pĩ:g] ‘swing’ (highlighted in green), but the participant incorrectly selected [pin] ‘drink.IMP’ (highlighted in red).	261

6.6	Results from Phase 3 imitations of the high-prestige voice by a single participant. The solid pink curve shows the participant's baseline f0 trajectory, averaged across vowel tokens from 25 words produced in Phase 1. The blue curve shows the high-prestige voice's f0 trajectory for the same set of words. The remaining curves show the participant's productions grouped by shadowing iteration (first, second, and third repetitions of each word), with f0 trajectories averaged across the 25 words within each iteration. For each token, f0 was measured at 10 equally spaced timepoints, and values are averaged at each time point.	262
6.7	Results from Phase 3 imitations of the low-prestige voice by a single participant. The solid pink curve shows the participant's baseline f0 trajectory, averaged across vowel tokens from 25 words produced in Phase 1. The yellow curve shows the low-prestige voice's f0 trajectory for the same set of words. The remaining curves show the participant's productions grouped by shadowing iteration (first, second, and third repetitions of each word), with f0 trajectories averaged across the 25 words within each iteration. For each token, f0 was measured at 10 equally spaced timepoints, and values are averaged at each time point.	264
6.8	Results from Phase 4 postproductions (dashed blue line) compared to the Phase 1 baseline productions (solid pink line). Shading around each line represents the 95% confidence intervals.	265

Abstract

From abstract to episodic: Representations in speech production

Jonathan Charles Paramore

Eliminativist usage-based and classical generative speech production models differ fundamentally in two respects: (i) the nature of speech representations and (ii) the locus of linguistic knowledge. In this dissertation, I ask what levels of abstraction and detail are present in speech representations and whether a lexicon of detailed speech memories is sufficient to explain a speaker’s linguistic knowledge of their language. I investigate these questions using phonetic, phonological, experimental, and computational evidence from two understudied Pakistani languages, Punjabi and Mankiyali.

First, I present evidence from Punjabi that lexical representations can be completely covert, containing information that is not present in any experienced realization of a lexical item. I then demonstrate that, given an appropriately structured learning architecture, a learner can acquire covert speech representations efficiently. At the same time, a pilot study provides preliminary evidence that talker-specific phonetic and socially indexed information from recent linguistic experience is retained and reflected in subsequent production. Crucially, the same lexical items used to motivate covert representations also appear to be represented at a richly detailed episodic level.

Turning to the locus of linguistic knowledge, I present acoustic data from a production experiment in Mankiyali, showing that the neutralization of a phonological contrast can be phonetically complete. I then use simulations of exemplar accumulation to examine whether a lexicon of detailed speech memories can account for

the stable realization of such patterns. In these simulations, once the production and similarity biases define a stable target for an acoustic parameter, introducing exemplars with different values cannot by itself shift that target. A persistent shift in a category’s optimal acoustic realization instead seems to require recalibration of the biases shaping production, motivating higher-order linguistic knowledge that operates beyond stored speech memories.

Taken together, these findings challenge both eliminativist usage-based models, which reduce linguistic knowledge to accumulated exemplars, and classical generative models, which exclude fine-grained phonetic and social detail from linguistic competence. I argue for a hybrid architecture in which speakers retain detailed episodic memories, maintain abstract categories that may diverge substantially from experience, and possess higher-order knowledge that regulates their joint influence on speech production.

Acknowledgments

This dissertation would not have been possible without guidance and help from a large number of people. Primary among those names is my advisor, Ryan Bennett. Over the past four years, Ryan spent over 150 hours meeting with me to discuss my research, and at least double that amount outside of our meeting times reading my manuscripts and offering detailed feedback that greatly improved my growth as a scholar and the outcome of the final dissertation. I am flabbergasted at how he has the capacity to produce such high-quality research, while at the same time offering so much of his time to his students. He is truly the prototype exemplar of an excellent advisor! I am lucky to have sat under him these past several years, and I hope to emulate his dedication with my own students.

I am also deeply grateful to Brett Hyde for seeing potential in my work early on that I didn't know was there and for showing me that "a day without linguistics is like a day without sunshine". Thank you as well to Rachel Walker, Jaye Padgett, and Douglas Bonett for giving me so much of their time and for being excellent teachers as well as researchers. I had an embarrassment of riches in excellent educators during my time at UC Santa Cruz, which has shaped both my writing and teaching for the better.

I am deeply grateful to Sadaf Munshi for introducing me and connecting me with

the communities I work with in Pakistan. I am also thankful for helpful feedback from Gaja Jarosz, Grant McGuire, and Amanda Rysling on different aspects of the content in this dissertation.

Several research assistants also expedited and improved the quality of the experiments outlined in the dissertation; thank you to Yuehan Cao, İkra Çavuşoğlu, Steffi Kim, Jimin Lee, Regina Lim, Adeline Sui, and Sabrina Swank.

I would also like to express deep gratitude to my friends and colleagues in Pakistan, who worked tirelessly to help me collect high-quality data and to make me feel at home in Pakistan. Thank you to Aurangzeb for inconveniencing himself so much to help me travel safely to and from Mansehra and for opening his home to me so often. Thank you to Amir Akhoulkar for spending dozens of hours helping me get around Islamabad and Punjab and for being a wonderful friend and colleague. Thank you as well to Gohar Rehman, Muhammad Riaz, and many others who aided me in my fieldwork on Punjabi and Mankiyali.

I am also deeply grateful to my family — Alexis, Bella, and Theophilus — for their unbending support in moving across the country, and being patient with me as I spent many long hours and less-than-lucrative years pursuing my PhD. Their companionship makes even the duller days sweet.

Above all, I pray this dissertation is pleasing to God, the Creator of language. Exploring the nature of language as an object of scientific study is, at its essence, a study of God's nature. By understanding and uncovering the beautiful patterns inherent in human language, we understand the nature of God more clearly and can worship him more truly. *Soli Deo Gloria*.

Chapter 1

Setting the Stage

1.1 Introduction

Generative (Chomsky and Halle, 1968) and usage-based (Bybee, 2001) theories of linguistic cognition share many fundamental assumptions about the nature of language. Perhaps most importantly, both schools of thought regard, as is a unifying theme across the cognitive sciences, that the faculty of language is, at its core, an information-processing system (Dawson, 2013; Marr, 1982). Especially relevant for this dissertation, all major theories of linguistic cognition agree that the speech production process minimally requires the activation of stored speech representations associated with a given semantic form to serve as targets for the precise articulations of a planned word.

As will be introduced in the following sections, the primary differences between generative and usage-based approaches to speech production lie not in the function of language, but rather in the substance of the mental representations that constitute linguistic knowledge and in the mechanisms by which these representations

map to the physical production of words.

1.2 The substance of speech representations

In the classic generative view (Chomsky and Halle, 1968), speech representations of morphemes are constructed from a small number of discrete symbolic units (phonemes) that define contrasts in a language. The substance of these representations is sparse and free of predictable information derivable from the linguistic context. In essence, speech representations are considerable abstractions from the phonetic realization of a word that a language user actually experiences. The mental representation for an English word like ‘prices’ [p^hˌaɪsɪz] would thus be /pˌaɪsɪz/, a sparse, redundancy-free, minimal specification of only those unpredictable aspects of the phonology and phonetics of ‘prices’ which must be memorized.

Conceiving of speech representations as redundancy-free abstract categories has several benefits. For example, despite the fact that the vowels in the English words ‘mean’ [mīn] and ‘cleat’ [k^hli?t] appear in distinct phonetic contexts — and are in some ways phonetically distinct as a result — their production is biased toward consistent, regular, and similar pronunciation along several phonetic dimensions, primarily the first and second formants. Without positing an abstract category for the sound /i/ of which both of these vowels are members, it is difficult to explain why their phonetic realizations do not diverge considerably over time (Wedel, 2007, 2012). In the same way, though the vowels in English ‘pin’ and ‘pen’ have merged in many dialects (e.g., Southern American English Brown 1990) due to their identical phonetic environments, the vowels remain both measurably and perceptually distinct in others (e.g., in the Northeastern US). The inconsistency of the ‘pin’/‘pen’ merger would defy explanation unless the vowels in ‘pin’ and

‘pen’ belong to two distinct abstract categories (/ɪ/ vs. /ɛ/), with speakers’ future productions biased toward resembling previous instantiations of that category (Wedel, 2007, 2012). Indeed, if the two vowels are not categorized differently in some way by speakers, it is unclear why the merger would not gain a foothold in *all* English dialects.

Another benefit of positing abstract representational categories is the explanation they provide for systematic Neogrammarian sound changes, in which a sound change sweeps through the entire lexicon, impacting all instances of the relevant abstract sound category, regardless of context or lexical frequency (Pierrehumbert, 2016). In Grimm’s law (Bloomfield, 1933), for instance, voiceless stops in Proto-Indo-European changed to voiceless fricatives in Proto-Germanic. This can be seen in (1) by comparing cognates from Latin and English. Latin’s Italic lineage split from the Germanic line before the sound change took place in Germanic, and English belongs to the Germanic branch. As is evident, voiceless stops remain intact in Latin, whereas they are systematically spirantized across contexts and words in English. While spirantization is phonetically motivated in some contexts (e.g., intervocally), other contexts in which spirantization arose lack a clear phonetic grounding (e.g., word-initially). Without an abstract category uniting the voiceless stops as a natural class across individual words and contexts, we might expect this sound change to only affect a subset of the vocabulary and/or stops, leaving low-frequency lexical items and non-phonetically motivated contexts behind. Positing a broad representational category, [VOICELESS STOP], on the other hand, explains the striking regularity of the pattern.

(1) Grimm’s Law (Bloomfield, 1933; Grimm, 1822)

Gloss	Latin	English	Relevant Processes
‘father’	[pater]	[fɑðə]	/p/ → /f/; /t/ → /θ/ (intervocalic voicing)
‘seven’	[septem]	[sɛvɪn]	/p/ → /f/ (intervocalic voicing)
‘foot’	[ped]	[fʊt]	/p/ → /f/
‘three’	[tres]	[θri]	/t/ → /θ/
‘head’	[kaput]	[hɛd]	/k/ → /h/
‘horn’	[kornu]	[hɒv.ɪn]	/k/ → /h/

In contrast to the sparse, redundancy-free, and phonetically abstract representations assumed in generative phonology, a broad body of experimental literature within the usage-based tradition has provided convergent evidence demonstrating that lexical representations do, in fact, include phonetically-detailed, predictable, and context-dependent information about the pronunciation of words and morphemes (e.g. Bybee 2001; Goldrick and Cole 2023; Johnson 1997, 2006; Pierrehumbert 2001, 2016, and references therein). For example, a classic study by Goldinger (1998) explored the roles of talker familiarity and word frequency in a shadowing task and found that participants were faster to repeat words out loud if they had already heard that word previously uttered by the same speaker. He took this result as evidence that words are stored in memory as sets of episodic *exemplars*, corresponding to phonetically-detailed memories of past experiences hearing each word, including detail about a speaker’s identity.

Goldinger also found that speakers’ own productions sometimes imitated fine-grained phonetic properties of the prompts they heard. Imitation effects were stronger with low-frequency words than high-frequency words. Goldinger proposed that this effect can be explained if speech production targets are deter-

mined by some form of averaging over the stored exemplars for each word. Such averaging predicts that imitation effects should be more evident for low-frequency words: since low-frequency words have fewer stored exemplars in memory, any single exemplar should have a relatively outsized influence on how that word is pronounced. High-frequency words have many stored exemplars and so should have relatively stable production targets and be less liable to influence from recently heard productions than low-frequency words. Given that speakers retain and are influenced by highly detailed phonetic and social information from their individual experience with language, proponents of usage-based theories contend that the substance of speech representations must be highly detailed and indexed to concrete linguistic episodes.

Despite their reliance on phonetically-detailed memory traces, usage-based models of speech representation are not incompatible with abstract representations — by which I mean symbolic structures that are not directly tied to the properties of individual tokens but instead serve to classify tokens into higher-order phonological categories and support generalizations across them (e.g. Goldinger 1998; Goldrick and Cole 2023). In this dissertation, I assume that abstract representational categories coexist with a network of highly detailed episodic exemplar memories in the mind, which together shape a speaker’s linguistic knowledge of individual sounds and lexical items, and by extension their speech production patterns. The primary question addressed in the first half of the dissertation is, how abstract can these higher-order representational categories be? Put another way, how far may the substance of abstract categories diverge from the actual, observed forms that belong to these categories? As Kenstowicz and Kisseberth (1977, p.2-4) note, understanding the level of abstraction which a representation of a morpheme or

lexical item can take is a foundational question for understanding the nature of linguistic competence.

In traditional generative frameworks, it is perfectly acceptable — and in some sense the norm — for stored representations of lexical forms (URs) to differ extensively from surface forms (SRs) (e.g. Kenstowicz and Kisseberth 1977). More specifically, classical generative phonology holds the rather strong position that a UR can be *completely covert*, exhibiting properties that never appear in *any* surface form or allomorph of that UR. For example, O’Hara (2017) discusses a stem-final [i]~[Ø] alternation in Klamath verbs, illustrated by forms like [ʔe:ᵛw-a] ‘is deep’ ~ [ʔe:ᵛi-tkʰ] ‘deep’. O’Hara argues that these alternations cannot be analyzed as the deletion or insertion of [i]. Instead, he proposes that stems like [ʔe:ᵛ(i)-] contain an underlying final /e/, e.g. /ʔe:ᵛe/. Stem-final /e/ deletes in non-initial syllables when possible, but otherwise raises to [i] when deletion would produce an illicit consonant cluster. The consequence of this analysis is that the proposed underlying forms of stems like /ʔe:ᵛe/ are *covert*: they contain a segment, /e/, which contains features that never appear in any surface allomorph, echoing a long line of analyses of ‘yer’ vowels in Slavic languages (e.g., Gouskova 2012).

The URs I describe as ‘covert’ have typically been called ‘abstract’ (e.g. Albright 2002; Hyman 2018; Kenstowicz and Kisseberth 1979; Kiparsky 1976; Krämer 2012; O’Hara 2017; Wang and Hayes 2025). However, I opt for the term ‘covert’ here to specify more precisely exactly what type of abstract representation I am concerned with: namely, representations which contain information that a language user never experiences, in any context, for a given lexical item. Analyses making use of covert URs involve a high level of abstraction: speakers must somehow arrive at

a stored lexical entry for a stem like /ʔe:we/ which is qualitatively distinct from any surface form of that morpheme they have ever actually heard or produced.

As various authors have argued, some form of ‘abstraction’ can result from statistical averaging over observed surface exemplars, as outlined above (e.g. Goldinger 1998; Goldrick and Cole 2023). However, it seems unlikely that any kind of computation over observed surface forms is sufficient, alone, to produce covert lexical representations like Klamath /ʔe:we/ → [ʔe:w(i)-].¹ To the extent that covert underlying representations like /ʔe:we/ can be justified, they motivate richer, more articulated mechanisms for arriving at URs than can be produced by pooling over surface forms.

The second chapter of this dissertation provides phonetic and phonological evidence from Pakistani Punjabi which motivates the need for covert lexical representations. I measure oro-nasal air pressure across three different types of vowels in Punjabi: oral vowels /V:/, contrastive nasal vowels / $\tilde{V}$:/, and vowels occurring before a tautomorphemic nasal consonant /V:N/ (pre-N vowels). The results indicate that Punjabi speakers consistently produce pre-N vowels with nasalization that is phonetically indistinguishable from contrastive nasal / $\tilde{V}$:/ vowels. However, pre-N vowels interact opaquely with a pattern of nasal harmony triggered by contrastive nasal / $\tilde{V}$:/ . For this reason, I argue that pre-N vowels in Punjabi are represented underlyingly as oral vowels, even though the vowels in question are non-alternating, and thus *always* phonetically nasal on the surface. Phonological opacity is crucial to the argumentation in this chapter: as already observed by Ettlinger (2008) and others, productive, general patterns of opacity are often difficult to model without the assumption of URs that are categorically distinct from

¹While not addressed in this dissertation, the need for composite URs to capture patterns of vowel reduction in Seediq (Kuo, 2023) presents similar difficulties for usage-based models.

any single SR, or averaged set of SRs (see McCarthy 2007a for related discussion).

Criticisms of abstract representations have also come from within the tradition of generative phonology. It has been proposed that the core functions of underlying representations — in particular, their role in determining unpredictable patterns of alternation across paradigms (Kenstowicz and Kisseberth, 1979; McCarthy, 2007a) — can be subsumed by surface-oriented generalizations, including surface-to-surface relationships between allomorphs of the same morpheme (e.g. Albright 2002; Allen and Becker 2015; Burzio 1996; Cole and Hualde 2011; Hayes 1999; Scobbie et al. 1996, and various others). However, the pattern of opacity I discuss in chapter 2 cannot be straightforwardly modeled through the use of surface correspondence, because the key sound sequence (derived $/V:N/ \rightarrow [\tilde{V}:N]$) is *morphologically simple*, and *non-alternating* in all of the forms considered: they have only a single surface form, and thus no possibility for producing opacity through relations between distinct surface forms (e.g. Benua 2000; Mascaró 2016; McCarthy 2005a). To the extent that surface-oriented approaches *can* account for this data, they do so by implementing a formal distinction between predictable vs. unpredictable phonetic detail, thus recapitulating a core function of abstract URs (e.g. Cole and Kisseberth 1994, 1995). More to the point, I also argue that such approaches ultimately require the same covert representations, $/V:N/ \rightarrow [\tilde{V}:N]$, that I argue for in this dissertation.

Together, the Punjabi facts presented in chapter 2 motivate a broader conclusion about the substance of abstract representational categories. While the discussion above has highlighted debates internal to both usage-based and generative traditions, the empirical issue at stake is not specific to any one framework. Instead, the results bear on a shared assumption of all *surface-oriented models*, whether

eliminativist usage-based approaches or generative theories that restrict the substance of abstract categories to surface-attested material: the assertion that lexical representations cannot contain covert structure. I argue that the Punjabi opacity pattern requires precisely the opposite, a stored representation that diverges categorically from any of its surface realizations.

Nevertheless, despite the analytical motivation for covert URs that I defend in chapter 2, recent work in phonological learnability has raised serious concerns about the viability of such high degrees of UR abstraction. Two main issues arise. First, if multiple URs successfully model the observed data, what mechanism ensures that the learner converges on the minimally abstract UR? That is, how is UR abstraction permitted only to the extent necessary (Prince and Smolensky, 2004)? Second, and perhaps more pressingly, if UR abstraction is unconstrained, the space of possible URs grows without bound, introducing a potentially intractable search space problem (Jarosz, 2019; Wang and Hayes, 2025).

While several learning algorithms have been proposed that minimize UR abstraction (O’Hara, 2017; Paramore, 2025), the issue of search space size brought on by permitting increased UR abstraction remains largely unresolved. To date, most learning models have simply stipulated a finite set of potential representations for the learner to consider (O’Hara, 2017; Paramore, 2025). Wang and Hayes (2025), on the other hand, demonstrate that at certain levels of UR abstraction, the space of potential URs grows too large, preventing the learner from converging on a correct solution. Given these results, Wang and Hayes (2025) propose that the boundary on permissible UR abstraction should be defined by learnability constraints. This conclusion places learnability-based limits on abstraction in direct tension with patterns in languages like the Punjabi nasality data discussed

in chapter 2 that appear to require highly abstract URs.

As a solution, in chapter 3, I introduce a novel learning algorithm—the UR Progressive Hypothesis Space Expansion (UR PHaSE) learner—which aims to reconcile these competing demands. The UR PHaSE learner solves the issues posed by unbridled representational abstraction without imposing an explicit upper bound on the degree of permissible abstraction. The central insight of the learner builds on work by Tesar (2014, 2016), which structures the UR search space so that, rather than evaluating an infinite set of potential URs in parallel, the learner partitions the space into ordered subsets and considers candidates incrementally based on disparity count, the number of featural changes in a UR→SR mapping. Beginning with fully concrete URs, the learner only expands the space to include increasingly abstract URs (i.e., those that introduce more disparities) when it fails to find an adequate learning solution at a lower level of abstraction. By simulating learning of the Punjabi nasality patterns discussed in chapter 2, I show in chapter 3 that covert representations are learnable, given the appropriate architecture.

1.3 The locus of linguistic knowledge

The distinct views of usage-based and generative theories regarding the substance of speech representations also give rise to different ways of handling contextual variation in the realization of individual sounds and lexical items. As noted above, the traditional generative view holds that a single abstract underlying representation is stored for each morpheme in a speaker’s lexicon, and this representation is stripped of all redundant phonetic information that is otherwise recoverable from the contexts in which that morpheme appears. To recover contextual information, abstract representations are filtered through a formal linguistic grammar that de-

fines the target realization of the morpheme (i.e., SR) based on its surrounding linguistic context. For example, when the abstract underlying representation for ‘prices’, /pɹaɪs-z/, is filtered through a native English speaker’s phonological grammar, various processes — such as the aspiration of voiceless stops in the onset of a stressed syllable, the devoicing of liquids that follow voiceless aspirated stops, and the insertion of a vowel to break up adjacent sibilants — are all applied to the UR to arrive at the target surface form, [p^hɹaɪsɪz].

The postulation of this type of formal grammar provides a clear explanation of why, for example, obstruents are systematically devoiced word-finally in languages like Russian (Matsui, 2015), and why phonological patterns like obstruent devoicing apply regularly across the lexicon: grammatical processes are applied uniformly and indiscriminately to underlying representations, thereby giving rise to consistent context-dependent phonological patterns. Under the classic generative view, then, the formal grammar is a core component of what it means for a language user to possess fluent knowledge of their language. Linguistic knowledge in the generative framework is therefore seen as the combined mastery over the lexicon of speech representations *and* the grammar that maps those representations to target forms.

In stark contrast, usage-based views of representations typically maintain that all information needed to account for regular phonological patterns across the lexicon is contained within the speech representations themselves. That is, because representations are envisioned as highly detailed memories of concrete experiences, usage-based theorists argue that a high-dimensional network of relations between these detailed memories allows for the *emergence* of consistent phonological patterns without a formal grammatical mechanism that imposes the patterns exter-

nally. Given that a speaker retains experienced forms in memory, interaction with individual words accumulates over time as a language user has more experiences with the word. Consequently, sound patterns that are repeated with high degrees of frequency across exemplar memories and lexical items become increasingly *entrenched* in a speaker’s memory. These highly practiced speech patterns then serve as schemas for future productions of learned words as well as for novel forms exhibiting similar phonetic content, thus biasing speakers toward systematically reproducing and continually reinforcing entrenched phonological patterns.

Thus, the usage-based approach explains the systematic phonological processes applied to the English word ‘prices’ quite differently than the generative approach. When intending to produce ‘prices’, all exemplar memories associated with the word — including lexically distinct memories that are either semantically related (e.g., ‘cost’, ‘pricing’, etc.) and/or contain similar phonetic content (e.g., ‘price’, ‘prize’, ‘plays’, etc.) — are activated. Connections between activated representations highlight the entrenched and highly practiced motor patterns, such that a speaker’s production of ‘price’ is influenced and shaped most strongly by the most densely embedded patterns. As a result, each production, though phonetically distinct from any previous production of the word, is drawn toward a highly similar pronunciation. Aspiration of word-initial /p/ in English, then, arises not because of a formal grammatical rule or constraint enforcing contextual aspiration, but from the influence of all previous productions and perceptions of ‘prices’ that exhibit aspiration, as well as the influence of productions and perceptions of paradigmatically and phonetically related exemplar memories that do as well.

In sum, a speaker’s linguistic knowledge, under eliminativist usage-based frameworks, is wholly contained within a lexicon comprised of a highly detailed network

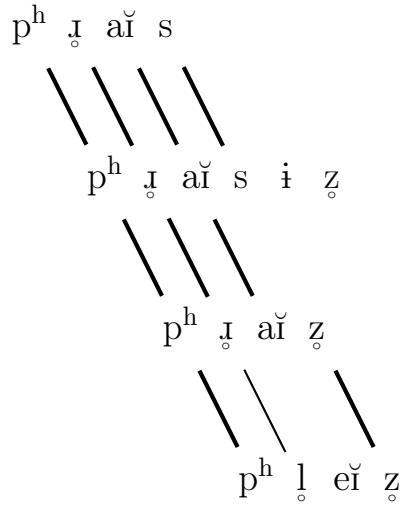


Figure 1.1: Lexical connections for ‘price’, ‘prices’, ‘prize’, and ‘plays’ (cf., Bybee 2001, p.22-23).

of episodic representations. That is, connections between these representations lead to the emergence of regular phonological processes and abstract representational categories, thereby *eliminating* the need to posit computation over symbolic representations (Goldrick and Cole, 2023, p.5). As a result, usage-based theories putatively obviate the need for linguistic knowledge that exists separate from the network of connections between exemplar memories.

Clearly, usage-based and generative views of the locus of linguistic knowledge are at odds with one another. The second half of the dissertation addresses this discrepancy, asking whether linguistic knowledge contained solely within the lexicon is sufficient to account for observed phonetic and phonological patterns. In chapter 4, I present phonetic and phonological data from oro-nasal airflow experiments on Mankiyali — an endangered language spoken by approximately 500 people in the mountains of northern Pakistan — which demonstrate that minimal pairs contrasting solely in vowel nasality are rendered homophonous when followed by a nasal consonant suffix. Crucially, I find that vowel nasality neutralization in

these contexts is not only phonologically complete but phonetically complete as well.

Previous work on phonetic neutralization has marshaled cases of phonetically *incomplete* neutralization in support of a network of related representational memories impacting future production of similar forms. For example, in German, vowels tend to be longer before voiced obstruents (e.g. Braunschweiler 1997; Du and Durvasula 2025). This effect seems to persist even when final devoicing has applied: the vowel in /ʁa:d/ → [ʁa:t] ‘wheel’ is slightly longer, phonetically, than the vowel in /ʁa:t/ → [ʁa:t] ‘council’. In usage-based models, the slight lengthening of the vowel in [ʁa:t] ‘wheel’ can be attributed to influence from episodic memories of the plural form [ʁɛ:d-ə] ‘wheels’, where the vowel is transparently lengthened before voiced [d]. No such influence exists for [ʁa:t] ‘council’, which has the plural form [ʁɛ:t-ə] ‘councils’ (see also Braver 2019; Kaplan 2017).

In chapter 5, I conduct a series of computational simulations of episodic memory accumulation in Mankiyali to show that cases of phonetically *complete* neutralization cannot be accounted for under usage-based frameworks without referencing linguistic knowledge that exists outside of the lexicon. Specifically, the strength of contextual and analogical biases that drive the categorically nasal realization of oral vowels before nasal suffixes in Mankiyali must be intentionally calibrated to produce a specific type of phonetic neutralization. That is, the strength (or weight) of the bias must be under speaker control to result in the variable realization of incomplete and complete neutralization. This kind of control is clearly a form of linguistic knowledge because the speaker must appropriately balance conflicting biases to achieve distinct phonetic realizations. Ultimately, I show in this chapter that, regardless of the input experienced by a speaker (i.e., the lexicon of

episodic memories), the calibration of biases is the sole driver of the stable state of the linguistic system.

Finally, chapter 6 returns to the substance of speech representations, presenting preliminary evidence from Pakistani Punjabi illustrating that speakers also retain phonetically and socially detailed episodic information in stored representations. In a pilot study examining the effects of repeated exposure to Pakistani Punjabi voices associated with different levels of social prestige, a participant showed asymmetric imitation patterns and a measurable post-exposure shift in f_0 toward that of the high-prestige voice. Although these findings are preliminary, they suggest that the same lexical items argued in chapter 2 to require covert abstract representations are also represented at a richly detailed episodic level. This result echoes a long line of work demonstrating that speakers encode highly detailed social and phonetic information in their speech representations.

Two primary findings emerge from the dissertation. First, abstract speech categories can be completely covert, exhibiting features that are never experienced by the language user. Second, some form of linguistic knowledge must exist beyond the lexicon. Chapter 6 complements these findings with preliminary evidence that words and sounds are simultaneously stored in representations containing detailed phonetic and social information. Taken together, these findings do not fit squarely within either the generative or the usage-based conception of speech production, thus giving rise to a number of questions: How do the different levels of speech representation interact? If abstract categories emerge over time from a language user's experience, what leads to the divergence of these categories from actual experience to the point where they become covert? What is the structure of the mechanism for linguistic knowledge that exists outside of these speech represen-

tations? How do these representations work together to shape actual productions of sounds and words?

These questions extend beyond the scope of the present dissertation. Chapter 7 synthesizes the findings presented here and argues that they motivate a hybrid architecture in which abstract symbolic categories coexist with richly detailed episodic memories and a higher-order mechanism that regulates their joint influence on speech production. Precisely how these components emerge and interact remains a question for future research.

Chapter 2

Covert URs in Pakistani Punjabi

In this chapter, I argue that abstract categories of speech can be completely *covert*, containing featural representations that are present underlyingly but are never experienced by the language user while speaking or listening. I support this claim with phonetic and phonological evidence from nasality patterns in Pakistani Punjabi.¹

This view of representations, though traditional in the generative sense, challenges the recent resurgence of surface-oriented models of linguistic representation. At least as early as Kruszewski (1883, p.51-52), before the Chomskian revolution that situated generative phonology as the dominant view in North American linguistics, speech representations were described as a collection of experienced memories that interact to form our representational knowledge of words and sounds. Today, usage-based theories abound, offering compelling evidence that speech representations contain redundant, socially structured information, and that much of the higher level linguistic structure argued to be innate by generative theorists can be

¹Much of the content in this chapter appears in Paramore and Bennett (submitted).

accounted for as emergent, arising from repeated patterns in stored speech experiences, rather than contained in a rich Universal Grammar predicated on symbolic representations.

Eliminativist usage-based theories advance the strong claim that the human cognitive system relies on broad domain-general cognitive faculties — such as rich memory capacity, pattern recognition, inference, and the ability to acquire expertise in precise motor control movements — to accomplish a wide array of complex tasks such as vision, walking, social interaction, *and language*. Under this view, language does not require specialized cognitive machinery unique to language, but instead reflects the application of general cognitive abilities to speech perception and production. As Bybee (2023, p.12) contends, linguistic representations can be constructed from these general capacities without recourse to innate, language-specific structure.

In this chapter, however, I argue that this view is insufficient. Not all phonological structure can be derived solely from averaging over one’s linguistic experience. In particular, some patterns require representational structure that go beyond the physical properties of experienced forms. Whether such structure is innate or emerges through mechanisms not reducible to experience alone remains an open question. What the evidence I provide here demonstrates is that linguistic representations can contain structure that is systematically absent from experience.

The empirical landscape from which I draw these conclusions is based on two experiments examining the phonetics and phonology of certain vowel nasality patterns in Pakistani Punjabi. Ultimately, these findings support a view of phonology in which abstract underlying representations (URs) can be completely covert, never surfacing in any allomorph without some kind of alteration. Because these

patterns involve non-alternating forms that opaquely interact with certain phonological processes, surface-oriented approaches to phonology — including eliminativist usage-based models that rely exclusively on observed exemplar memories to form lexical representations — struggle to account for them.

2.1 Punjabi

Punjabi (alternatively spelled Panjabi) is an Indo-Aryan language spoken by more than 100 million people worldwide (Hussain et al. 2019). Most Punjabi speakers live in Pakistan (78 million), where Punjabi is by far the most widely spoken first language. Additionally, the Indian states of Punjab, Haryana, and Himachal Pradesh are home to about a quarter of the total native Punjabi speakers (33 million) (Bashir and Connors 2019). Punjabi is considered to possess a high degree of dialectal diversity, with the varieties spoken in Pakistan often grouped together and termed Western Punjabi, and the varieties spoken in India termed Eastern Punjabi. Linguistic studies exploring Punjabi have largely focused on Eastern dialects, but a few descriptions of Pakistani varieties have been published in recent years (e.g., Baart 2014; Hussain et al. 2019). The data used in this chapter focuses on Punjabi as spoken in Pakistan. As of the summer of 2023, all 36 speakers who participated in the two experiments described here live in one of the twin cities of Pakistan, Rawalpindi and Islamabad. A significant number (56%) were born within 25 miles of the twin cities, with the rest scattered throughout the Pakistani Punjab province, in and around the areas of Gujrat and Sialkot (19%), Faisalabad (14%), Lahore (6%), and Multan (5%). Though Western Punjabi dialects vary widely from town to town, apart from the two speakers from Lahore, the nasalization results from the present study displayed a remarkable uniformity

across speakers.

The Punjabi vowel inventory is provided in Figure 2.1. As is evident, all long vowels contrast in nasality, but Bashir and Connors (2019, p.45) note that this contrast is confined to morpheme-final syllables, as shown in (1).

		Front			Central			Back		
		short	long	nasal	short	long	nasal	short	long	nasal
High	tense		i:	ĩ:					u:	ũ:
	lax	ɪ						ʊ		
Mid	tense		e:	ẽ:					o:	õ:
	lax		ɛ:	ẽ:	ə				ɔ:	õ:
Low	tense									
	lax								ɑ:	ã:

Figure 2.1: The Punjabi vowel inventory (Shackle 2003)

Crucially, a coda nasal consonant neutralizes the nasality contrast such that no minimal pairs of the shape [CV:N]-[C $\tilde{V}$:N] exist. This neutralization likely arises because all [V(:)N] vowels are said to be characterized by a high degree of nasality (hence [tã:n] ‘melody’ is attested, but hypothetical *[tɑ:n] is not). This account is supported by the results of Zahid and Hussain (2012), who conducted an acoustic study of vowel nasality in Pakistani Punjabi. While their study consisted of only three speakers, each uttering 12 repetitions of various short and long vowels from both nasal vowel types, Zahid and Hussain did not find a statistical difference in nasality between contrastive nasal [$\tilde{V}$:] vowels and pre-N [$\tilde{V}$:N] vowels in terms of their A1-P0 and A1-P1 values, measurements reported to correlate highly with nasality (Chen 1997; Styler 2017).²

²A1 refers to the amplitude of the highest harmonic in the first formant, which is typically dampened during the production of nasal vowels. P0 and P1, on the other hand, refer to the first and second ‘nasal poles’ respectively. These nasal poles are additional resonances that arise from the coupling of the nasal and oral cavities during the production of a nasal vowel. Accordingly, A1-P0 and A1-P1 serve as acoustic measures of these nasality resonances, normalized by A1 to control for variation in utterance intensity and/or recording conditions.

(1) Contrastive vowel nasality in Punjabi

- | | | | | | |
|----|----------|-----------------|-----|-----------|-----------------|
| a. | [tɑ:] | ‘warmth, fever’ | vs. | [tã:] | ‘that; so that’ |
| b. | [pɑ:vẽ:] | ‘cot leg’ | vs. | [tʃɑ:vẽ:] | ‘pumicestone’ |

In addition to the nasality contrast on long vowels and the neutralization of this contrast before nasal consonants, Punjabi also employs a process of nasal harmony, in which nasality spreads leftward from a nasal vowel, affecting all glides (/j/ and /v/) and vowels unless an intervening non-glide consonant blocks the spread. For example, words with a stem-final oral vowel (e.g., [sɑ:] ‘breath.SG’ and [mətʃi:] ‘fish.SG’) undergo nasal harmony when followed by the underlyingly nasal plural suffix, /-ã:/, as shown in (2a-b).³ However, when a non-glide consonant intervenes between the stem-final oral vowel and the plural suffix, harmony is blocked, as in (2c). Importantly, descriptions of the nasal harmony process do not clarify whether the pre-N [Ṽ:N] vowel in (2d-e)⁴ also instigates the spread of nasality, which serves as the central question of the second experiment discussed in this chapter (section 2.3).⁵

(2) Nasal harmony in Punjabi (Bashir and Connors, 2019; Bhatia, 1993)

- | | | | | |
|----|-------------|---|---------------------|------------------|
| a. | /sɑ:-ã:/ | → | [sã:vã:] | ‘breath-PL’ |
| b. | /mətʃi:-ã:/ | → | [mətʃi:ã:] | ‘fish-PL’ |
| c. | /ʃɑ:k-ã:/ | → | [ʃɑ:xã:] | ‘branch-PL’ |
| d. | | | [ã:vã:m]/[ɑ:vã:m]? | ‘general public’ |
| e. | | | [sĩjã:ŋ]/[si:jã:ŋ]? | ‘recognition’ |

There are two possibilities with regard to the participation of [Ṽ:N] vowels in nasal harmony. First, and perhaps most straightforwardly, if predictably nasal pre-

³A glide is epenthesized intervocalically in these cases to avoid vowel hiatus. An alternative analysis, in which the plural marker is underlyingly /-vã:/ and the glide deletes before stops, is ruled out because stop-glide clusters are independently licit (e.g., [tɑ:kətvər] ‘powerful’).

⁴I omit the underlying representations of these pre-N forms here because the following experiments aim to investigate them.

⁵/ɑ:vɑ:m/ is historically complex (/ɑ:v-ɑ:m/ ‘PL-common man’) but may not be synchronically analyzable. In any case, [ã:m] is non-alternating under either morphological analysis, which is the relevant point here.

N $[\tilde{V}:N]$ vowels are indeed phonetically indistinguishable from contrastive nasal $/\tilde{V}:/$ vowels and if nasal harmony in Punjabi is computed over surface phonetic representations, it would be logical to assume that both vowel types instigate the nasal harmony process in a similar fashion. After all, no nasality difference exists on the surface that distinguishes the two vowels.

On the other hand, it is possible that, despite being phonetically indistinguishable from contrastive nasal $/\tilde{V}:/$ vowels in nasality, non-alternating $[\tilde{V}:N]$ vowels do not instigate nasal harmony in the same way because they are phonologically distinct: specifically, because they are underlyingly oral, with predictable nasality provided on the surface by the phonological grammar.

This conclusion — which I argue in favor of — would support the existence of *covert URs*. Despite speakers consistently producing non-alternating $[\tilde{V}:N]$ vowels as categorically nasal, these vowels can be analyzed as oral in their URs $/V:N/$, given that nasality is predictable in this context. This distinction between underlyingly nasal $/\tilde{V}:/$ and underlyingly oral $/V:N/ \rightarrow [\tilde{V}:N]$ can be leveraged to explain the uneven application of harmony: only underlying, contrastive nasal $/\tilde{V}:/$ is a harmony trigger. Without positing a difference in the underlying representations of contrastive nasal $/\tilde{V}:/$ vowels and non-alternating $[\tilde{V}:N]$ nasal vowels in (2), it is difficult to explain why the two vowel types behave differently in their interaction with leftward nasal spread.

Crucially, $[\tilde{V}:N]$ vowels in words like those in (2c-d) do not exhibit morphological alternations in which the vowel occurs without the following nasal consonant. This means there are no surface forms (apart from those involving harmony) which directly and transparently demonstrate that these $[\tilde{V}:N]$ vowels are underlyingly distinct from contrastive nasal $/\tilde{V}:/$ vowels (i.e. there are no alternations like

/ɑ:va:=m/ → [ɑ:vã:=m] vs. [ɑ:va:=C] for these forms). As a result, the only direct, observed information speakers have regarding the nasality of pre-nasal [ĩ:N] vowels in words like ‘citizens’ and ‘recognition’ (2c-d) is that they are consistently realized as [+NAS]. Given that speakers do not have any direct surface evidence that [ĩ:N] vowels are underlyingly oral, surface-oriented models of speech production lack the necessary mechanisms to distinguish predictable pre-N vowels in [ĩ:N] from underlyingly contrastive nasal /ĩ:/ vowels.

2.2 Experiment 1: How are pre-N vowels realized?

Experiment 1 measured the phonetic realization of Punjabi non-alternating [ĩ:N] vowels and compared their realization to contrastive nasal /ĩ:/ and oral /V:/ vowels in the language. As mentioned above, Zahid and Hussain (2012) analyzed [ĩ:N] vowels for three Punjabi speakers and did not discover a statistical difference between their realization and the realization of contrastive nasal /ĩ:/ vowels with respect to acoustic correlates of vowel nasality. Thus, the first experiment aimed to replicate their findings using more speakers, and oro-nasal air pressure measurements rather than other properties of the acoustic signal that correlate with nasality.

2.2.1 Participants

Twenty native Punjabi speakers participated in the study (eleven men and nine women) ranging in age from 22-79 ($\mu = 39.7$; median = 33.5). All participants reported using Punjabi frequently throughout their daily lives in home, social, and academic settings. Participants were literate in Shahmukhi, a Persio-Arabic script used by most Western Punjabi speakers. Besides Punjabi, all participants also

spoke Urdu, the national language of Pakistan, and many were fluent in English. While almost all participants currently reside in Rawalpindi or Islamabad as of this writing, the cities in which they grew up span the Punjab Province. Ten speakers grew up in and around Rawalpindi, four in Gujrat, two in Sialkot, one in Sargodha, two near Toba Tek Singh, and one in the far south of Punjab in Sadiqabad. Despite the expansive distances between the hometowns of the participants, the results for Experiment 1 were strikingly consistent across speakers.

2.2.2 Stimuli and Design

Stimuli consisted of 63 monosyllabic tokens separated into three conditions: oral /V:/ (CV:), contrastive nasal / $\tilde{V}$:/ (C $\tilde{V}$:), and pre-N [$\tilde{V}$:N] (CV:N). A sample of words from each condition is given in (3). The complete list of tokens is given in Figure 2.8 of the Appendix. Pakistani Punjabi speakers use the Shahmukhi orthography — a Perso-Arabic script — for writing and reading the language. Nasal consonants are written with discrete symbols (𑂀 /m/, 𑂁 /n/, 𑂂 /ŋ/). Long vowels also have their own symbols, but glides in Shahmukhi share the grapheme of their corresponding vowel. For example, 𑂣 may represent /j/ or /i:/ and 𑂤 may represent /v/, /u:/, /o:/, or /ɔ:/. In addition, Shahmukhi is a partial abjad, such that short vowels are generally not indicated in the writing system. Finally, nasality on long vowels is indicated with a *nun gunnah* 𑂥, which is a similar character to the Shahmukhi /n/ 𑂁, but without the *nunqta*, or dot. However, the distinction in orthography between /n/ and nasality on vowels collapses when the nasal vowel is not the word final segment; in these cases, the *nun gunnah* is written with a *nun*, making it indistinguishable from /n/ in the orthography when not word-final, as shown for the nasal vowel in ‘swing’ in (3).

(3) Experiment 1 Conditions

oral V: (26)	contrast. nas. $\tilde{V}$: (20)	pre-N V: (17)
تھا $t^h\alpha$: ‘COP.PST’	تھال $t^h\tilde{\alpha}$: ‘place, room’	تھان $t^h\alpha:n$ ‘bolt of cloth’
سو $s\alpha$: ‘hundred’	سون $s\tilde{\alpha}$: ‘sleep.IMP’	سون $s\alpha:\eta$ ‘rainy season’
بی bi : ‘seed’	پینگ $p\tilde{i}:g$ ‘swing’	بین $bi:n$ ‘instrument’

Tokens from the **CV:** condition were used to establish a baseline measurement of orality, so no nasal segments were included in these forms. Tokens from the **C $\tilde{V}$:** condition had contrastive nasal $/\tilde{V}:/$ vowels and served as a baseline for nasality; to avoid any coarticulatory effects, nasal consonants were excluded from all tokens in this condition. Finally, the **CV:N** condition consisted of monomorphemic tokens, all with a nasal coda. Nasal onsets were also excluded from this condition to avoid the effects of carryover coarticulation.

Whereas the oral **CV:** and contrastive nasal **C $\tilde{V}$:** conditions provide clear baselines of orality and nasality, the **CV:N** condition allowed for a comparison between the realization of pre-N $[\tilde{V}:N]$ vowels and the other two conditions. Whenever possible, near-minimal pairs or triplets across the three conditions were used.

As is well known, differences in vowel height correlate with differences in nasality (Bell-Berti 1993; Delvaux et al. 2008; Henderson 1984). To control for this, a balanced range of vowel qualities was included in each condition as well, as shown by the contingency table in (4).

(4) Distribution of vowel quality across conditions in Experiment 1

		Condition			Total Items
		oral CV:	contrast. nas. C $\tilde{V}$:	pre-N CV:N	
Vowel Quality	i:	4	1	2	7
	u:	4	5	3	10
	e:	2	1	0	3
	o:	4	3	2	9
	ɔ:	2	1	2	5
	ɑ:	10	9	8	27
Total Items		26	20	17	63

2.2.3 Procedure

All participants were recorded individually in a sound-attenuated room at a university in Rawalpindi, Pakistan. A native Punjabi research consultant — Amir Adkuokhar — provided participants with a consent form in Punjabi detailing what the experiment involved. They were told that the experiment measured air pressure variations from the nose and mouth to study the Punjabi sound system.

After signing the consent form and filling out a demographic questionnaire, the research consultant introduced participants to the equipment and experimental procedure. All instructions were given in Punjabi. A dual chamber oro-nasal air-pressure mask connected to two separate transducers produced by Glottal Enterprises⁶ was used to measure air pressure from the nose and mouth during speech at a sampling rate of 11,025 Hz using Audacity. The silicone mask was constructed with a bifurcation to separate the oral and nasal chambers, allowing for simultaneous, time-aligned air-pressure measurements in each channel. These signals were used to compute changes in nasalance over time, defined as the ratio of nasal air pressure (A^n) to the total amount of oral (A^o) and nasal air pressure across both chambers: $\frac{A^n}{A^n + A^o}$. The mask was also equipped with a handle which

⁶Specifications for the product can be found at glottal.com/Oro_Nasal_Mask.html.

participants used to raise and lower the mask to and from their face during the recording. Before the recording began, participants completed a test trial, in which they uttered several words into the mask to ensure they both understood the procedure and held the mask firmly against their face during productions to prevent airflow leakage.

At this point, the recording started, and participants were presented with the 63 monosyllabic tokens in a randomized order using a Python Graphical User Interface. Participants were told to read each item four times while holding the mask firmly against their face: twice slowly and twice using fast/rapid speech. Speech rate was self determined by speakers. However, I monitored productions during the experiment, reminding speakers to maintain the target speech rate patterns if they deviated noticeably from it. The experiment included speech rate as a condition to test whether vowel duration modulates the degree of nasality on /V:N/ vowels (Solé 1992, 1995). As shown by the lmer model in Table 2.1, the intended speech rate manipulations were successful: vowels in tokens produced with instructions to speak slowly were approximately 82 milliseconds longer than those produced with instructions to speak quickly.

Predictor	β (ms)	95% CI	t	p
Intercept (fast speech)	205	[186, 224]	20.74	< .001 ***
slow speech	82	[72, 91]	17.14	< .001 ***

Table 2.1: LMER model summarizing the effect of speech rate manipulations on vowel duration: `lmer(VDuration~SPEECHRATE + (1 + SPEECHRATE|speaker) + (1|word))`. The inclusion of vowel quality as a fixed effect did not improve fit ($\chi^2(5) = 2.53$, $p = .77$), so it was excluded from the model.

Tokens were presented in the Shahmukhi script with English translations below and instructions repeated at the bottom of the screen. If a word was mispro-

nounced, the research consultant instructed participants to repeat all four iterations of the token. Recording sessions lasted between 8 and 16 minutes.

2.2.4 Analysis

Recordings were hand-annotated with segment boundaries using Praat (Boersma and Weenink 2015). Details of the annotation procedure are outlined in the Appendix.

After annotation, a Praat script iterated through the annotated files, converting the oral and nasal air pressure channels to intensity tracks. A series of R scripts extracted mean nasalance values ($\frac{A^n}{A^n + A^o}$) over 11 time-normalized windows for each segment from these intensity tracks. The values were then saved in a data frame for subsequent analysis.

2.2.5 Results

Data from nineteen of the twenty speakers that participated in the experiment is included in the results presented here, summing up to 4,761 tokens (63 tokens x 4 repetitions x 19 speakers - 27 low-quality tokens) in the analysis. One speaker, MA0, was removed from the analysis because they produced a large number of tokens in which vowels from the **C $\tilde{V}$** and **CV:N** conditions exhibited abnormally low nasalance, essentially the same as vowels in the oral **CV** condition.

Tokens were analyzed using R (RCoreTeam 2016). To visually compare the relative nasality between the three conditions, the vowel segment was extracted from each token, and nasalance — defined as the proportion of total air pressure present in the nasal cavity — was used to fit three generalized additive model (GAM) curves, one for each vowel type, to the data using ggplot2 (Wickham

2016). No statistical difference in nasalance was found between tokens uttered in slow speech and tokens uttered in fast speech, so speech rate was not included as a predictor in the model. This fact is evident from comparing the two low-pass filtered nasalance traces in Figure 2.2 for the word /bi:n/ ‘musical instrument’. The same speaker uttered both tokens in short succession, but the words differed in speech rate; the vowel of the first repetition (#7) was 389 ms in duration, and the vowel of the second repetition (#8) was 191 ms in duration. Despite this substantial durational difference, nasalization is produced categorically across the entire span of both CV:N vowels. This indicates that the degree of nasality on the CV:N vowel is not proportionally modulated based on speech rate differences (Solé 1992, 1995).

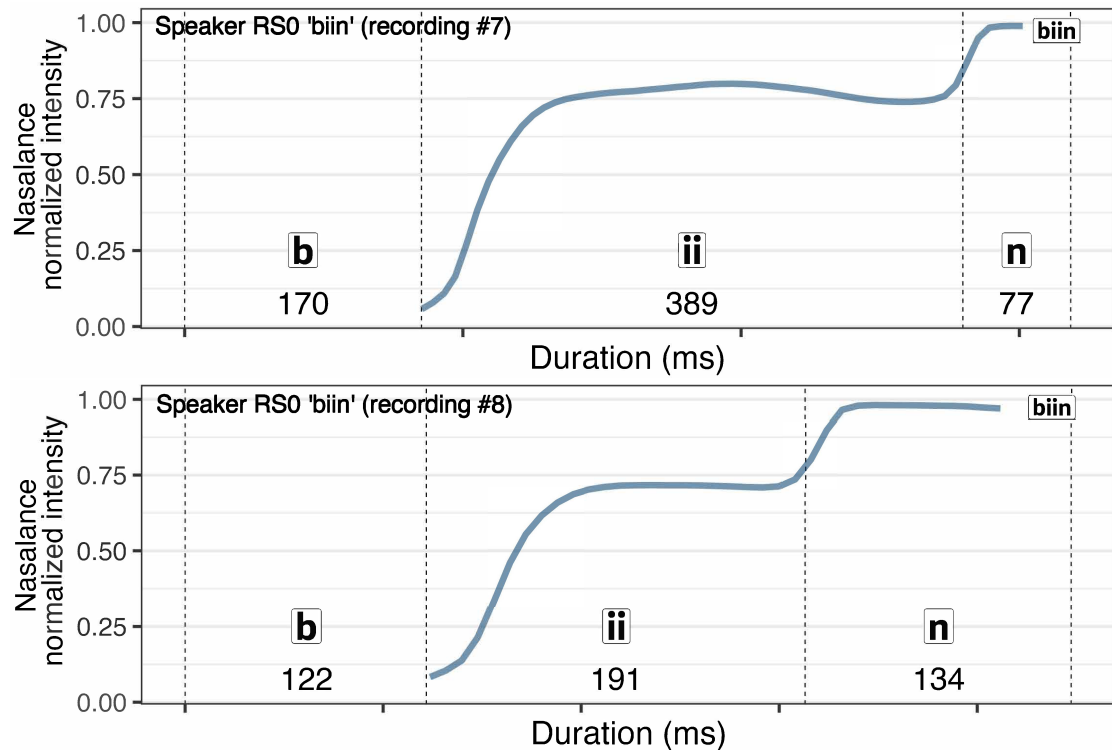


Figure 2.2: Nasalance traces of the word /bi:n/ ‘musical instrument’ uttered by the same speaker in quick succession that differ only in the speech rate. The traces show the normalized nasalance amplitude across the length of the word.

To confirm that nasalance is not modulated by speech rate for CV:N vowels, I conducted a Pearson’s correlation test to examine the relationship between the two variables. Because speech rate varies systematically across speakers, vowel durations were z-scored within speaker (i.e., each token’s duration was centered and scaled by that speaker’s mean and standard deviation). Across speakers, raw vowel durations ranged from 50 - 658 ms ($\mu = 248$, $\sigma = 74$), providing an ample range to evaluate rate-induced differences in nasality. The results ($r(1290) = .09$, $p < .001$ ***), visualized by the scatterplot in Figure 2.3, show a statistically significant positive correlation between speech rate and mean nasalance for CV:N vowels. However, Pearson correlation values less than .3 are typically regarded as *practically* insignificant, regardless of whether they reach statistical significance (Bonett, 2021, vol.2, p.3).

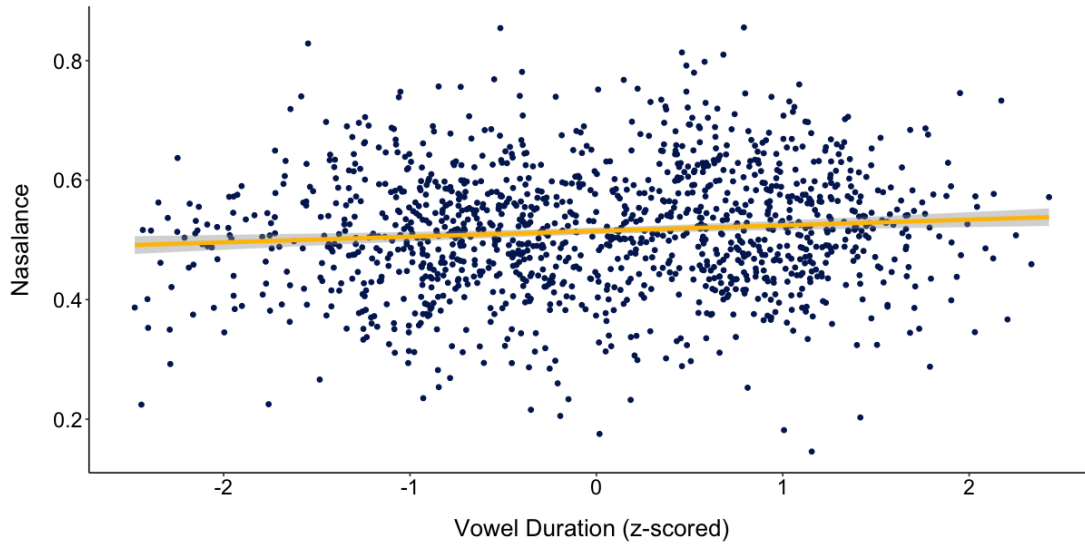


Figure 2.3: Scatterplot showing the .09 Pearson Correlation value between speech rate (by-speaker z-scored vowel duration) and mean nasalance for all CV:N vowels produced by nineteen speakers in Experiment 1.

Figure 2.4 displays plots of the three GAM curves depicting mean nasalance across

all tokens and speakers. As expected, nasalance for oral **CV:** vowels remains steady and noticeably lower than the other two vowel types at about 40% across the length of the oral vowel. For the contrastive nasal $/\tilde{V}:/$ vowel condition (**C $\tilde{V}$:**), on the other hand, nasalance begins at about 32% at the onset of the vowel and rises steadily until reaching a plateau-like position at about 55-60% nasalance a third of the way through the vowel. This proportion of nasalance is maintained throughout the remainder of the vowel’s duration. As for the nasalance of $[\tilde{V}:N]$ vowels in the **CV:N** condition, they follow a very similar pattern to contrastive nasal $/\tilde{V}:/$ vowels, in which nasalance at the onset of the vowel sits at approximately 32%, followed by a steady rise in nasalance until a plateau is reached at about 55-60% nasalance, which is maintained through the offset of the vowel.

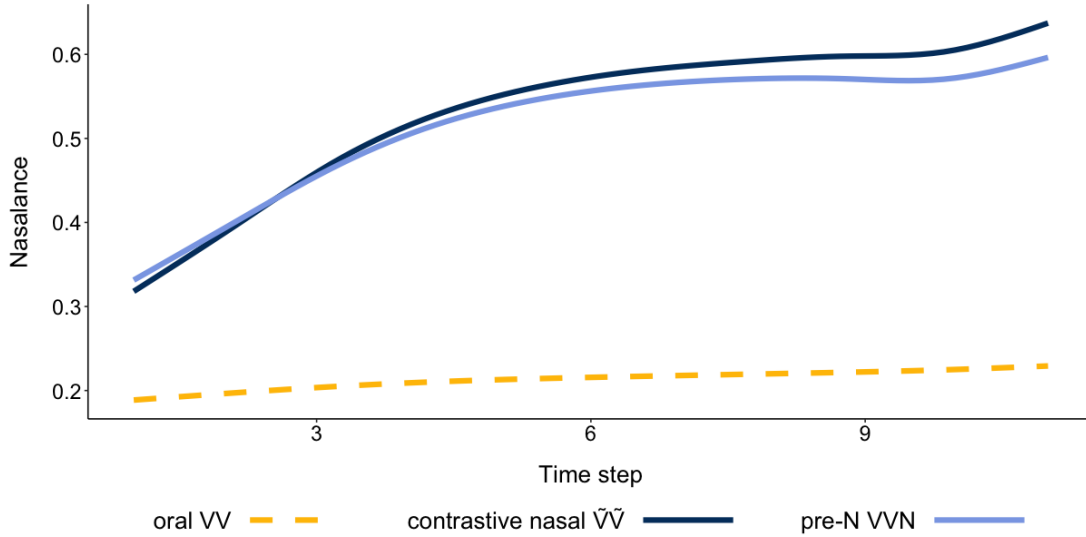


Figure 2.4: GAM curves separated by condition showing mean vowel nasalance across 11 normalized timesteps for all monosyllabic tokens from Experiment 1 for nineteen speakers.

To quantitatively compare the nasalance between **C $\tilde{V}$:** and **CV:N** vowels, a linear mixed-effects regression (LMER) model was fit to the data using the `lmer()` func-

tion from the lme4 package (Bates et al., 2025) with mean vowel nasalance as the dependent variable and CONDITION ($\text{C}\tilde{\text{V}}\text{:}$ and $\text{CV}\text{:N}$), TIMESTEP (beginning, middle, final), the CONDITION*TIMESTEP interaction, and VOWELQUALITY included as the four fixed effects. Oral vowels were omitted as an additional level of CONDITION in the model to allow for a more straightforward comparison between $\text{CV}\text{:N}$ and $\text{C}\tilde{\text{V}}\text{:}$ vowels. Moreover, rather than include all eleven steps at which nasalance was measured as separate levels of the TIMESTEP factor, the steps were binned into three different levels, with steps 1-4 recoded as the first 1/3 of the vowel, steps 5-8 recoded as the second 1/3, and steps 9-11 recoded as the final 1/3. VOWELQUALITY was included as a predictor to account for the known differences in nasality across different vowel qualities mentioned above. A likelihood ratio test showed that the addition of VOWELQUALITY as a fixed effect significantly improved the fit ($\chi^2(5) = 28.83$, $p < .001$ ***), so it was retained in the model. By-word random intercepts were included, as were by-speaker random intercepts and random slopes for CONDITION.

The results of the LMER model are provided in Table 2.2. The main effects of TIMESTEP and VOWELQUALITY were effect (aka sum) coded, and the main effect of CONDITION was dummy (aka treatment) coded. Consequently, the intercept of the model (β_0) refers to the mean nasalance of contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowels, averaged across all levels of the TIMESTEP and VOWELQUALITY factors. The $\text{CV}\text{:N}$ coefficient (β_1) is the value of primary interest in the model, quantifying the estimated difference between the two vowel types. The coefficients for TIMESTEP and VOWELQUALITY— β_2 - β_{10} in Table 2.2 — represent deviations of the listed factor level from the average nasalance of contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowels, controlling for the other predictors. The final three coefficients (β_{11} - β_{13}) reflect the

CONDITION*TIMESTEP interaction, capturing potential differences in the temporal trajectory of nasalance between contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ and $\text{CV}\text{:N}$ vowels.

The CONDITION*TIMESTEP interaction was significant at the beginning and end of the vowel, showing that $\text{CV}\text{:N}$ vowels are slightly more nasalized than contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ in the beginning third of the vowel and slightly less nasalized in the final third. Nevertheless, while the interaction is statistically significant, the predicted differences in nasalance at the vowel onset and offset are very small on the nasalance scale ($\pm .017$ nasalance or 1.7% percentage differences). These localized deviations across conditions are consistent with small phonetic differences across contexts (e.g., non-identical onsets and the lack of a coda in most $\text{C}\tilde{\text{V}}\text{:}$ tokens). Moreover, the large dataset provides sufficient statistical power for the model to detect such subtle effects. Crucially, incorporating the interaction does not change the overall pattern; the two conditions exhibit highly similar nasalance trajectories across the vowel, and the main effect of condition remains substantively interpretable in light of these small timestep interactions. Similarly, the main effect of VOWELQUALITY was significant for four of the six vowels included in the study, reaffirming that different vowels exhibit differing degrees of nasalization, controlling for other factors.

With the CONDITION*TIMESTEP interaction and VOWELQUALITY effects noted above, the primary takeaway from the model for our purposes is located in the first two rows. Specifically, the estimated difference of .016 nasalance between the contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowel and the pre-N $\text{CV}\text{:N}$ vowel does not reach statistical significance. This means that the null hypothesis that the two vowels are indistinguishable in terms of their nasalance values cannot be ruled out.

Predictor	β	95% CI	t	p
Intercept ($\text{C}\tilde{\text{V}}\text{:}$)	.541	[.514, .568]	39.55	< .001 ***
$\text{CV}\text{:N}$	-.016	[-.039, .007]	-1.36	.18
Beg	-.114	[-.117, -.112]	-91.54	< .001 ***
Mid	.040	[.038, .043]	32.29	< .001 ***
Fin	.074	[.071, .077]	51.04	< .001 ***
u:	.038	[.016, .059]	3.44	.002 **
i:	.052	[.021, .082]	3.34	.002 **
o:	-.046	[-.071, -.022]	-3.66	< .001 ***
ɔ:	-.009	[-.040, .021]	-.60	.55
e:	-.012	[-.061, .037]	-.47	.64
ɑ:	-.022	[-.04, -.004]	-2.44	.02 *
$\text{CV}\text{:N}:\text{Beg}$	.017	[.013, .021]	9.23	< .001 ***
$\text{CV}\text{:N}:\text{Mid}$	-.002	[-.005, .002]	-.88	.38
$\text{CV}\text{:N}:\text{Fin}$	-.015	[-.019, -.011]	-7.27	< .001 ***

Table 2.2: LMER model summarizing the effect of CONDITION ($\text{C}\tilde{\text{V}}\text{:}$, $\text{CV}\text{:N}$), TIMESTEP (beginning, middle, final), the CONDITION*TIMESTEP interaction, and VOWELQUALITY (ɑ:, ɔ:, o:, e:, u:, i:) on mean nasalance for nineteen speakers: $\text{lmer}(\text{nasalance} \sim \text{CONDITION} * \text{TIMESTEP} + \text{VOWELQUALITY} + (1 + \text{CONDITION} | \text{speaker}) + (1 | \text{word}))$.

Importantly, while no effect of CONDITION was found to alter the mean nasalance of the two vowels, the results do not allow us to conclude that the nasalance of contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ and pre-N $\text{CV}\text{:N}$ vowels are equivalent. To do so, an equivalence test must be conducted to determine whether we can accept the null hypothesis that the mean nasalance of the two vowel types is the same. Within the framework of classical statistics, a test of equivalence relies on determining a region of practical equivalence (-h to h) for a specific research question and examining whether or not the confidence interval for the difference in means between two groups falls within that region (Bonett 2021; Lakens et al. 2018; Wellek 2010). If the confidence interval for the difference is wholly contained within that interval, the hypothesis that a practical difference exists between the two groups with respect to the dependent variable can be rejected. For this experiment, the de-

pendent variable is mean nasalance, and the two groups of interest are the vowels in the $\text{C}\tilde{\text{V}}\text{:}$ and $\text{CV}\text{:N}$ conditions.

To determine a region of practical equivalence (*RoPE*) for mean nasalance in Punjabi, I first grouped all contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ tokens from Experiment 1 by speaker and vowel quality and calculated the mean and standard deviation of segmental nasalance within these groups. Computing by-speaker nasalance means separately for each vowel quality controls for inherent differences in nasalance across vowel heights. For each speaker, the mean nasalance and standard deviation was then averaged across vowel qualities to obtain standard nasalance distributions. Finally, the *RoPE* was defined as ± 2 standard deviations of nasalance for the speaker with the narrowest variation. Because all vowels in the contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ condition in this experiment are phonologically contrastive nasal vowels, any variation in nasalance on these vowels within an individual speaker can be treated as phonologically negligible. Setting the *RoPE* to capture 95% of tokens ($\pm 2\text{SDs}$) produced by the speaker with the strictest distribution in nasalance therefore establishes a conservative threshold below which variation can be interpreted as within-category phonetic noise, rather than a cross-category, intentional phonetic or phonological manipulation. Using this method, I estimated the *RoPE* for nasalance on contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowels in Punjabi to be $\pm .061$.

Figure 2.5 shows the proposed *RoPE* along with the nasalance data from vowels in the contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ condition from all speakers in Experiment 1. The yellow points on the plot represent the mean nasalance for contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowels for each speaker from Experiment 1, with the tails indicating $\pm$ two standard deviations (capturing 95% of productions) from that mean. Speakers are ordered from left to right by magnitude of their standard deviation in mean nasalance.

When a listener normalizes expected nasalance levels for a given speaker, any nasalance value within two standard deviations of that speaker’s typical nasalance level for a $\text{C}\tilde{\text{V}}\text{:}$ vowel can probably be safely interpreted as nasal. The magnitude of the suggested *RoPE* of $\pm.061$ shown in blue can be interpreted in a similar way.

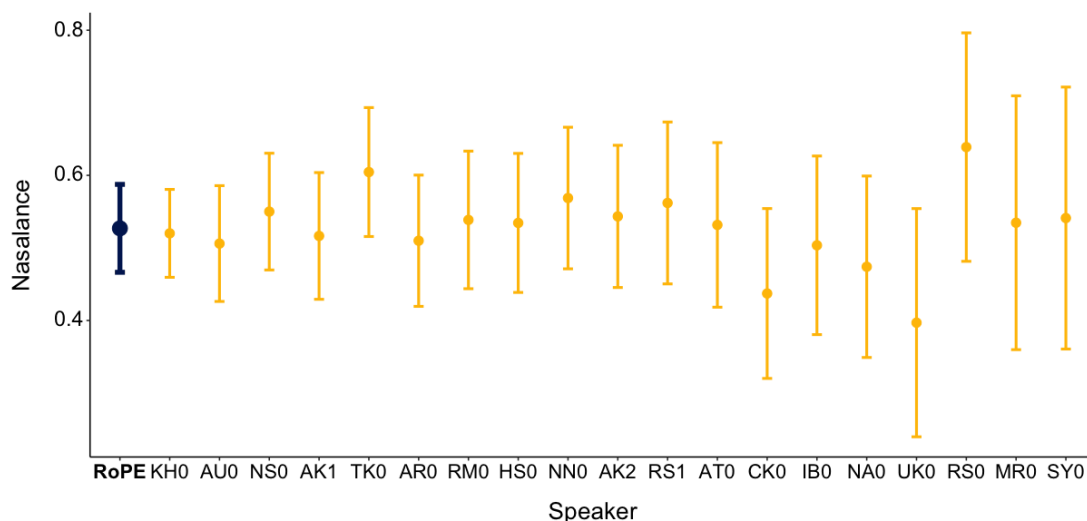


Figure 2.5: Dot and whisker plot showing mean nasalance of $\text{C}\tilde{\text{V}}\text{:}$ vowels from Experiment 1. The yellow dots and tails represent the mean $\text{C}\tilde{\text{V}}\text{:}$ vowel nasalance for each speaker and $\pm$ two standard deviations around this mean nasalance value respectively. Speakers are ordered from left to right by magnitude of standard deviation. The blue point shows the mean nasalance across all speakers, and its tails represent the proposed *RoPE* of $\pm.061$.

Reexamining the 95% CI for the estimated difference in nasalance between pre-N $\text{CV}\text{:N}$ and contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ in row 2 of Table 2.2, which compares nasalance in $\text{CV}\text{:N}$ vowels to nasalance in $\text{C}\tilde{\text{V}}\text{:}$ vowels, we see that the entire confidence interval ($[-.039, .007]$) falls well within the bounds of the proposed *RoPE* ($\pm.061$). Therefore, it seems appropriate to conclude that the mean nasalance of the two vowel types is practically equivalent. This allows us to assert that pre-N $\text{CV}\text{:N}$ vowels surface as phonologically $[+\text{NAS}]$ in the same way as contrastive nasal $\text{C}\tilde{\text{V}}\text{:}$ vowels in Punjabi.

To sum up, the results of Experiment 1 confirm and extend the findings from the acoustic analysis in Zahid and Hussain (2012): pre-N $[\tilde{V}:N]$ vowels are not meaningfully different from contrastive nasal $/\tilde{V}:/$ vowels with respect to their phonetic nasality.⁷

2.3 Experiment 2: Do pre-N vowels trigger nasal harmony?

Experiment 1 found that contrastive nasal $/\tilde{V}:/$ and non-alternating pre-N $[\tilde{V}:N]$ vowels are practically and statistically equivalent with respect to the realization of vowel nasalance. Experiment 2 explores how these two vowel types interact with the phonological process in Punjabi of leftward nasal spread. Specifically, Experiment 2 is designed to test if statistical differences exist between contrastive nasal $/\tilde{V}:/$ and non-alternating pre-N $[\tilde{V}:N]$ with respect to their ability to trigger nasal harmony in Punjabi.

The results of Experiment 1 help frame two potential hypotheses regarding the interaction of surface $[\tilde{V}:N]$ vowels with nasal harmony. The first hypothesis asserts that, because contrastive nasal $/\tilde{V}:/$ and non-alternating $[\tilde{V}:N]$ vowels appear to be identical in their phonetic nasality, they should *behave* identically in terms of triggering nasal harmony.

The abstract underlying representations provided by generative phonology also raise a second possibility. Vowel nasality in $[\tilde{V}:N]$ is predictable rather than con-

⁷The discussion here assumes that – because they are produced with statistically indistinct nasality – $C\tilde{V}:$ and $CV:N$ vowels are *perceptually* indistinguishable as well. Nevertheless, the production studies in this chapter cannot answer questions about perception, so future research is needed to determine whether or not contrastive nasal $C\tilde{V}:$ vowels are perceptually distinct from pre-N $CV:N$ vowels to listeners. See also Beddor and Krakow (1999).

trastive in Punjabi. As such, surface $[\tilde{V}:N]$ could be analyzed as underlyingly $/V:N/$, with an oral vowel; predictable nasality would then be supplied by a rule or other grammatical device. This analysis produces a distinction between contrastively nasal vowels $/\tilde{V}:/$ — which would be underlyingly nasal — and underlyingly oral vowels in $/V:N/$.

Thus, a second hypothesis is that contrastive nasal $/\tilde{V}:/$ and pre-N $[\tilde{V}:N]$ vowels might behave *differently* with respect to triggering harmony because their representational status is distinct. Specifically it could be that only underlying, contrastive nasal $/\tilde{V}:/$ triggers harmony, while underlying oral $/V:N/ \rightarrow [\tilde{V}:N]$ does not.

Importantly, as mentioned above, forms with contextually nasal vowels $/V:N/ \rightarrow [\tilde{V}:N]$ in these experiments are all non-alternating (e.g., $/ta:n/$ ‘melody’ and $/xu:n/$ ‘blood’). There is thus no surface evidence that these vowels are anything other than nasal. For this reason, hypothesis 2 is only statable in frameworks which either (i) allow for covert URs which are qualitatively distinct from any observed surface form(s), or (ii) invoke other mechanisms for distinguishing between contrastive vs. predictable vowel nasality.

In sections 2.4.2.1 and 2.4.2.2 I argue that these results are difficult to accommodate in surface-oriented models couched in either usage-based or generative frameworks.

2.3.1 Stimuli, design, and analysis

Sixteen additional native Punjabi speakers (eleven men and five women) who did not participate in Experiment 1 took part in Experiment 2. Ages ranged

from 18 to 43 ($\mu = 28.2$; median = 27). Similar to the first experiment, stimuli were sourced from three main vowel types: oral $/V:/$, contrastive nasal $/\tilde{V}:/$, and non-alternating pre-N $[\tilde{V}:N]$ vowels. Here, however, all words were either di- or trisyllabic, and the shape of the final two syllables was controlled, such that every word contained a vowel–glide–vowel sequence (VGV) across the last two syllables. This construction made it possible to examine whether the three vowel types differ in their interactions with nasal harmony.

Tokens from all three vowel conditions were produced with an enclitic CV postposition beginning with an oral stop (e.g. $/\alpha:v\alpha:m=d\partial/$ ‘general public=GEN’). In addition to these three conditions, words ending in a contrastive nasal $/\tilde{V}:/$ vowel followed by a nasal-initial NV postposition were included as a fourth condition. Bashir and Connors (2019, p.244) state that, in Punjabi, grammatical postpositions immediately follow the constituent they modify, phonologically integrating with it and lacking independent stress or prosodic constituency. By including a condition in which contrastive nasal $/\tilde{V}:/$ vowels are immediately followed by a postposition beginning with a nasal consonant, the contextual environment of $/\tilde{V}:/$ more closely mimics that of $[\tilde{V}:N]$ vowels. This facilitates a more direct comparison of nasalance patterns between the two vowel types (see section 2.3.4 for a detailed discussion).

Altogether, four distinct conditions with 6-10 tokens each were included in the experiment, as shown in Table 2.3. For the conditions with oral vowels and pre-N vowels, the tokens do not alternate between CV and NV postpositions. Accordingly, I omit explicit reference to the postposition in the condition labels, referring to them as **oral** [VGV] and **pre-N** [VG $\tilde{V}$ N] respectively.

Notably, the label for the **pre-N** [VG $\tilde{V}$ N] condition reflects a surface

form in which the pre-N vowel is nasalized, but nasal harmony does not extend to the preceding segments. This label anticipates the results presented in section 2.3.2.4, where I show that nasal harmony does not apply to pre-N forms.

By contrast, for the two contrastive nasal conditions — $[\tilde{V}\tilde{G}\tilde{V}=NV]$ and $[\tilde{V}\tilde{G}\tilde{V}=CV]$ — I retain the enclitic in the condition labels to distinguish between them. The labels of the two contrastive nasal conditions also encode nasal harmony, in anticipation of the results in sections 2.3.2.2 and 2.3.4.

Twenty-seven filler items were randomly dispersed throughout the token set during the experiment. Complete glosses of all tokens included in the experiment are provided in Figure 2.9 in the Appendix.

oral [VG̃V]		pre-N [VG̃VN]	
/sət-ɑːjiː də/	‘27’	/dʒəvɑːŋ də/	‘a youth’
/ɔːliː-jaː də/	‘pious men’	/ədʒəvɛːŋ də/	‘omum seed’
/həvɑː də/	‘air’	/siːjaːŋ də/	‘recognition’
/tʃɔː-viː də/	‘24’	/tɛːjaːŋ də/	‘attention’
/pɑːvɛː də/	‘cot leg’	/lɛː-jaːŋ dɛ/	‘taking’
/beːvɑː də/	‘widow’	/ɑːvɑːm də/	‘general public’
/dəvɑː də/	‘medicine’		
/tɑːvuː də/	‘uncle’		
/vəkʰɑːvɛː də/	‘boasting’		
/møkʰiːjaː də/	‘chief’		

contrastive nasal $[\tilde{V}\tilde{G}\tilde{V}=NV]$		contrastive nasal $[\tilde{V}\tilde{G}\tilde{V}=CV]$	
/tʃɑːvẽː nʊ/	‘pumicestone’	/tʃɑːvẽː də/	‘pumicestone’
/uː-vẽː nʊ/	‘in that way’	/uː-vẽː də/	‘in that way’
/dɔː-vãː nʊ/	‘both sides’	/dɔː-vãː paːse/	‘both sides’
/kiː-vẽː nʊ/	‘somehow’	/kiː-vẽː də/	‘somehow’
/ɛː-vẽː nʊ/	‘in vain’	/ɛː-vẽː də/	‘in vain’
/mətʃiː-jãː nɪ/	‘fish-PL’	/mətʃiː-jãː də/	‘fish-PL’
/tiːvĩː nɪ/	‘woman’	/tiːvĩː də/	‘woman’
/sɑː-vãː nɪ/	‘breaths’	/sɑː-vãː də/	‘breaths’

Table 2.3: Experiment 2 tokens.

Both the **oral** [VG $\tilde{V}$] and **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions provide established baselines for an oral vowel with no nasality in the signal and a nasal / $\tilde{V}$:/ vowel that triggers nasal harmony, respectively. The **pre-N** [VG $\tilde{V}$ N] condition, on the other hand, is the condition of interest. Importantly, [V:N] sequences from the **pre-N** [VG $\tilde{V}$ N] condition never straddle a morpheme boundary, so Punjabi speakers lack experience with alternations that reveal the oral status of these pre-N vowels when not followed by a nasal consonant. Thus, assuming similar results to Experiment 1, in which pre-N [V:N] vowels are phonetically indistinguishable from contrastive nasal / $\tilde{V}$:/ vowels in terms of nasality, [V:N] vowels in these tokens should always be realized as [+NAS].

The primary focus here is whether pre-N vowels trigger nasal harmony to the same extent as contrastive nasal / $\tilde{V}$:/ vowels. Section 2.3.2.1 discusses the apparent optionality of nasal harmony in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition. Section 2.3.2.2 analyzes tokens with harmony in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition. Section 2.3.2.4 shows that pre-N [V:N] vowels do not trigger nasal harmony, despite being as strongly nasalized as contrastive nasal / $\tilde{V}$:/ vowels (section 2.3.2.3).

2.3.2 Results

A total of 2,048 potential tokens (16 speakers x 4 repetitions x 32 words) resulted from the recordings. Unfortunately, the recording for one speaker (MI0) showed little to no differences between the **oral** [VG $\tilde{V}$] and **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions, so the data from this participant was removed. In addition, 53 more tokens across the remaining fifteen speakers were removed due to poor audio quality. After removing these words, a total of 1,867 tokens remained in

the analysis.

All speakers in Experiment 2 produced tokens from the **oral** [VG_V] and the **contrastive nasal** [ṼG̃Ṽ=CV] and [ṼG̃Ṽ=NV] conditions uniformly, but productions of the **pre-N** [VG̃ṼN] condition fell into two distinct patterns that require individual attention. All fifteen speakers lacked nasal harmony across the preceding V:G sequence in the **pre-N** [VG̃ṼN] condition (section 2.3.2.4). Thirteen speakers categorically nasalized the pre-N [Ṽ:N] vowel immediately preceding the nasal consonant (section 2.3.2.3). However, the remaining two speakers systematically lacked categorical nasality on all pre-N [Ṽ:N] vowels, instead demonstrating a pattern of anticipatory coarticulation. I discuss these two speakers in the Appendix, where I speculate that their productions arise from dialectal differences; their data is excluded from the analysis in the main text.

2.3.2.1 Optionality of Nasal Harmony

Before examining the results in more detail, one other fact about nasal harmony in the **contrastive nasal** [ṼG̃Ṽ=CV] and [ṼG̃Ṽ=NV] conditions must be considered. Across all speakers and words in the experiment, it was evident that nasal harmony on the penultimate, pre-glide stem vowel only sometimes applied: for 44% of tokens in the contrastive nasal conditions (n=399), nasal harmony was not triggered. These 399 tokens are broadly distributed across speakers and words rather than concentrated in a small subset of the data. As I am unaware of a grammatical description that reports optionality or variability for nasal harmony in Punjabi, this result was unexpected.

Figure 2.6 illustrates with two productions of /u:uẽ: də/ ‘in that way’ by speaker AA1, uttered within one second of each other. Examining the nasal channels

in the figure for both productions, nasalance on the penultimate, pre-glide stem vowel exhibits categorical nasalization across the entire [VG̃V] sequence for recording #159 (top panel). For recording #160 (bottom panel), on the other hand, nasalance remains at the level of an oral vowel for the entirety of the pre-glide vowel’s duration and then rises rapidly at the onset of the glide.

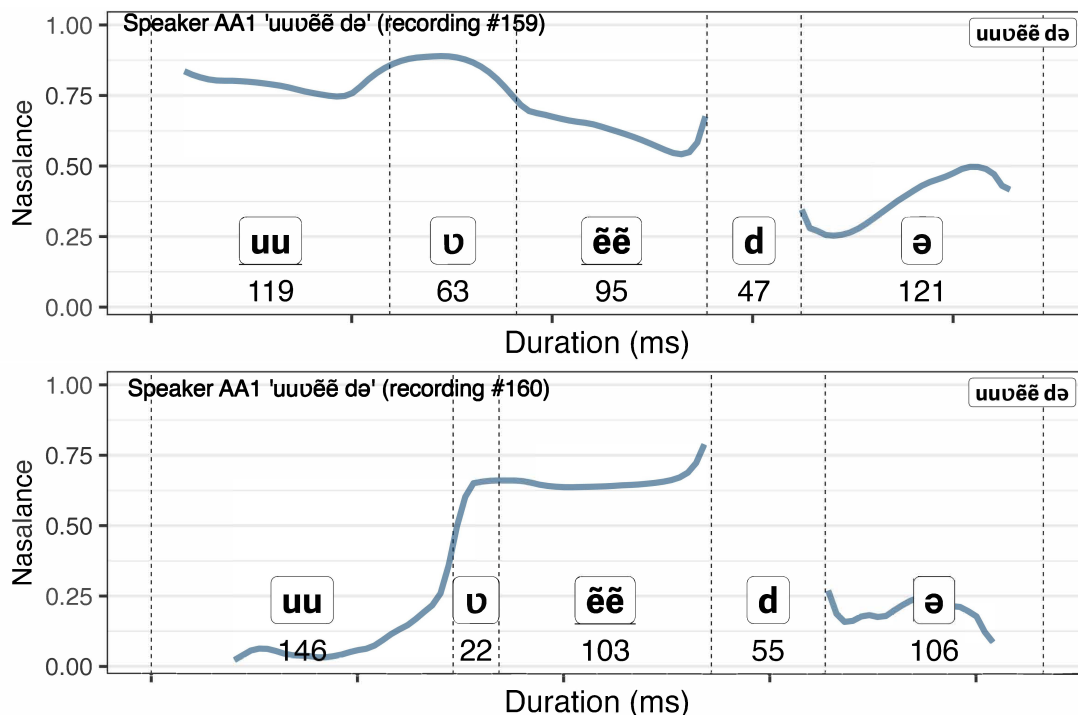


Figure 2.6: Nasalance traces for two different productions of /u:ṽe: də/ ‘in that way’ by the same speaker in quick succession that differ only in the application of nasal harmony. The y-axis corresponds to nasalance, calculated in Praat by taking the normalized proportion of nasal air pressure to the total amount of air pressure in the vocal tract.

Importantly, this variable application of the nasal harmony process was present for all speakers and words in both the **contrastive nasal** [ṼG̃Ṽ=CV] and [ṼG̃Ṽ=NV] conditions. Moreover, harmony was virtually absent in the **pre-N** [VG̃ṼN] condition, as I discuss below.⁸ These observations indicate that nasal harmony in

⁸Across the dataset, a small number of tokens in the **pre-N** [VG̃ṼN] condition showed un-

Punjabi is only optionally triggered by contrastive nasal / $\tilde{V}$:/ vowels.

Because Experiment 2 aims to examine whether pre-N [$\tilde{V}$:N] vowels differ from contrastive nasal / $\tilde{V}$:/ vowels in their ability to trigger nasal harmony in the language, all tokens that clearly lacked harmony in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] and [$\tilde{V}\tilde{G}\tilde{V}$ =NV] conditions were removed to prevent the obfuscation of the harmony pattern. This was accomplished for each speaker by removing tokens from these conditions in which the penultimate (pre-glide) stem vowel exhibited a mean nasalance clustering with the mean nasalance of penultimate pre-glide stem vowels in the **oral** [VG $\tilde{V}$] condition for that speaker. For example, Figure 2.7 shows a density plot depicting the concentration of mean nasalance for penultimate pre-glide stem vowels in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] and [$\tilde{V}\tilde{G}\tilde{V}$ =NV] conditions for speaker AW0. As is evident, the distribution is bimodal, with the large majority of penult pre-glide stem vowels exhibiting a mean nasalance of approximately .51-.53. Conversely, another smaller concentration of penultimate vowels cluster around a mean nasalance of roughly .25. The vertical line shows the approximate cut in which tokens with mean nasalance to the left of the line were removed and tokens with mean nasalance to the right of the line retained. No tokens from the **pre-N** [VG $\tilde{V}$ N] condition were removed in this process. Thus, the comparison between contrastive nasal / $\tilde{V}$:/ vowels and pre-N [$\tilde{V}$:N] vowels in this study includes only those tokens in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] and [$\tilde{V}\tilde{G}\tilde{V}$ =NV] conditions in which the nasal harmony process *does* apply. That is, tokens like recording #160 (bottom panel) in Figure 2.6 above were removed.

usually high nasalance across the full [V:G $\tilde{V}$:N] sequence. Similarly, a few utterances in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition displayed a final stem vowel – an underlying, contrastive nasal vowel – that nonetheless exhibited nasalance levels characteristic of an oral vowel. Given how few tokens show these patterns, I interpret them as performance artifacts rather than the intentional implementation of a particular nasalance pattern as such. None of these tokens were removed from the dataset during analysis.

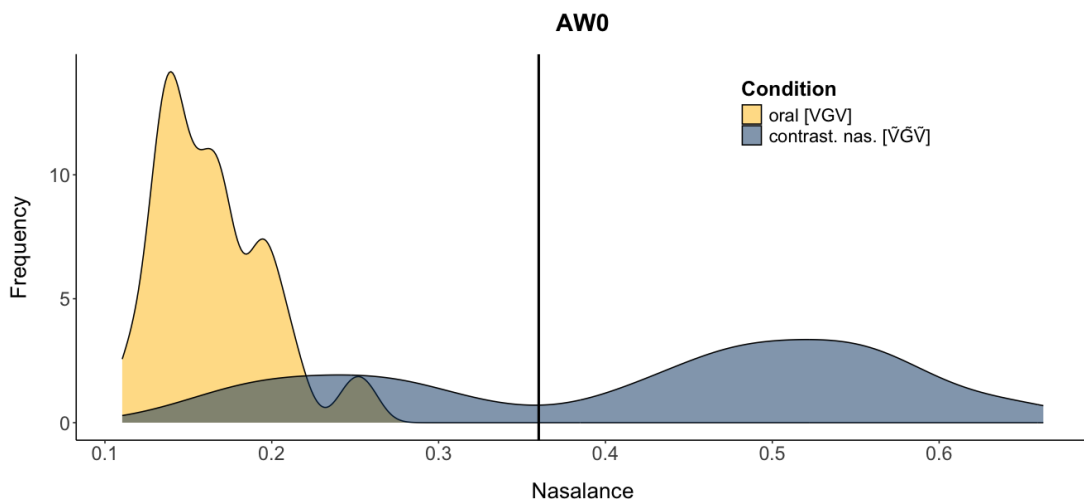


Figure 2.7: Density plot showing the distribution of mean nasalance for penultimate pre-glide stem vowels for speaker AW0. **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}]$ includes penultimate pre-glide stem vowels from both the $[\tilde{V}\tilde{G}\tilde{V}=CV]$ and $[\tilde{V}\tilde{G}\tilde{V}=NV]$ conditions. The plot shows a bimodal distribution for penultimate stem vowels preceding contrastive nasal $/\tilde{V}:/$, with two distinct peaks around .25 and .52. The vertical line marks the cutoff point for speaker AW0, in which tokens with a mean nasalance to the left of the line were removed from the analysis. The multimodal distribution in the oral condition reflects inherent nasalance differences across vowel qualities (Bell-Berti, 1993).

2.3.2.2 Contrastive nasal vowels and nasal harmony

Figure 2.8 displays GAM curves fit to nasalance at each of the 33 timesteps across the [VGv] sequence for the **oral** [VGv] and **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ conditions, for all fifteen speakers.⁹ As expected, the **oral** [VGv] condition exhibits a relatively stable nasalance at about 20% across the entire [VGv] sequence. Tokens from the **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ condition, on the other hand, show an increase in nasalance from the onset of the penultimate, pre-glide stem vowel until

⁹Because nasalance measurements were taken at 11 equidistant timesteps for each segment, Figure 2.8 and other similar figures reporting nasality across [VGv] sequences involve some distortion in time. For instance, glides were typically much shorter than either of the vowels in any given token, and so the time steps during [G] in Figure 2.8 are not strictly comparable to the time steps for the adjacent vowels [V].

a plateau of nasalance is reached about a third of the way through the penultimate pre-glide stem vowel, and maintained through the end of the word. This pattern is qualitatively consistent with a process of nasal harmony triggered by the final / $\tilde{V}$:/ vowel and targeting both the glide and the penultimate pre-glide stem vowel. The initial slope in nasalance across the first third of the penultimate pre-glide stem vowel is likely due to the transition out of an oral onset consonant. In sum, the GAM curves highlight a clear and relatively stable distinction in nasalance between the oral [VG $\tilde{V}$] and contrastive nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions across the length of the [VG $\tilde{V}$] sequence.

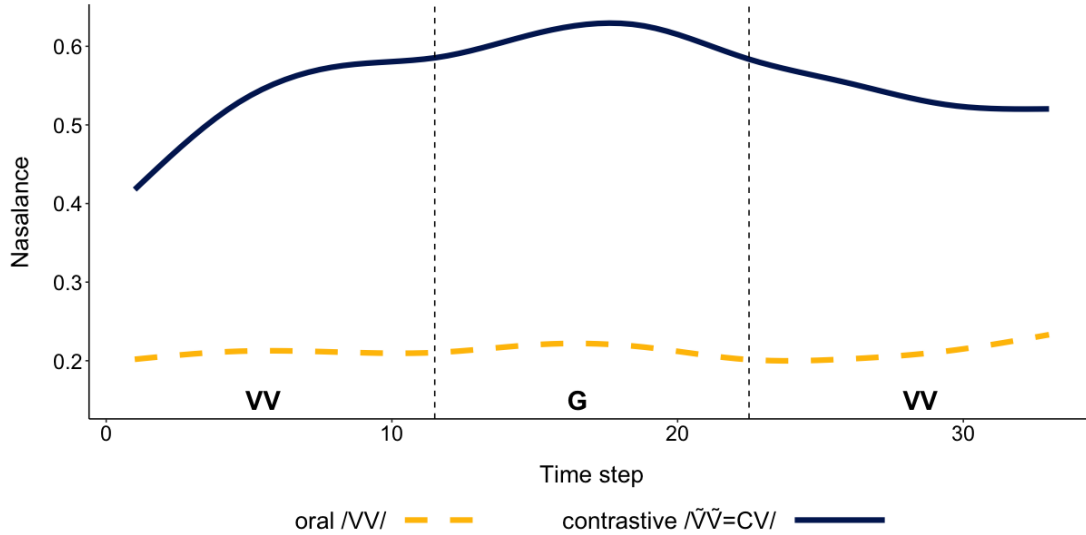


Figure 2.8: GAM curves for the oral [VG $\tilde{V}$] and contrastive nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions in Experiment 2. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG $\tilde{V}$] sequence for all 15 speakers in the experiment. Vertical dashed lines represent segment boundaries.

2.3.2.3 Nasality of pre-N vowels

The crucial condition in this experiment is **pre-N** [VG $\tilde{V}$ N]. The GAM curves in Figure 2.9 show that nasalance on the penultimate, pre-glide stem vowel does not diverge from the oral [VG $\tilde{V}$] condition with respect to nasalance until the

latter portion of the pre-glide stem vowel, at which point a cline-like increase in nasalance begins and continues through the onset of the final pre-N vowel. Strikingly, nasalance on the final vowel itself in the **pre-N** [VG $\tilde{V}$ N] condition resembles the level of nasalance found on the final vowel in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition through most of its duration. A faceted visualization of these results is provided in Figure 2.5 of the Appendix, demonstrating the stability of this pattern across the thirteen participants.

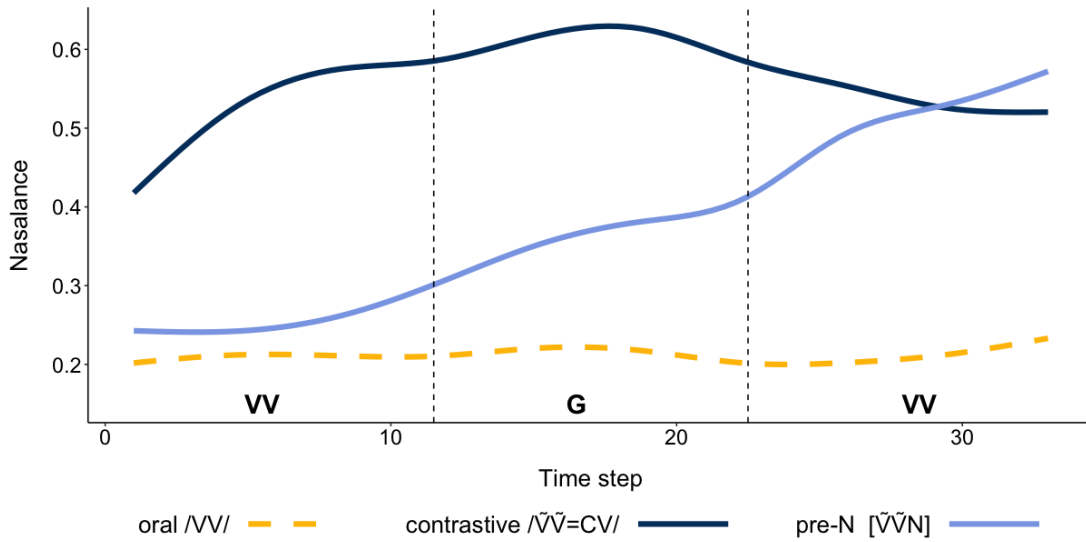


Figure 2.9: GAM curves for the oral [VG $\tilde{V}$], contrastive nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV], and **pre-N** [VG $\tilde{V}$ N] conditions in Experiment 2. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG $\tilde{V}$] sequence for the 13 speakers in the experiment who categorically nasalized the stem-final vowel in the **pre-N** [VG $\tilde{V}$ N] condition. Dashed vertical lines represent segment boundaries.

To confirm that pre-N [$\tilde{V}$:N] vowels are not different in nasalance to contrastive nasal vowels, I ran a linear mixed effects model (table 2.4) comparing nasalance on the final stem vowel in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] and **pre-N** [VG $\tilde{V}$ N] conditions. **CONDITION** (2 levels) was included as a dummy-coded main effect, and – due to the visual change in nasalance over time for both target vowels in

Figure 2.9 – TIMESTEP (3 levels) and its interaction with CONDITION were also included as fixed effects. As before, TIMESTEP was effect-coded, and its three levels roughly correspond to the first third, middle third, and final third of the final vowel in the [VGV] sequence. In addition, including an effect-coded main effect of VOWELQUALITY (4 levels) resulted in a significantly better fit to the data ($\chi^2(3) = 15.17, p = .002^{**}$), so it was retained in the model. By-word random intercepts were included, along with by-speaker random intercepts and random slopes for condition. All levels of the effect-coded main effects are included in the model output displayed in Table 2.4 for clarity.

With a closer inspection of the final three rows in Table 2.4, it is clear that the interaction between CONDITION and TIMESTEP is significant, with the nasalance of the final stem vowel in **pre-N** [VG $\tilde{V}$ N] diverging significantly from the final **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] vowel at each portion of the vowel. In the first third of the vowel, the final stem vowel in **pre-N** [VG $\tilde{V}$ N] was less nasalized, but in the final two thirds it was significantly more nasalized compared to that of the final stem vowel in **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV]. Given these significant interaction effects and their nontrivial effect sizes, comparisons between the two vowel types must be analyzed at each time step level, rather than examining the main effect of vowel type as a whole.

Predictor	β	95% CI	t	p
Intercept ([ṼG̃Ṽ=CV])	.539	[.485, .593]	19.72	< .001 ***
[VG̃VN]	.019	[-.031, .069]	.75	.46
Beg	.021	[.018, .025]	12.53	< .001 ***
Mid	-.006	[-.009, -.002]	-3.21	.001 **
Fin	-.016	[-.02, -.012]	-8.26	< .001 ***
ii	.132	[.072, .191]	4.33	.002 **
ee	-.007	[-.048, .035]	-.32	.76
εε	-.047	[-.111, .018]	-1.41	.19
ɑ:	-.044	[-.077, -.012]	-2.67	.02 *
[VG̃VN]:Beg	-.075	[-.079, -.07]	-31.97	< .001 ***
[VG̃VN]:Mid	.017	[.013, .022]	7.38	< .001 ***
[VG̃VN]:Fin	.058	[.052, .063]	22.03	< .001 ***

Table 2.4: Results of the LMER model – `lmer(nasalance~CONDITION*TIMESTEP + VOWELQUALITY + (1 + CONDITION|speaker) + (1|word))` – estimating nasalance across the final vowel for 13 speakers who categorically nasalize pre-N [Ṽ:N] vowels. `CONDITION` was dummy coded, and effect-coded fixed effects include `TIMESTEP` and `VOWELQUALITY`.

In an effort to compare nasalance values between the two conditions more closely, I established a new Region of Practical Equivalence (*RoPE*) for contrastive nasal vowels in the **contrastive nasal** [ṼG̃Ṽ=CV] condition in the same way as described for Experiment 1, arriving at a value of $\pm .051$. For the beginning third of the vowel, the lower bound of the 95% CI in Table 2.4 extends well below $-.051$, suggesting that pre-N [Ṽ:N] vowels are significantly less nasalized than final stem vowels in the **contrastive nasal** [ṼG̃Ṽ=CV] condition at this time juncture in the vowel. As I argue in the next section, this difference in nasalance is likely a consequence of the fact that nasal harmony applies in contrastive /V:G̃Ṽ:/ $\rightarrow$ [Ṽ:G̃Ṽ:], but not in pre-N /V:GṼ:N/ $\rightarrow$ [Ṽ:GṼ:N]. If the glide is fully nasalized in [Ṽ:G̃Ṽ:] but not in [Ṽ:GṼ:N], then observance of an oral-to-nasal transition between the beginning of the final vowel and the preceding glide in [Ṽ:GṼ:N] is not unexpected.

Moreover, while the 95% CI for the middle third of final stem vowels in **pre-N** [VG̃VN] in Table 2.4 falls within the *RoPE*, the 95% CI for the final third of the vowel falls entirely *above* the *RoPE*. This means that final stem vowels in **pre-N** [VG̃VN] are significantly more nasalized than the final stem vowels in **contrastive nasal** [ṼG̃Ṽ=CV] in this region. This nasality difference is consistent with the influence of the following context: **pre-N** [Ṽ:N] vowels precede a nasal consonant, which maximizes nasalance, whereas the final stem vowels in **contrastive nasal** [ṼG̃Ṽ=CV] precede an oral stop, which typically exhibits minimal nasalance.

Taking the effects of context into account, it seems reasonable to conclude that the final stem vowels in the **contrastive nasal** [ṼG̃Ṽ=CV] and **pre-N** [VG̃VN] conditions are both targeted as [+NAS], not only in monosyllabic tokens (experiments 1) but in longer words as well. Apparent differences between them arise from different patterns of local, nasal coarticulation on the preceding glide and following consonant.

2.3.2.4 Nasal Harmony

Having confirmed that the final stem vowels in non-alternating **pre-N** [VG̃VN] and **contrastive nasal** [ṼG̃Ṽ=CV] are both targeted as [+NAS] in the tokens used in Experiment 2, it is now possible to quantitatively examine whether the two vowel types differ in how they trigger nasal harmony.

From a cursory examination of the GAM curves in Figure 2.9 above, it seems clear that the two conditions interact in a qualitatively distinct manner with the nasal harmony process. To corroborate these facts quantitatively, I ran two separate models to examine nasalance on the glides and the penultimate pre-glide stem vowels in the [VG̃V] sequence, respectively.

The results of the model comparing the glides in the **contrastive nasal** [ṼG̃Ṽ=CV] and **pre-N** [VG̃VN] conditions are given in Table 2.5. The data in this model is limited to the thirteen speakers who categorically nasalize final stem vowels in the **pre-N** [VG̃VN] condition; see the Appendix for discussion of the remaining two speakers.

The model's intercept is the stem-final vowel in the **contrastive nasal** [ṼG̃Ṽ=CV] condition. Contrastive nasal /Ṽ:/ vowels were included in the model in order to compare the glides in the two conditions to a segment that is clearly categorically [+NAS]. Doing so provides a clearer picture as to whether final stem vowels in **pre-N** [VG̃VN] are triggering nasal harmony.

The glide in the **contrastive nasal** [ṼG̃Ṽ=CV] condition is significantly more nasalized than the contrastive nasal /Ṽ:/ vowel that follows it (as is evident in Figure 2.9). The increased nasalance on the glide in **contrastive nasal** [ṼG̃Ṽ=CV] relative to the final stem nasal vowels is likely due to the fact that glides have narrower constrictions in the oral cavity than vowels. All else being equal, narrowed oral constrictions will force a greater proportion of overall airflow through the nasal cavity (e.g. Bell-Berti 1993, Delvaux et al. 2008, p.595). In any case, these results provide clear support for the claim that the glide in /V:G̃Ṽ:=CV/ → [Ṽ:.G̃Ṽ:=CV] strings is targeted for nasal harmony, triggered by the contrastive nasal /Ṽ:/ vowel in the stem-final syllable.

For glides in the **pre-N** [VG̃VN] condition, on the other hand, mean nasalance falls significantly below the mean nasalance of contrastive nasal /Ṽ:/ vowels in the **contrastive nasal** [ṼG̃Ṽ=CV] condition (by 17.1%). The 95% confidence interval for the difference falls entirely outside (and below) the proposed *RoPE* (± 0.051), confirming they are less nasalized than final stem vowels in **contrastive**

nasal [ŃĜŃ=CV]. I therefore conclude that, unlike glides preceding contrastive nasal /Ń:/ vowels, glides preceding a non-alternating pre-N [Ń:N] vowel are not targets of nasal harmony.

Predictor	β	95% CI	t	p
Intercept (Final /Ń:/ in [ŃĜŃ=CV])	.535	[.477, .593]	18.20	< .001 ***
glide (in [ŃĜŃ=CV])	.065	[.047, .082]	7.23	< .001 ***
glide (in pre-N [VGŃN])	-.171	[-.232, -.109]	-5.42	< .001 ***

Table 2.5: Results of the LMER model estimating mean nasalance of the glide in the [VGŃ] sequence for 13 speakers. CONDITION was included as a dummy-coded fixed effect. The final /Ń:/ in the **contrastive nasal** [ŃĜŃ=CV] condition serves as the intercept in the model, and the estimates for the other predictors describe the predicted deviation from the intercept: `lmer(nasalance~CONDITION + (1 + CONDITION|speaker) + (1|word))`.

I now turn to a model comparing nasalance of the penultimate, pre-glide stem vowels in the **pre-N** [VGŃN], **oral** [VGŃ], and **contrastive nasal** [ŃĜŃ=CV] conditions. The results of this model are summarized in Table 2.6. Here, penultimate pre-glide oral vowels in the **oral** [VGŃ] condition serve as the intercept: the goal is to determine whether penultimate pre-glide stem vowels in the **pre-N** [VGŃN] condition significantly differ from those in the **oral** [VGŃ] condition.

The model coefficients indicate that the penultimate pre-glide stem vowel in the **contrastive nasal** [ŃĜŃ=CV] condition exhibits about 38% more nasalance than in the **oral** [VGŃ] baseline. In the **pre-N** [VGŃN] condition, the penultimate vowel does not significantly differ from the oral baseline vowel. However, I computed a *RoPE* of $\pm .033$ for penultimate oral vowels, and the upper bound of the 95% CI for the oral-[Ń:N] difference extends outside this region. Thus, while a distinction in nasalization between penultimate pre-glide stem vowels in the **pre-N** [VGŃN] and **oral** [VGŃ] conditions cannot be ruled out, the estimated difference

in nasalization from the **oral** [VG $\tilde{V}$] condition is about 31% smaller than the difference between the oral baseline and the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition.

Predictor	β	95% CI	t	p
Intercept (penult oral [VG $\tilde{V}$])	.188	[.122, .253]	5.64	< .001 ***
penult V: (pre-N [VG $\tilde{V}$ N])	.071	[-.001, .144]	1.94	.07
penult V: (nasal [$\tilde{V}\tilde{G}\tilde{V}$ =CV])	.378	[.307, .450]	10.38	< .001 ***

Table 2.6: Results of the LMER model estimating mean nasalance of the penultimate, pre-glide stem vowel in the [VG $\tilde{V}$] sequence for 13 speakers who categorically nasalize pre-N [$\tilde{V}$:N] vowels. **CONDITION** (penultimate stem vowel (**oral** [VG $\tilde{V}$], **pre-N** [VG $\tilde{V}$ N], **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV]) was included as a dummy-coded fixed effect. The intercept is the penultimate stem vowel in the **oral** [VG $\tilde{V}$] condition: $\text{lmer}(\text{nasalance} \sim \text{CONDITION} + (1 + \text{CONDITION}|\text{speaker}) + (1|\text{word}))$.

2.3.2.5 Summary of results for Experiment 2

To summarize, the stem-final vowels in the **pre-N** [VG $\tilde{V}$ N] and **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions are both targeted as [+NAS], replicating results from Experiment 1. Despite the similar surface nasality on the final stem vowels in **pre-N** [VG $\tilde{V}$ N] and **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] conditions, pre-N [$\tilde{V}$:N] and contrastive nasal / $\tilde{V}$:/ vowels triggered distinct patterns of nasalization across the preceding [VG] sequence. The segments preceding the contrastive nasal / $\tilde{V}$:/ vowel were significantly nasalized across the entire span, to a similar degree as the contrastive nasal / $\tilde{V}$:/ vowel itself. This confirms previous impressionistic descriptions that contrastive nasal / $\tilde{V}$:/ vowels in Punjabi trigger leftward nasal harmony across [VG] sequences, affecting both preceding vowels and glides (albeit optionally).

Conversely, [VG] sequences in the **pre-N** [VG $\tilde{V}$ N] condition exhibited a different pattern of nasality. The immediately preceding glide was significantly less nasal-

ized than in the **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ condition; in contrast, glides preceding the final contrastive nasal $/\tilde{V}:/$ in **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ exhibited similar nasalance to contrastive nasal $/\tilde{V}:/$. This is consistent with the claim that glides are nasalized by harmony when they precede contrastive nasal $/\tilde{V}:/$, but not when they precede pre-N $[\tilde{V}:N]$.

For the **pre-N** $[VG\tilde{V}N]$ condition, nasalance is low on the glide, and apparently even lower on the penultimate pre-glide stem vowel (figure 2.9). In other words, nasality steadily declines in strength for segments which are further away from the $[\tilde{V}:N]$ sequence, which I analyze as the only phonologically nasalized segments in the $[V:G\tilde{V}:N]$ string (setting aside the nasal consonant).

Nasalance on the penultimate pre-glide stem vowel in **pre-N** $[VG\tilde{V}N]$ was only slightly elevated compared to the nasalance on the penultimate pre-glide stem vowel in the **oral** $[VG\tilde{V}]$ condition. Thus, the final stem vowel in **pre-N** $[VG\tilde{V}N]$ is itself phonologically $[+NAS]$, but does not trigger nasal harmony across the preceding glide and penultimate, pre-glide stem vowel. Any vestigial nasality on the glide and pre-glide vowel appears to reflect lower-level, local coarticulation for nasality rather than nasal harmony as such.

In the following two sections, I evaluate alternative analyses of the nasality patterns observed in Experiment 2. These analyses attribute the distinct nasality patterns in the **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ and **pre-N** $[VG\tilde{V}N]$ conditions, not as the differential application of nasal harmony, but instead to either stress differences between the two conditions (section 2.3.3) or a fixed nasality window spanning different portions of the $[VG\tilde{V}]$ sequence in each condition (section 2.3.4). Ultimately, I show that neither account is plausible.

2.3.3 Nasalization Reflects Harmony, not Coarticulation

A potential confound in my interpretation of nasalance values in the **contrastive nasal** [ṼG̃Ṽ=CV] vs. **pre-N** [VG̃VN] conditions comes from stress placement. Stress has been described as weight sensitive in Punjabi, though in somewhat inconsistent terms (Bashir and Connors 2019; Bhatia 1993; Dhillon 2007; Hussain et al. 2019; Shackle 2003). Both syllables in the stem of the **contrastive nasal** [ṼG̃Ṽ=CV] condition are bimoraic, whereas, for **pre-N** [VG̃VN], the penultimate syllable in the stem is bimoraic, while the final syllable of the stem is at least plausibly trimoraic. These differences in syllable weight could affect stress placement across conditions. If so, based on some prior descriptions of weight-sensitive stress in Punjabi (e.g. Bashir and Connors 2019), we might expect penultimate stress on [Ṽ:G̃Ṽ:], and final stress on [V:G̃Ṽ:N].

Stress has also been found to correlate with increased nasality in some contexts and languages (e.g. Avelino et al. 2020; Krakow 1989; Zellou 2022 and references therein). This at least raises the possibility that the increased nasalance on the glide and penultimate pre-glide vowel in the **contrastive nasal** [ṼG̃Ṽ=CV] condition may not represent nasal harmony, but could instead result from long-distance coarticulation for nasality between the stressed penultimate syllable and the final contrastive nasal vowel.¹⁰

Since increased vowel duration is found to be a correlate of stress in many of the world’s languages (van Heuven and Turk, 2021), we can try to assess this claim by comparing the mean duration of the penultimate and final stem vow-

¹⁰To my knowledge, reported effects of stress on nasal coarticulation involve local interactions between [NV] and [VN] sequences, rather than the kind of longer-distance effects I am entertaining here. I nonetheless consider this scenario in the interest of ruling it out as a possible counter-explanation, particularly since nasal coarticulation is known to occur over a wide timespan in some languages (e.g. Pouplier et al. 2024a).

els across the two conditions of most interest, **contrastive nasal** [ṼḠṼ=CV] vs. **pre-N** [VGṼN]. The penultimate and final stem vowels in the **contrastive nasal** [ṼḠṼ=CV] condition were 106 ms and 103 ms in mean duration, respectively, compared to a mean duration of 65 ms for the penultimate vowel and 151 ms for the final vowel in the **pre-N** [VGṼN] condition. These measurements are at least consistent with the claim that **contrastive nasal** [ṼḠṼ=CV] and **pre-N** [VGṼN] differ with respect to stress placement, stress being penultimate in the former case, and final in the latter. However, Bhatia (1993, p.343) offers an impressionistic description that vowel duration is unaffected by stress placement, which complicates a straightforward interpretation of these results.

Regardless, even if stress is responsible for vowel length differences between the two conditions, this does not entail that we should see the drastic nasality differences reported here. In at least some languages where stress conditions greater nasal coarticulation, there is a positive correlation between the duration of the stressed vowel and the degree of nasalance on that vowel (e.g. discussion in Zellou 2022, p.21, 35). However, no such correlation, positive or negative, is present in the data presented here. Considering all forms in the **contrastive nasal** [ṼḠṼ=CV] and [ṼḠṼ=NV] conditions (including those that do not exhibit categorical nasal harmony), a Pearson's correlation analysis ($r(723) = -.05$, $p = .22$) between penultimate stem vowel duration and mean nasalance on the penultimate stem vowel did not detect a significant effect (Figure 2.10), and the magnitude of the correlation coefficient is quite small.

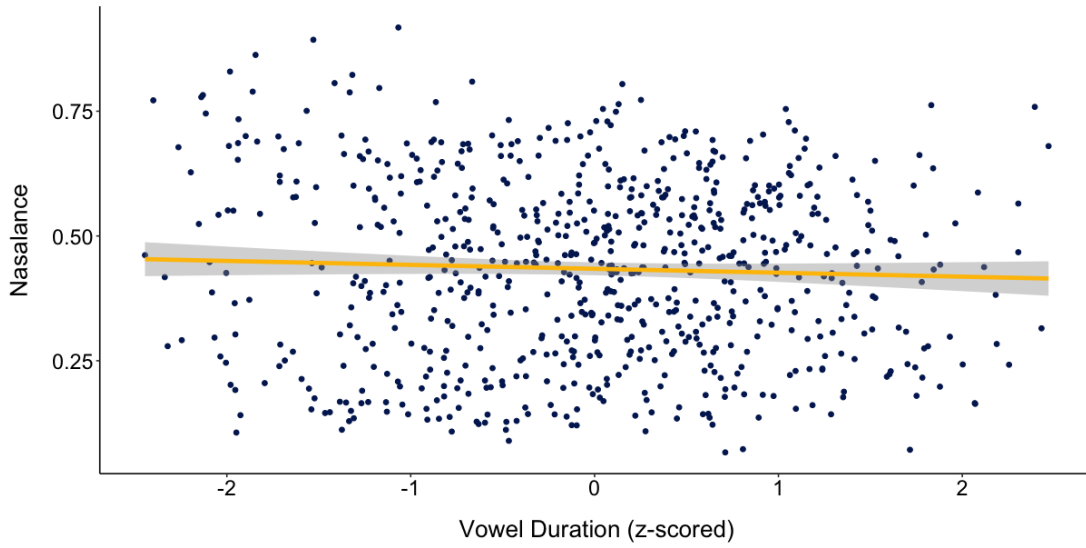


Figure 2.10: Scatterplot showing the estimated $-.05$ Pearson Correlation value between penultimate vowel duration (using by-speaker z-scored values) and mean nasalance in the **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ and $[\tilde{V}\tilde{G}\tilde{V}=NV]$ conditions.

If the nasal harmony pattern observed in the two contrastive nasal conditions reflected long-distance coarticulation conditioned by stress, we would expect at least *some* association between nasalance and vowel duration, be it positive or negative. The lack of such a correlation argues against treating nasal harmony as a phonetic process of nasal coarticulation, rather than a phonological process of nasal spreading. (I return to this point in section 2.3.4.)

In sum, the incongruent descriptions of Punjabi stress in prior literature, the uncertainty about the phonetic correlates of stress in the language, and the finding that nasalance on the penultimate pre-glide vowel in **contrastive nasal** $[\tilde{V}\tilde{G}\tilde{V}=CV]$ and $[\tilde{V}\tilde{G}\tilde{V}=NV]$ conditions is insensitive to duration all argue against re-analyzing nasal harmony as a stress-sensitive process of long-distance nasal coarticulation. While further work on the phonetics of stress in Punjabi would of course be welcome, at present I see no reason to treat nasality on the penults of

contrastive nasal / $\tilde{V}$:/ as anything other than the result of a categorical process of phonological nasal harmony (see also section 2.3.4).

2.3.4 Harmony does not Reflect a Fixed Nasalization Window

I have argued that nasality spreads leftward to the penultimate pre-glide stem vowel in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition in Punjabi, but not in **pre-N** [VG $\tilde{V}$ N]. I have further argued that spreading is best construed as a phonological harmony process which propagates the abstract feature [+NAS] leftward from contrastively nasal vowels in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition, but not from predictable, contextually nasal vowels in the **pre-N** [VG $\tilde{V}$ N] condition.

An alternative analysis of this data might be possible in which nasalization in **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] owes to a coarticulatory window in which the velum is lowered for a more or less fixed period of time. Assume that in both [$\tilde{V}$:. $\tilde{G}\tilde{V}$:] and [V:. $\tilde{G}\tilde{V}$:N] there is a *fixed period* of nasality anchored at the right edge of the word (see e.g. Beddor 2007). If so, we might expect nasality to begin later in [... V:. $\tilde{G}\tilde{V}$:N] than in [... $\tilde{V}$:. $\tilde{G}\tilde{V}$:], because the additional, word-final nasal segment in [... V:. $\tilde{G}\tilde{V}$:N] would cause the overall ‘window’ of nasality to shift rightward. This difference in the timing of phonetic nasality could give the impression of a difference in the application of harmony, without implicating a harmony process as such.

There are at least three arguments against this analysis. The first comes from the optionality of nasal harmony in Punjabi (section 2.3.2). Leftward nasal spreading in **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] often yields nasalized [$\tilde{V}$:. $\tilde{G}\tilde{V}$:], but outcomes

in which harmony fails to apply [V:G $\tilde{V}$:] are also frequent across all items and speakers. If there is a fixed period of velar lowering in all words that end in a nasal segment, then the non-application of nasal harmony in **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] is quite puzzling. On the other hand, if nasal harmony is an abstract phonological process, then categorical variation between nasalized [$\tilde{V}$: $\tilde{G}\tilde{V}$:] and non-nasalized [V:G $\tilde{V}$:] can be straightforwardly analyzed as optional rule application (see also Plug et al. 2019).

Second, in section 2.3.3 we observed that there is no correlation between the duration of the penultimate pre-glide vowel in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] and [$\tilde{V}\tilde{G}\tilde{V}$ =NV] conditions and the degree of nasalance on that vowel. This result is again surprising if harmony owes to a coarticulatory window of approximately fixed duration: in that scenario, we should see a decrease in nasalance as the pre-glide vowel grows longer, and more distant from the phonetic source of nasality at the right edge of the word in [$\tilde{V}$: $\tilde{G}\tilde{V}$:] (e.g. Solé 1992, 1995).

Finally, a third argument against this approach comes from the analysis of nasalization in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =NV] condition. If contrastive nasal / $\tilde{V}$:/ vowels are responsible for triggering nasal harmony — understood as a phonological process — the presence of a following /=NV/ postposition should not affect the application of harmony in any way in a [VG $\tilde{V}$] sequence. Conversely, if the consistent nasalance in the **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =CV] condition is due to the anchoring of a fixed window of nasalization to the right edge of the rightmost nasal segment, we would expect this window to shift to the right to align with the initial nasal consonant of the postposition. In this scenario, nasalization in **contrastive nasal** [$\tilde{V}\tilde{G}\tilde{V}$ =NV] should mirror the nasalance pattern of the **pre-N** [VG $\tilde{V}$ N] condition: there should be no evidence of harmony.

The GAM curve for all fifteen speakers for the **contrastive nasal** [̃Ǿ̃Ǿ̃=NV] condition is provided in Figure 2.11, alongside the baseline **oral** [VG̃V] and **contrastive nasal** [̃Ǿ̃Ǿ̃=CV] conditions for comparison.

As is evident from Figure 2.11, the **contrastive nasal** [̃Ǿ̃Ǿ̃=NV] condition shows a high degree of nasalance across the entire [VG̃V] sequence, closely resembling the nasalance contour for the **contrastive nasal** [̃Ǿ̃Ǿ̃=CV] condition (oral postposition). The slight difference in nasality at the very end of [̃Ǿ̃Ǿ̃=CV] vs. [̃Ǿ̃Ǿ̃=NV] undoubtedly reflects local coarticulation with the following /C/ in **contrastive nasal** [̃Ǿ̃Ǿ̃=CV] vs. /N/ in [̃Ǿ̃Ǿ̃=NV].

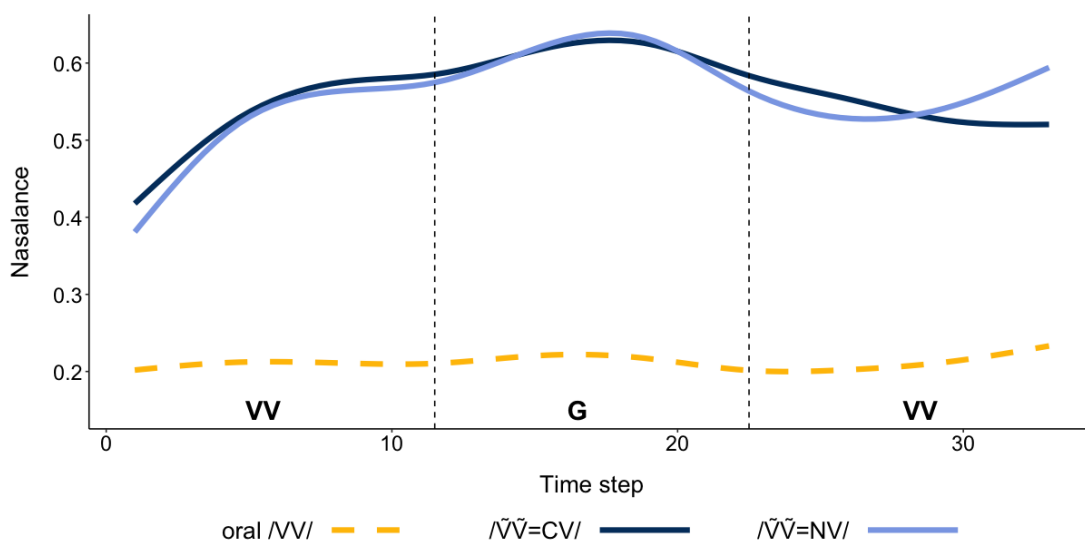


Figure 2.11: GAM curves comparing the [̃Ǿ̃Ǿ̃=NV] condition to the **oral** [VG̃V] and [̃Ǿ̃Ǿ̃=CV] conditions. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for each condition for all fifteen speakers.

A lmer model estimating the difference between the penultimate pre-glide stem vowels in the **contrastive nasal** [̃Ǿ̃Ǿ̃=CV] and [̃Ǿ̃Ǿ̃=NV] conditions in Table 2.7 shows that the mean nasalance of the penultimate pre-glide stem vowel in [̃Ǿ̃Ǿ̃=NV] is not significantly different from the mean nasalance of the penultimate

pre-glide stem vowel in $[\tilde{V}\tilde{G}\tilde{V}=CV]$.

Predictor	β	95% CI	t	p
Intercept (Penult V: in $[\tilde{V}\tilde{G}\tilde{V}=CV]$)	.545	[.465, .625]	13.28	< .001 ***
Penult V: in $[\tilde{V}\tilde{G}\tilde{V}=NV]$	-.046	[-.142, .050]	-.94	.36

Table 2.7: Results of the LMER model estimating mean nasalance of the penultimate vowel in the $[VGV]$ sequence for 15 speakers. `CONDITION` with two levels (penultimate vowel ($/\tilde{V}:CV/$) and penultimate vowel ($/\tilde{V}:NV/$)) was included as a dummy-coded fixed effect: `lmer(nasalance~CONDITION + (1 + CONDITION|speaker) + (1|word))`.

I conclude that the results of Experiment 2 cannot be analyzed by assuming that a nasalization window of approximately fixed length is anchored at the right edge of the rightmost nasal segment. The results are most straightforwardly consistent with a phonologically-controlled, optional harmony process which spreads nasality leftward from contrastive nasal vowels $/\tilde{V}:/$, but not from derived, predictable pre-N $[\tilde{V}:N]$ vowels.

2.3.5 Summary of Experiment 2

This experiment sought to uncover how pre-N $[\tilde{V}:N]$ vowels interact with nasal harmony in Punjabi. While the results of this experiment and Experiment 1 confirm that pre-N $[\tilde{V}:N]$ vowels show phonetically indistinguishable nasality from underlyingly nasal $/\tilde{V}:/$ vowels for most speakers, section 2.3.2.2 showed they do not trigger nasal harmony like contrastive nasal $/\tilde{V}:/$ vowels. In the following section I argue that this phonological difference between contrastive nasal $/\tilde{V}:/$ and predictably nasal $[\tilde{V}:N]$ vowels is best analyzed by assuming that surface $[\tilde{V}:N]$ reflects an underlying oral vowel, $/V:N/$ — in other words, a covert UR.

2.4 Discussion

2.4.1 Analyzing $[\tilde{V}:N]$ vowels in Punjabi with abstract URs

The above experiments established several facts about non-alternating $[\tilde{V}:N]$ vowels in Punjabi. First, these vowels are phonologically nasal on the surface. Their realization mirrors the categorical nasality of contrastive nasal $/\tilde{V}:/$ vowels in both experiments, and changes in speech rate do not impact the degree of nasalization on pre-N $[\tilde{V}:N]$ vowels. Nonetheless, predictably nasal vowels in pre-N $[VG\tilde{V}N]$ differ from contrastive nasal vowels in $[\tilde{V}\tilde{G}\tilde{V}=CV]$ and $[\tilde{V}\tilde{G}\tilde{V}=NV]$ in failing to trigger leftward nasal harmony: $/V:G\tilde{V}:/ \rightarrow [\tilde{V}:\tilde{G}\tilde{V}:]$, but $/V:GV:N/ \rightarrow [V:.G\tilde{V}:N]$, $*[\tilde{V}:\tilde{G}\tilde{V}:N]$ (Experiment 2).

The interaction between nasal harmony and pre-N $[\tilde{V}:N]$ vowels can be handled straightforwardly within any framework that (i) permits covert URs, and (ii) has mechanisms for generating counterfeeding opacity. Here, I provide an analysis using ordered rewrite rules (e.g. Chomsky and Halle 1968), but various other approaches are also viable (See chapter 3 for a novel approach to acquiring covert URs of this kind). I gloss over the optionality of nasal harmony, but this simplification has no consequences for the overall analysis.

The derivation in (5) produces an opaque interaction between nasal harmony and the contextual nasalization of pre-N vowels in $/V:N/ \rightarrow [\tilde{V}:N]$ (which I dub ‘assimilation’ in (5)). In (5a), the word-final nasal vowel $/\tilde{a}:/$ is present underlyingly. Harmony is the first rule to apply in the derivation, so it can only be sensitive to underlying structure. As such, harmony is triggered by the final $/\tilde{a}:/$ in $/sa:\tilde{u}\tilde{a}:/ \rightarrow [s\tilde{a}:\tilde{u}\tilde{a}:]$, as desired.

(5) Counterfeeding between nasal harmony and nasal assimilation in Punjabi

	a. /sɑ:vã:/	b. /ɑ:vɑ:m/
Harmony:	$[-\text{CONS}] \rightarrow [+ \text{NAS}] / __\text{X}_0 [-\text{CONS}] \text{V}_{[+ \text{NAS}]}$	sã:ũã: -
Assimilation:	$\text{V} \rightarrow [+ \text{NAS}] / __\text{C}_{[+ \text{NAS}]}$	- a:vã:m
	[sã:ũã:]	[ɑ:vã:m]

In (5b), the $[\tilde{\text{V}}:\text{N}]$ vowel is underlyingly oral, /ɑ:vɑ:m/. Since harmony is the first rule to apply, it cannot ‘know’ that the final vowel of /ɑ:vɑ:m/ will eventually be nasalized by the assimilation rule in (5) — harmony is only sensitive to underlying nasality, not derived, predictable nasality. As such, harmony fails to apply to /ɑ:vɑ:m/. The underlying form /ɑ:vɑ:m/ subsequently undergoes assimilation, producing the surface form [ɑ:vã:m]. This is a textbook case of counterfeeding opacity: assimilation applies ‘too late’ in the derivation to feed nasal harmony (e.g. Baković 2011; Baković and Blumenfeld 2024; Kiparsky 1976; Wiese 2000).

To reiterate a crucial aspect of the analysis, underlying representations like /ɑ:vɑ:m/ are *covert* in Punjabi, because pre-N vowels in non-alternating morphemes never actually surface as oral. The use of covert URs is a centerpiece of the analysis in (5): it is the key to distinguishing between derived (predictable) and contrastive (unpredictable) nasality for the purposes of determining when nasal harmony applies.

With that said, although I analyze all Punjabi pre-N vowels before a tautosyllabic nasal coda as underlyingly oral vowels that are systematically neutralized to nasal vowels, this analysis is similar to, but distinct from, the kind of ‘absolute neutralization’ rejected by Kiparsky (1982, p.127-135). Kiparsky defines absolute neutralization as a situation in which phonological distinctions “are never realized

on the phonetic surface ...regardless of environment” (Kiparsky, 1982, p.127). For example, (Kiparsky, 1982, p.133) notes that Chomsky and Halle (1968) analyze English words like ‘tolerance’, ‘competence’, etc. as ending in a glide-like phoneme with an e-quality, / ϵ /, that is always deleted. Because this phoneme never surfaces in any environment in English, such an analysis constitutes absolute neutralization. By contrast, the analysis of Punjabi pre-N vowels proposed here does not involve absolute neutralization: oral vowels surface freely in the language (e.g., [bɔ:t] ‘story’), and are only neutralized before nasal consonants.

To demonstrate that the pre-N nasal assimilation process in Punjabi is not a static phonotactic generalization fossilized in the words examined in this study, we can compare the pre-N forms from Experiment 2 — all of which lack a morpheme boundary between the pre-N vowel and following nasal consonant — with forms containing an underlyingly oral vowel that appear before a nasal consonant suffix. Although these forms were not analyzed as a primary condition in Experiment 2, all participants produced the words [ve:jɔ:ɳ] ‘to marry-NOML’ and [ge:jɔ:ɳ] ‘to know-NOML’ under the same elicitation procedure. In these forms, the vowel in the stem is oral in the absence of the nasal consonant suffix, but surfaces as nasal when the suffix is present. This is confirmed by the GAM curves in Figure 2.12, in which the underlyingly oral vowels followed by a nasal consonant suffix exhibit strikingly similar patterns to the tautomorphemic pre-N vowels from Experiment 2: they are systematically nasalized across the pre-N vowel itself, but they do not trigger nasal harmony across the preceding glide and vowel. This systematic nasalization of underlyingly oral vowels in combination with the lack of nasal spreading provides additional evidence that non-alternating pre-N vowels are underlyingly oral, and that their consistent surface nasalization reflects an active phonotactic alternation

in Punjabi.

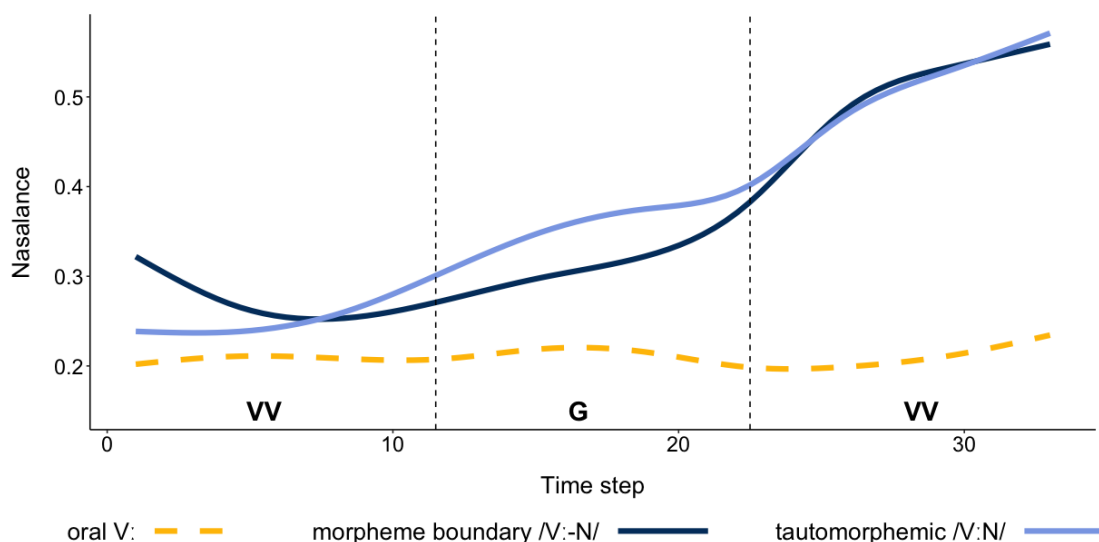


Figure 2.12: GAM curves comparing the tautomorphemic **pre-N** [VGVN] condition from Experiment 2 to forms with an underlying oral vowel followed by a nasal consonant suffix: morphemic boundary / $\tilde{V}$:-N/ ($n = 104$). Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VGV] sequence for each condition for the thirteen speakers from Experiment 2 that categorically nasalize pre-N vowels.

2.4.2 Punjabi nasal opacity: a challenge for surface-oriented frameworks

As demonstrated in the previous section, frameworks that permit covert structure offer a straightforward explanation for the nasal opacity patterns observed in Pakistani Punjabi. By contrast, theories that either reject URs altogether or tightly tether them to surface realizations struggle to account for these patterns. Such theories include eliminativist exemplar models (e.g. Bybee 2001; Johnson 1997, 2006; Pierrehumbert 2001) and surface-oriented generative approaches, in which symbolic URs are permitted but must correspond transparently to one of their surface forms (e.g. Albright 2002; Allen and Becker 2015; Burzio 1996;

Cole and Hualde 2011; Hayes 1999; Scobbie et al. 1996; Wang and Hayes 2025). Although these approaches differ in the representational format they assume, they are unified by the claim that all linguistic generalizations must be recoverable from surface distributions — that is, they reject covert representations.

In the subsections that follow, I consider how each type of surface-oriented model might attempt to explain the Punjabi data and show that any theory lacking covert representations faces serious difficulty in capturing the opacity effects documented here.

2.4.2.1 Opacity in exemplar models

A primary concern of generative phonology has been the modeling of productive, general alternations in phonological systems. Consider the famous ‘Bach test’ of Halle (1978): asked to pluralize the proper name *Bach* [bax], English speakers will uniformly provide [bax-s], as opposed to any of the other plural forms conceivably suggested by the vocabulary of English (*[bax-z], *[bax-ɪz], *[bax-ɹz], *[bax-ən], etc.). This is a productive pattern of generalization inasmuch as (i) speakers may have never heard, or themselves produced, the plural of *Bach* before, and (ii) the velar fricative [x] does not normally occur in English, so speakers have little or no direct evidence about how to pluralize forms ending in [x]. Traditionally, productive phonological generalization of this type has been modeled with abstract, feature based rules or other formal mechanisms (e.g. the voicing assimilation rule [-SON] → [αVOI]/[-SON, αVOI]__# ; Halle 1978).

Usage-based models have not been equally concerned with modeling productive phonological generalizations and alternations of this type, though work in this vein does exist (see e.g. Goldrick and Cole 2023; Kaplan 2015; Kirchner 2011; Kirchner

et al. 2010; Pierrehumbert 2006, 2016).¹¹ This is not to suggest that usage-based models lack mechanisms for productive generalization (as Pierrehumbert 2016 emphasizes). I am simply noting that the issue has received less attention in the usage-based theory literature than in generative phonology. Relatedly, the precise mechanisms that determine productive generalization in usage-based models remain a somewhat open question (e.g. how, exactly, ‘similarity’ is calculated between novel and stored forms for the purposes of pattern extension; how statistical regularities are actually computed for the purposes of learning phonotactics, etc. (e.g. Albright and Hayes 2003; Becker et al. 2017, 2011, 2012; Daland et al. 2011; Du and Durvasula 2025; Gorman 2013; Gouskova and Becker 2013)).

Perhaps more to the point, I am unaware of any attempt within the usage-based theory literature to model regular, opaque interactions like the counterfeeding interaction between nasal assimilation and nasal harmony in Punjabi discussed in (5) (see also Ettlinger 2008, 2009; Kawahara 2001). For that reason, my assessment of usage-based theory and related models in this section is based on my own understanding of the framework, rather than any specific, previous proposals. My focus is on ‘eliminativist’ usage-based theories — that is, theories which dispense with discrete, symbolic abstraction of the speech stream (Goldinger, 1998; Goldrick and Cole, 2023).

Several mechanisms have been proposed for producing some form of abstraction in a usage-based theory. The first mechanism is at the very core of the theory: the storage of many, phonetically-detailed memory traces for individual lexical items. If production and perception are mediated by some form of averaging over these

¹¹An infamous exception is the ‘English past-tense debate’; see e.g. Pinker (2002); Seidenberg and Plaut (2014). Note that this is a debate about morphological patterning (with phonological conditioning, Albright and Hayes 2003), where here I am concerned with purely phonological patterns.

traces, the result is abstraction: any single production target, or perceptual classification, will reflect the aggregate structure of the exemplar cloud, rather than any single, specific, stored concrete exemplar (e.g. Goldinger 1998; Pierrehumbert 2003, 2001, 2002; Wedel 2004, 2006). This is a species of abstraction.

As I have argued above, surface, phonetic nasality is functionally equivalent in contrastive nasal / $\tilde{V}$:/ and pre-N [$\tilde{V}$:N] in Punjabi. Consequently, no form of averaging over stored exemplar memories will produce an abstract difference between contrastive nasal / $\tilde{V}$:/ and pre-N [$\tilde{V}$:N] as such. I conclude that this first mechanism for abstraction in a usage-based theory — averaging over stored exemplar clouds — doesn’t provide any insight into why contrastive nasal / $\tilde{V}$:/ triggers harmony, but pre-N [$\tilde{V}$:N] does not.

A second mechanism for producing abstraction in usage-based theories comes from phonetic influence between morphologically and/or semantically related forms. For example, consider the well-known case of incomplete neutralization in German (see Du and Durvasula 2025 for references and recent discussion). In German, vowels tend to be longer before voiced obstruents (e.g. Braunschweiler 1997). This effect seems to persist even when final devoicing has applied: the vowel in / ʁa:d / $\rightarrow$ [ʁa:t] ‘wheel’ is slightly longer, phonetically, than the vowel in / ʁa:t / $\rightarrow$ [ʁa:t] ‘council’. In usage-based models, the slight lengthening of the vowel in [ʁa:t] ‘wheel’ can be attributed to influence from the plural form [ʁɛ:d-ə] ‘wheels’, where the vowel is transparently lengthened before voiced [d]. No such influence exists for [ʁa:t] ‘council’, which has the plural form [ʁɛ:t-ə] ‘councils’ (see also Braver 2019; Kaplan 2017).

Incomplete neutralization is a kind of phonetic opacity: singular [ʁa:t] ‘wheel’ opaquely inherits the vowel length of plural [ʁɛ:d-ə] ‘wheels’, where lengthening

is phonetically transparent. (This bears a close resemblance to the generative mechanism of OUTPUT-OUTPUT FAITHFULNESS (Benua, 2000; Braver, 2019), discussed in section 2.4.2.2.) The question naturally arises whether connections between morphologically related forms can also be leveraged to model the opaque non-application of harmony in forms ending in $[\tilde{V}:N]$.

I believe the answer is no: the requisite morphological relations are simply not present for the $[\tilde{V}:N]$ items in this study. Words like $[\partial d\zeta\partial v\tilde{a}:n]$ ‘omum seed’ are monomorphemic roots. To force the opaque non-application of harmony in $[\partial d\zeta\partial v\tilde{a}:n]$ through phonological or phonetic connections with morphologically related forms, there must be an item like $[\partial d\zeta\partial v\alpha:]$, in which harmony *transparently* fails to apply, because there is no nasal vowel to trigger it. However, no such form exists, because $[\partial d\zeta\partial v\tilde{a}:n]$ is a non-decomposable root: it always ends in $[\tilde{V}:N]$ on the surface. (I elaborate on this point in more detail in section 2.4.2.2.)

For suffixed forms like $[\text{ɛ}:-j\tilde{a}:n]$ ‘take-INF.OBL’, one could appeal to influence from a non-nasalized, related form of the stem vowel ($[\text{ɛ}:-j\alpha:]$ ‘take-IMP’) to explain why harmony fails to apply in $[\text{ɛ}:-j\tilde{a}:n]$. However, this explanation also falls short, because it equally — and again, incorrectly — predicts that nasal harmony should be inhibited in suffixed forms containing a *contrastive* nasalized vowel in the suffix, such as $[\text{m}\partial\text{t}fii-v\alpha:]$ ‘fisherman’ vs. $[\text{m}\partial\text{t}f\tilde{i}:-j\tilde{a}:]$, $*[\text{m}\partial\text{t}fi:-j\tilde{a}:]$ ‘fish-PLURAL’.

A third potential mechanism for producing opacity in $/V:GV:N/ \rightarrow [V:G\tilde{V}:N]$, $*[\tilde{V}:\tilde{G}\tilde{V}:N]$ is to invoke influence between *phonologically* related forms. Research in usage-based theories has sometimes assumed that phonologically similar forms can influence each other, through their gestalt similarity, regardless of their morphological (non-)relatedness (e.g. Bybee 2001, p.22). Concretely, if pre-N $/V:GV:N/ \rightarrow [V:G\tilde{V}:N]$ is influenced by *oral* $[V:GV:]$, that influence could in-

hibit harmony in the penult of [V:G $\tilde{V}$:N], in order to maintain consistency in the nasality of the penultimate pre-glide stem vowel.

The obvious question here is why pre-N /V:GV:N/ $\rightarrow$ [V:G $\tilde{V}$:N] should be influenced by oral [V:GV:], specifically. Indeed, forms with contrastive nasal vowels, [V:G $\tilde{V}$:], are at least equally plausible candidates for influencing pre-N [V:G $\tilde{V}$:N] through some notion of overall phonological similarity. Privileging oral [V:GV:] for this role is not well-motivated. Furthermore, it simply sneaks covert underlying representations in through the back door, by implicitly admitting that pre-N /V:GV:N/ $\rightarrow$ [V:G $\tilde{V}$:N] is in some sense more closely related to oral [V:GV:] than to contrastively nasal [V:G $\tilde{V}$:] (see also McCarthy 1999 on sympathy theory).

Fourth, it is possible that the distinction between pre-N /V:GV:N/ $\rightarrow$ [V:G $\tilde{V}$:N] and contrastive [V:G $\tilde{V}$:] is simply an artifact of historical change. Presumably, pre-N assimilation developed *after* the emergence of nasal harmony, thus yielding an opaque synchronic interaction between the two processes. Speakers might simply memorize forms like [mətʃi:~jã:] ‘fish-PLURAL’ vs. [lɛ:jã:n] ‘take-INF.OBL’. This would absolve the theory of any need to model the interaction between nasal harmony and local nasal assimilation.

The problem with this analysis is a familiar one: it fails to account for the productivity and generality of both nasal harmony and local nasal assimilation. As far as I am aware, these are completely regular processes in the varieties of Punjabi analyzed here. Through informal wug testing carried out with two native Punjabi speakers (Muhammad Riaz and Amir Adkuokhar), novel forms mirroring the structure of the Punjabi words in Experiment 2 impressionistically follow the same nasality patterns found on real Punjabi words in this paper: contrastive nasal vowels trigger harmony ([V $\tilde{G}\tilde{V}$]), but pre-N vowels do not ([VG $\tilde{V}$ N]). Thus,

I feel confident in assuming that the differences between these two nasal vowels would evince the same phonological behavior even when produced in completely novel forms. Something must ensure this; simple rote memorization will not do the job.

Lastly, the patterns of nasal harmony described in this paper could be accounted for with a more complex co-occurrence restriction on nasality, $*[\tilde{G}\tilde{V}:N]$, which bans a nasalized glide before a nasalized (long) vowel, followed by a nasal consonant. The trigram constraint $*[\tilde{G}\tilde{V}:N]$ straightforwardly excludes unattested $*[\tilde{V}:\tilde{G}\tilde{V}:N]$ sequences while allowing $[\tilde{V}:\tilde{G}\tilde{V}:]$. Furthermore, it does so with reference to surface forms alone: this renders the lack of $*[\tilde{V}:\tilde{G}\tilde{V}:N]$ sequences transparent, rather than opaque, and thus vitiates the need for abstract underlying representations. Within an usage-based framework — and likewise within symbolic theories that reject abstract URs not faithfully realized in some surface form (e.g., Albright 2002) — a constraint like $*[\tilde{G}\tilde{V}:N]$ could plausibly emerge from distributional learning over the Punjabi lexicon, based on surface forms alone.

However, two central issues have been raised in the literature concerning the viability of such complex co-occurrence constraints. The first is computational: the cost of learning constraints over segmental sequences quickly becomes intractable as the length of those sequences increases (Hayes and Wilson, 2008, 391-2). Constraints are stated over windows of length n , where n is some number of segments or other symbols (e.g. $*[g\#]$, $*[VtV]$, etc.). As n increases, the hypothesis space of possible constraints expands exponentially (the exact rate of growth depends on how many natural classes constraints may refer to). This makes it increasingly difficult for a learner to efficiently navigate the constraint space as constraints expand their segmental scope. For example, Hayes and Wilson (2008) argue that

constraints on segment combinations must generally be limited to $n = 2$; they relax this to $n = 3$ for constraints which refer only to the major class features [syllabic], [consonantal], and [sonorant]. The constraint $*[\tilde{G}\tilde{V}:N]$ of course goes beyond these limitations.

The second, and perhaps more pressing concern is empirical. Allowing the learner to posit highly specific co-occurrence restrictions based solely on attested patterns in the lexicon makes it difficult—if not impossible—to distinguish between principled phonotactic gaps and accidental ones (Chomsky and Halle (1965); Hayes and White (2013); Wilson and Gallagher (2018)). Principled gaps are those which native speakers judge to be ill-formed in their language, whereas accidental gaps are unattested but fully acceptable to speakers. A statistical learner, lacking a mechanism for making this distinction, would be just as likely to form constraints that exclude acceptable but unattested forms as it would be to discover true phonological generalizations.

To test the extent to which such overfitting might occur in Punjabi, I examined phonotactic patterns in the Pakistani Punjabi corpus, *pnbTenTen* (Jakubíček et al. (2013)), a 2.8 million-word corpus compiled in 2017 from 5,351 webpages written in Shahmukhi, the standard orthography for Pakistani Punjabi. The corpus contains approximately 100,000 unique words of varying frequencies. Using a custom Python script, I extracted the set of 3-gram sequences absent from the corpus by determining the complete set of unique Shahmukhi characters in the corpus and transliterating those characters into their corresponding natural class category: nasal consonant (N), liquid (L), obstruent (O), long oral vowel (V:), or long nasal vowel ($\tilde{V}$:). Several complications involved in transliterating Shahmukhi to broad natural classes led me to exclude glides and short vowels from the analysis. First,

Shahmukhi is a partial abjad, meaning that short vowels are written using optional diacritics, which are generally omitted from everyday writing and inferred from context. As a result, consistent transliteration of short vowels is not possible. For example, the conjunction [pər] ‘but, yet’ is most often written with two characters in Shahmukhi, omitting the short vowel: ڀڙ. The script transliterated this word as <#pr#>, with no short vowel. Furthermore, glides in Shahmukhi are not represented by a unique orthographic symbol but are instead written with the same symbol as their corresponding vowel. For example, ڄ may represent /j/ or /i:/ and ڀ may represent /v/, /u:/, /o:/, or /ɔ:/ . The exact context for each symbol-to-sound mapping is complex and not well documented for Shahmukhi. Consequently, I chose to transcribe these symbols consistently as vowels. Given these constraints, my analysis of the corpus does not accurately analyze potential phonotactic 3-gram sequence gaps containing glides, short vowels, or vowel hiatus. Thus, our search for phonotactic gaps is limited to 3-gram sequences containing individual long vowels and non-glide consonants.

After transliterating the corpus into broad natural classes, I generated all possible 3-gram sequences for the five target natural classes mentioned above and extracted the set of unattested sequences. As expected, many missing sequences reflect principled phonotactic restrictions – e.g. the absence of [#NN] sequences, which reflects an illicit word-initial consonant cluster of two nasals. However, other gaps are less easily explained. One particularly striking case is the complete absence of [L $\tilde{V}$:L] sequences (liquid-nasal vowel-liquid). Punjabi readily allows [LV:L] sequences with oral vowels – for example, [la:l] ‘red, ruby’ appears 634 times in the corpus. Additionally, two native speaker consultants (Muhammad Riaz and Amir Adkuokhar) independently provided judgements of well-formedness

for [l̃a:l], indicating — when asked whether [l̃a:l] could be a Punjabi word — that it is fully acceptable, though unattested. Moreover, smaller bigram [L̃V:] (type frequency 1,602, 3% of words) and [Ṽ:L] (type frequency 8, .01% of words) sequences are attested in the corpus, so less complex constraints based on simpler bigram co-occurrence restrictions cannot account for the gap. Another example is the absence of [Ṽ:L̃V:]. Given that [V:L̃V:] sequences are extremely common (type frequency 7,134, 11.6% of words), and that nasal vowels separated by other consonants are attested (e.g., type frequency of 7 for [Ṽ:ÕV:], .01% of words), this gap likewise appears to be accidental rather than principled.

While the phonotactic analysis here provides an illustration of how permitting complex co-occurrence constraints can lead to the grammatical prohibition of accidental gaps, previous work has demonstrated that this is a general issue for any approach that permits these types of constraints (Chomsky and Halle, 1968; Hayes and White, 2013; Wilson and Gallagher, 2018). In any case, the analysis here shows that a statistical learner limited to making generalizations over observed forms might posit a constraint like *[L̃V:L] or *[Ṽ:L̃V:] simply because no such sequence occurs in the lexicon. This is clearly an undesirable outcome — an example of a complex co-occurrence constraint banning a form that is well-formed for native speakers, despite its absence. While it is theoretically possible to model Punjabi’s nasal harmony patterns using surface-based generalizations alone, doing so comes at a significant cost. The computational demands of learning high-order co-occurrence restrictions are substantial, and, more importantly, this approach runs the risk of overfitting the lexicon, thereby failing to distinguish between accidental and principled gaps and resulting in the mischaracterization of speakers’ phonological competence.

In sum, in order to model the Punjabi data presented here, usage-based theories must be augmented with some mechanism(s) that allow speakers to arrive at covert underlying representations. I believe this is a matter of principle, rather than a consequence of the particular assumptions discussed above: it follows, more or less logically, from the surface-oriented character of all contemporary eliminativist exemplar frameworks.

2.4.2.2 Opacity in surface-oriented generative models

In the generative tradition, it has sometimes been suggested that underlying representations can be dispensed with entirely, their functions being subsumed by surface-to-surface relationships between morphologically related surface allomorphs, or other purely surface-oriented generalizations (e.g. Albright 2002; Allen and Becker 2015; Burzio 1996; Cole and Hualde 2011; Hayes 1999; Scobbie et al. 1996; Wang and Hayes 2025, and various others; cf. Bermúdez-Otero 2018; Bowers 2015; Hyman 2018; Krämer 2012 and references there). I illustrate with a case of counterfeeding opacity in Berbice Dutch Creole (Dow, 2014).

In Berbice Dutch Creole, nasal consonants generally assimilate in place of articulation to following oral stops. Additionally, stem-final vowels may delete when followed by a suffix. These two processes interact opaquely, in a counterfeeding relationship, as can be seen from derivations like /nimi-tɛ/ → [nim-tɛ] ‘know-ANTERIOR’ (6).

(6) Counterfeeding opacity in Berbice Dutch Creole

		/nimi-tɛ/
Place assimilation:	$C[+NASAL] \rightarrow [\alpha PL] / __\ [\alpha PL, -CONT, -SON]$	–
Vowel deletion:	$V \rightarrow \emptyset / __\]_{STEM-SUFFIX}[$	[nim-tɛ]
		[nim-tɛ]

As in the rule-based analysis of Punjabi nasal harmony (5), the rule-based analysis of Berbice Dutch Creole hinges on a distinction between underlying /N/+stop clusters (which undergo place assimilation), and derived /N/+stop clusters (which do not assimilate).¹²

An alternative, surface-oriented analysis of these facts is also possible, which does not make crucial reference to underlying representations. Following Benua (2000), we may assume that OUTPUT-OUTPUT FAITHFULNESS constraints enforce similarity between morphological bases and complex forms built on those bases. OO-FAITH constraints evaluate similarity along specific phonological dimensions: the constraint OO-IDENT-PLACE (7b) specifically bans changes in place of articulation between a morphological base like [nimi] and a corresponding inflected form like [nim-tɛ] (7). Importantly, OO-FAITH constraints refer only to the *surface* forms of bases and their morphological derivatives.

The simplified tableau in (7c) shows how output-oriented faithfulness relations can produce counterfeeding opacity without drawing a distinction between underlying and derived forms. In short, the regular process of nasal place assimilation is inhibited in [nim-tɛ] because it would produce a phonological discrepancy between the base [nimi] and the derived form *[nin-tɛ]. The input to phonological evaluation is a concatenation of independently observed surface allomorphs, [nimi] and [-tɛ], rather than an abstract UR.

(7) Counterfeeding opacity in Berbice Dutch Creole as output-oriented faithfulness

¹²The Berbice Dutch Creole facts are more nuanced than the text here implies. For example, nasal place assimilation is in fact optional in forms like [nim-tɛ]~[nin-tɛ] ‘know-ANTERIOR’, and stem-final vowel deletion shows evidence of morphological idiosyncrasy. I present a simplified version of the Berbice Dutch Creole data for expository purposes; see Dow (2014) for more details.

- a. NT PLACE ASSIMILATION: assign one violation for every nasal consonant followed by an oral stop, such that the nasal and stop differ in their place of articulation.
- b. OO-IDENT-PLACE: assign one violation for every morphologically complex form $\mathbb{B}\text{-}\mathbb{A}$, where $\mathbb{B}$ is a morphological base and $\mathbb{A}$ an affix, in which the $\mathbb{B}$ substring of $\mathbb{B}\text{-}\mathbb{A}$ contains a segment which differs in place of articulation from the corresponding segment in the base form $\mathbb{B}$.

	[nimi-tɛ] / BASE: [nimi]	OO-IDENT-PLACE	NT PLACE ASSIMILATION
c. a.	nin-tɛ	*! W	L
b.	nim-tɛ		*

However, it is well-known that output-oriented faithfulness does not provide a general analysis of phonological opacity (e.g. Bermúdez-Otero 2011; Mascaró 2016; Trommer 2013; Wolf 2008). OO-FAITH constraints can only produce opacity in morphologically complex forms, by demanding that phonological similarity to a morphological base take precedence over general phonotactic constraints (7).

In morphologically simple forms — that is, roots — there is no base to be faithful to, and thus no mechanism for generating opacity. This is precisely the situation presented by the opaque non-application of nasal harmony in forms like [dʒəvã:ŋ], *[dʒẽũã:ŋ] ‘youth, soldier’ in Punjabi. The root [dʒəvã:ŋ] has no morphological base to be faithful to; hence, output-oriented mechanisms for producing opacity fall short of explaining why [dʒəvã:ŋ] fails to undergo harmony. This ‘missing base’ problem is not specific to OO-FAITH: it generalizes to all phonological theories which attempt to generate opaque phonological interactions by means of surface relationships between morphologically related forms.

A conceptually distinct, surface-oriented approach to opacity in nasal spreading is provided by OPTIMAL DOMAINS THEORY (ODT) (Cole and Kisseberth 1994, 1995). Briefly summarizing, in ODT, harmony domains are specified in surface,

output forms. Surface-oriented constraints require all segments in a harmony domain to share the same value for a particular feature. The distributions of harmony domains themselves are also specified by surface-oriented constraints.

Cole and Kisseberth (1995) provide an analysis of nasal harmony in Máihĩki (formerly, Orejón) which is strikingly similar to the pattern of nasal harmony I describe for Punjabi (though see Sylak-Glassman et al. 2014 for a different description of the core facts). Nasality spreads rightward from nasal vowels in the initial syllable, the only position in which vowel nasality is contrastive. Harmony may spread through the glide [j], which I assume is nasalized, following Cole and Kisseberth (1995). Harmony is illustrated in (8a), in which I indicate nasal harmony domains by means of parentheses.

(8) Nasal harmony in Máihĩki

- a. Long-distance rightward nasal spreading from contrastive / $\tilde{V}$ /:

/sěje/ $\rightarrow$ (sějě) ‘species of bird’

- b. Only local rightward nasal spreading from nasal consonants:

/naji/ $\rightarrow$ (nã)ji ‘grandson’

However, words beginning in a nasal stop show only local nasalization of the following vowel, without longer-distance harmony (8b).¹³ Cole and Kisseberth (1995) analyze this contrast by assuming that contrastively nasal vowels initiate harmony domains that must spread as far to the right as possible (8a), while nasal consonants initiate harmony domains that only need spread as far as the rightmost syllable boundary (8b).

¹³Sylak-Glassman et al. (2014) argue that post-nasal vowels are oral rather than nasal in Máihĩki; this does not affect the discussion.

Crucially, Cole and Kisseberth (1995) must *also* invoke covert underlying representations for this analysis to function as intended. Roots like [nãji] ‘grandson’ are non-alternating: they are always realized with an initial nasal vowel. Lexical representations like /nãji/, which are faithful to the invariant surface realization [nãji], incorrectly predict nasal harmony in [nãji], *[nãjĩ] on analogy with [sẽjẽ], in which nasal harmony is initiated by an underlying nasal vowel (8a). The only tractable lexical representation for ‘grandson’ must therefore be the abstract, covert UR /naji/ — despite the fact that this morpheme never surfaces as such (i.e. the initial vowel is always nasal on the surface). We conclude that optimal domains theory also requires covert underlying representations.

A similar surface-oriented analysis might be developed from perceptually grounded theories of harmony, such as the motivation behind weak triggers of harmony outlined in Walker (2005, 2014). Under this analysis, contrastive nasal / $\tilde{V}$:/ vowels in Punjabi might be treated as weakly cued triggers: because nasality is realized only on the vowel, harmony is employed to enhance the perceptual recoverability of the nasal feature by extending it across additional vowels and glides. Pre-N vowels, on the other hand, occur adjacent to a nasal consonant, which buttresses the percept of the pre-N vowel’s nasality with an additional, highly salient nasal cue. Harmony is therefore unnecessary in pre-N [$\tilde{V}$:N] forms because nasality cues are sufficiently salient on the nasal consonant without it. Local vowel nasalization could help cue consonant nasality, but further spreading is not necessary.

While such an analysis is conceptually distinct from ODT, it encounters the same representational quandary. Speakers must still distinguish nasal vowels with contrastive nasality from nasal vowels with predictable nasality adjacent to a nasal consonant. This distinction is not available from the surface phonetic realization

of the vowels alone, since they are phonetically equivalent. Thus, a weak-trigger analysis cannot eliminate the need for a representational distinction between the two vowel types. If pre-N $[\tilde{V}:N]$ vowels are represented simply as nasal vowels, nothing prevents them from patterning with contrastive nasal vowels as harmony triggers. Conversely, if they are exempted from triggering harmony because their nasality is treated as distinct in some way, then speakers must encode precisely this distinction. Since the two vowel types are phonetically equivalent in their realization, that distinction cannot be read directly from the surface signal.

2.5 Conclusion

I have presented original phonetic data from Punjabi to establish the novel descriptive generalization that nasal harmony interacts opaquely with local nasal assimilation in $/V:N/$ sequences. I have further argued that modeling this opaque interaction requires reference to abstract underlying forms that never phonetically surface — forms I dub *covert URs*.

Covert URs allow for a straightforward distinction between underlying nasal vowels (which trigger harmony) and phonologically derived nasal vowels (which do not). Drawing this distinction *without* covert URs proves difficult. Many pre-N vowels in $/V:N/ \rightarrow [\tilde{V}:N]$ sequences are in non-alternating roots and affixes. This means that the vowels in question are *always* nasal $[\tilde{V}:N]$ on the surface. The lack of alternations in nasality presents a challenge for distinguishing non-alternating, contrastive nasal vowels $/\tilde{V}:/$ (which are unpredictably nasal) from non-alternating, derived nasal vowels $[\tilde{V}:N]$ (which are predictably nasal) by reference to surface forms alone.

I am broadly sympathetic to the goals of usage-based frameworks, and the empirical findings that motivate these and related frameworks. The question is not whether surface-oriented models of phonological computation are justified; the question is whether they are *sufficient*. I contend that an adequate theory of phonology will require mechanisms for learners to arrive at abstract lexical representations that depart from observed surface forms in filtering out at least some surface-predictable information. This conclusion, while traditional, strikes me as surprising in a modern context in which grammatical abstraction has been downplayed in favor of sophisticated, surface-oriented models of phonological learning and generalization.

Appendix

Experimental items

Condition 1 - CV:	Condition 2 - C $\tilde{V}$:	Condition 3 - CV:N
bɑ:t ‘story’	bā: ‘arm’	pa:n ‘Betel nut chewing’
bi: ‘seed’	bā:k ‘early teen’	bi:n ‘a musical instrument’
bo: ‘smell’		
pi:d ‘pain’	pī:g ‘swing’	
po: ‘12th Punjabi month’	pō:tʃ ‘arrival’	ko:n ‘who’
da: ‘strategy (wrestling)’		da:m ‘price, money’
da:k ‘post’	dā:g ‘big stick’	
ɖe:d ‘one and a half’		
do: ‘two’		do:n ‘ropes on a knit bed’
dʒi: ‘family member’		
	dʒō: ‘barley’	
dʒo:g ‘coming together’	dʒō:k ‘leach’	
	dʒū: ‘louse’	dʒu:n ‘June’
tɑ: ‘warmth, fever’	tā: ‘that; so that’	tɑ:n ‘melody’
tʰɑ: ‘COP.PST’	tʰā: ‘place, room’	tʰɑ:n ‘bolt of cloth’
tɔ:k ‘yoke, iron collar’		tɔ:ŋ ‘neck’
tʃɑ: ‘desire, wish’	tʃʰā: ‘shade’	tʃʰɑ:n ‘non-refined flour’
ru:p ‘beauty’	rū: ‘cotton’	lu:n ‘salt’
sa: ‘breath’		
se:k ‘warmth’	sē:k ‘termite’	
si: ‘sew.IMP’		si:n ‘ploughing’
sɔ: ‘hundred’	sō: ‘sleep.IMP’	sɔ:ŋ ‘rainy season’
su:d ‘profit, interest’	tū: ‘you’	
	vā:g ‘like, similar’	va:n ‘strings of a knit bed’
ga:l ‘abuse’	gā: ‘cow’	
ka:dʒ ‘button-hole’	kā: ‘crow’	kʰá:n ‘a proverb’
ku:k ‘melodious sound’	kū:dʒ ‘black sparrow’	
kʰu: ‘a well’	kʰū:dʒ ‘a corner, angle’	xu:n ‘blood’
xɑ:b ‘dream’		xantʃɑ: ‘corner, trick’

Table 2.8: Experiment 1 tokens, consisting of 63 monomorphemic words separated into three groups: oral vowels (CV:), nasal vowels (C $\tilde{V}$:), and pre-N vowels (CV:N).

Oral Condition - [VG̃V]			
sət-ɑːjiː	seven-twenty	ɔːliː-jɑː	pious man-PL
həʊɑː	air	tʃɔː-viː	four-twenty
pɑːʊeː	cot leg	bɛːʊɑː	a widow
dəʊɑː	medicine, remedy	tɑːʊuː	father's oldest brother
mʊk ^h iːjɑː	chief	ʊək ^h ɑːʊeː	boasting

Nasal Conditions - [ṼG̃Ṽ=CV/NV]			
tʃãːvẽː	pumicestone	ũː-ũẽː	that-way
dõː-ʊãː	two-way	kĩː-ũẽː	what-way
ẽː-ũẽː	this-way	mətʃĩː-jãː	fish-PL
sãː-ũãː	breath-PL	tĩːũĩː	a woman, wife

/VːN/ Condition - pre-N [VG̃ṼN]			
dʒəʊãːŋ	youth, soldier	ədʒəʊãːŋ	omum seed
sɪjãːŋ	recognition	tɛːjãːŋ	attention
lɛːjãːŋ	take-INF.OBL	ɑːʊãːm	general public

Table 2.9: Experiment 2 tokens, consisting of 24 di- and trisyllabic words separated into three groups: oral [VG̃V], contrastive nasal /Ṽː/, and pre-N vowels /VːN/.

Annotation Procedure & Excrescent nasals

Two undergraduate research assistants and I completed all annotations. I verified the quality of all annotations. Boundaries between oral consonants and vowels were marked at the onset/offset of high amplitude in the oral channel, along with the onset of increased formant activity in the spectrogram. The boundary between vowels and the onset of nasal consonants was marked at the attenuation of higher formant activity coupled with a rapid decline in waveform amplitude in the oral channel. A sample annotation is provided in Figure 2.13.

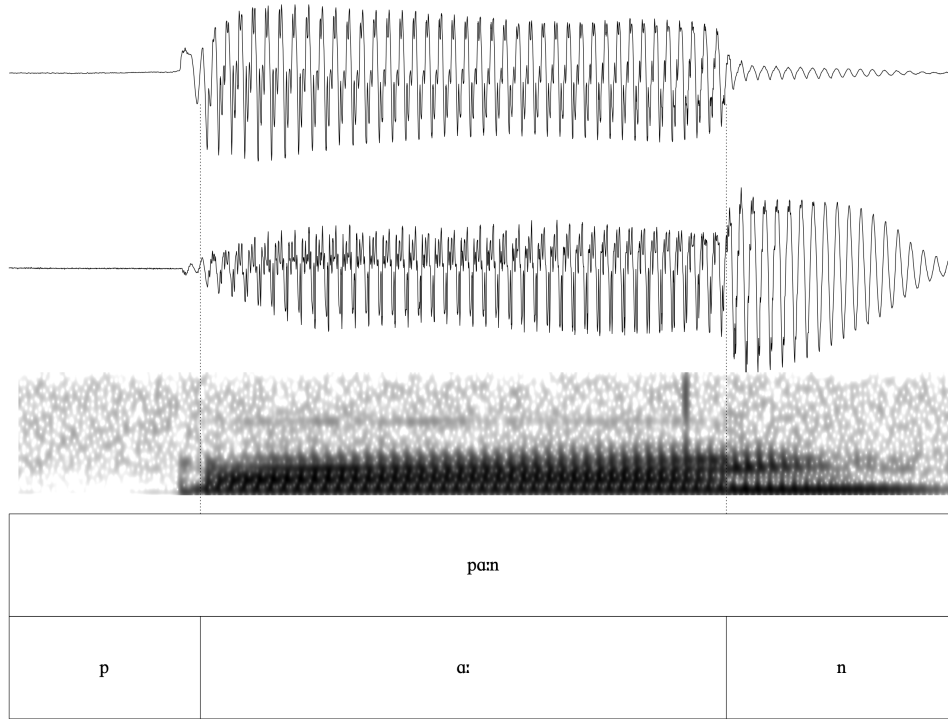


Figure 2.13: Spectrogram and waveform showing a sample annotation for /pa:n/ for speaker AR0. The top waveform shows amplitude for the oral channel, and the bottom waveform shows amplitude for the nasal channel.

For almost all of the $C\tilde{V}:$ tokens with oral stops in the coda position, oral occlusion for the coda consonant was achieved before the offset of articulation of nasality for the $/\tilde{V}:/$ vowel. This resulted in the physical presence of an intrusive nasal stop for a short period between the gestures of the vowel and the coda (see e.g. Bennett et.al. to appear; Desmeules-Trudel and Brunelle 2018; Lapierre 2023). Due to the minimal amount of oral air expelled during the production of nasal consonants, nasality increased rapidly at the offset of the contrastive nasal $/\tilde{V}:/$ vowel when these intrusive nasal consonants were present. To avoid corrupting the reported nasality of the $/\tilde{V}:/$ vowel, intrusive nasals were annotated as part

of the consonant segment whenever present. Importantly, it is clear from other phonotactic restrictions in Punjabi that this nasal stop is indeed intrusive and not present in the phonology. Specifically, Bhatia (1993) reports that NC coda clusters are illicit in the language when following a long vowel. A representative example of these intrusive nasal consonants is evidenced by the spectrogram and waveform in Figure 2.14.

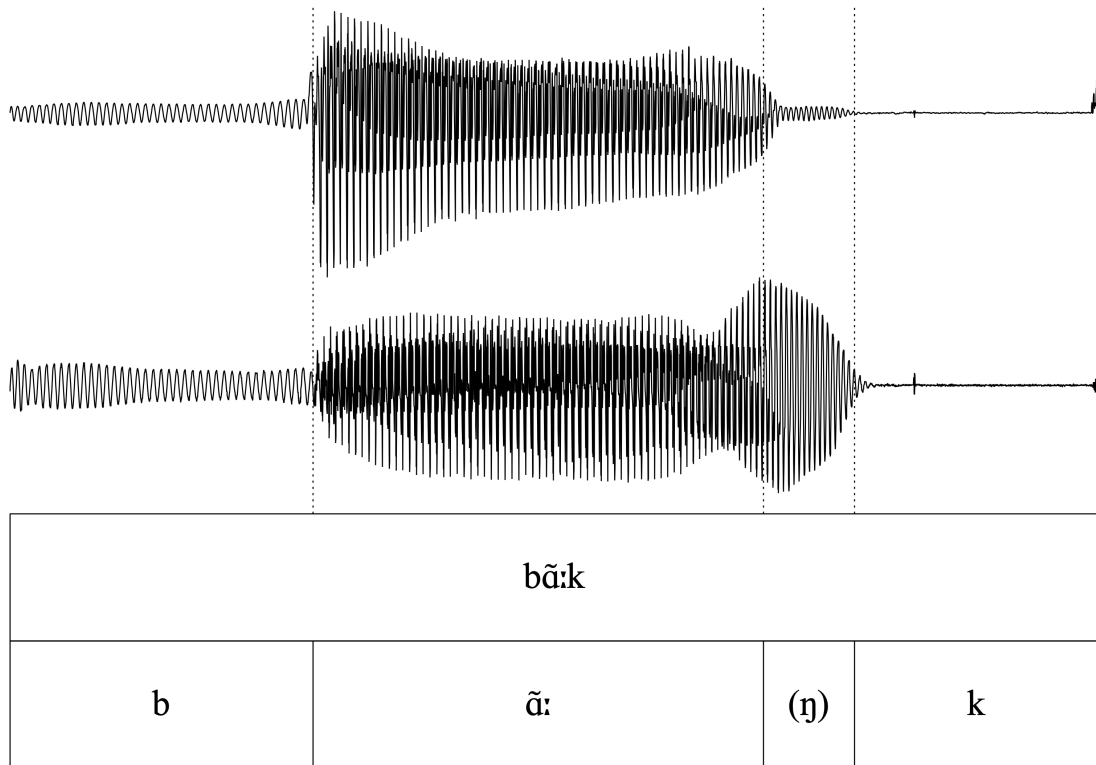


Figure 2.14: Waveforms showing the intrusive [ŋ] in /bã:k/ for speaker AK2. The top and bottom waveforms correspond to the oral and nasal channels respectively.

By-Speaker Harmony Patterns

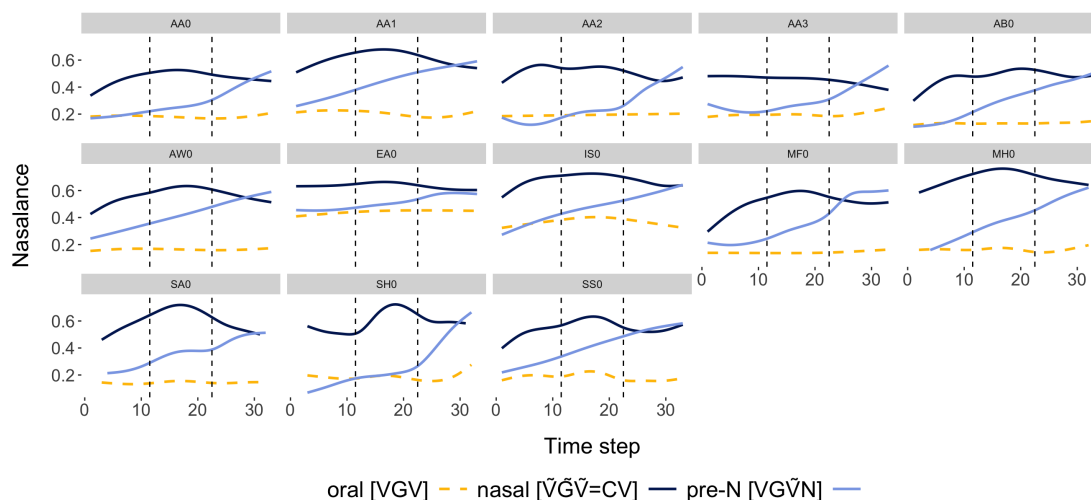


Figure 2.15: GAM curves for the oral [VG̃V], contrastive nasal [ṼG̃Ṽ=CV], and pre-N [VG̃ṼN] conditions, separated by speaker. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for the thirteen speakers that exhibit categorical nasality on final stem vowels in pre-N [VG̃ṼN].

Nasalization in Lahori Punjabi

In section 2.3.2 I noted that two speakers in Experiment 2 produced /V:N/ vowels with anticipatory nasal coarticulation rather than categorical nasality. GAM curves across [VG̃V] sequences are displayed in Figure 2.16 for these two speakers. As can be seen by the curves, nasalance across the penultimate pre-glide stem vowel and glide for the pre-N [VG̃ṼN] condition is proportionally similar to the oral [VG̃V] condition, but a cline-like increase in nasalance begins near the onset of the final stem vowel in pre-N [VG̃ṼN], only nearing the nasalance level of the contrastive nasal [ṼG̃Ṽ=CV] condition at the offset of the [VG̃V] sequence. This pattern suggests that the pre-N /V:N/ vowel is not intentionally targeted as [+NAS] by these speakers, which is quite distinct from all other speakers in

either Experiment 1 or Experiment 2. Crucially, however, the two speakers that exhibited this pattern grew up in Lahore, a large city of about 11 million people located on the eastern border of Pakistan, and no other speakers in either experiment grew up in this region. While further research is needed, I speculate that the dialect of Punjabi in Lahore displays a distinct pattern of nasality on /V:N/ vowels, such that they do not surface as [+NAS] vowels but instead are coarticulatorily nasalized as a result of the following nasal consonant. I leave it to future research to explore the veracity of the assumption that Lahori Punjabi does indeed represent a distinct dialect in terms of its nasality processes.

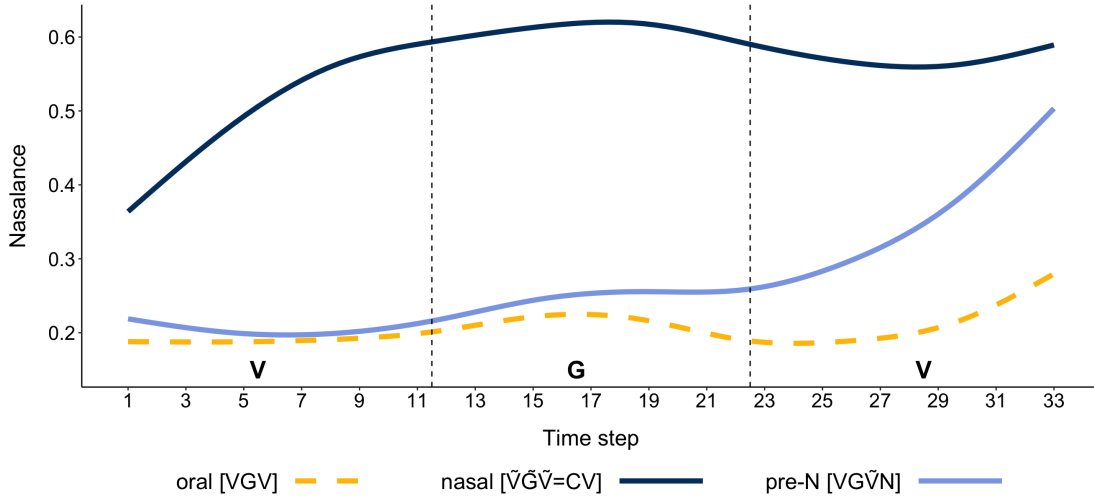


Figure 2.16: GAM curves for the oral [VG̃V], contrastive nasal [ṼG̃Ṽ=CV], and pre-N [VG̃ṼN] conditions. Curves visualize the mean vowel nasalance at 33 normalized timesteps across the [VG̃V] sequence for the two speakers that exhibit coarticulatory nasality on pre-N /V:N/ vowels.

Chapter 3

Learning Abstract Structure

As argued in the previous chapter, the behavior of non-alternating pre-N /V:N/ vowels in Punjabi motivates covert phonological structure. Usage-based approaches that reject such a high degree of abstraction have difficulty providing substantive explanations for the nasality patterns of the language. Nevertheless, despite the analytical motivations for assuming abstract URs, recent work on phonological learnability has raised serious concerns about the viability of such high degrees of UR abstraction. First, if multiple URs successfully model the observed data, what mechanism ensures that the learner converges on the minimally abstract UR? That is, how is UR abstraction permitted only to the extent necessary (Prince and Smolensky, 2004)? Second, and perhaps more pressingly, if UR abstraction is unconstrained, the space of possible URs grows without bound, introducing a potentially intractable search space problem (Jarosz, 2019; Wang and Hayes, 2025).

While several learning algorithms have been proposed that minimize UR abstraction within Maximum Entropy frameworks (O’Hara, 2017; Paramore, 2025), the

issue of search space size brought on by permitting increased UR abstraction remains largely unresolved. To date, most learning models have simply stipulated a finite set of potential URs for the learner to consider (O’Hara, 2017; Paramore, 2025). Wang and Hayes (2025), on the other hand, demonstrate that at certain levels of UR abstraction, the space of potential URs grows too large, preventing the learner from converging on a correct solution. Given these results, they propose that the boundary on permissible UR abstraction should be placed at the point at which learning fails. This conclusion places learnability-based limits on abstraction in direct tension with patterns in languages like the Punjabi data from chapter 2 that appear to require highly abstract URs.

This chapter introduces a novel learning algorithm—the **UR Progressive Hypothesis Space Expansion (UR PHaSE) learner**—which aims to reconcile these competing demands. The UR PHaSE learner solves both of the above issues posed by permitting abstract structure without imposing an explicit upper bound on the degree of abstraction. The central insight of the learning model builds on work by Tesar (2014, 2016), which structures the UR search space so that, rather than evaluating an infinite set of potential URs in parallel, the learner partitions the space into ordered subsets and considers candidates incrementally based on disparity count. Beginning with fully concrete URs, the learner only expands the space to include increasingly abstract URs when it fails to find an adequate learning solution at a lower level of abstraction.

Using the vowel nasality patterns in Pakistani Punjabi discussed in the previous chapter, I show that an appropriately structured hypothesis space of potential URs coupled with a serial search algorithm renders covert structure learnable without needing to provide external stipulations on the UR candidate set.

Section 3.1 outlines the conceptual and mathematical properties of the UR PHaSE learning architecture, and in Section 3.2 I apply the UR PHaSE learner to the Punjabi data from the previous chapter, demonstrating its successful acquisition of minimally abstract covert URs. In sections 3.3, 3.4, and 3.5, I explore three solutions proposed in the literature for learning covert URs and demonstrate how they fail to successfully acquire the correct covert structure for underlyingly oral pre-N /V:N/ $\rightarrow$ [$\tilde{V}$:N] vowels in Punjabi. Finally, Section 3.6 considers implications of these results as well as directions for future learnability research.

3.1 The UR PHaSE Learner

In this section, I introduce a novel learning algorithm — the UR Progressive Hypothesis Space Expansion (UR PHaSE) Learner — to address the computational challenges arising from including highly abstract representations in morphophonemic learning. The proposed learner structures the UR hypothesis space by disparity distance and considers potential URs in batches, beginning with fully concrete URs, only expanding the UR candidate space if the current set of UR candidates fails to meet a predetermined minimum likelihood threshold. When expanding the UR candidate set, the learner uses markedness violation profiles to identify features that are potentially mis-specified underlyingly and limits the generation of new URs to changes of those feature values. Overall, the learner inherently restricts abstraction to cases where introducing it demonstrably improves likelihood, while avoiding issues associated with the exhaustive search of an unbounded hypothesis space.

The UR PHaSE Learner integrates Maximum Entropy (MaxEnt) learning (Hayes and Wilson, 2008) with expectation-maximization (EM) learning (Jarosz, 2006a,

2015; Wang and Hayes, 2025) to jointly identify optimal URs for all morphemes in a given dataset and the constraint weights used in UR→SR mappings. Learning proceeds in four steps: (1) phonotactic learning to establish a best initial guess of constraint weights, (2) n-disparity UR expectation maximization learning, in which candidate URs with up to n disparities are evaluated; (3) Minimum Likelihood Threshold evaluation, which assesses whether the n-disparity UR candidate sets from step (2) sufficiently explain the observed data, and (4) generation of n+1 disparity UR candidates, which expands the UR candidate space when the current UR candidates fail the minimum likelihood threshold evaluation in step (3). Steps (2)-(4) are repeated until the Minimum Likelihood Threshold is satisfied for all morphemes.

Two properties distinguish the UR PHaSE learner from prior learning models: the progressive expansion of UR candidates and the establishment of a Minimum Likelihood Threshold at which the learner is satisfied with the success of a given UR in modeling the data and ceases to look for more optimal solutions. In the remainder of this section, I successively unpack the details of the four steps in the learner.

3.1.1 Step (1): Phonotactic Learning

The UR PHaSE learner first acquires constraint weights to model phonotactic patterns before proceeding to consider URs for individual morphemes. Not only does this two-stage learning approach approximate how children first acquire phonotactic patterns before becoming morphologically aware (Hayes, 2001; Jusczyk et al., 1994), but it also provides an initial set of markedness constraint weights that act as input to the morphological learning stage in step (2). This ‘initial guess’

at markedness constraint weights in the phonotactic learning stage provides the learner with a head start in finding the correct constraint weights during the morphological learning stage.

Phonotactic learning here follows the standard error-driven learning approach to phonotactic learning undertaken by MaxEnt learning models, similar to Hayes and Wilson (2008). The UR PHaSE learner takes as input a list of surface forms and pre-defined markedness constraints initialized at a neutral weight (e.g., set at 50 with constraint bounds of $[0, 100]$). Using gradient descent, the learner adjusts constraint weight values to minimize the objective function made up of two components: the negative log likelihood of the observed data and an L2 Gaussian prior (cf., O’Hara 2017) that favors high-weighted markedness constraints, thus increasing grammar restrictivity. The output of the phonotactic stage is an array of optimized constraint weights that capture the surface phonotactic patterns in the data while maximally satisfying the prior’s preference for a restrictive grammar.

3.1.2 Step (2): n-disparity UR EM Learning

After the phonotactic stage, learning progresses to a morphologically-aware stage in which the UR PHaSE learner jointly acquires a probability distribution over the set of possible URs and the optimal constraint weights that maximize the likelihood of the surface data. To do so, Expectation Maximization (EM) learning is employed to acquire UR probability distributions and constraint weights (Dempster et al., 1977). Jarosz (2006a,b, 2009, 2010, 2015) demonstrates that EM learning within a probabilistic OT framework can successfully learn a broad range of hidden linguistic structure, including URs, and the EM learning principles

developed in that work provide the inspiration for the approach taken here.

Morphologically-aware learning involves the simultaneous acquisition of URs for individual morphemes and the constraint weights that map those URs to their observed surface realizations (SRs) across contexts. This creates a classic chicken-and-egg problem. In order to correctly identify the UR for a morpheme that maximizes the likelihood of its SRs, the constraint weights governing the UR→SR mappings must already be known. At the same time, in order to arrive at the optimal constraint weights that maximize the likelihood of the UR→SR mappings, it is also crucial to know a morpheme’s UR. Consequently, learning both the constraint weight parameters and the UR distributions simultaneously is quite difficult. The EM algorithm provides a solution to this issue by breaking the learning problem into two iterated steps.

In the expectation step (E-step), the constraint weights are held constant while the posterior probability distribution for a UR candidate space is calculated for each morpheme, in essence calibrating UR probabilities to maximize the observed data given the current grammar. In the maximization step (M-step), the posterior probabilities calculated in the E-step are held constant while the constraint weights are adjusted to maximize the likelihood of the observed data. The two steps are applied in alternation until the model converges on a (locally) optimal combination of UR probability distributions and constraint weight values.

3.1.2.1 Expectation Step (E-step)

The goal of the E-step is to update prior beliefs about a hypothesis given new evidence. For an individual morpheme, the initial hypothesis of the learner is that the probability of candidate URs is equivalent across the entire set for that

morpheme. The initial new evidence is comprised of the combination of constraint weights learned during the phonotactic stage and the observed SRs of the morpheme.

Under these conditions, the learner computes posterior probabilities over the space of candidate URs for each morpheme, updating the probability that each UR is the correct UR for a given SR, using Bayes' Theorem in the equation in (1). When a morpheme has multiple allomorphs, these probabilities are conditioned jointly across SRs.

- (1) Calculation of posterior UR probabilities using Bayes' Theorem

$$\underbrace{\mathbb{P}(UR_i | SR_j)}_{\textit{Posterior}} = \frac{\overbrace{\mathbb{P}(SR_j | UR_i)}^{\textit{Likelihood}} \overbrace{\mathbb{P}(UR_i)}^{\textit{Prior}}}{\underbrace{\sum_{UR_k \in S} \mathbb{P}(SR_j | UR_k) \mathbb{P}(UR_k)}_{\textit{Marginal}}}$$

Bayes' theorem provides a principled mechanism for updating the learner's uncertainty over which UR is the correct representation for a given morpheme. With a set of observed SRs and fixed constraint weights, the posterior probability of a candidate UR can be computed by multiplying the joint likelihood of observing all SRs of a morpheme given that UR by the prior probability assigned to the UR, and then normalizing by the marginal probability of the surface forms across all candidate URs for the morpheme. As is known, given a candidate UR and a set of constraint weights, the output of a MaxEnt grammar is a probability distribution over possible SRs. Thus, for each candidate UR, both the likelihood term and the marginal probability can be computed straightforwardly.

The UR's prior probability is initially uniform across candidate URs before any

learning has taken place and is re-estimated in the M-step as the average posterior responsibility assigned to that UR across the SRs of the morpheme. Altogether, the E-step leverages prior beliefs about UR probabilities coupled with UR→SR likelihood calculations to recalibrate the learner’s expectation about the optimal UR for each morpheme.

Crucially, the size of a morpheme’s UR candidate space is what typically results in failures during UR learning (Wang and Hayes, 2025, p.33-35). In EM learning, each UR candidate’s probability for a given morpheme must be calculated, which means that the candidate space must be finite to be effectively traversed by the learner. As increasingly abstract URs are permitted, the hypothesis space size explodes, making traversal across the candidate set costly, leading to learning failures (Jarosz, 2019, p.78-79).

One of the main insights of the UR PHaSE learner is that the UR hypothesis space is structured based on the number of disparities in a UR→SR map for each morpheme (cf., Tesar 2014, 2016). This structure is illustrated by the lattice for a hypothetical morpheme with surface forms [i] and [ĩ] and possible disparities for the features [FRONT], [HIGH], and [Rd] in Figure 3.1. As Tesar (2014, p.21) defines it, a featural disparity in a UR→SR mapping occurs when corresponding input-output segments differ in their value for a feature.¹

The position of a UR in a morpheme’s UR hypothesis space is determined by the disparity distance between that UR and the closest SR. URs that are realized faithfully in at least one allomorph count as 0-disparity URs and reside at the top of the UR space. 1-disparity URs, on the other hand, can be reached by flipping

¹I set aside disparities caused by insertion and deletion here, but future work should incorporate them into the learner.

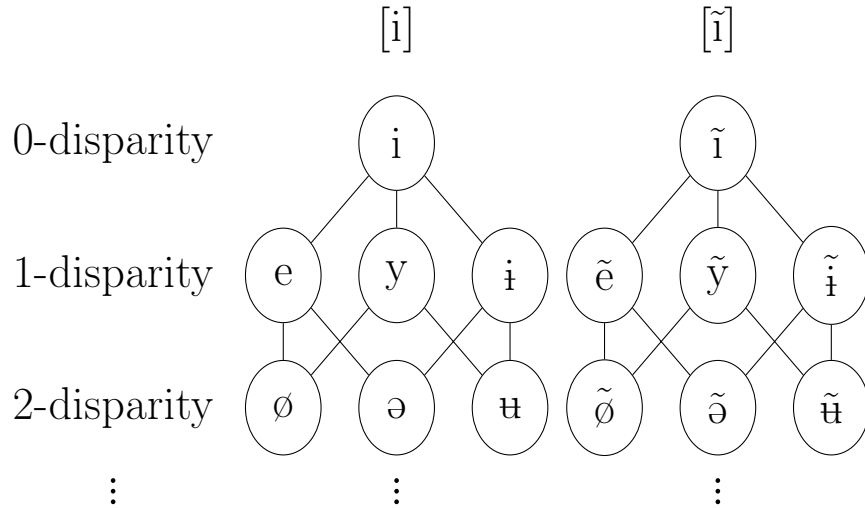


Figure 3.1: Lattice for a structured search space of candidate URs for a single morpheme realized contextually as [i] and [ĩ]. Visualized disparities only include the features [HIGH], [RD], and [FRONT].

exactly one feature value in one segment for at least one of the SRs. 2-disparity URs are those that can be reached by flipping exactly two feature values—either in one segment or across two segments—for at least one of the SRs. URs with more than two disparities (not depicted in Figure 3.1) are positioned on the appropriate disparity level.

Importantly, while the feature flips resulting in the level 2 candidate URs in Figure 3.1 are implemented on a single segment, this does not have to be the case. If a morpheme is constructed of multiple segments, there must be exactly two combined feature flips across all segments for a UR to be considered a 2-disparity UR.

Underlying the motivation to structure the UR candidate space based on disparities is the assumption that learners inherently disprefer differences in UR→SR mappings (Prince and Smolensky, 2004). That is, the learner assumes morphemes

will be represented concretely unless evidence suggests otherwise. The framework of the UR PHaSE learner incorporates this assumption into its algorithm by considering potential UR candidates in batches, beginning with all 0-disparity candidates and only considering higher-disparity candidates if the likelihood of the data remains low using URs with the current number of permitted disparities.

Note that nothing constrains the number of possible disparities a candidate UR can take, which means the entire hypothesis space can be quite large for any given morpheme. Nevertheless, the UR PHaSE learner’s algorithm takes advantage of the structure of the space to avoid searching it exhaustively. For example, rather than needing to search the entire UR candidate space in Figure 3.1, the UR PHaSE learner will initially include only those URs on the 0-disparity level ($/i/$ and $/\tilde{i}/$) in the UR candidate space when computing posterior probabilities in its first pass through step (2) of the learning process. URs positioned at higher-disparity levels are only considered in later EM learning rounds if, in step (3), the 0-disparity UR candidates are not found to predict the observed data at a high enough likelihood.

As will be shown in section 3.2, this approach allows for efficient UR learning, thereby enabling the acquisition of abstract representations posited to account for the behavior of pre-N $/V:N/$ in Pakistani Punjabi.

3.1.2.2 Maximization Step (M-step)

After defining the UR candidate space for each morpheme and calculating the posterior probability distributions of each UR candidate space, the learner moves to the M-step, seeking to maximize the likelihood of the observed data (Jarosz, 2006a,b). To do so, the learner updates the constraint weights ($\mathbf{w}$), using gradient descent to minimize an objective function (2) made up of the negative log likeli-

hood of all n observed SRs and an L2 Gaussian prior that maximizes grammar restrictivity by preferring high-weighted markedness constraints and low-weighted faithfulness constraints (O’Hara, 2017; Paramore, 2025).²

(2) Objective function during M-step of morphophonemic learning

$$\text{NLL} = -\ln \left[\prod_{i=1}^n (\mathbb{P}[SR_i \mid (\mathbf{w}, \pi_{post})]) \right]$$

$$\text{L2 Prior} = \sum_{w_i \in \mathbf{w}} \frac{(w_i - c_i)^2}{\sigma_i^2}$$

Once the optimal constraint weights are discovered in the M-step, the learner returns to the E-step, updating these posterior probability distributions based on the newly optimized constraint weights. Iteration between these steps continues until the learner converges on stable constraint weight values and UR probability distributions.

3.1.3 Step (3): Minimum Likelihood Threshold Evaluation

Convergence on a set of optimal UR distributions and constraint weights in the EM learning step is not equivalent to solving the learning problem. Rather, it means that the learner has arrived at a locally optimal solution with the current UR candidates under consideration for each morpheme, but it does not mean this solution is adequate. To determine adequacy, step (3) evaluates the likelihood of each individual SR in the observed data using the UR probabilities and constraint weights learned in step (2). If all SRs exceed a .95 likelihood threshold, the learner — satisfied that the current URs generate the observed data with sufficiently high probability — stops its search for a more optimal model. If, however, any single

²See Hayes and Wilson (2008) for a detailed explanation of the gradient descent process.

SR exhibits less than .95 likelihood (setting aside cases of variation), the learner has failed to meet the Minimum Likelihood Threshold and continues iterating through the learning pathway.

Establishing a Minimum Likelihood Threshold at which the learner is satisfied with its learned representations is a defining feature of the UR PHaSE learner. Typically, morphophonological learning algorithms do not settle on a solution that is *good enough*, but instead seek to maximally optimize the hypothesis. While this *maximal optimization* strategy has the potential to yield an analysis with a higher likelihood than that of the UR PHaSE learner, two main problems plague it.

First, as noted above, finding the globally optimal solution is not possible given a nonfinite hypothesis space, so these learners must constrain the hypothesis space in some way to effectively examine it. In-so-doing, these approaches stipulate—without any internal motivation—that certain UR candidates are excluded from consideration, which ultimately amounts to accepting a hypothesis that does not consider all possible candidates. The UR PHaSE learner, on the other hand, provides a principled reason for excluding potential UR candidates. Namely, once the likelihood threshold is reached, the learner deems the predictive capabilities of its current URs sufficient and calls off the search for a better alternative.

A second problem with the maximal optimization learning strategy lies in its decidedly nonhuman approach to decision making. Rather than examine all possible candidates in a hypothesis space exhaustively, humans almost always settle on a solution once it meets some threshold that satisfies their expectations (Simon, 1955, 1956). Indeed, decision-making theorists have argued since the mid-twentieth century that the limited cognitive capacities and time constraints with which humans approach learning problems prevent them from searching for

the maximally optimal solution. For example, when searching for the best price on a used car, a person cannot exhaustively examine all possible used cars for sale to get the best deal, but instead must settle on price that meets or exceeds their acceptability threshold of a *good deal*. Given this, it is likely that language acquisition is handled similarly.

Of course, it is certainly true that acceptability thresholds vary from individual to individual and decision to decision. In this sense, the .95 likelihood threshold utilized here serves as a stand-in for individual variation of what a satisfactory solution may be.

3.1.4 Step (4): UR Candidate Expansion

If any surface form falls below the .95 threshold in step (3) using n -disparity URs, the UR PHaSE learner will expand the UR hypothesis space for all morphemes by introducing $n+1$ -disparity UR candidates. Crucially, even at low-disparity levels, the unrestricted generation of UR candidates can cause the hypothesis space to grow too large to support effective learning.

To avoid this combinatorial explosion, the UR PHaSE learner does not generate all logically possible $n+1$ -disparity URs. Instead, it exploits information already present in the grammar to selectively incorporate new URs. Specifically, only features associated with markedness constraints that (i) are violated by at least one surface representation and (ii) currently exhibit low weights (i.e., non-maximal) are eligible to be flipped to generate new UR candidates.

This targeted strategy equips the learner with a principled means of expanding the UR search space. Rather than indiscriminately introducing all higher-

disparity candidates, the learner proposes only those URs that introduce additional UR→SR mapping candidates that make violations of these markedness constraints informative for learning. These competitors force the grammar to actively suppress illicit outcomes, thereby providing the learner with evidence for revising the relevant constraint weights. Moreover, because constraints with high weights are already active in shaping surface forms, introducing URs that flip features tied to those constraints offers no additional learning benefits. By contrast, generating URs that differ along features governed by low-weighted constraints creates an opportunity for the learner to reassess whether those constraints should change in strength.

In this way, the *selective* introduction of UR candidates at a single level of increased abstraction provides the learner with additional evidence to revise the grammar, while keeping the hypothesis space tractable.

3.1.5 Iteration Until Convergence

After the UR candidate space is expanded in step (4), the UR PHaSE learner repeats the learning process in steps (2)-(4) until all morphemes exceed the Minimum Likelihood Threshold evaluation in step (3).

3.2 The UR PHaSE Learner and Punjabi pre-N vowels

This section demonstrates that the level of UR abstraction required for pre-N vowels in Punjabi is learnable under the architecture of the UR PHaSE learner. The code and data used in this section are available at github.com/joncparamore

/UR_PHaSE_Learner.

Surface forms exemplified by the five word categories in Table 3.1 served as the input data for the learning model. The entire dataset is provided in the appendix. Six words were included from each of the five categories, resulting in a total of thirty forms. Along with the observed surface form of each word, the learner was also provided with morpheme boundary information.³ Categories in Table 3.1.i-ii demonstrate the nasal harmony pattern triggered by contrastive nasal vowels, and the patterns in Table 3.1.iii-iv illustrate that vowel nasality is contrastive. Finally, the category in Table 3.1.v demonstrates that pre-N vowels are nasalized but do not trigger harmony.

i.	[CV]	stem1	ii.	[C $\tilde{V}$ - $\tilde{G}\tilde{V}$]	stem1-suff1
iii.	[C $\tilde{V}$]	stem2	iv.	[C $\tilde{V}$ - $\tilde{G}\tilde{V}$]	stem2-suff1
v.	[CVG $\tilde{V}$ N]	stem3			

Table 3.1: Punjabi surface form categories

3.2.1 Step (1): Phonotactic Learning

The ten constraints listed in Table 3.2, along with their definitions (provided in the appendix), were also supplied to the learner. All constraint weights were initialized at the midpoint value of 50 at the outset of step (1). ID[NAS]/_V[-NAS] is a unique constraint, in that it is a *contextual faithfulness* constraint (Hauser and Hughto, 2020) that penalizes changes to a segment’s [NAS] feature value when directly preceding a vowel specified as [-NAS] in the *input*. As is well known, parallel versions of Optimality Theory are unable to capture many opaque phonological processes (McCarthy, 2000, 2007b), including the opaque counterfeeding pattern

³See Wang and Hayes (2025) for a method of learning morpheme boundaries in morpho-phonemic learning.

Constraint	Type	initial w	final w
SPRD-L[NAS]	mark.	50.00	0.13
*NASOBS	mark.	50.00	100.00
*NASG	mark.	50.00	0.47
*NASV	mark.	50.00	0.00
*VN	mark.	50.00	100.00
ID[NAS]	faith.	50.00	0.00
IDFIN[NAS]	faith.	50.00	0.00
ID[NAS]/_V[-NAS]	contfaith.	50.00	0.00
ID[RD]	faith.	50.00	0.00
*LOWRD	mark.	50.00	100.00

Table 3.2: Constraint weights with their initial and optimized weights in the phonotactic stage.

in which pre-N vowels are nasalized on the surface but do not trigger harmony in Punjabi. To address this, contextual faithfulness constraints reference a segment’s input context to explain why vowels and glides preceding pre-N vowels do not undergo nasalization.

As will become clear later, the two constraints governing the [RD] feature are included to demonstrate the UR PHaSE learner’s selectiveness based on markedness constraint violations in step (4) of the learning process.

The weights learned in the phonotactic stage are provided in the final column of Table 3.2 and serve as the input to the morphologically-aware learning stage that follows. From a cursory examination of the learned weights, it is clear that the phonotactic grammar correctly bans nasal obstruents, oral vowels before nasal consonants, and low round vowels — none of which occur in Punjabi — by maximizing the associated markedness constraints. Also, because the grammar does not consider URs during the phonotactic stage, all faithfulness constraints were

reduced to zero to allow for a maximally restrictive grammar. *NASG must have a low weight during the phonotactic stage because nasal glides are phonotactically licit in Punjabi (e.g., [sã:ũã:] ‘breaths’). Likewise, because pre-N forms do not exhibit nasal harmony (e.g., [tavã:n] ‘penalty’), SPRD-L[NAS] must also be assigned a low weight in the phonotactic stage. Otherwise, the grammar would impose nasal harmony in all applicable environments, including in pre-N forms.

3.2.2 Step (2): 0-disparity UR EM Learning

After phonotactic learning, step (2) uses the Expectation-Maximization (EM) algorithm to search for the combination of URs and constraint weights that maximize the likelihood of the observed surface forms. As a reminder, the algorithm alternates between an E-step, in which the current constraint weights and UR probability distributions are used to compute posterior UR probabilities for each morpheme, and an M-step, in which constraint weights are updated to maximize the likelihood of the thirty observed surface forms given the derived posterior UR distributions.

Although the learner optimizes UR probability distributions for all morphemes across the dataset, I report updates for only three representative word forms here: [sã:] ‘breath’, [sã:ũã:] ‘breath-PL’, and [tavã:n] ‘penalty’. Together, these forms are sufficient to diagnose whether the learner has acquired (i) nasal harmony triggered by contrastive nasal / $\tilde{V}$:/ vowels and (ii) the absence of nasal harmony when vowel nasality is derived in pre-N contexts.

In this round of EM learning, the UR PHaSE learner constructs the UR candidate space for each morpheme based solely on the surface alternations it has encountered. This is because the UR PHaSE learner inherently restricts abstraction by

considering URs in batches according to their disparity count. As such, in the first round of EM learning, only 0-disparity URs are considered in an attempt to acquire the patterns using exclusively concrete URs. The 0-disparity UR candidates for the morphemes ‘breath’ and ‘penalty’ (Table 3.3) are all identical to at least one SR of their respective morpheme.

The results of EM learning using only 0-disparity URs are given in Table 3.3. Note that the final posterior UR probabilities the learner converged on remain unchanged from their initial uniform distributions. Additionally, the likelihood of observing [ta:vã:n] ‘penalty’ is 1.0, indicating that nasal harmony triggered by non-alternating /V:N/ vowels is correctly not expected. For the two realizations of ‘breath’ [sa:]~[sã:ũã:], however, the learner is unable to correctly model nasal harmony given the current UR options, so instead optimizes the likelihood as much as possible by distributing UR probabilities evenly across the two candidate URs to capture the fact that sometimes ‘breath’ contains a nasal vowel and other times it does not. This results in approximately 50% likelihood of observing the correct nasality patterns of ‘breath’.

The ultimate reason for the learning failure is analytically straightforward: Given only concrete forms as UR candidates, the learner cannot find constraint weights that induce nasal harmony in ‘breath’ while simultaneously suppressing it in the pre-N form ‘penalty’. Crucially, this outcome represents the optimal solution available under a purely concrete UR hypothesis space: constraint reweighting alone cannot produce productive nasal harmony without simultaneously overgenerating harmony in pre-N contexts.

Constraint	Type	initial w	final w
SPRD-L[NAS]	mark.	0.13	62.92
*NASOBS	mark.	100.00	100.00
*NASG	mark.	0.47	60.24
*NASV	mark.	0.00	23.16
*VN	mark.	100.00	100.00
ID[NAS]	faith.	0.00	83.44
IDFIN[NAS]	faith.	0.00	89.46
ID[NAS]/_V[-NAS]	contfaith.	0.00	100.00
ID[RD]	faith.	0.00	6.29
*LOWRD	mark.	100.00	100.00

Morpheme	UR candidate	Prior	Posterior	SR	Likelihood
‘breath’	/sa:/	0.5	0.5	[sa]	0.50
	/sã:/	0.5	0.5	[sã:ũã:]	0.48
‘penalty’	/tavã:n/	1.0	1.0	[tavã:n]	1.00

Table 3.3: Constraint weights, UR probabilities, and SR likelihoods after 0-disparity EM learning.

3.2.3 Step (3): Minimum Likelihood Threshold Eval. for 0-disparity URs

As is already evident from the sampling of SRs in Table 3.3 above, the six singular surface forms with an underlying oral vowel in the stem (like [sq] ‘breath’) and the six surface forms with those same stems in the plural (like [sã:ũã:] ‘breath-PL’) all fail to meet the Minimum Likelihood Threshold of .95 when using only 0-disparity URs.

3.2.4 Step (4): UR Candidate Expansion

Unsatisfied with this learning result, the UR PHaSE learner proceeds to consider more abstract 1-disparity URs as potential additions to the UR candidate space. Importantly, 1-disparity URs are generated for all morphemes the learner encounters, not only those that fail to meet the likelihood threshold. Expanding the hypothesis spaces of all morphemes prevents the learner from cementing the 0-disparity analysis into part of the lexicon while allowing abstraction elsewhere, which can distort subsequent EM updates.

For instance, at the end of EM learning with concrete 0-disparity URs in section 3.2.2, the learner is unable to simultaneously model the presence of nasal harmony in alternations like [sa~sã:ũã:] ‘breath’~‘breaths’ and the lack of harmony in pre-N forms like [tavã:n] ‘penalty’. As a compromise, the learner lowers the weights of the constraints enforcing harmony and, in doing so, predicts the absence of harmony in pre-N forms with perfect probability, as shown in Table 3.3. If only the candidate spaces of morphemes with low likelihood were expanded, the learner would preserve the incorrect URs assigned to pre-N forms during the 0-disparity stage, thereby preventing it from ever arriving at the correct abstract, underlyingly

oral URs for pre-N forms.

In generating 1-disparity URs, the learner first examines the violation profiles of all markedness constraints with respect to the observed surface forms at the conclusion of 0-disparity learning (Table 3.3). If a markedness constraint exhibits a non-maximal weight (here, <100) and is violated at least once by one of the observed SRs, the feature associated with that markedness constraint becomes eligible to be flipped to construct new 1-disparity URs. A violated markedness constraint that fails to reach its upper bound indicates that weight re-calibration alone cannot cleanly explain the relevant patterns across SRs under current UR assumptions. That is, intermediate weights for markedness constraints signal uncertainty/inconsistency, which may be resolved by considering other candidate URs. The learner, therefore, treats such constraints with intermediate weights as evidence that the UR may be mis-specified along the associated featural dimension and permits increasing UR abstraction along this dimension in an effort to test whether a more abstract UR yields higher likelihood.

In fact, only SPRD-L[NAS], *NASG, and *NASV satisfy both criteria, because they have non-maximal weights but are violated by at least one of the observed forms. This, therefore, renders [NAS] as the only feature eligible to be flipped to generate new 1-disparity candidate URs. Crucially, *LOWRD both reaches the upper weight bound of 100 and is never violated by any of the observed SRs. Consequently, the learner determines that introducing UR→SR mappings with a disparity in [RD] is unlikely to be informative for learning because the observed surface forms provide no evidence that the [RD] feature is active in the pattern being learned. In this way, the learner selectively generates candidate URs that have the potential to improve learning.

Of particular importance for the present dataset, most other markedness constraints behave similarly to *LOWRD in not meeting both criteria to indicate to the learner that the associated feature should be added to the set of features used to generate new candidate URs. This is because the observed surface forms are relatively limited in their marked features.⁴ As a result, the expansion of the UR candidate space with the addition of 1-disparity URs will be relatively modest in the present case, only introducing URs that flip the [NAS] feature on eligible segments.

Table 3.4 shows the updated potential UR candidate set for each morpheme after adding 1-disparity URs to the hypothesis spaces of the morphemes ‘breath’ and ‘penalty’. Notice that the two morphemes differ sharply in how their UR candidate spaces expand at the 1-disparity level. For ‘breath’, the candidate space remains unchanged because flipping the [NAS] feature on the vowel yields URs that already belong to the 0-disparity level. Furthermore, the nasalized obstruent [ṣ̃] is unattested in Punjabi as a phoneme or allophone, leaving the learner without a phonological category that could support representations such as /ṣ̃ɑː/ or /ṣ̃ɑː/.⁵

In contrast, the UR candidate space for the morpheme ‘penalty’ expands from a single concrete UR to four potential URs. By flipping the [NAS] feature on all segments for which a corresponding segmental category exists in Punjabi, the learner introduces a small but analytically consequential set of 1-disparity URs. Specifically, the expanded UR space includes the analytically correct UR with the

⁴A crucial assumption here is that most sets of surface forms will only violate a handful of markedness constraints associated with different features. Future research is needed to determine whether this assumption is borne out cross-linguistically and how learning is affected by datasets that violate several families of markedness constraints.

⁵The assumption that URs must be composed of segments attested in the language is not crucial to the proposed learner. While excluding unattested segments reduces the size of the UR hypothesis space, allowing such segments would not necessarily render the space intractable.

Morpheme	UR Candidate Space			
‘breath’	/sa:/		/sã:/	
‘penalty’	/tavã:n/	/ta:vã:n/	/tã:vã:n/	/ta:va:n/

Table 3.4: UR candidates for ‘breath’ and ‘penalty’, expanded to include all possible 1-disparity URs generated by flipping the [NAS] feature.

underlyingly oral pre-N vowel, /ta:va:n/, while remaining tightly constrained and computationally tractable.

3.2.5 Step (2): 1-disparity UR EM Learning

Equipped with the newly expanded hypothesis spaces, the learner returns to step (2) to reinstantiate EM learning, the results of which are shown in Table 3.5. As is evident, the combination of learned weights and UR probability distributions enables the learner to predict the observed SRs with high degrees of likelihood. This is achieved by assigning all of the UR probability for ‘breath’ to oral /sa:/ and deriving nasality on the stem by way of nasal harmony triggered by the vowel in the plural suffix. Conversely, the learner also correctly assigns all of the probability to the abstract UR /ta:va:n/ for ‘penalty’, thereby enabling the pre-N vowel to surface as nasal (due to highly weighted *VN) while not predicting nasal harmony on these forms (due to high weight on the contextual faithfulness constraint, ID[NAS]/_V[-NAS]).

3.2.6 Step (3): Minimum Likelihood Threshold Eval. for 1-disparity URs

With the addition of 1-disparity URs, the learner converged on a combination of constraint weights and UR probabilities that yielded a likelihood greater than .95

Constraint	Type	initial w	final w
SPRD-L[NAS]	mark.	0.13	48.45
*NASOBS	mark.	100.00	100.00
*NASG	mark.	0.47	48.45
*NASV	mark.	0.00	0.00
* _{VN}	mark.	100.00	100.00
ID[NAS]	faith.	0.00	0.00
IDFIN[NAS]	faith.	0.00	100.00
ID[NAS]/_V[-NAS]	contfaith.	0.00	100.00
ID[RD]	faith.	0.00	6.28
*LowRD	mark.	100.00	100.00

Morpheme	UR candidate	Prior	Posterior
‘breath’	/sa:/	0.5	1.00
	/sã:/	0.5	0.00
‘penalty’	/ta:rvã:n/	0.25	0.00
	/ta:va:n/	0.25	1.00
	/ta:ũã:n/	0.25	0.00
	/tã:ũã:n/	0.25	0.00

SR	Likelihood
[sa:]	1.00
[sã:ũã:]	0.96
[ta:rvã:n]	1.00

Table 3.5: Constraint weights, UR probabilities, and SR likelihoods after 1-disparity EM learning.

for each SR, thus successfully acquiring a grammar and lexicon that accounts for the data. As a result, the learner ceases its search for a more optimal solution.

3.3 The Free Ride Learning Algorithm

McCarthy (2005b) offers the “free ride” solution to the learnability of covert URs, in which he argues that learners arrive at non-alternating covert URs via analogy with similar forms in the language that exhibit unfaithful surface alternations. Formally, if an underlying representation contains the feature value, $[\alpha F]$, but $[\alpha F]$ always surfaces as $[\beta F]$ (i.e. $[\alpha F]$ is covert), $[\alpha F]$ can be acquired by generalizing from another morphological form that contains $[F]$ that does exhibit morphophonemic surface alternations between $[\alpha F]$ and $[\beta F]$. The process responsible for the surface variation in the alternating form is generalized to the form that shows no surface alternation because the learner generalizes the unfaithful mapping, $/\alpha F/ \rightarrow [+ \beta F]$, from the alternating form to the non-alternating form. In this way, the non-alternating form takes a “free ride” in the unfaithful mapping from UR $\rightarrow$ SR with the forms that plainly exhibit the derivation.

One example McCarthy cites is a coalescence process that conglomerates two adjacent vowels with differing heights into a long mid vowel in Sanskrit. The Sanskrit data in (3) illustrates the phenomenon. In both examples, a low vowel and a high vowel coalesce into a single long mid vowel. Many Sanskrit forms with long mid vowels demonstrate surface alternations that confirm this coalescence process. However, some long mid vowels in the language lack such alternations.⁶

⁶Though McCarthy is clear to point out that he has no direct evidence that non-alternating long mid vowels are underlyingly diphthongs in Sanskrit, Lloret and Pons-Moll (2016) provide evidence from vowel epenthesis in Majorcan Catalan in which learners seem to “undo” non-alternating epenthetic vowels, consistent with free-ride learning of URs. Nevertheless, for the sake of simplicity, I use the Sanskrit coalescence example to introduce the FRLA framework.

By applying the Free-Ride Learning Algorithm (FRLA), McCarthy contends that learners analyze these non-alternating mid vowels as diphthongs in their underlying forms despite the complete absence of evidence that these non-alternating forms undergo the coalescence process.

- (3) Sanskrit coalescence (McCarthy, 2005b; Whitney, 1889)
- | | | |
|--------------------------|------------------------|------------------------|
| /tava i ndra/ | tave ɪ ndra | ‘for you, Indra (VOC)’ |
| /hita u padaifah/ | hito ʊ pade:fah | ‘friendly advice’ |

The FRLA successfully accounts for cases in which the underlying representation of some morpheme never surfaces in any environment. For example, in Sanskrit, an adjacent high-low vowel pair is never realized in any position, thereby allowing for an analysis in which learners make a blanket generalization that all high-low vowel pairs are transformed into long mid vowels when URs are filtered through the phonological grammar. The benefit of positing that the learner prefers a grammar of Sanskrit in which *all* mid vowels are derived from underlying diphthongs is that it explains the absence of short mid vowels [e o] in Sanskrit. That is, the analysis allows for a *restriction* on the number of potential surface forms with a general prohibition against all mid vowels (e.g., *MID), only permitting mid vowels that are derived by the coalescence process. The result is that the predicted forms in a grammar which does not exercise a general prohibition against mid vowels makes up a proper superset of those in which mid vowels are prohibited, with the unattested *[e] as the permissible element in less restrictive grammar that does not intersect with the more restrictive grammar. In sum, while both grammars successfully generate the Sanskrit vowel system, a less restrictive grammar is inferior because it predicts more unattested surface forms (Prince and Tesar, 2004; Tesar, 2014).

For cases of contextual neutralization, where two sounds are contrastive and there-

fore realized faithfully in non-neutralizing positions, McCarthy is clear to point out that the FRLA cannot arrive at a grammar to account for the data because it would be *too* restrictive. In Punjabi, for example, oral and nasal vowels are clearly contrastive in the language, as evidenced by the (near) minimal pairs in (4i-vi). Nevertheless, this contrast is neutralized before nasal consonants such that vowels are always realized as nasal on the surface, as in (4vii-viii).

(4) Contrastive vowel nasality and pre-N neutralization in Punjabi

i.	[sɑ:]	‘breath’	ii.	[tʃ ^h ɑ:]	‘shade’
iii.	[sɔ:]	‘hundred’	iv.	[sõ:]	‘sleep’
v.	[se:k]	‘warmth’	vi.	[sẽ:k]	‘termite’
vii.	*[sim]		viii.	[sĩm]	‘ploughing’

Accounting for the inability of pre-N /V:N/ vowels to trigger nasal harmony by positing that learners give /V:N/ vowels a free ride on some process that nasalizes oral vowels is untenable because not all nasal vowels can be derived from oral vowels in Punjabi. As McCarthy (2005b, 34) concedes, “taking a free ride on a neutralization process is a bad choice”. Therefore, the free ride hypothesis cannot account for the learnability of non-alternating /V:N/ vowels as underlyingly oral.

What follows demonstrates the failure of the FRLA to arrive at a grammar that successfully learns covert /V:N/ URs in Punjabi. To employ the FRLA, McCarthy utilizes the biased multi-recursive constraint demotion learning procedure (BMCD) outlined in Prince and Tesar (2004), in which multiple inputs are considered simultaneously to find a grammar. Under BMCD, constraints are ranked one by one; this process is iterative until either every constraint has been ranked or no grammar can be found to account for the data. If any losing candidates are ruled out by the ranking of a constraint, they are considered “already accounted for” in the subsequent loops through the learning procedure. Because BMCD stipulates that more restrictive grammars are favorable, the procedure prefers to rank

markedness constraints over faithfulness constraints whenever possible. Thus, when both a markedness and faithfulness constraint are rankable, the markedness constraint is chosen first. Importantly, a constraint is only rankable if no remaining “unaccounted-for” loser is preferred to a winner for any of the inputs being considered.

Definitions for the constraints relevant for learning the Punjabi nasality pattern are provided in (5). Most constraints are straightforward, but one is worthy of further elaboration. ID[NAS]/_V[-NAS] is a unique constraint in that it is a contextual faithfulness constraint (Hauser and Hughto, 2020) that penalizes changes to a segment’s [NAS] feature value when directly preceding a vowel specified as [-NAS] in the *input*. As is well known, parallel versions of Optimality Theory are unable to capture many opaque phonological processes (McCarthy, 2000, 2007b), including the opaque counterfeeding pattern in which pre-N vowels are nasalized on the surface but do not trigger harmony in Punjabi. To address this, contextual faithfulness constraints reference a segment’s input context to explain why vowels and glides preceding pre-N vowels do not undergo nasalization.

(5) Constraint Definitions

- i. SPRD-L[NAS] (Walker, 2003, p.47)
For every occurrence of a [+NAS] feature in a prosodic word, if that [+NAS] feature is dominated by some segment, assign a violation for every segment to the left of that segment in the prosodic word that does not dominate the [+NAS] feature.
- ii. *NASOBS (Walker, 2003, p.51)
Assign a violation for every obstruent that dominates a [+NAS] feature.
- iii. ID[NAS]
For every segment, *A*, assign a violation if the output value for the [NAS] feature dominated by *A* does not match the input value for the [NAS] feature dominated by *A*.
- iv. IDFIN[NAS]
For every segment, *A*, assign a violation if the output value for the

[NAS] feature dominated by A does not match the input value for the [NAS] feature dominated by A in the final syllable of a prosodic word.

v. ID[NAS]/_V[-NAS]

Let A be a segment that occurs before an oral vowel, __V[-NAS], in the input. Assign one violation if the output correspondent of A does not have the same specifications for [NAS] as A .

The support tableau (Tesar and Prince, 2003) in (6) shows the relevant inputs and their constraint violations.^{7,8} The tableau represents the onset of phonotactic learning, at which point no patterns have been learned, so constraints remain unranked. A crucial aspect of the tableau is that the learner makes the initial simplifying assumption that all surface forms, including [sĩ:n], are faithful realizations of their underlying representations. As mentioned previously, previous work suggests that children seem to be sensitive to phonotactics prior to constructing a lexicon of morphemic representations and identifying alternations (Hayes, 2001; Hayes and Wilson, 2008; Jusczyk et al., 1994). As will be shown, however, when the learner is exposed to novel forms and morphophonemic alternations, this assumption of faithful mapping must be revised, along with an additional round of BMCD to account for the new forms.

According to BMCD, when presented with an input and potential outputs, the learner first attempts to rank any markedness constraint that does not favor a loser. *NASOBS is the only markedness constraint that meets this criterion at this stage, so it is ranked at the top of the constraint set.

⁷I assume here that the contextual faithfulness constraint, ID[NAS]/_V[-NAS], behaves like standard faithfulness constraints under BMCD. More specifically, if both ID[NAS]/_V[-NAS] and a markedness constraint are both rankable, the markedness constraint will be ranked first. However, nothing in the present analysis hinges on this assumption.

⁸*VN, the markedness constraint penalizing oral vowels before nasal consonants is typically called upon to ensure /V:N/ → [Ṽ:N] (Cohn, 1993). However, the same effect of precluding oral vowels before nasal consonants is achieved in the current Punjabi dataset by ranking SPRD-L[+NAS] over ID[NAS]. Therefore, I omit *VN in the following analysis in an attempt to reduce the number of constraints — and thus complexity — of the present analysis.

(6) Unranked Punjabi nasality constraints before phonotactic learning

Lexicon	cands.	IdFIN[NAS]	Id[NAS]/_V[-NAS]	SPRD-L[NAS]	Id[NAS]	*NASOBS
/sa:/	 sa:					
	sã:	*W		*W	*W	
	šã:	*W	*W		*W*	*W
/tʃʰã:/	 tʃʰã:			*		
	tʃʰa:	*W		L	*W	
	ṭʃʰã:			L	*W	*W
/sĩ:n/	 sĩ:n			*		
	si:n			**W	*W	
	šĩ:n			L	*W	*W

As displayed by the greyed-out rows in the support tableau in (7), *NASOBS rules out any candidate that nasalizes an obstruent. Thus, the greyed-out candidates are now considered accounted for and need not be considered for subsequent iterations through this round of BMCD.

(7) Support Tableau after first ranking pass through BMCD

Lexicon	cands.	*NASOBS	IdFIN[NAS]	Id[NAS]/_V[-NAS]	SPRD-L[NAS]	Id[NAS]
/sa:/	 sa:					
	sã:		*W		*W	*W
	šã:	*W	*W	*W		*W*
/tʃʰã:/	 tʃʰã:				*	
	tʃʰa:		*W		L	*W
	ṭʃʰã:	*W			L	*W
/sĩ:n/	 sĩ:n				*	
	si:n				**W	*W
	šĩ:n	*W			L	*W

After *NASOBS has been ranked, the learner searches for other rankable markedness constraints (constraints that favor winners over all unaccounted-for losers). SPRD-L[NAS] is the only remaining markedness constraint, but it cannot be ranked

at this point because it favors the losing candidate, $[t^h\alpha:]$, over $[t^h\tilde{\alpha}]$ for the input $/t^h\tilde{\alpha}/$. Consequently, the learner moves on to search for a rankable faithfulness constraint. Importantly, the BMCD favors ranking faithfulness constraints that will allow a markedness constraint to be ranked on the next ranking pass (see Prince and Tesar 2004 for a detailed explanation of how this selection process plays out). ID[NAS] permits SPRD-L[NAS] to be ranked on the next pass, so it is ranked next.

After ranking ID[NAS], all remaining losing candidates are accounted for, as illustrated in (8). Because no more unaccounted-for losing candidates remain, the markedness constraint, SPRD-L[NAS], is now free to be ranked next, followed by the two remaining unranked faithfulness/contextual faithfulness constraints, IDFIN[NAS] and ID[NAS]/_V[-NAS], which are ranked at the bottom of the hierarchy.

(8) Support Tableau after second ranking pass through BMCD

Lexicon	cands.	*NASOBS	ID[NAS]	IDFIN[NAS]	ID[NAS]/_V[-NAS]	SPRD-L[NAS]
$/s\alpha:/$	 $s\alpha:$					
	$s\tilde{\alpha}:$		*W	*W		*W
	$\tilde{s}\tilde{\alpha}:$	*W	*W*	*W	*W	
$/t^h\tilde{\alpha}/$	 $t^h\tilde{\alpha}:$					*
	$t^h\alpha:$		*W	*W		L
	$\tilde{t}^h\tilde{\alpha}:$	*W	*W			L
$/\tilde{s}\tilde{i}n/$	 $\tilde{s}\tilde{i}n$					*
	$si:n$		*W			**W
	$\tilde{s}\tilde{i}n$	*W	*W			L

The final constraint ranking after the first round of BMCD concludes is given in (9) and successfully captures the phonotactic data presented to the learner thus far.

(9) Punjabi nasality constraint rankings after phonotactic learning

*NASOBS » ID[NAS] » SPRD-L[NAS] » IDFIN[NAS], ID[NAS]/_V[-NAS]

Nevertheless, the BMCD process is not complete because the current grammar does not account for unfaithful mappings generated by alternations, such as /sa:-vã:/ → [sã:ũã:] ‘breath-PL’. As the learner is exposed to surface alternations like [sa:] ~ [sã:], it begins to become morphologically aware, and the unlearned forms demonstrating the alternation are added to the support tableau in (10). Here, exposure to the alternation [sa:] ~ [sã:ũã:] provides evidence that the stem vowel in ‘breath’ is underlyingly oral and undergoes predictable nasalization when followed by a contrastive nasal vowel with no intervening non-glide consonants, as in the plural form [sã:ũã:] ‘breaths’. With the addition of /sa:-vã:/, a new round of BMCD ensues to account for the novel data.

(10) Initial support Tableau for morphologically aware learner

Lexicon	cands.	*NASOBS	ID[NAS]	SPRD-L[NAS]	IDFIN[NAS]	ID[NAS]/_V[-NAS]
/sa:/	☞ sa:					
	sã:		*W	*W	*W	
	šã:	*W	*W*		*W	*W
/tʃʰã:/	☞ tʃʰã:			*		
	tʃʰa:		*W	L	*W	
	ṭʃʰã:	*W	*W	L		
/sĩ:n/	☞ sũ:n			*		
	si:n		*W	**W		
	šĩ:n	*W	*W	L		
/sa:-vã:/	☞ sã:ũã:		**	*		
	sa:vã:		L	**W*		
	šã:ũã:	*W	***W	L		*W
	sa:va:		*L	L	*W	

Rather than providing an individual support tableau to demonstrate each step of the BMCD process, the final support tableau after all constraints are ranked

is provided in (11). The ranking process for this round of BMCD is described in what follows. *NASOBS is still the only markedness constraint that does not predict the emergence of a suboptimal candidate, so it is ranked first. As a result, almost half the losing candidates are accounted for. After that, the only remaining markedness constraint — SPRD-L[NAS] — cannot yet be ranked because it favors an unaccounted-for loser ($t_j^h\alpha:$) over a desired winner ($t_j^h\tilde{\alpha}:$). Consequently, the faithfulness constraint that would allow SPRD-L[NAS] to be ranked on the following iteration, IDFIN[NAS], is ranked next, followed by SPRD-L[NAS], which rules out the remaining losing candidates. Finally, the remaining unranked faithfulness constraints, ID[NAS] and ID[NAS]/_V[-NAS], are ranked at the bottom of the constraint hierarchy.

(11) Support Tableau for morphologically aware learner after final pass

Lexicon	cands.	*NASOBS	IDFIN[NAS]	SPRD-L[NAS]	ID[NAS]/_V[-NAS]	ID[NAS]
/sɑ:/	 sɑ:					
	sã:		*W	*W		*W
	šã:	*W	*W		*W	*W*
/t ^h ã:/	 t ^h ã:			*		
	t ^h ɑ:		*W	L		*W
	ṭ ^h ã:	*W		L		*W
/sĩ:n/	 sĩ:n			*		
	si:n			**W		*W
	šĩ:n	*W		L		*W
/sa:-vã:/	 sã:ũã:			*		**
	sa:vã:			**W*		L
	šã:ũã:	*W		L	*W	***W
	sa:va:		*W	L		*L

The updated constraint rankings are provided in (12) and illustrate the constraint reordering required to successfully capture the data when the learner is exposed

to surface alternations.

(12) Punjabi nasality constraint rankings after morphologically aware learning

*NASOBS » IDFIN[NAS] » SPRD-L[NAS] » ID[NAS]/_V[-NAS], ID[NAS]

Crucially, instances of non-alternating pre-N /V:N/ vowels occurring in environments in which contrastive / $\tilde{V}$:/ vowels trigger leftward nasal harmony have not been introduced to the learner (e.g., [ɑ:vũ:m] ‘general public’). If they had, the learner would not have been able to converge on a grammar that accounts for all the data. This is because the learner still assumes faithful mappings from UR to SR unless alternations exist that prove otherwise. For /sɑ:/ ‘breath’, the surface alternation between nasality and orality provides the learner with sufficient evidence to learn a non-faithful UR in the case of /sɑ:-vũ:/ ‘breath-PL’. For forms like [ɑ:vũ:m] ‘general public’, on the other hand, the learner has no exposure to alternations in which the /V:N/ vowel surfaces as oral. Consequently, the learner would incorrectly assume a faithful mapping from UR to SR and cannot learn otherwise. Put differently, unless the learner is encoded with a mechanism that allows it to arrive at the conclusion that non-alternating pre-N /V:N/ vowels are underlyingly oral, the Punjabi nasalization patterns cannot be acquired.

This is where the Free Ride Learning Algorithm comes into play. Because the underlying oral vowel in /sɑ:-vũ:/ surfaces as nasal, the learner will attempt to generalize that unfaithful mapping so that all surface nasal vowels, even those that lack alternations, take the free ride and are derived from underlying oral vowels generally, with the trigger of the mapping being a separate issue. The impetus for the learner’s generalization originates from the bias of BMCD to arrive at the most restrictive grammar possible. A grammar that derives all surface nasal vowels from underlying oral vowels is more restrictive because it generates less potential surface forms by precluding oral vowels from faithfully surfacing.

In Punjabi, using the FRLA principles, the alternation-supported unfaithful mapping ($/\alpha:/ \rightarrow [\tilde{\alpha}:]$) is generalized to all surface nasal vowels and can, in theory, predict the different nasal harmony patterns between $[\alpha:v\tilde{\alpha}:m]$ and $[s\tilde{\alpha}:\tilde{v}\tilde{\alpha}:]$. Using evidence of this unfaithful mapping, the FRLA copies the support tableau in (11), locates all instances of ($/\alpha:/ \rightarrow [\tilde{\alpha}:]$), and changes the underlying form to match the alternation-supported unfaithful mappings ($/\alpha:/ \rightarrow [\tilde{\alpha}:]$). The FRLA then updates the faithfulness evaluations accordingly. In essence, the learner avails itself of the information it has gleaned from the alternation-supported unfaithful mapping in the oral vowel of $/s\alpha:-v\tilde{\alpha}:/ \rightarrow [s\tilde{\alpha}:\tilde{v}\tilde{\alpha}:]$ to apply the unfaithful map to the lexical entries for all surface nasal vowels in order to increase restrictivity.

The step where the FRLA fails is illustrated by the tableau in (13). The fatal error occurs with instances in which vowels are unpredictably nasalized, as in $[t^h\tilde{\alpha}:]$ or $[s\tilde{\alpha}:\tilde{v}\tilde{\alpha}:]$. Given the current constraint set, desired optimal candidates with contrastive nasal vowels like $[t^h\tilde{\alpha}:]$ or $[s\tilde{\alpha}:\tilde{v}\tilde{\alpha}:]$ are harmonically bounded by the faithful oral candidates highlighted in grey in (13). Consequently, observed forms with unpredictable nasal vowels can never be realized regardless of rankings.

- (13) FRLA fails to simultaneously generate faithful $/\alpha:/ \rightarrow [\alpha:]$ and unfaithful $/\alpha:/ \rightarrow [\tilde{\alpha}]$

Lexicon	cands.	*NASOBS	IDFIN[NAS]	ID[NAS]/_V[-NAS]	SPRD-L	ID[NAS]	*ORAL
/sq:/	sq: sq:						*
	sā:		*W		*W	*W	L
	ṣā:	*W	*W	*W		*W*	L
/tʃ ^h α:/	tʃ^hā: tʃ ^h ā:		*		*	*	
	tʃ ^h α:		L		L	L	*W
	ṭʃ ^h ā:	*W	*	*W	L	**W	
/si:n/	sī:n sī:n				*	*	
	si:n				**W	L	*W
	ṣi:n	*W		*W	L	**W	
/sq:va:/	sā:ũā: sā:ũā:		*	*	*	***	
	sq:ũā:		*	L	**W*	*L	*W
	ṣā:ũā:	*W	*	*	L	****W	
	sq:va:		L	L	L	L	*W*
/α:va:m/	α:vā:m α:vā:m				**	*	*
	α:va:m				***W	L	**W
	ṭā:ũā:m			*W	L	**W*	L

Even if an additional constraint that dispreferred oral vowels, *ORAL, were added to the constraint set to resolve the harmonic boundedness of $[tʃ^h\tilde{\alpha}]$ and $[s\tilde{\alpha}:\tilde{\upsilon}\tilde{\alpha}]$, the FRLA could not converge on a grammar. This is because for $/sq:/$ ‘breath’ to be realized faithfully, ID[NAS] would have to outrank *ORAL. In direct contradiction, *ORAL would need to outrank ID[NAS] for $/tʃ^h\alpha:/$ ‘shade’ to be realized unfaithfully. In short, the attempt of the FRLA to arrive at a grammar in which $/V:N/$ vowels are underlyingly oral will never succeed because it is too *restrictive*.

In short, while the Free Ride Learning Algorithm successfully models the acquisition of covert representations when the surface realizations of a covert UR are consistently derivable from that UR across all contexts, the approach fails to acquire covert URs arising from contextual neutralization, as in Punjabi.

3.4 The MDL Evaluation Metric

Another framework with the potential to acquire non-alternating covert URs like /V:N/ vowels in Punjabi is the Minimum Description Length (MDL) metric outlined in Rasin and Katzir (2018). Nevertheless, while the MDL framework supplies an avenue for learning the Punjabi nasalization pattern discussed in this paper, I discuss two unresolved issues with the approach. First, although the MDL metric provides a criterion for evaluating competing grammars, successful acquisition also depends on a procedure for generating and searching those grammars. Rasin and Katzir (2016) implement a concrete search procedure using simulated annealing, but they explicitly make no cognitive claims about this component of the learner. Consequently, while the MDL metric is expected to favor the analytically preferred Punjabi grammar when the relevant alternatives are compared, it remains unclear whether a learner could efficiently and reliably discover that grammar within a substantially broader hypothesis space.

Second, a potentially undesirable assumption surrounding the universal nature of the phonological grammar that must be accepted for the MDL metric to function correctly could present issues in other aspects of learning that are too costly to accept. Specifically, rather than assuming that the universal set of segments is entirely present for speakers of every language so that inputs are unconstrained — i.e., Richness of the Base (Prince and Smolensky, 2004) — the MDL metric restricts this set to segments that exhibit positive evidence for their inclusion. This means that inputs, and by extension, underlying representations, are constrained to be made up of this set of segments.

Rasin and Katzir (2016, 2018, 2020) apply the MDL evaluation metric to a learner

whose grammar comprises a lexicon, a language-specific alphabet (Σ), and a ranked constraint set. Furthermore, the MDL learner simultaneously attempts to (1) maximize grammar simplicity by minimizing the grammar’s length in bits ($|G|$) and increase restrictivity by minimizing $|D:G|$, which is the length of the instructions (in bits) required to generate the data (D) using G . This selection process by the learner is equivalent to maximally reducing the sum of $|G|$ and $|D:G|$ in terms of encoding size. The MDL evaluation metric is provided in (14).

(14) MDL EVALUATION METRIC (Rasin and Katzir, 2018, 3)

If G and G' can both generate the data D , and if $|G|+|D:G| < |G'|+|D:G'|$, prefer G to G' .

What this means for learning covert URs is that, if a grammar that includes covert URs decreases the sum of $|G|+|D:G|$, the covert UR can be learned. Consider the three surface forms in Punjabi — $[s\alpha:]$ ‘breath’, $[s\tilde{\alpha}:-\tilde{u}\tilde{\alpha}:]$ ‘breath-PL’, and $[a:v\tilde{a}:m]$ ‘general public’ — as a test case to examine whether the inclusion of a covert UR is preferred by the MDL metric. The alphabet, lexicon, and constraint set for a complex grammar that only permits surface forms to serve as URs are given in (15). In order to prevent changes from $UR \rightarrow SR$, a generic faithfulness constraint, FAITH, that prohibits differences between the two levels of representation must outrank the relevant markedness constraints in order to generate the nasality pattern in Punjabi. To account for the nasalization differences in the morpheme $s\alpha:-$ when it is and is not followed by a morpheme that includes a contrastive nasal $/\tilde{V}:/$ vowel, the complex grammar must include both alternations of the morpheme in the lexicon. In addition, the glide that precedes the contrastive $/\tilde{V}:/$ vowel in the inflectional plural morpheme $[\tilde{u}\tilde{\alpha}:]$ must be added to Σ . If the data included in the MDL evaluation metric were expanded to include the entire Punjabi lexicon, the same types of additions to the Lexicon and Σ would need to

be considered for every alternating form, resulting in a substantially larger $|G|$. Moreover, by including both $/sa:-/$ and $/sã:-/$ in the Lexicon, $|D:G|$ increases because it predicts both $[sa:ũã:]$ and $[sã:ũã:]$ as potential surface forms.

(15) Complex Grammar of Punjabi $/V:N/$ vowels

- i. $\Sigma = \{s, a:, \tilde{a}:, m, v, \tilde{v}\}$
- ii. Lexicon = $\{sa:-, sã:-, -ũã:, a:vã:m\}$
- iii. CON: FAITH » ID[NAS] » ID[NAS]/_V[-NAS] » SPRD-L[NAS] »
*NASOBS

Alternatively, the inclusion of covert URs into the grammar significantly reduces the combined size of the alphabet and lexicon, as shown in (16). This reduction arises in two ways. First, by analyzing the $/V:N/$ vowel in $[a:vã:m]$ as underlyingly oral, the alternation in vowel nasality for $[sa:]$ vs. $[sã:ũã:]$ can be explained in terms of phonological processes rather than additions to the lexicon. That is, nasal harmony is triggered by the underlying contrastive $/\tilde{V}:/$ vowel in the plural morpheme $/-vã:/$ but not by the derived nasality of the $/V:N/$ vowel in $/a:vã:m/$, as discussed in section 2.4.1. By attributing the alternation to a phonological process, there is no longer a need to include $/sã:-/$ in the lexicon, thereby reducing $|G|$. Second, because nasalized glides are not contrastive in Punjabi, only occurring in the predictable environment before underlying $/\tilde{V}:/$ vowels, $|G|$ can be further reduced by removing $/\tilde{v}/$ from Σ and attributing surface $[\tilde{v}]$ in forms like $[sã:ũã:]$ to the harmony process. As long as the instructions required to implement the constraint ranking to capture the data do not counteract the reduced computational load on the lexicon and alphabet by increasing $|G| + |D:G|$, the grammar will be learned. Given that the complete number of alternations like $/sa:-/$ vs. $/sã:-/$ explained by the harmony process is quite expansive, the resulting reduction in the alphabet and lexicon is likely to outweigh any additional cost

associated with encoding the phonological analysis.

- (16) Simple Grammar of Punjabi /V:N/ vowels with covert URs
- i. $\Sigma = \{s, \alpha:, \tilde{\alpha}:, m, v\}$
 - ii. Lexicon = $\{s\alpha:, -v\tilde{\alpha}:, \alpha:v\alpha:m\}$
 - iii. CON: *NASOBS, ID[NAS]/_V[-NAS] » SPRD-L[NAS] » ID[NAS]
» FAITH

In sum, a grammar of Punjabi nasalization without covert URs for non-alternating /V:N/ vowels is expected to have a substantially greater description length than one containing covert URs, because both Σ and the Lexicon must be enlarged to account for the data.

The comparison above suggests that, when the space of potential grammars is restricted to these two analyses, the MDL metric favors the grammar containing covert URs. It does not, however, demonstrate that a learner would discover this grammar within a broader hypothesis space. As mentioned above, Rasin and Katzir (2016) provide a simulated-annealing procedure for searching grammars that differ in their lexicons, constraints, and rankings, but they present this procedure as an implementation choice rather than a cognitively motivated account of search. In a realistic learning environment, the analytically preferred grammar must be located within a much larger space of grammars differing in their lexicons, alphabets, constraints, and rankings. It therefore remains to be established whether this procedure, or another independently motivated search procedure, could efficiently and reliably recover the proposed Punjabi grammar from a realistic dataset. Thus, although the MDL metric provides a criterion for evaluating competing grammars, it does not by itself resolve the computational problem of efficiently searching the hypothesis space.

In addition to this search space issue, the MDL metric also requires the abandon-

ment of Richness of the Base (RotB, Prince and Smolensky 2004), a theory in which the inventory of potential inputs is universal. Without restricting inputs, and thereby potential URs, to the combination of features that actually surface in a language, the alphabet (Σ) would be identical across potential grammars. Rasin and Katzir (2020) explain in great detail how the consequences of RotB within the MDL learning framework lead to a pathological prediction they refer to as the *Extensional Restriction Problem*. The explanation is quite involved and the details intricate, so I refer the interested reader to Rasin and Katzir (2020). The general issue arises from cases in which a linguistic element, x , never appears in a specific context, C , in a language, thereby precluding cases of x in C in the lexicon. However, RotB requires that potential inputs include x in C , thereby providing potential for novel forms with x in C to enter the lexicon at a later date. Furthermore, assuming constraints are acquired and not universal, the MDL learner has no positive evidence in the lexicon to motivate the added bitlength to include instructions applying a constraint to block x in C . Thus, a learner could fully acquire this hypothetical language without the systematic knowledge that x cannot appear in C . The deleterious result is that the learner would fail to rule out future novel forms with x in C . In sum, assuming RotB within MDL learning leads to a problem in which speakers who systematically judge a form with x in C as illicit have no mechanism to arrive at this judgment. However, precluding x from Σ — jettisoning RotB in the process — avoids the problem entirely.

Still, while restricting the potential symbols used for encoding items in the lexicon is useful in the above case, rejecting a universal inventory of inputs has other unintended consequences, two of which I summarize here. First, including a rich base as a theoretical necessity limits the locus of variation across languages to

constraint reranking, thereby restricting the theoretical apparatus that can influence a grammar to constraints. If, on the other hand, Σ were language-specific, a grammar could be influenced by changing the inventory of inputs and/or changing the ranking of constraints. A byproduct of allowing two levers to exist that both control variation — one through limiting permissible inputs and another through constraint ranking — is a duplication of efforts and a decrease in grammar restrictivity. For example, eliminating coda consonants in a language by restricting inputs to exclude them rather than relying on a constraint like NoCoda is capable of accounting for language patterns that lack codas, but in principle, this grammar does not rule out illicit forms, it merely avoids the problem by ensuring the grammar doesn't encounter them. Furthermore, if both the input constraint and the OT constraint are employed to deal with the issue, this simply results in duplication (cf., Kenstowicz and Kisseberth 1977). Conversely, Richness of the Base requires the grammar to explicitly rule out all potential illicit forms, thereby making explicit restrictions on what forms are grammatical and what forms are not, all while doing so with restrictions on surface forms.

Beyond theoretical restrictivity and the duplication problem, Smolensky (1996) also demonstrates potential learnability issues that arise if Richness of the Base is eschewed. Consider a hypothetical language that only permits CV syllable structure. As Smolensky demonstrates in detail, if the input of this language is restricted to only include CV syllables, an unmarked structure, no evidence exists to define the correct constraint rankings. That is, both NoCoda » FAITH and FAITH » NoCoda effectively result in limiting the language to CV surface syllables; the constraint ranking is irrelevant. Consequently, when a loanword with a coda consonant, CVC, is encountered, the grammar may yield the wrong

result. Moreover, if one argues that a CVC loanword is prevented from entering the input by some type of constraint used to shape the input, this results in the duplication problem mentioned above, where the mechanisms restricting the input overlap with the mechanisms that exist in the grammar to accomplish the same thing.

The debate over the need to embrace a Rich Base is still undecided (cf. Gouskova 2023), but if the potentially adverse effects associated with repudiating the theory are correct, a learning approach to covert URs is needed that does not require its rejection.

3.5 MaxLex

Another approach aimed at alleviating the computational load brought on by permitting covert representations is outlined in O'Hara (2017) with the use of a Maximum Entropy grammar called MaxLex. The basic learning procedure taken by MaxLex is similar to other MaxEnt learning models (e.g. Hayes and Wilson 2008; Pater et al. 2012). Two general stages characterize the process. In the first stage, the learner is oblivious to morphological alternations and paradigmatic relations, and, as a consequence, the identity of underlying representations and mappings from those representations to surface realizations is not considered. Instead, the learner has been confronted with a wealth of linguistic data and focuses on acquiring fluency in language-specific phonotactics, an aspect of the grammar that remains unchanged regardless of what the underlying forms turn out to be. For example, in Punjabi, a learner can arrive at the conclusion that oral vowels are prohibited before nasal consonants (i.e., *VN is highly weighted) even if it is unclear whether the underlying form of /V:N/ vowels is phonologically oral or

nasal. As such, the learner assumes at this stage that all observed surface forms are faithfully mapped from their corresponding UR. The focus of the phonotactic learning stage, then, is acquiring the appropriate constraint weighting relationships to appropriately describe the phonotactics of the language.

In the second stage of learning, MaxLex becomes morphologically aware, understanding that words are constructed from morphemes, and those morphemes sometimes appear in phonologically distinct ways, depending on the context. For instance, during the phonotactic stage, the learner assumes that the URs for [sa:] ‘breath’ and [sã:ũã:] ‘breath-PL’ are identical to their surface forms. In the morphologically aware stage, however, the learner understands that the same morpheme for ‘breath’ occurs in both words and seeks to assign a single UR that can map to both of the observed forms. As such, the learner is confronted with a more complex learning problem in which it must work to determine what combination of constraint weights and underlying forms maximizes the likelihood of observing the data to which it has been exposed.

A crucial aspect of the morphologically aware learning stage that is different for MaxLex than other MaxEnt learners is the way in which abstraction is curtailed in the choice of potential URs. A standard approach in the learning literature in the morphologically aware stage is to simply provide the learner with a finite set of potential URs to entertain, and this is the approach MaxLex takes as well. However, while this approach does not explain how the list of URs will be compiled, organized, and searched efficiently, MaxLex can constrain abstraction in covert URs through a built-in feature-economic component of the algorithm. Specifically, MaxLex is equipped with an L2 Gaussian prior that creates a preference for using constraints active elsewhere in the grammar to account for abstract phonological

patterns rather than altering the weight of novel constraints to accomplish the same task. To increase grammar restrictivity, the prior inherently favors markedness constraints with maximum weight (the upper bound for constraint weights is set at 100 for all simulations in this chapter) and faithfulness constraints with a minimum weight (a lower bound of 0). If, however, the language data fed to the learner indicates that different constraint weights would improve the success of the grammar in modeling the data, these biases can be overcome. Thus, if a faithfulness constraint is given a non-zero weight to model some phonotactic pattern in the first stage of language learning, that same constraint will be preferred over a novel constraint with a zero weight to model another pattern concerning underlying forms, assuming both constraints can account for the observed data equally well. This preference to use the already-active faithfulness constraint falls out from the fact that the MaxLex prior seeks to minimize changes in constraint weights from their initial values. Because of this, O’Hara argues — as will be shown in the following analysis of an $[i] \sim [\emptyset]$ alternation in Klamath — that a segment’s UR is naturally restricted in the degree of abstraction to representations that surface in other positions.

I demonstrate the success of MaxLex in learning covert structure here by examining a $[i] \sim [\emptyset]$ alternation in Klamath verb stems, which, as O’Hara (2017) shows, capitalizes on a faithfulness constraint that is active in another area of the grammar to account for the alternation. Nevertheless, that same learning process used to constrain UR abstraction is unavailable for non-alternating $/V:N/$ vowels in Punjabi because changes in faithfulness constraint weights do not need to differ between minimally abstract and increasingly abstract URs to account for the nasalization pattern. The consequence of this is that using the MaxLex learner,

the surface form of the /V:N/ vowel can be derived equally well from both a restrictedly abstract oral vowel and increasingly abstract alternatives which never surface in Punjabi in any context. In what follows, I first provide an overview of O’Hara’s (2017) explanation of MaxLex’s success in restricting abstraction in cases like Klamath, where an alternation for a given morpheme exists, and a covert UR is needed to accurately analyze the data. I then proceed to show how MaxLex fails to restrict UR abstraction in the case of Punjabi /V:N/ vowels, which completely lack alternations.

3.5.1 MaxLex and Klamath

To demonstrate the effectiveness of MaxLex, O’Hara (2017) analyzes a vowel alternation process in Klamath verbs from Barker (1964), exemplified by the forms in (17). The first verb stem in (17i) illustrates the behavior of a typical verb stem, which is realized identically regardless of whether it is followed by the indicative suffix, /-a/, or the present perfect progressive suffix, /-tk^h/. For the class of stems in (17ii-iv), however, the vowel [i] is present stem-finally when followed by /-tk^h/ but not when followed by /-a/.

(17) [i]~[∅] alternation in Klamath verbs

stem			-a suffix	-tk ^h suffix
i.	wen-	‘wear out’	[wena]	[wentk ^h]
ii.	ʔe:ɰ(i)-	‘is deep’	[ʔe:ɰa]	[ʔe:ɰitk ^h]
iii.	q ^ʔ e:l ^ʔ (i)-	‘act silly’	[q ^ʔ e:l ^ʔ a]	[q ^ʔ e:l ^ʔ itk ^h]
iv.	ntew ^ʔ (i)-	‘break’	[ntew ^ʔ a]	[ntewitk ^h]

The two potential concrete URs for the verb stems in (17ii-iv) include a consonant-final stem (e.g., /ʔe:ɰ-/) with [i]-epenthesis before the /-tk^h/ suffix or an i-final stem (e.g., /ʔe:ɰi-/) with [i]-deletion before /-a/. O’Hara provides compelling evidence to demonstrate that neither concrete UR is acceptable given the phonology

of the language, and I summarize those arguments here, setting aside a number of details.

An [i]-epenthesis analysis of these verbs — in which the UR of the stems in (17) would end in a consonant (e.g., /ʔe:w-/) — falls short for two reasons. First, the standard epenthetic vowel is [a] in Klamath, so it would be difficult to explain why this class of verbs idiosyncratically epenthesizes [i] instead. Second, for certain verbs showing the alternation (e.g., /ntew²(i)-/ in 17iv), the stem-final consonant /w/ exhibits glottalization only when [i] is not present. In Klamath, prevocalic glottal stops (i.e., occurring in the onset) undergo a regular process of coalescence with the preceding consonant, resulting in a glottalized version of that consonant on the surface: [ntew²a]. When glottal stops occur postvocally (i.e., as codas), on the other hand, they delete. As O'Hara notes, if verb stems like the one in (17iv) lacked an underlying stem-final vowel, and [i] was epenthesized before /-tk^h/, the glottal stop would be in the onset position after [i] epenthesis, so we would incorrectly expect to see the glottal stop coalesce with the preceding [w]: *[ntew²itk^h]. The upshot of these facts is that some underlying vowel must be present in the final syllable of stems for the verbs in (17ii-iv), specifically preceding the glottal stop in forms like (17iv), where its presence causes the glottal stop to surface in coda position and therefore delete.

Nevertheless, the additional Klamath verb forms in (18) indicate that the vowel present in the UR is not [i]. Specifically, the verbs in (18) are realized with a stem-final [i] before /-tk^h/, just like the verbs in (17ii-iv), but when these stems are followed by the indicative /-a/, [a] deletes and [i] is maintained on the surface, unlike the verb forms in (17ii-iv). Because [i] always surfaces for the verb forms in (18), it is reasonable to conclude that the vowel is present in the UR. Consequently,

the verb forms exhibiting the [i]~[ø] alternation in Klamath cannot possess an underlying /i/, or else these verbs would delete the [a] in the indicative form rather than the stem-final [i].

(18) Klamath verbs with underlying stem-final /i/

	stem	-a suffix	-tk ^h suffix
i.	stupw _i - ‘has first menstruation’	[stupw _i]	[stupwitk ^h]
ii.	sla:m [?] i- ‘is a widower’	[sla:m [?] i]	[sla:m [?] itk ^h]

Because neither of the potential concrete URs is viable, a covert UR could lead to a satisfactory analysis. O’Hara stipulates a set of two potential covert URs for this purpose. The first potential UR utilizes a stem-final /e/ (e.g., /ʔe:we-/). Because /e/ occurs in word-initial syllables in Klamath, the faithfulness constraint preserving its height (ID[HIGH]) must have a non-zero weight in the language. However, [e] never surfaces in the context in which the [i]~[ø] alternation occurs, so it is a covert UR in the context of interest. Moreover, /e/ is a minimally abstract UR because it surfaces in initial syllables in Klamath and only differs from the surface form [i] by a single feature. Conversely, positing a UR like /ɪ/ for these verb stems (e.g., /ʔe:wi-/) is more abstract than /e/ because it never surfaces anywhere in Klamath. O’Hara demonstrates that both /e/ and /ɪ/ model the [i]~[ø] alternation in Klamath equally well, but they require slightly different constraint weights to do so.

3.5.1.1 Phonotactic Learning in Klamath

The constraints necessary for deciding between the two potential covert URs are given in (19). Most of the constraints are quite standard, but PHONOTAC in (19vi) is used by O’Hara as a utilitarian constraint that penalizes illicit consonant clusters like [Ctk^h]. ID[HIGH] and ID[HIGH]_{σ1} are necessary for an analysis with

/e/ as the covert UR, and ID[ATR] is included to consider whether the covert UR of the troublesome alternation is /ɪ/.

(19) OT Constraints for analysis of Klamath [i]~[ø] alternation

- i. ID[HIGH]
For every segment, *A*, assign a violation if the output value for the [HIGH] feature of *A* does not match the input value for the [HIGH] feature of *A*.
- ii. ID[HIGH]_{σ1}
For every segment in the initial syllable of a prosodic word, *A*, assign a violation if the output value for the [HIGH] feature of *A* does not match the input value for the [HIGH] feature of *A*.
- iii. ID[ATR]
For every segment, *A*, assign a violation if the output value for the (ATR) feature of *A* does not match the input value for the (ATR) feature of *A*.
- iv. MAXV
For every vowel, *A*, assign a violation if there is not a corresponding
- v. *MID
Assign a violation for every mid-vowel that occurs in the output.
- vi. PHONOTAC
Assign a violation for word-final [Ctk^h] clusters.
- vii. *_I
Assign a violation for every [ɪ] vowel that occurs in the output.

The three weighting relationships provided in (20) must be true to model the phonotactics of Klamath, regardless of whether the covert UR is /e/ or /ɪ/. Because [e] is prohibited in non-initial syllables in Klamath, *MID must outweigh both MAXV and ID[HIGH]. This ensures non-initial [e] either deletes or raises when deletion would generate a phonotactically impermissible cluster. Additionally, because [e] surfaces faithfully in the initial syllable, the sum of the weights for ID[HIGH] and ID[HIGH]_{σ1} must outweigh *MID.⁹ This is a crucial weighting

⁹Note that [e] could also surface faithfully in the initial syllable if ID[HIGH]_{σ1} by itself outweighed *MID. Nevertheless, O'Hara (2017) outlines in detail two reasons why the L2 Gaussian prior prefers to spread the work across both the general faithfulness constraint (ID[HIGH]) and the context-specific faithfulness constraint (ID[HIGH]_{σ1}). In a nutshell, models like MaxLex would have difficulty increasing the weight of the specific faithfulness constraint without also increasing the weight of the general faithfulness constraint (which exhibits a superset of the violations).

that must be true regardless of which UR is chosen and is a major factor for why MaxLex prefers /e/ over /ɪ/, as will be discussed below. Specifically, ID[HIGH] must be active (i.e., possess a non-zero weight) in Klamath to enable /e/ to surface faithfully in the initial syllable. Finally, because /ɪ/ never surfaces anywhere in Klamath, *_I is undominated and the relevant faithfulness constraint, ID[ATR], is not necessarily active in the language and is consequently driven to a zero weight to increase grammar restrictivity. O’Hara demonstrates that these necessary weight relationships are effectively learned by MaxLex in the Phonotactic learning stage. Again, these weights must hold true regardless of which UR is ultimately acquired by the algorithm.

(20) Necessary weight relationships to model Klamath phonotactics

- i. $w(*\text{MID}) > w(\text{ID}[\text{HIGH}]), w(\text{MAXV})$
- ii. $w(\text{ID}[\text{HIGH}]) + w(\text{ID}[\text{HIGH}]_{\sigma 1}) > w(*\text{MID})$
- iii. $w(*\text{I}) > w(\text{ID}[\text{ATR}])$

3.5.1.2 Morphologically Aware Learning in Klamath

During the second learning stage, in which the learner becomes morphologically aware, further refinement of the constraint weights to learn the Klamath alternation pattern is dependent on the choice of UR. For /e/ to be chosen as the optimal UR and model the data effectively, the two additional weighting relationships in (21) must be learned. First, to compel /e/ to delete in non-initial syllables when phonotactically permissible (e.g., /ʔe:we-a/ → [ʔe:wa]), ID[HIGH] must outweigh MAXV. Otherwise, MAXV would prevent deletion and cause /e/ to surface as [i] instead. Second, because /e/ raises to [i] when deletion would generate an il-

Additionally, because the prior uses the sum of squares of weights to minimize distances between the initial and final weights of constraints, spreading a weight increase across two constraints will decrease the penalty compared to a single constraint increasing alone (e.g., for most values, $(\frac{a}{2})^2 + (\frac{a}{2})^2 < a^2$).

licit consonant cluster (e.g., /ʔe:we-tk^h/ → [ʔe:witk^h]), the sum of the weights of MAXV and PHONOTAC must outweigh ID[HIGH].

(21) Necessary weight relationships for /e/ to be covert UR

- i. $w(\text{ID}[\text{HIGH}]) > w(\text{MAXV})$
- ii. $w(\text{MAXV}) + w(\text{PHONOTAC}) > w(\text{ID}[\text{HIGH}])$

Conversely, for /I/ to be chosen as the correct UR and model the data correctly, the additional constraint weights in (22) must be learned. When ID[ATR] outweighs MAXV, /I/ deletes when phonotactically licit (e.g., /ʔe:wɪ-a/ → [ʔe:wɑ]). Furthermore, for /I/ to become [+ATR] and surface as [i], the sum of the weights for MAXV and PHONOTAC must together outweigh ID[ATR].

(22) Necessary weight relationships for /I/ to be covert UR

- i. $w(\text{ID}[\text{ATR}]) > w(\text{MAXV})$
- ii. $w(\text{MAXV}) + w(\text{PHONOTAC}) > w(\text{ID}[\text{ATR}])$

Both sets of weighting relationships in (21) and (22) are learnable and can be acquired simultaneously without contradiction. Using a standard MaxEnt model, a learning problem arises in which the learner has no way of discriminating between the two URs and thus cannot appropriately restrict UR abstraction in preferring /e/ over /I/.

This is where the preference of MaxLex in its L2 Gaussian prior to use already-active faithfulness constraints comes in. A key difference between the required weighting relationships in (21) and (22) is the use of constraints that are already active to model the phonotactics of the language. In the phonotactics stage, the weight of ID[HIGH] must combine with ID[HIGH]_{σ1} to outweigh *MID, as shown in (20ii). The result of this is that ID[HIGH] already possesses a non-zero weight entering the morphologically aware learning stage. Thus, the net change in weight required for ID[HIGH] to outweigh MAXV in (21i) is less than if ID[HIGH] possessed

a zero weight. The same is not true of ID[ATR]. It is completely inactive in the phonotactics of Klamath and is thereby assigned a zero weight during the phonotactic stage. For /ɪ/ to delete in phonotactically licit conditions, then, ID[ATR] must increase to outweigh MAXV. Decisively, this increase is greater in magnitude than the increase in weight for ID[HIGH] required for /e/ to delete in the same context.

For example, if ID[HIGH] was already weighted at 45 to account for the surfacing of syllable-initial /e/, and only needed to increase to 50 to model the [i]~[ø] alternation, the magnitude is only 5. Conversely, since ID[ATR] is inactive in the phonotactics of the language, it would need to increase by a magnitude of 50 to achieve the same results.

MaxLex capitalizes on this difference in magnitude changes of the constraint weights because of the components of its learning algorithm in the morphologically aware stage. The objective function for the morphologically aware learning stage in MaxLex is constructed from the negative log-likelihood of the data plus the value of the MaxLex L2 Gaussian Prior.

Crucially, if the negative log likelihood were the only component of the objective function, it would not be able to discriminate between a restrictedly abstract UR that already appears in the language like /e/ and increasingly abstract URs like /ɪ/ which do not appear in the language because both URs equally violate their respective faithfulness constraints. This is because both URs are equally effective at maximizing the probability of the observed data, assuming equality of different faithfulness violations.

With the addition of the L2 Gaussian prior, MaxLex prefers to minimize faithful

constraints and maximize markedness constraints. As shown in (23), for each constraint, the prior is calculated by squaring the difference between the weight of the constraint needed to model the data (w_i) and the weight of the bias for that constraint (c_i) and dividing it by a scaling factor (σ_i^2). The bias pushes faithfulness constraints to 0 and markedness constraints to 100, which increases the restrictivity of the grammar. The central takeaway of including the prior in the calculation of the objective value is that the more constraint weights vary from their biased weight, the greater the penalty on the objective function.

(23) MaxLex Objective Function in Morphologically Aware Stage

$$\mathcal{O}_{Lex} = \text{NLL} + \underbrace{\sum_{w_i \in \mathbf{w}} \frac{(w_i - c_i)^2}{\sigma_i^2}}_{\text{L2 Gaussian prior}}$$

where:

w_i = weight of i th constraint

c_i = bias of i th constraint (0 for faithfulness, 100 for markedness)

σ_i^2 = scaling factor to increase/decrease prior importance

A consequence of including the prior in the objective function is that MaxLex will inherently prefer /e/ over /ɪ/ when choosing a covert UR for the relevant verb forms. This falls out from the fact that the weights of the two faithfulness constraints that the learner must change have different initial weights entering the morphologically aware learning stage. Because Klamath utilizes vowel height contrasts like /e/ vs. /i/, ID[HIGH] possesses a non-zero weight. On the other hand, ID[ATR] possesses a zero weight because Klamath exhibits *no* (surface) ATR contrasts. Consequently, the penalty added to the objective function by the prior will be larger when changing the weight of ID[ATR] from zero to outweigh MAXV than when changing the weight of ID[HIGH] from a non-zero number to

outweigh MAXV. That is, for /ɪ/ to be the preferred UR, not only would ID[HIGH] need to be non-zero to account for the surface realization of /e/ in initial syllables, but ID[ATR] would also need to increase to model the [i]~[ø] alternation. Since only having a single faithfulness constraint deviate from the faithfulness bias of 0 is less costly, /e/ will be preferred by way of the L2 prior.

As O’Hara delineates in detail in his computational proof, MaxLex takes advantage of these faithfulness constraint weight differences when deciding upon the optimal covert UR. Though both URs model the data equally well, /e/ is preferred by MaxLex because using it as the UR for these verb forms requires only a single faithfulness constraint to deviate from the bias, thereby comparatively minimizing the L2 prior. In other words, it minimizes the penalty introduced by the prior compared to the penalty introduced by using /ɪ/ as the covert UR. Thus, when two covert URs model a set of data equally well, the MaxLex L2 prior has an emergent preference for the covert UR that uses faithfulness constraints already deviating from their bias.

3.5.2 MaxLex and Punjabi /V:N/ vowels

Nevertheless, when applied to the Punjabi nasalization data presented in this paper, MaxLex fails to prefer the restrictedly abstract UR over increasingly abstract alternatives. The ensuing discussion highlights this point. As a reminder, long vowels in Punjabi contrast in nasality, but this contrast is neutralized before a nasal consonant (e.g., [tɑ:] ‘warmth’ vs. [tã:] ‘that’ but [tã:n] ‘melody’ vs. *[tɑ:n]). Additionally, Punjabi exhibits a process of nasal harmony, in which contrastive nasal / $\tilde{V}$:/ vowels trigger the leftward spread of nasalization, with glides and vowels participating and other consonants acting as blockers, as shown

in (24i). /V:N/ vowels, on the other hand, surface as categorically nasalized and phonetically indistinguishable from contrastive nasal / $\tilde{V}$:/ vowels, but they do not trigger nasal harmony (24ii).

(24) Nasal Harmony in Punjabi

- | | | | | |
|-----|-----------|---|-----------|-------------|
| i. | /sa:-vã:/ | → | [sã:ũã:] | ‘breath-PL’ |
| ii. | /ta:va:n/ | → | [ta:vã:n] | ‘penalty’ |

In attempting to learn the nasalization patterns in Punjabi, MaxLex begins in the same way as it did for Klamath, with an initial phonotactic learning stage. The observed data fed to the learner is given in (25). The forms in (25i-iv) show that underlyingly oral vowels are nasalized via nasal harmony when the appropriate suffix is attached (in this case, the plural marker). The forms in (25v-viii) are show the learner that a nasality contrast exists for vowels. Finally, the forms in (25ix-x) provide the learner with examples of the underapplication of nasal harmony for non-alternating /V:N/ vowels.

(25) Punjabi surface forms fed to MaxLex

- | | | | | | |
|------|-----------|-------------|-------|------------|--------------|
| i. | [sa:] | ‘breath’ | ii. | [ʊfɑ:] | ‘morning’ |
| iii. | [sã:ũã:] | ‘breath-PL’ | iv. | [ʊfã:ũã:] | ‘morning-PL’ |
| v. | [gã:] | ‘cow’ | vi. | [tʃʰã:] | ‘shade’ |
| vii. | [gã:ũã:] | ‘cow-PL’ | viii. | [tʃʰã:ũã:] | ‘shade-PL’ |
| ix. | [ta:vã:n] | ‘penalty’ | x. | [prəvã:n] | ‘accepted’ |

Individual Python scripts were developed for the phonotactic learning stage and morphologically aware learning stage in this section to carry out the computational optimizations. These scripts are available at https://github.com/joncpamore/Learning_Covert_URs. The constraints used in the computational learner are provided in Table 3.6 with the initial weights set at 50, along with the weights acquired in the phonotactic learning stage in the rightmost column. Most of the constraints are standard, but a few important factors merit discussion. First, as noted above, MaxLex is equipped with a bias that pushes faithfulness constraints

to 0 and markedness constraints to 100, unless evidence in the data strongly suggests otherwise. This bias is included to maximize the grammar’s restrictivity.

Constraints	Constraint type	initial w	final w
ID[NAS]	faithfulness	50.00	51.37
IDFIN[NAS]	faithfulness	50.00	44.83
SPRD-L[NAS]	markedness	50.00	92.83
*NASOBS	markedness	50.00	100.00
*NASG	markedness	50.00	99.48
ID[NAS]/_V[-NAS]	context. faith.	50.00	100.00
*VN	markedness	50.00	100.00
ID[RD]	faithfulness	50.00	0.00
*LOWRD	markedness	50.00	100.00

Table 3.6: Constraint weights after phonotactic learning with MaxLex

Another important note is the inclusion of ID[RD] and *LOWRD in the constraint set. To test whether MaxLex selects minimal abstraction for non-alternating pre-N /V:N/ morphemes in Punjabi, I provide the learner with two potential covert URs to choose between. The restrictedly abstract and intuitively most appealing covert UR for a non-alternating /V:N/ vowel like [ã:] in [ta:vã:n] is /a:/. /a:/ possesses an identical feature set to [ã:] except for one disparity: nasality. Because nasality is the key underlying feature that results in distinct harmony patterns for pre-N /V:N/ and contrastive nasal / $\tilde{V}$:/ vowels, it makes sense for nasality to be the only feature that changes between the UR and SR of /V:N/ vowels. With that said, MaxEnt learning algorithms like MaxLex do not contain an inherent mechanism to act upon this sensible conclusion. Instead, the learner is free to choose any covert UR that models the data and minimizes changes in constraint weights from their biases, regardless of whether there are one or fifty feature disparities in the UR→SR mapping. To simplify the discussion, I provide MaxLex with

one additional potential covert UR, /ɒ:/ . Just like its unrounded counterpart /ɑ:/ , the low round back vowel /ɒ:/ is quite similar to its corresponding SR, [ã:], except it contains *two* disparities rather than one: nasality and roundedness. Importantly, any increasingly abstract UR (e.g., diacritics) would suffice in the following discussion, but /ɒ:/ is an especially good candidate because it is *more* abstract than /ɑ:/ (/ɒ:/ never surfaces in Punjabi and has more disparities in the input-output mapping) but only minimally so. Thus, /ɒ:/ serves as a stand-in for any overly abstract covert UR that needs to be ruled out, and if /ɒ:/ is ruled out, potential URs with greater disparities are ruled out along with it.¹⁰

The weights acquired in the phonotactic learning stage of MaxLex demonstrate three phonotactic restrictions in Punjabi that must hold regardless of the particular UR chosen for pre-N /V:N/ vowels. First, low round vowels never surface in Punjabi, so *LOWRD is undominated and ID[RD] is inactive and set to 0. As shown by the tableau in (26), this weighting relationship appropriately unrounds all inputs with low round /ɒ:/ with a probability of 1.0.

(26) Low round vowels never surface

/sɒ:/		*LOWRD 100.00	ID[RD] 0.00	$\mathcal{H}$	$\tilde{\mathcal{P}}$
a.	↗ sɑ:		-1	0	1.0
b.	sɒ:	-1		-100	$4e^{-44}$

Another phonotactic restriction MaxLex acquires is the absolute ban on nasal obstruents in Punjabi. To accomplish this, *NASOBS must outweigh SPRD-L[NAS], as in (27).

¹⁰Note that a covert UR like the *nasalized* low back round vowel /õ:/ only has a single disparity in its map to [ã:] (roundedness), so it is equivalent to /ɑ:/ in its disparity differences. However, just like the concrete UR /ã:/ fails to model the lack of harmony triggered by VN vowels in Punjabi, any other nasal vowel runs into the same issue.

Constraints	Constraint type	initial w	final w
ID[NAS]	faithfulness	51.37	3.36
IDFIN[NAS]	faithfulness	44.83	99.96
SPRD-L[NAS]	markedness	92.83	5.65
*NASOBS	markedness	100.00	100.00
*NASG	markedness	99.48	0.19
ID[NAS]/_V[-NAS]	context. faith.	100.00	100.00
*VN	markedness	100.00	100.00
ID[RD]	faithfulness	0.00	0.00
*LOWRD	markedness	100.00	100.00

UR	$\mathcal{P}$
/tɑ:vã:n/	1.0

Table 3.7: Constraint weights and UR probabilities with concrete URs only

However, using only concrete URs prevents the morphologically aware learner from successfully acquiring the appropriate constraint weights and results in an almost zero probability of learning the correct nasalization pattern of forms with /V:N/ vowels. This is exemplified by the tableau in (29). Because the URs for both non-alternating /V:N/ and contrastive nasal / $\tilde{V}$:/ vowels are identical, MaxLex cannot correctly learn the pattern. When presented with /tɑ:vã:n/, the learner incorrectly assigns almost all the probability to the candidate that exhibits nasal harmony.

(29) Failure of Concrete URs to model Punjabi nasalization

/tɑ:vã:n/	*VN 100.00	SPRD-L[NAS] 5.65	*NASG 0.19	ID[NAS] 3.36	$\mathcal{H}$	$\tilde{\mathcal{P}}$
a. tɑ:vã:n		-3			-16.95	0.012
b. tɑ:va:n	-1	-4		-1	-125.96	$6e^{-50}$
c. tã:ũã:n		-1	-1	-2	-12.56	0.988

Up to this point, the learning process has followed the same general pattern as the Klamath data. The phonotactic patterns were learned, and using a concrete UR for /V:N/ vowels resulted in a failure to accurately predict the observed data.

Now, just as for Klamath, if a small set of covert URs is specified, MaxLex can successfully model the data. The results of the morphologically aware learning stage with concrete / $\tilde{\alpha}$:/, analytically preferred oral / α :/, and increasingly abstract / $\mathfrak{v}$:/ included as potential URs for forms with / $V:N$ / vowels are provided in Table 3.8. Here, the final constraint weights are quite similar to the weights when concrete URs were the only potential option, but the inclusion of the covert representations as potential URs for forms with / $V:N$ / vowels allows MaxLex to accurately model the data, with a .98 total probability of observing the correct surface forms for all words fed to the learner.

However, while MaxLex is successful in modeling the data with the inclusion of these two covert URs, it is unsuccessful in discriminating between them, instead assigning an equal 0.5 probability to both covert URs. In other words, the MaxLex prior could not distinguish between a restrictedly abstract single-disparity UR like / α :/ and a more abstract multi-disparity UR like / $\mathfrak{v}$:. The reason for this is that changes in constraint weights from the phonotactic to the morphologically-aware learning stage are identical regardless of which covert UR is used. To permit the nasal harmony pattern in forms with contrastive / $\tilde{V}$:/ vowels like [s $\tilde{\alpha}$: $\tilde{u}$ $\tilde{\alpha}$:] ‘breaths’, the weight of both ID[NAS] and *NASG need to lower so that their combined sum is less than the weight of SPRD-L[NAS]. This change holds regardless of whether the UR for the / $V:N$ / vowel in [t α : $\tilde{u}$ $\tilde{\alpha}$:n] is / α :/ or / $\mathfrak{v}$:. Additionally, ID[RD] — the faithfulness constraint associated with the increasingly abstract UR, / $\mathfrak{v}$:/ — remains at 0 without any pressure to increase. This is different from the Klamath data because the [i]-[$\emptyset$] alternation demonstrates that some faithfulness constraint must outweigh MAXV to prevent /e/ from simply deleting in all non-initial syllables. Here, however, no alternation exists for morpheme-internal / $V:N$ /

vowels, so faithfulness constraints are not driving their surface realization. In cases like Punjabi, then, when an alternation does not exist but a covert UR is still needed, the MaxLex prior fails to restrict abstraction because minimally abstract URs like /ɑ:/ and increasingly abstract URs like /ɒ:/ do not require distinct constraint weights to accurately model the data.

Constraints	Constraint type	initial w	final w
ID[NAS]	faithfulness	51.37	0.07
IDFIN[NAS]	faithfulness	44.83	100.00
SPRD-L[NAS]	markedness	92.83	5.42
*NASOBS	markedness	100.00	100.00
*NASG	markedness	99.48	0.02
ID[NAS]/_V[-NAS]	context. faith.	100.00	100.00
*VN	markedness	100.00	100.00
ID[RD]	faithfulness	0.00	0.00
*LOWRD	markedness	100.00	100.00

<i>UR</i>	$\mathcal{P}$
/tɑ:vɑ:n/	0.5
/tɑ:vɒ:n/	0.5
/tɑ:vã:n/	0.0

Table 3.8: Constraint weights and UR probabilities with abstract URs included

Moreover, setting aside the problem of appropriately minimizing abstraction in cases like Punjabi, the MaxLex learning model does not address the second issue that arises when covert representations are permitted among the set of potential URs for morphemes. Specifically, allowing abstraction introduces a search space problem: the learner is faced with a potentially unbounded set of candidate URs, yet lacks the resources to exhaustively evaluate this space in order to identify the optimal representation. Because each additional abstract feature or segmental distinction can generate more logically possible URs, the number of candidates expands combinatorially as the learner permits increasingly abstract representations. As a result, exhaustively evaluating every potential UR rapidly becomes computationally intractable. Taken together, these considerations show that re-

lying on an L2 prior that favors grammar restrictivity is insufficient on its own as a learning solution for covert URs.

3.6 Discussion & Future Research

This chapter introduced the UR Progressive Hypothesis Space Expansion (UR PHaSE) learner, a learning architecture designed to reconcile the empirical need for abstract URs with the learnability concerns raised by permitting increased UR abstraction. Rather than evaluating an effectively unbounded UR hypothesis space in parallel, the learner structures that space by disparity count and searches it serially, beginning with fully concrete (0-disparity) URs and expanding to include more abstract candidates only when the current level fails to meet a Minimum Likelihood Threshold. I applied the UR PHaSE learner to the Pakistani Punjabi vowel nasality patterns outlined in chapter two, which require abstract URs to account for pre-N vowels. I showed that 0-disparity EM learning predictably fails to acquire the Punjabi nasality patterns because concrete URs cannot simultaneously induce harmony in contrastive-nasal contexts while suppressing it in pre-N contexts. The learner then selectively expands the UR candidate spaces using information already present in the learned grammar to generate only those higher-disparity URs that are plausibly informative, keeping hypothesis space growth tractable. More specifically, when learning fails at lower levels of abstraction, the learner uses the constraints that highlight uncertainty/inconsistency in the grammar (i.e., constraints with non-maximal weights) to collect a set of features that could be changed in the representations to resolve this uncertainty. With this targeted expansion, the learner successfully converges on the minimally abstract URs required by the Punjabi pattern, thereby

demonstrating that restrictedly abstract URs are learnable when hypothesis space expansion is progressive, structured, and threshold-governed.

In addition, I surveyed current learning models proposed in the literature to model UR abstraction. In each case, I showed that the model either fails to restrict unnecessary abstraction, structure learning for efficient search of the UR candidate space for an optimal learning solution, or both.

Wang and Hayes (2025) argue that permitting highly abstract URs can render morphophonemic learning intractable, and they therefore propose that abstraction should be constrained to the level at which morphophonemic learning is feasible. In their view, the boundary for permissible abstraction is set by the learner’s ability to successfully search the resulting hypothesis space: once candidate UR sets become sufficiently large or complex, convergence failures and computational complexity place an effective ceiling on abstraction.

The UR PHaSE learner developed in this chapter, however, shows that constraints on UR abstraction need not be imposed as an external stipulation. Building on the conception of structuring UR hypothesis spaces by disparity count proposed by Tesar (2014, 2016), the proposed learner avoids evaluating an unbounded set of candidates in parallel by generating and considering URs in batches, based on their disparity distance. This architecture naturally restricts abstraction because minimally abstract URs are always preferred if they are sufficient. Moreover, more abstract URs are introduced only when the learner has explicit evidence that additional abstraction is required. At the same time, the learner addresses search space concerns presented by UR abstraction by expanding the hypothesis space in a targeted way, using markedness violation profiles and constraint weight information to restrict which features can be flipped when generating new

UR candidates, only selecting features associated with constraints that reveal uncertainty. Non-maximally weighted constraints that are violated by at least one observed form reveal uncertainty because the data exerts competing pressures on those constraints. On the one hand, violations of the constraint push its weight downward (e.g., pre-N forms violating $\text{SPRD-L}[\text{NAS}]$); on the other hand, alternations that appear to be governed by the same constraint push its weight upward (e.g., vowels that are otherwise oral becoming nasalized before suffixes with nasal vowels). As the Punjabi nasality data shows, this combination allows the learner to arrive at the highly abstract URs needed for pre-N vowels without incurring the search problems that unrestricted abstraction introduces.

Several directions for future work follow from these results. First, the present chapter demonstrates success on a single pattern that motivates UR abstraction, Pakistani Punjabi vowel nasality. The learner ought to be tested on additional case studies that have been argued to require abstract URs (e.g., O’Hara 2017 and Wang and Hayes 2025). Second, because the UR PHaSE learner does not impose an explicit upper bound on abstraction, it predicts that if a pattern truly requires 2-disparity or 3-disparity URs, the learner will eventually incorporate those URs into the candidate space. At the same time, after the learner has climbed deep enough into the disparity ladder, the UR candidate space may still grow too large, ultimately yielding prohibitive compute time or convergence failures. If so, then we should not expect to find stable phonological systems in which very high degrees of $\text{UR} \rightarrow \text{SR}$ mapping disparities are analytically necessary. Exploring where this boundary lies could clarify how far abstraction can plausibly extend under realistic learning constraints and why, over time, learners tend to reanalyze highly abstract URs as concrete (Kiparsky, 1973; Kuo, 2024a,b).

In sum, the UR PHaSE learner offers a principled way to reconcile abstract UR analyses with learnability. By structuring the UR hypothesis space and searching it serially, the UR PHaSE learner both restricts abstraction to the extent necessary and preserves tractability, thereby removing the need for an externally imposed ceiling on UR abstraction.

Appendix

Category	Surface Forms					
CV	[sa:]	[do:]	[to:]	[pe:]	[ge:]	[si:]
CV-G $\tilde{V}$	[sã:ũã:]	[dõ:ũã:]	[tõ:ũã:]	[pẽ:ũã:]	[gẽ:ũã:]	[sĩ:ũã:]
C $\tilde{V}$	[gã:]	[tʃʰã:]	[kõ:]	[ʒã:]	[kã:]	[põ:]
C $\tilde{V}$ -G $\tilde{V}$	[gã:ũã:]	[tʃʰã:ũã:]	[kõ:ũã:]	[ʒã:ũã:]	[kã:ũã:]	[põ:ũã:]
CVG $\tilde{V}$ N	[ta:ʋã:n]	[prəʋã:n]	[te:ja:n]	[ʒəʋã:n]	[a:ʋã:m]	[si:ja:n]

Table 3.9: Surface forms provided to the UR PHaSE Learner.

(30) Constraint Definitions

- i. SPRD-L[NAS] (cf. Walker, 2003, p.47)
For every occurrence of a [+NAS] feature in a prosodic word, if that [+NAS] feature is dominated by some segment, assign a violation for every segment to the left of that segment in the prosodic word that does not dominate the [+NAS] feature.
- ii. *NASOBS (Walker, 2003, p.51)
Assign a violation for every obstruent that dominates a [+NAS] feature.
- iii. *NASG (Walker, 2003, p.51)
Assign a violation for every glide that dominates a [+NAS] feature.
- iv. *NASV (Walker, 2003, p.51)
Assign a violation for every vowel that dominates a [+NAS] feature.
- v. *VN
Assign a violation for every vowel that dominates a [-NAS] feature when directly preceding a nasal consonant.

- vi. ID[NAS]
For every segment, A , assign a violation if the output value for the [NAS] feature dominated by A does not match the input value for the [NAS] feature dominated by A .
- vii. IDFIN[NAS]
For every segment, A , assign a violation if the output value for the [NAS] feature dominated by A does not match the input value for the [NAS] feature dominated by A in the final syllable of a prosodic word.
- viii. ID[NAS]/_V[-NAS]
Let A be a segment that occurs before an oral vowel, $_{_}V$, in the input. Assign one violation if the output correspondent of A does not have the same specifications for [NAS] as A .
- ix. ID[RD]
For every segment, A , assign a violation if the output value for the [RD] feature dominated by A does not match the input value for the [RD] feature dominated by A .
- x. *LOWRD
Assign a violation for every vowel that dominates a [RD] feature and a [LOW] feature simultaneously.

Chapter 4

Phonetically Complete

Neutralization in Mankiyali

A fundamental premise of eliminativist usage-based models of speech production is that no formal grammar is necessary to explain the systematic application of phonological processes across a language's lexicon (Bybee, 2001, p.6-8). Instead, these processes are said to *emerge* over time due to the high degree of practiced repetition of specific patterns across episodic memories/exemplars in the lexicon. In other words, grammatical structure is embedded in the lexical representations themselves, and the permeation of a specific pattern across exemplars is argued to arise from frequency effects and results in the strengthening of this pattern over time and the subsequent *entrenchment* of that pattern in the lexicon through usage. Moreover, as novel forms are acquired, these highly practiced speech acts emerge naturally in the same phonetic contexts in which they occur in already-acquired forms. In short, the notion of a formal grammar through which underlying representations are filtered to arrive at target surface forms is rejected

wholesale under these types of usage-based models.¹

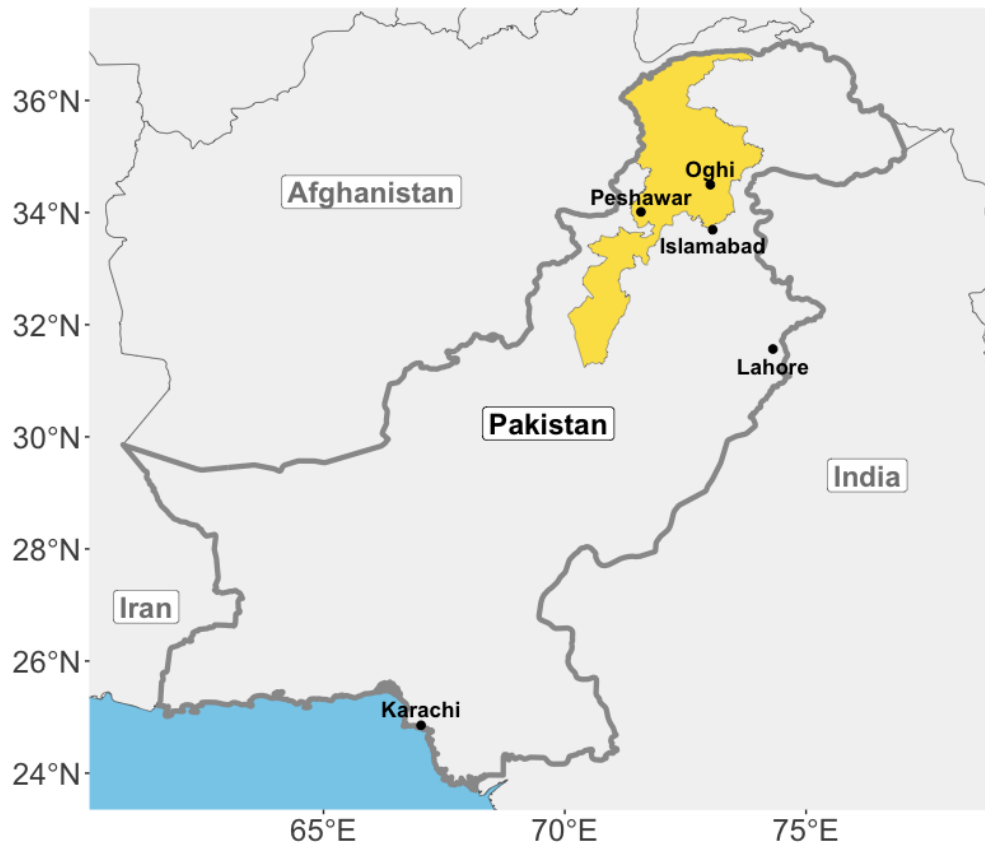
In this chapter, I provide evidence for a case of phonetically complete neutralization of vowel nasality in Mankiyali, which challenges the premise that repetition, frequency effects, and entrenchment alone are sufficient to capture the observed phonetic patterns in the lexicon. More specifically, I argue that the existence of processes that exhibit phonetically complete neutralization of a sound contrast substantiates the presence of linguistic knowledge external to lexical representations themselves. The main goal of the chapter is to provide empirical evidence for a case of phonetically complete neutralization of vowel nasality in Mankiyali and provide a conceptual explanation for why eliminativist usage-based models struggle to account for it. In chapter 5, I build upon the empirical findings outlined here, discussing simulations of exemplar memory accumulation over time for a single speaker to show that eliminativist usage-based models — which *eliminate* computation over representations to generate production targets — face difficulties in explaining the occurrence of both phonetically complete and phonetically incomplete neutralization.

4.1 Mankiyali

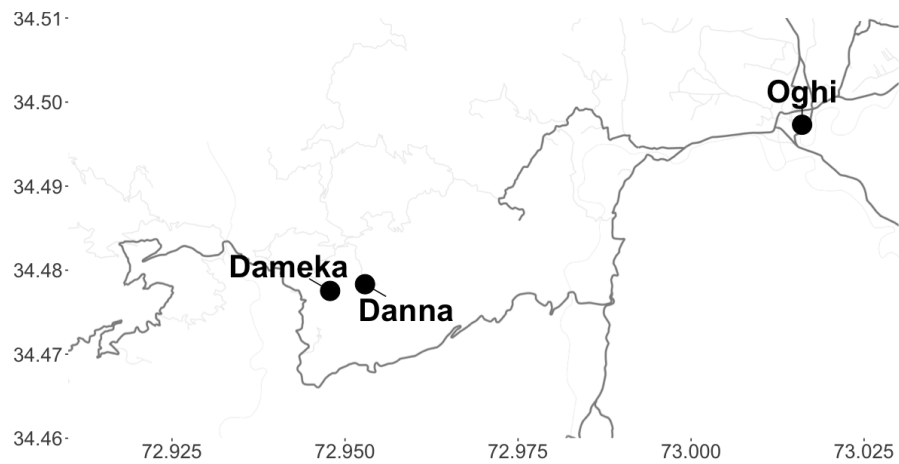
Mankiyali is an endangered Indo-Aryan language spoken by about 500 people in Mansehra district, Khyber Pakhtunkhwa Province, Pakistan. As shown by the maps in Figure 4.1, Khyber Pakhtunkhwa Province, highlighted in yellow, is situated in the northwest region of Pakistan. Mankiyali is spoken in the villages of Danna and Dameka, located about 5 km outside Oghi Tehsil on an iso-

¹These models do, however, require some higher-order categorization strategy that facilitates relations between instances of the same sound/word/phrase across exemplar memories. How the notion of ‘sameness’ is determined under these models is not entirely clear.

lated mountain. All Mankiyali speakers are at least trilingual, frequently using Hindko (Northwest Indo-Aryan) — the dominant language of Mansehra — and Urdu (Central Indo-Aryan) — the national language of Pakistan — in addition to Mankiyali. Many are also fluent in Pakistani Pashto (Eastern Iranian), and a small number of men speak English and/or Gojri (Central Indo-Aryan). While exhibiting many features common to Indo-Aryan languages, Mankiyali is mutually unintelligible with any other language.



(a) Pakistan and Khyber Pakhtunkhwa Province



(b) Villages of Danna and Dameka, where Mankiyali is spoken

Figure 4.1: Maps showing the location of Mankiyali-speaking villages in Pakistan. The yellow region is Khyber Pakhtunkhwa Province. Mankiyali is spoken in two small villages near Oghi Tehsil in Mansehra District, Danna and Dameka.

Until recently, no formal writing system existed for Mankiyali, but most male speakers are fully literate in Urdu. As such, the Perso-Arabic script is typically adapted and used as a medium for informal written communication among the community. The Mankiyali vowel inventory is provided in Figure 4.2. As is evident, five of the vowel qualities exhibit a nasality contrast in long vowels.

		Front			Central			Back		
		short	long	nasal	short	long	nasal	short	long	nasal
High	tense	i	i:	ĩ:				u	u:	ũ:
	lax	ɪ	ɪ:	ĩ:						
Mid	tense		e:						o:	õ:
	lax							ʌ		
Low	tense									
	lax								ɑ:	ã:

Figure 4.2: The Mankiyali vowel inventory (Paramore, 2021; Paramore and Aurangzeb, 2025)

Examples of (near) minimal pairs of nouns contrasting in vowel nasality are provided in (1). These nouns are in the unmarked nominative case.

- (1) Contrastive vowel nasality in Mankiyali
- a. /d̪i:/ ‘giant.NOM’ vs. /d̪ĩ:/ ‘pine firewood.NOM’
 - b. /gɑ:/ ‘cow.NOM’ vs. /gã:/ ‘village.NOM’
 - c. /survi:/ ‘big needle.NOM’ vs. /tʃuv-ĩ:/ ‘squash-M.PL.NOM’

When followed by a nasal consonant suffix, this vowel nasality contrast is neutralized, as shown by the examples in (2), with the genitive feminine singular case marker, /-ŋ/. Native speaker consultants consistently judge vowels occurring before nasal consonant suffixes as categorically nasalized. This holds true regardless of whether the vowel is underlyingly oral or nasal. In what follows, I describe two experiments measuring vowel nasality patterns in Mankiyali. The results demonstrate that the neutralization of vowel nasality before nasal consonant suffixes is both phonologically complete (section 4.2.5) and phonetically complete (sections

4.2.6 and 4.3). The existence of phonetically complete neutralization of this kind lends support to the claim that linguistic knowledge is not wholly contained within lexical representations.

(2) Neutralization of vowel nasality contrasts in Mankiyali

- a. /dʲɪ:-ŋ/ → [dʲĩ:-ŋ] ‘giant-GEN.F.SG’
- b. /dʲĩ:-ŋ/ → [dʲĩ:-ŋ] ‘pine firewood-GEN.F.SG’
- c. /gɑ:-ŋ/ → [gã:-ŋ] ‘cow-GEN.F.SG’
- d. /gã:-ŋ/ → [gã:-ŋ] ‘village-GEN’
- e. /su:vĩ:-ŋ/ → [su:vĩ:-ŋ] ‘big needle-GEN.F.SG’
- f. /tʃuv-ĩ:-ŋ/ → [tʃu:vĩ:-ŋ] ‘squash-M.PL-GEN.F.SG’

4.2 Experiment 1: Is nasality neutralized on the vowel?

Experiment 1 aimed to compare the realization of vowel nasality on contrastive oral /V:/ and nasal / $\tilde{V}$:/ vowels when followed by the feminine singular genitive suffix, /-ŋ/. While nasalance is a primary correlate of nasality, Shosted (2015) and others have shown that vowel nasality sometimes covaries with other acoustic properties such as tongue height (F1 modulation) and breathiness. As such, four phonetic measures (nasalance, vowel breathiness, duration, and F1 modulation) were extracted from the experimental data to examine the phonetic realization of vowel nasality for oral and nasal vowels in various contexts.

4.2.1 Participants

Twenty male native Mankiyali speakers participated in the experiment. The ages ($\mu = 33$, $\sigma = 12$), languages spoken, birthplace, and current location of residence were recorded for each participant before the experiment began. All participants were born in and have lived in Danna most of their lives. Every participant

reported fluency in Hindko and Urdu in addition to Mankiyali, and many also speak Pakistani Pashto and Punjabi. A few speakers speak English and/or Gojri as well. Due to cultural customs and low literacy rates among women in the region, only men were available to participate in the study.

4.2.2 Stimuli and Design

61 monosyllabic words sourced from five conditions were elicited. Two baseline conditions established the endpoints of the [NAS] feature contrast: underlying oral /CV:/ and underlying nasal /C $\tilde{V}$:. To avoid carryover coarticulation effects, items in the two baseline conditions contained no nasal consonants in any position. To test whether vowel nasality is phonetically neutralized before nasal suffixes, a subset of words from the baseline oral and nasal conditions were produced with the feminine singular genitive nasal suffix /-ŋ/, yielding the critical conditions /CV:-N/ (neutralized oral) and /C $\tilde{V}$:-N/ (neutralized nasal). A fifth condition comprised monomorphemic words with pre-N vowels (/CV:N/). In all three coda-nasal conditions, items with nasal onsets were excluded to avoid carryover coarticulation.

Table 4.1 summarizes the distribution of words and expected token counts across conditions; complete glosses are provided in Table 4.10 of the appendix. As the data was originally collected to describe nasal patterns in pre-N vowels without regard to their underlying status, there is an imbalance in token count and vowel qualities across the neutralized oral /CV:-N/ and neutralized nasal /C $\tilde{V}$:-N/ forms. This imbalance is addressed in the analysis of the results.

Condition	# Words	Expected # lapel tokens	Expected # mask tokens
CV:	20	800	1,600
C $\tilde{V}$:	19	760	1,520
CV:-N	4	160	320
C $\tilde{V}$:-N	7	280	560
CV:N	11	440	880
Total	61	2,440	4,880

Table 4.1: Distribution of words, expected lapel tokens (20 speakers x 2 repetitions x # of words), and expected air pressure mask tokens (20 speakers x 4 repetitions x # of words) used in experiment 1, separated by condition.

4.2.3 Procedure

All recording sessions took place in a quiet home in the village of Danna. As in the Punjabi experiments from chapter 2, a native speaker consultant — Aurangzeb Mankiyali — with knowledge of the experimental goals presided over the entire experiment, interacting with participants only in Mankiyali. At the start of each session, participants walked through a consent form with the consultant, in which they were told that the experiment measured air pressure variations from the nose and mouth to study how different words are pronounced in Mankiyali. After consent was given, the native-speaker consultant introduced participants to the recording and air pressure equipment. A Glottal Enterprises dual-chamber oro-nasal air pressure mask equipped with a handle and connected to two separate transducers was used to measure air pressure from the nose and mouth separately at a sampling rate of 11,025 Hz. In addition, an Audio-Technica AT831 cardioid condenser lapel microphone connected to a Zoom H5 Handy Recorder was attached to the upper collar of participants’ attire to obtain high-quality acoustic recordings along with the air pressure data obtained from the airflow mask.

After familiarizing themselves with the equipment and completing a short test

trial to ensure clear understanding of the task, and that the mask was held firmly against the face during speech to prevent airflow leakage, the experiment began. Participants were exposed to tokens in the Urdu script in a randomized order using a Python GUI. Each token was produced six times in the following order: twice without the mask to obtain clear acoustic recordings with the lapel microphone, twice slowly into the mask and twice using rapid speech into the mask to test whether speech rate impacted the degree of nasality on neutralized vowels (Solé, 1992, 1995). If a word was misread or any directions not followed, the native-speaker consultant directed participants to repeat all six iterations of the token. Recording sessions lasted approximately 10-20 minutes.

The native-speaker consultant presided over the entire experiment, providing instructions and ensuring native pronunciations throughout. I was present for all experiment sessions to ensure recording equipment was operating properly and that the experimental procedure was followed.

4.2.4 Analysis

Data processing to obtain nasalance values ($\frac{A^n}{A^n + A^o}$) followed the same general process as in the Punjabi experiment outlined in chapter 2. All recordings obtained from the nasal airflow mask were hand-annotated by three undergraduate research assistants; I reviewed all annotations to ensure consistency. A Praat script was then used to iterate through the files and convert the oral and nasal air pressure channels to intensity tracks, and subsequently extract nasalance values over 11 time-normalized windows for each segment.

While nasalance, as defined in this dissertation, acts as a relatively straightforward primary correlate to vowel nasality, previous work has shown that the acoustic

underpinnings highlighting the contrast between nasal and oral vowels are more dynamic than a simple linear correlation between velopharyngeal port opening and vowel type (Shosted, 2015). Indeed, previous work on the oral-nasal vowel contrast in languages such as French (Carignan, 2017) and Hindi (Shosted et al., 2012) has shown that nasal vowels may be produced with changes in the first formant (i.e., tongue height) depending on the underlying quality of the vowel.

In addition, breathiness and nasality have been shown to share similar acoustic and perceptual properties, even to the point that the two phonological features are sometimes derived from one another diachronically (Garellek et al., 2016; Styler, 2017). Synchronically, Garellek et al. (2016) found that breathiness covaries with vowel nasality in some languages (e.g., Luchun Hani), such that nasal vowels are consistently more breathy than oral vowels, controlling for vowel quality. This covariation between breathiness and nasality likely arises because one of the primary acoustic effects of velopharyngeal coupling (velum lowering) is a dampened first formant in relation to the first or second harmonic (often called the nasal pole because of its increased amplitude during the production of nasal sounds). Similarly, the increased proportion of vocal fold abduction and acoustic coupling with the sub-glottal cavity during breathy vowels also leads to a weaker F1 and a stronger first or second harmonic. Because of this, speakers sometimes produce nasal vowels with additional breathiness to enhance the phonetic percept of nasality.

Given the preceding discussion, in addition to recording nasalance measures, three other acoustic properties of the vowels were also analyzed: formant values, duration, and voice quality information for each vowel. To obtain formant measures, the lapel recordings were time-aligned with the airflow mask TextGrids, and a

separate Praat script iterated over the files, extracting formant values over 10 time-normalized windows for each vowel. PraatSauce (Kirby and Puggaard-Rode, 2025) was used to obtain spectral information to examine H1*-H2*, a common acoustic measure used to evaluate vowel breathiness (Chai and Garellek, 2022; Garellek, 2019; Garellek et al., 2016), though recent work has critiqued H1*-H2* in favor of residual H1 (Brinkerhoff, 2025; Brinkerhoff and McGuire, 2025; Chai and Garellek, 2022). While F1 modulation proved to covary with nasality (section 4.2.6.2), no differences in breathiness were found between baseline oral /CV:/ and nasal /C $\tilde{V}$:/ vowels in the Mankiyali data. There were also no duration differences between baseline oral /CV:/ and nasal /C $\tilde{V}$:/ vowels. As a result, only results for F1 modulation are presented.

4.2.5 Vowel Nasality is Phonologically Neutralized

Diagnostic evidence for the phonological neutralization of a phonemic contrast can take several forms. For example, Braver and Kawahara (2016) use minimal word restrictions in Japanese to show that a vowel lengthening process is phonologically complete. Similarly, Du and Durvasula (2022, 2025), in their analysis of lexical tones in Huai'an Mandarin, demonstrate that a high-falling tone undergoes tone sandhi to a low-rising tone in certain contexts. Crucially, this process is shown to be phonologically complete because derived low-rising tones feed subsequent sandhi processes in the same way as an underlying low-rising tone. Finally, Cohn (1993, 2006) argues that phonological processes alter the phonetic targets of specific sounds. Under this view, if a feature value is phonologically neutralized, it should exhibit categorical realization — i.e., a clear phonetic target of the new feature value implemented across the relevant segment — rather than gradient

variation.

Along these same lines, Solé (1992, 1995) used the interaction of speech rate and proportion of nasality on vowels in English as a diagnostic of the phonological implementation of nasality. She found that as vowel duration increased in English pre-nasal positions, the proportion of the vowel that was nasalized remained constant. She took this to suggest that nasalization of pre-nasal vowels in English is phonological, since a constant proportion of nasality implies intentional implementation rather than purely phonetic anticipatory coarticulation, which would predict an inverse correlation between nasality and vowel length.²

In Mankiyali, it is unclear whether any phonological processes are triggered by nasal vowels. Similarly, I have not uncovered any phonological constraints in the language which would shed light on whether a vowel is phonologically targeted as nasal or not. As such, I attempt to show here the phonological neutralization of vowel nasality with two forms of evidence. First, the application of nasality on underlying oral vowels that occur before nasal consonants is categorical and insensitive to speech rate, which, following arguments from Solé (1992, 1995), indicates that nasality is intentionally implemented by Mankiyali speakers. Second, in forms in which the pre-N vowel and following nasal consonant are tautomorphic, no minimal pairs based on vowel nasality exist. Moreover, judgments from native speakers suggest that all pre-N vowels are nasalized.

If the contrast in vowel nasality in Mankiyali is phonologically neutralized before nasal suffixes, we should expect that speakers target underlying oral vowels as

²In later work, Solé (2007) revises this analysis, arguing instead that the distinction between stable proportional nasality across speech rates and fixed nasality irrespective of speech rate reflects a distinction between speaker-controlled nasality (rather than phonological implementation as such) and mechanically induced nasality arising from transitions from an oral to a nasal segment.

[+NAS] in this context. Moreover, the implementation of nasality should largely be impervious to changes in speech rate. On the other hand, if the presence of nasality on underlying oral vowels is merely a phonetic effect of anticipatory nasalization, we might expect the degree of nasalance to decrease as the vowel lengthens. Figure 4.3 shows a scatterplot visualizing the relationship between vowel duration and mean vowel nasalance for all oral vowel tokens from experiment 1. The yellow dots plot the relationship of mean vowel nasalance and vowel duration for oral /CV:/ vowels in the bare nominative form (without a nasal suffix). As evidenced by the yellow regression line, increased speech rate exhibits a small correlation with increased nasalance in these completely oral forms: $r(304) = .16$, $p = .006^{**}$. However, Pearson correlation values less than .3 are generally treated as practically insignificant, regardless of whether they reach statistical significance (Bonett, 2021, p.3-5), which unsurprisingly indicates no meaningful correlation exists between vowel length and nasalance for bare nominative /CV:/ forms.

The blue dots and regression line in Figure 4.3 depict the relationship between mean vowel nasalance and vowel duration for neutralized oral /CV:-N/ vowels. The typical mean nasalance of these vowels is clearly elevated compared to that of the completely oral forms, and the blue regression line shows that the nasalance of neutralized oral /CV:-N/ vowels is immune to speech rate changes: $r(310) = .06$, $p = .29$. The consistent and elevated nasalance across speech rates means that the implementation of nasality is categorical in neutralized oral /CV:-N/ forms, suggesting phonological neutralization. In other words, when underlying oral /V:/ vowels are followed by a nasal suffix, speakers intentionally produce them as categorically [+NAS].

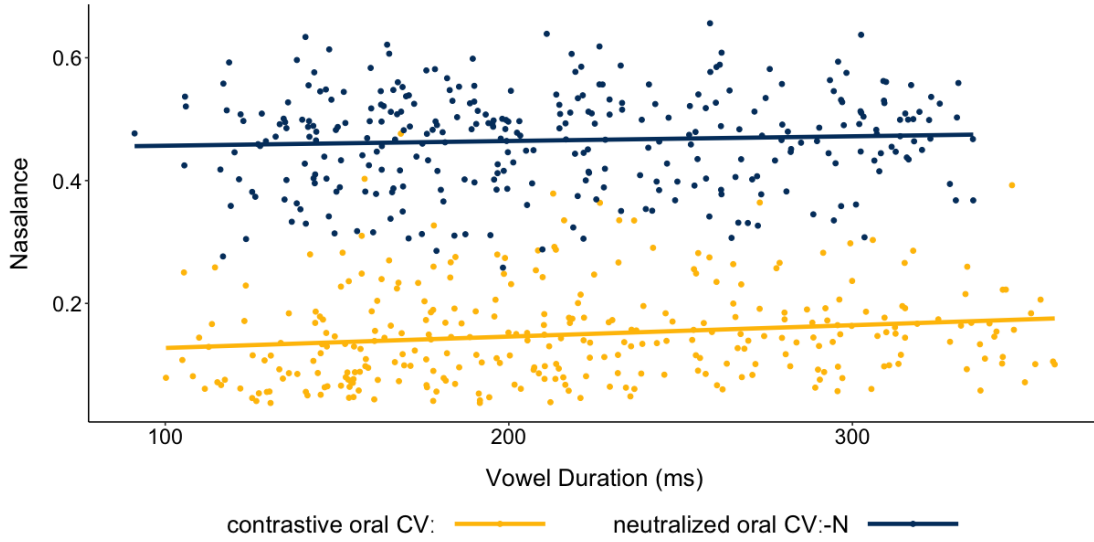


Figure 4.3: Scatterplot showing the relationship between vowel duration (ms) and mean vowel nasalance for the oral /CV:/ ($n = 306$; yellow dots and regression line) and neutralized oral /CV:-N/ conditions ($n = 312$; blue dots and regression line). Lemmas included in the plot are /ba:(-ŋ)/, /kjo:(-ŋ)/, /po:(-ŋ)/, and /sa:(-ŋ)/. For both conditions, tokens greater than two standard deviations from a speaker’s mean vowel duration were discarded, resulting in the removal of 3 tokens from the /CV:/ condition set and 12 tokens from the /CV:-N/ condition set.

Further evidence that underlyingly oral vowels undergo phonological neutralization in Mankiyali is the level and trajectory of nasalance traces of oral vowels followed by a nasal suffix. Figure 4.4 shows GAM curves for all neutralized oral /CV:-N/ words (dashed light blue lines) from experiment 1. A qualitative examination of the plot reveals that these words exhibit nasalance levels and trajectories that categorically group with the level and trajectories of words with contrastive nasal /CṼ:/ vowels (yellow lines). That is, oral vowels occurring before nasal codas in Mankiyali are intentionally targeted as [+NAS] in all of these forms, and the implementation is categorical.

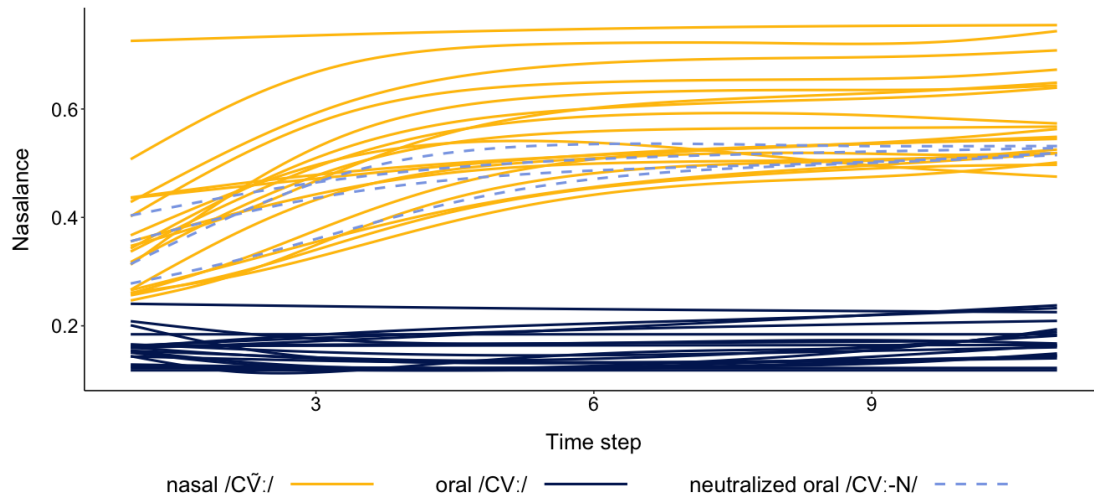


Figure 4.4: GAM curves separated by word and condition showing mean vowel nasalance across 11 normalized timesteps from experiment 1 for the nasal /CṼ:/, oral /CV:/, and neutralized oral /CV:-N/ conditions.

In short, the consistent increase in nasality and the categorical nature of nasalance trajectories for neutralized oral /CV:-N/ forms suggests that their nasalization is not merely a product of coarticulation. Rather, these vowels are intentionally produced as phonologically [+NAS].

Nevertheless, while I have tried to establish that vowel nasality is phonologically neutralized before nasal consonant codas, it is still possible that this phonological neutralization is phonetically incomplete. In the next section, I demonstrate that phonologically neutralized oral vowels lack any residual phonetic traces that would reveal their underlying lexical identity. This *phonetically complete neutralization* is surprising under usage-based models of speech representation, since the influence on production of a word from semantically and phonologically related forms is expected to result in phonetically incomplete neutralization, as it does in many other cases (see section 4.4).

4.2.6 Vowel Nasality is Phonetically Neutralized

4.2.6.1 Nasalance

Of the 4,880 potential tokens measuring nasalance in experiment 31 low-quality trials were removed, resulting in a total of 4,849 tokens included in the analysis. Figure 4.5 shows GAM curve visualizations — grouped by condition — of the vowels tokens uttered in experiment 1. Since vowel height has been shown to covary with nasality in both oral and nasal vowels (Bell-Berti, 1993), the data is partitioned into three separate plots for high, mid, and low vowels respectively.

The vowel qualities of tokens from the neutralized oral /CV:-N/ condition were all either low or mid vowels, whereas vowel qualities of tokens from the neutralized nasal /C $\tilde$ V:-N/ condition were all either mid or high vowels. While this unbalanced vowel quality is accounted for in the statistical analysis using linear mixed effects modeling (Baayen et al., 2008; Speelman et al., 2018), including the full range of vowel qualities in the plots distorts the reality of the similarities in nasalance traces across conditions, so I chose to partition the visualizations by vowel quality to remedy this.

As shown by the plots, when vowel quality is controlled for, the contrast in nasality between underlying contrastive oral and nasal vowels qualitatively vanishes in the context of a nasal consonant suffix. Importantly, the two vowel types do not converge at a midpoint in nasalance between contrastive nasal and oral vowels. Instead, the neutralized oral /CV:-N/ vowel adopts a similar nasalance trajectory as underlying nasal /C $\tilde$ V:/ and /C $\tilde$ V:-N/ vowels across its entire span.

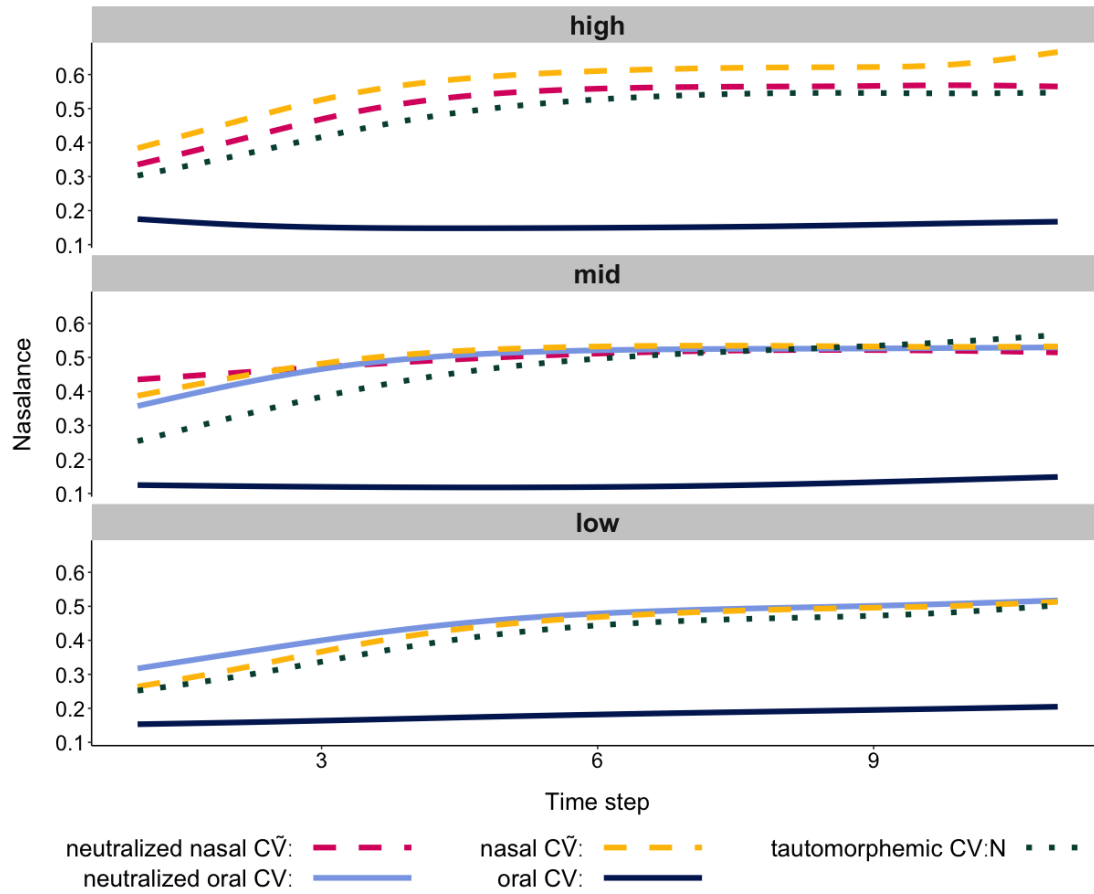


Figure 4.5: GAM curves separated by condition showing mean vowel nasalance across 11 normalized timesteps for all tokens in experiment 1, grouped by vowel height and condition. Underlyingly oral vowels are represented by the solid lines, underlyingly nasal vowels by the dashed lines, and vowels followed by a tautomorphic nasal coda with a dotted line. The pink lines depict nasalance trajectories for neutralized nasal $/C\tilde{V}:-N/$ vowels, and the light blue line shows trajectories for neutralized oral $/C\tilde{V}:-N/$ vowels with a nasal coda.

To quantitatively assess whether neutralized oral $/C\tilde{V}:-N/$ vowels undergo complete phonetic neutralization to $[+NAS]$, I fit a linear mixed effects model to all tokens in the underlying nasal $/C\tilde{V}:/$ ($n = 1520$), neutralized nasal $/C\tilde{V}:-N/$ ($n = 541$), and neutralized oral $/C\tilde{V}:-N/$ ($n = 315$) conditions. `CONDITION`, `VOWEL QUALITY`, `TIMESTEP`, and the interaction between `CONDITION` and `TIMESTEP` were included as fixed effects. By-speaker random intercepts and random slopes

for CONDITION were included as random effects. For the fixed effects, CONDITION was dummy coded, while VOWEL QUALITY and TIMESTEP were effect (sum) coded. Consequently, the model intercept (β_0) represents the vowel nasalance of contrastive nasal /C $\tilde{V}$:/ vowels, averaged over VOWEL QUALITY and TIMESTEP levels. The remaining fixed effect coefficients (β_1 - β_{10}) capture the estimated deviation from baseline contrastive nasal /C $\tilde{V}$:/ vowels for each predictor. Finally, the interaction terms (β_{11} - β_{13}) indicate how much neutralized oral /CV:-N/ vowels vary from the vowel nasalance of underlying nasal /C $\tilde{V}$:/ vowels at each specified time step, adding to the main effect size of vowel condition (β_1).

Results of the model (Table 4.2) show that the interaction terms between CONDITION and TIMESTEP are significant, indicating that nasalance trajectories across the span of the vowel differ between underlying nasal /C $\tilde{V}$:/ and neutralized oral /CV:-N/ vowels. Specifically, neutralized oral /CV:-N/ vowels are estimated to be slightly *more* nasalized at the onset of the vowel than underlying nasal /C $\tilde{V}$:/ vowels (.016). In the final 2/3 of the vowel, however, underlying nasal /C $\tilde{V}$:/ vowels are slightly more nasalized. Nevertheless, the interaction effect sizes are quite constrained, with 95% CIs remaining within $\pm .021$ nasalance across all time steps. Thus, it is safe to examine the main effect between the two vowels without interaction effects obscuring the results.

Setting aside the main effects of TIMESTEP and VOWEL QUALITY, the portion of the model most relevant to the present discussion is located in the top three rows. Rows 1-3 show that neutralized oral /CV:-N/ vowels do not differ significantly from underlying nasal /C $\tilde{V}$:/ vowels. In contrast, the estimated nasalance of neutralized *nasal* /C $\tilde{V}$:-N/ vowels in row 3 *were* significantly different, with a mean nasalance about .045 *less than* underlying nasal /C $\tilde{V}$:/ vowels. This suggests

that, controlling for TIMESTEP and VOWEL QUALITY, neutralized oral /CV:-N/ vowels exhibit a slightly *higher* degree of nasalance than neutralized nasal /C $\tilde{V}$:-N/ vowels.

Predictor	β	95% CI	t	p
Intercept (nasal C $\tilde{V}$:-)	.532	[.508, .557]	42.29	< .001 ***
neutralized oral CV:-N	-.005	[-.049, .040]	-.20	.84
neutralized nasal C $\tilde{V}$:-N	-.045	[-.078, -.013]	-2.74	.01 *
Beg	-.083	[-.085, -.081]	-79.87	< .001 ***
Mid	.034	[.032, .036]	32.97	< .001 ***
Fin	.049	[.046, .051]	38.59	< .001 ***
u:-	.036	[.008, .063]	2.53	.02 *
i:-	.125	[.100, .151]	9.68	< .001 ***
ɪ:-	-.050	[-.076, -.025]	-3.87	< .001 ***
o:-	-.018	[-.042, .007]	-1.44	.16
ɑ:-	-.093	[-.120, -.066]	-6.71	< .001 ***
neut. oral CV:-N: Beg	.016	[.011, .021]	6.07	< .001 ***
neut. oral CV:-N: Mid	-.007	[-.012, -.002]	-2.82	.005 **
neut. oral CV:-N: Fin	-.008	[-.015, -.002]	-2.66	.008 **

Table 4.2: LMER model summarizing the effects of CONDITION, (3 levels), TIMESTEP (3 levels), VOWEL QUALITY (5 levels), and the interaction between CONDITION and TIMESTEP on nasalance for all vowel tokens in experiment 1: `lmer(nasalance~CONDITION*TIMESTEP + VOWELQUALITY + (1 + CONDITION|speaker) + (1|word))`.

In order to examine the relationships between the two neutralized vowel conditions more closely, I extracted a pairwise comparison from the LMER model in Table 4.2, using the `emmeans()` package in R (Lenth, 2024), which provides estimated marginal means across conditions, adjusting for the other model predictors. As can be seen in Table 4.7, the pairwise comparison between the nasalance of neutralized oral /CV:-N/ vowels and neutralized nasal /C $\tilde{V}$:-N/ vowels is insignificant. In other words, underlying oral vowels were not found to be meaningfully different in their nasalance from underlying nasal vowels when a nasal coda immediately follows each vowel.

Comparison	β	95% CI	t	p
(neut. oral CVɿ-N) – (neut. nasal CṼɿ-N)	.041	[-.024, .105]	1.51	.39

Table 4.3: Pairwise comparison showing the effect of vowel condition on nasalance. Results were extracted using `emmeans()`, which averages over other predictors in the model: `emmeans(lmer-exp1, ~ condition)`.

With that said, the absence of a significant difference between neutralized oral /CVɿ-N/ vowels and nasal /CṼɿ:/ and /CṼɿ-N/ vowels in experiment 1 does not entail that the vowel types are equivalent in their nasalance. Rather, it simply means that we lack evidence to conclude they are different. To formally establish equivalence – and thereby test for phonetically complete neutralization of vowel nasality – we can establish a *Region of Practical Equivalence* (*-h to h*) for the nasalance of underlying nasal vowels in Mankiyali. Within this region, differences in mean nasalance can be treated as random variation within the same category. After defining the *RoPE*, the 95% confidence interval for differences in nasalance between neutralized oral and underlying nasal vowels can be evaluated against it. If the CI falls entirely within the *RoPE*, we can conclude that the two vowel types are practically equivalent in terms of nasalance.

To determine a *RoPE* for mean nasalance of nasal vowels in Mankiyali, I grouped all underlying nasal /CṼɿ:/ tokens from experiment 1 by speaker and vowel quality, then calculated the mean and standard deviation of each group. Computing by-speaker nasalance means separately for each vowel quality controls for inherent differences in nasalance across vowel heights. For each speaker, the mean nasalance and standard deviation was then averaged across vowel qualities to obtain by-speaker standard nasalance distributions. Using these measures, I defined the *RoPE* as the range covering 95% of nasal /CṼɿ:/ nasalance values (± 2 SDs) for the speaker with the tightest clustering of nasalance values across tokens. Because

all vowels in the underlying nasal $/C\tilde{V}:/$ condition are phonologically contrastive nasal vowels, any within-speaker variation in nasalance can be treated as random noise. Setting the *RoPE* equivalent to the speaker with the least variability's standard deviation thus provides a stringent threshold, within which variation can be interpreted as within-category random phonetic noise rather than a cross-category phonological distinction. Figure 4.6 shows the mean nasalance and ± 2 SDs of nasal $/C\tilde{V}:/$ vowels for each speaker in experiment 1. Speaker SR0 exhibited the least amount of variation, with 2 SDs of .078, so a conservative estimate of the *RoPE* in this experiment is $\pm .078$.

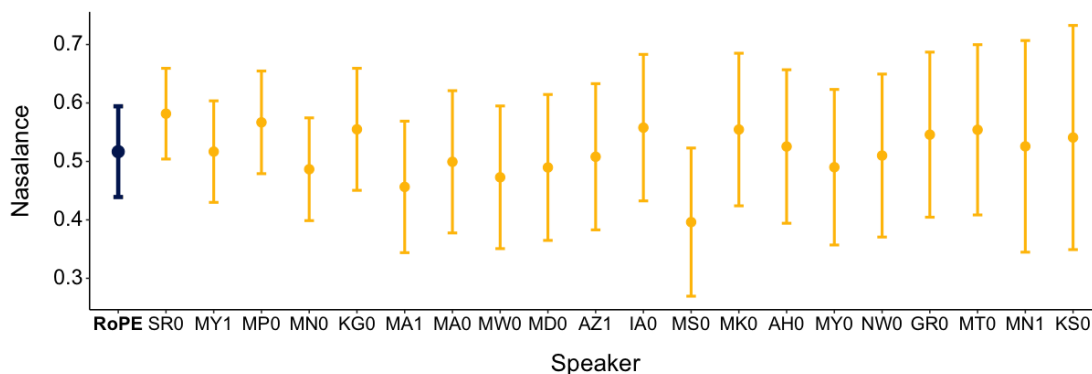


Figure 4.6: Dot and whisker plot showing mean nasalance of $/C\tilde{V}:/$ vowels from experiment 1. Yellow dots indicate each speaker's mean $/C\tilde{V}:/$ vowel nasalance, with whiskers showing $\pm$ two standard deviations (95% of values) around that mean. Speakers are ordered from left to right by magnitude of standard deviation. The blue dot marks the mean nasalance across all speakers, and its whiskers represent the proposed *RoPE* of $\pm .078$ for underlying nasal $/C\tilde{V}:/$ vowels, equivalent to the 2 SD range of the speaker with the narrowest nasalance distribution (SR0).

Reevaluating the 95% confidence interval for the estimated difference in nasalance between underlying nasal $/C\tilde{V}:/$ vowels and neutralized oral $/CV:-N/$ vowels (row 2 in Table 4.2), it is clear that the entire interval of $[-.049, .040]$ falls within the $\pm .078$ *RoPE*. Moreover, examining the 95% CI for the estimated difference in

nasalance between neutralized oral /CVɹ:-N/ vowels and neutralized nasal /C $\tilde{V}$ ɹ:-N/ vowels in Table 4.7, the upper bound of the 95% CI (.105) extends beyond the *RoPE*. This deviation reflects the possibility that neutralized nasal /C $\tilde{V}$ ɹ:-N/ vowels are, if anything, slightly *less* nasalized than neutralized oral /CVɹ:-N/ vowels.³ Taken together, these results support the conclusion that the neutralization of nasalance between oral and nasal vowels is phonetically complete on the vowels themselves.

4.2.6.2 F1 Modulation

As Shosted (2015) observed, contrastive nasal vowels are not simply [+NAS] oral vowels. In other words, the contrast between oral and nasal vowels does not reduce to the lowering of the velum to couple the nasal tract with the oral tract during vowel production. Instead, other modifications of the vocal tract commonly covary with velum lowering and serve as secondary correlates to nasality. In this section, I explore whether residual differences in tongue height (F1) persist between neutralized oral /CVɹ:-N/ and underlying nasal /C $\tilde{V}$ ɹ:/ and neutralized nasal /C $\tilde{V}$ ɹ:-N/ vowels.

Carignan (2017) examined the differences in the realization of minimal pairs contrasting in nasality in Southern French and found that mid front nasal / $\tilde{e}$ / covaried with a systematically lower tongue height in the beginning portion of the vowel than mid front oral / ϵ /. Similarly, the mid back nasal / $\tilde{o}$ / exhibited a lower tongue height in the first half of the vowel but raised dramatically in the second half of

³In discussing phonetic neutralization, Du and Durvasula (2025) note that cases of *over-neutralization* — in which the phonetic realization of the relevant feature in a neutralized segment is realized to a greater degree than in a segment that is underlyingly specified for that feature — have never been observed. While the results here are non-significant, it is nevertheless intriguing that the underlyingly oral vowels show, if anything, a greater degree of nasality than their nasal counterparts in the same context.

the vowel to be higher than mid back oral /o/. When comparing low nasal /ã/ to low oral /a/, on the other hand, he found that the two low vowels shared a similar F1 at the onset of the vowel, but the low oral vowel maintained a stable tongue height throughout the time course of the vowel. Conversely, the tongue height of the low nasal vowel raised over time to be significantly higher than its oral counterpart. In short, vowels with different heights exhibited distinct dynamic changes in tongue height (F1) as a function of nasality. Mid nasal vowels tended to begin with a lower tongue height compared to their oral counterparts and raised later in the vowel. Low nasal vowels were initiated with similar tongue heights to low oral vowels, but their tongue height raised significantly over time.

Interestingly, Shosted et al. (2012) examined the articulatory and acoustic differences between nasal and oral vowels in Hindi and found a difference in F1 modulation between non-low front and back vowels. Front nasal vowels exhibited a higher tongue height (lower F1) compared to their oral counterparts, but back nasal vowels showed a lower tongue height (higher F1). Considering the results of both the Hindi and Southern French studies, it is clear that while the direction of F1 modulation in nasal vowel production may be language-specific, dynamic changes in F1 are a consistent correlate to nasality.

Figure 4.7 shows the vowel space in the F1/F3 — F2/F3 plane, with the mean formant ratio values extracted from the midpoint of each token averaged by condition across all speakers in the study (dynamic changes in F1 are examined below). F1/F3 and F2/F3 ratios were used rather than mean F1 and F2 values alone because F3-normalized values have been shown to successfully remove cross-speaker variation (Monahan and Idsardi, 2010), thus providing an effective means of formant normalization. From a cursory inspection of the vowel space, base-

line nasal / $\tilde{V}:$ / vowels (yellow) are generally pronounced with the tongue slightly lower and further back than their oral / $V:$ / counterparts (dark blue). Crucially, the neutralized oral / $CV:-N$ / (light blue), neutralized nasal / $C\tilde{V}:-N$ / (pink), and tautomorphemic / $CV:N$ / (green) vowels cluster with the baseline nasal vowels in the vowel space.

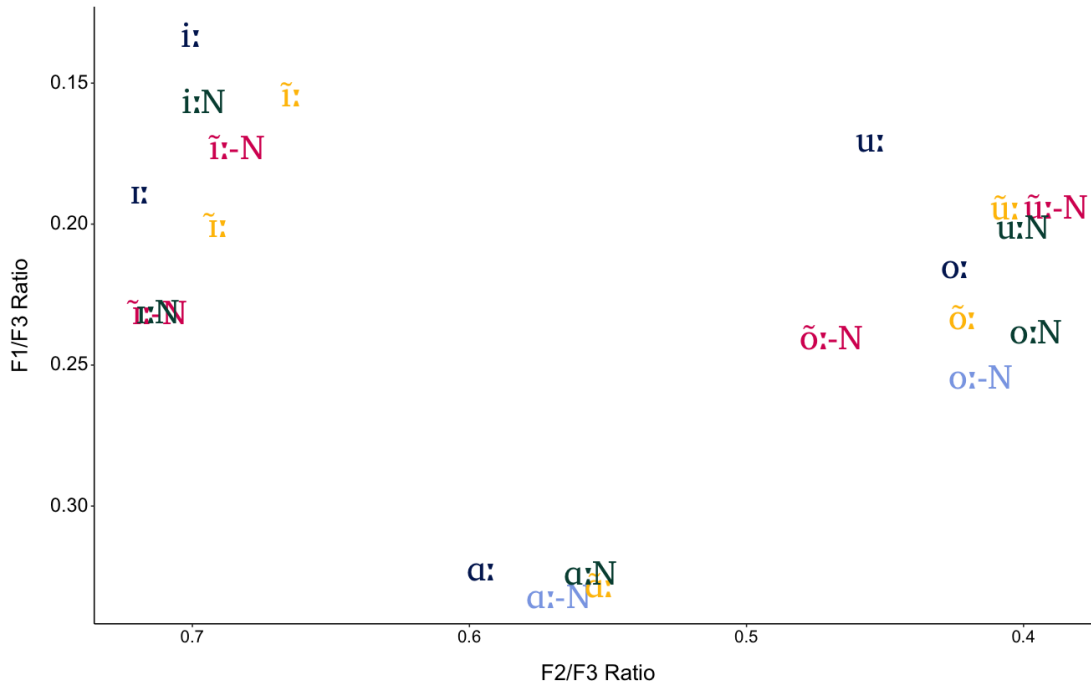


Figure 4.7: F1/F3 — F2/F3 vowel space showing the mean F1/F3 and F2/F3 ratio values calculated from the midpoint of all vowels, averaged across speakers, and separated by condition: oral / $CV:$ / (dark blue), nasal / $C\tilde{V}:$ / (yellow), neutralized oral / $CV:-N$ / (light blue), neutralized nasal / $C\tilde{V}:-N$ / (pink), and tautomorphemic / $CV:N$ / (green).

To examine potential differences between neutralized oral / $CV:-N$ / and neutralized nasal / $C\tilde{V}:-N$ / vowels, I first examined differences in the baseline oral and nasal vowels and then compared those differences to the realization of the neutralized forms. Because the mid back / $o:$ / vowel is the only vowel present in all four conditions, the present analysis only compares formant differences for this

vowel. In line with previous literature, the mid back /o:/ vowel showed no distinction in tongue backness (F2/F3 ratio), so the remainder of this section focuses on potential differences in tongue height (F1/F3 ratio).⁴

To evaluate the potential for dynamic changes in F1 over time, formant values were extracted from each /o:/ token at 10 normalized time windows across the length of each vowel token. Figure 4.8 shows GAM curves fit to the 10 timepoints for all F1/F3 values for the bare nominative oral /CV:/ and nasal /C $\tilde{V}$:/ conditions. The plot shows that, at the endpoints of the vowels, no tongue height distinction exists between oral and nasal mid back /o:/. However, nasal / $\tilde{o}$:/ exhibits a lower tongue height (larger F1/F3 ratio) compared to oral /o:/ throughout the middle portion of the vowel. This is good evidence that F1 is recruited as an additional correlate to nasality in Mankiyali, at least for mid back vowels, with the dynamic change over time found in other languages also present in Mankiyali.

⁴Interestingly, while the mid back /o:/ vowel showed no significant changes in F2/F3 ratio in relation to nasality, the vowel space in Figure 4.7 highlights that nasal vowels are produced with a lower F2/F3 ratio compared to oral vowels for most of the vowel qualities in Mankiyali. As such, future research exploring whether F2 modulation correlates with nasality in the language would be welcomed.

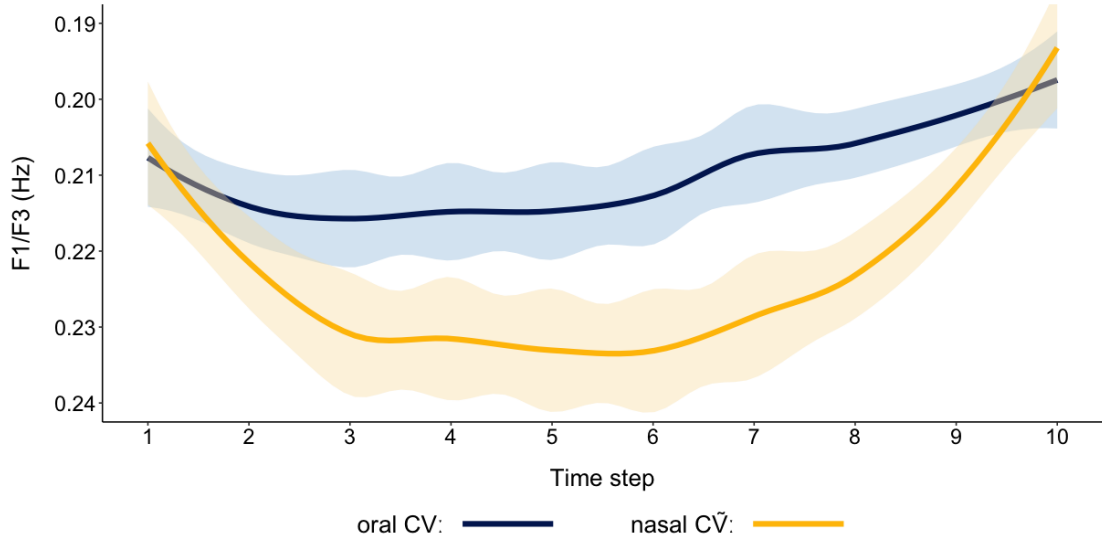


Figure 4.8: Generalized Additive Model (GAM) curves of mean F1/F3 ratio across the middle 10 timesteps for tokens with the low back mid vowel /o:/ or /õ:/ in Experiment 1. The blue line represents oral /Co:/ vowels and the yellow line represents nasal /Cõ:/ vowels. Shaded region shows the 95% confidence interval around the mean.

7

While oral /o:/ correlates with a raised tongue height compared to nasal /õ:/ in the bare nominative forms, this distinction is neutralized before the genitive suffix, /-ŋ/, as shown in Figure 4.9. Interestingly, the tongue height of neutralized oral /CV:-N/ vowels is visually lower (larger F1/F3 ratio) than the tongue height of either underlying nasal /Cõ:/ or neutralized nasal /Cõ:-N/ vowels. The markedly raised F1/F3 values in the later portions of the vowel for the two neutralized conditions undoubtedly reflects the transition into the retroflex nasal consonant coda, /-ŋ/, since retroflexion significantly lowers F3 (Hamann, 2003).

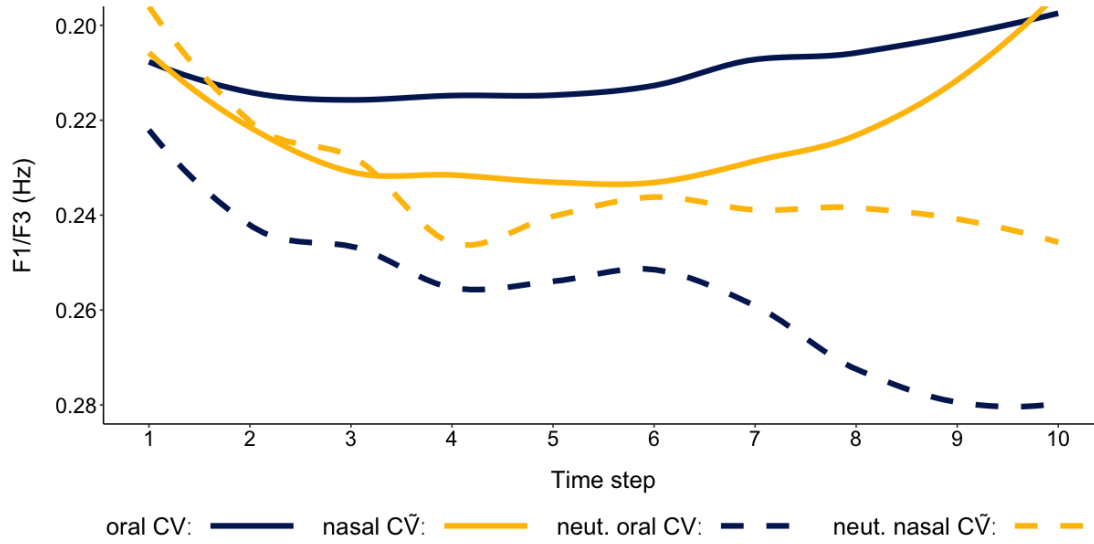


Figure 4.9: Generalized Additive Model (GAM) curves of mean F1/F3 ratio across 10 timesteps for tokens with the low back mid vowel /o:/ or /õ:/ in Experiment 1.

Table 4.4 summarizes a linear mixed effects model fit to the F1/F3 ratio data, with `CONDITION` as the sole fixed effect. By-speaker random intercepts and random slopes for `CONDITION` were included as random effects. A likelihood ratio test comparing this model to one including `TIMESTEP` (steps 4-7) and its interaction with `CONDITION` as additional fixed effects did not significantly improve model fit ($\chi^2(9) = 4.08$, $p = .91$), so `TIMESTEP` was excluded from the analysis. The oral /CV:/ /o:/ condition was also removed to allow for better assessment of potential differences between the other three conditions.

As is evident from the results in Table 4.4, neither of the neutralized vowel conditions were found to exhibit significantly different F1/F3 ratio values.

Predictor	β	95% CI	t	p
Intercept (nasal Cõ:)	.232	[.205, .259]	16.76	< .001 ***
neut. oral Co:-N	.023	[-.005, .051]	1.6	.16
neut. nasal Cõ:-N	.008	[-.027, .043]	.46	.67

Table 4.4: LMER model summarizing the effects of CONDITION, (4 levels) on F1/F3 ratio for all mid low back vowel tokens uttered by all 20 speakers in experiment 1: LMER(F1/F3~CONDITION + (1 + CONDITION|speaker) + (1|word)).

To examine whether the neutralization of vowel height differences between oral and nasal vowels was phonetically complete in the neutralized oral /CV:-N/ condition, I established a *RoPE* for F1 in a manner parallel to that of section 4.2.6.1. For each speaker, I calculated the mean and standard deviation of the F1/F3 ratio for all nasal /õ:/ vowels in the /CṼ:/ condition. The *RoPE* was then defined using the speaker with the narrowest two-standard-deviation interval in F1/F3 ratio variation, which in this case was $\pm .029$.

Comparing this *RoPE* to the 95% CIs in Table 4.4, the upper bound of the 95% for the estimated difference of neutralized oral /CV:-N/ vowels from underlying nasal /CṼ:/ vowels extends beyond the *RoPE*, while the lower bound falls within it. Thus, it is clear that, if anything, neutralized oral /CV:-N/ vowels are ‘over-neutralizing’ in tongue height.

Taken together, the analysis of changes in F1/F3 in this section and nasalance in section 4.2.6.1 provide evidence that the phonological neutralization of vowel nasality on oral vowels occurring before nasal suffixes is phonetically complete. Neutralized oral /CV:-N/ vowels showed no meaningful differences in either nasalance or F1 modulation compared to underlying nasal /CṼ:/ and /CṼ:-N/ vowels.

4.3 Experiment 2: Are there anticipatory coarticulation differences?

While the phonetic analysis in the previous section showed that vowel nasality is completely neutralized on the vowels themselves before nasal suffixes in Mankiyali, it does not provide sufficient evidence that the neutralization is phonetically complete in these words. Indeed, in previous work demonstrating the incomplete neutralization of phonologically neutralized contrasts, residual surface differences revealing the underlying contrast have been discovered on adjacent segments. For instance, in the classic example of German word-final devoicing (Port and O’Dell, 1985; Roettger et al., 2014), not only did underlying voiced obstruents in the word-final position of forms like /bad/ maintain differences in closure voicing duration and post-burst release aspiration length compared to underlying voiceless obstruents, but the different coarticulatory effects of voiced and voiceless obstruents on preceding vowels persisted as well. Specifically, Port and O’Dell (1985) found that pre-stop vowels in words like /bad/ → [baɖ] ‘bathroom’ were significantly and noticeably longer than vowels before underlying voiceless codas like /bat/ → [bat] ‘ask’. These vowel length differences consistently correlated with the underlying contrast in obstruent voicing, providing evidence for phonetically incomplete neutralization of stop voicing in the form of coarticulatory information.

It is well known that in many languages, segments preceding nasal consonants and vowels exhibit anticipatory nasalization, in which the velum is lowered well before the onset of a [+NAS] segment in anticipation of the production of the nasal sound (Beddor and Krakow, 1999). In some cases like English (Poupier et al., 2023), anticipatory nasalization commences well before the [+NAS] segment, resulting in

long-distance anticipatory nasal coarticulation. It is possible, then, that similar to the case of German word-final devoicing, the phonetic remnants revealing the lexical identity of the vowel nasality contrast in Mankiyali manifest, not on the vowels themselves, but on preceding segments exhibiting anticipatory nasalization. To explore this possibility in Mankiyali, experiment 2 examined nasalance on segments preceding oral and nasal vowels to see if any lingering phonetic effects hinting at the underlying nasality contrast were preserved.

4.3.1 Stimuli, Design, and Analysis

Sixteen additional male Mankiyali speakers who did not participate in experiment 1, took part in the second experiment. Ages ranged from 24 to 57 ($\mu = 40.3$, $\sigma = 9.1$). All participants spoke Urdu and Hindko in addition to Mankiyali, and some also spoke Punjabi, Pakistani Pashto, English, and/or Arabic. All speakers were born in Danna and resided there at the time of the experiment.

Stimuli were sourced from four conditions: baseline oral $/V:/$, baseline nasal $/\tilde{V}:/$, neutralized oral $/V\text{:-}N/$, and neutralized nasal $/\tilde{V}\text{:-}N/$. Here however, all words were at least disyllabic, containing a vowel–glide–vowel ($V\text{:}GV\text{:}$) sequence in the final two syllables of the word. The penultimate pre-glide vowel in the $/V\text{:}GV\text{:}/$ sequence was always oral and the final vowel was either oral or nasal, depending on the experimental condition. The experiment consisted of 38 unique word forms across the four conditions: 11 from oral $/V:/$, 11 from neutralized oral $/V\text{:-}N/$, 8 from nasal $/\tilde{V}:/$, and 8 from neutralized nasal $/\tilde{V}\text{:-}N/$. Thirty-six filler items were also dispersed throughout the experiment. The complete list of words from experiment 2 with their glosses is provided in Table 3.

- (3) Mankiyali experiment 2 items and conditions

oral /V:/ (11)	neutralized oral /V:-N/ (11)
sa:vu: ‘powerful’	sa:vu:-ŋ ‘powerful-GEN.F.SG’
bi:va: ‘marriage’	bi:va:-ŋ ‘marriage-GEN.F.SG’
pa:-vi: ‘table leg-PL’	pa:-vi:-ŋ ‘table leg-PL-GEN.F.SG’
si:va: ‘more’	si:va:-ŋ ‘more-GEN.F.SG’
tʃi:-ve: ‘early birth-PL’	tʃi:-vi:-ŋ ‘early birth-PL-GEN.F.SG’
sa:-vi: ‘green/blue-PL’	sa:-vi:-ŋ ‘green/blue-PL-GEN.F.SG’
bu:va: ‘ploughing’	bu:va:-ŋ ‘ploughing-GEN.F.SG’
su:vi: ‘big needle’	su:vi:-ŋ ‘big needle-GEN.F.SG’
ha:va: ‘air’	ha:va:-ŋ ‘air-GEN.F.SG’
se:vu: ‘apple’	se:vu:-ŋ ‘apple-GEN.F.SG’
tɑ:vu: ‘competitor’	tɑ:vu:-ŋ ‘competitor-GEN.F.SG’
nasal /Ṽ:/ (8)	neutralized nasal /Ṽ:-N/ (8)
li:-vã: ‘I mud’	li:-vã:-ŋ ‘I will mud’
ta:-vã: ‘I heat up’	ta:-vã:-ŋ ‘I will heat up’
tʃu:-vĩ: ‘squash-PL’	tʃu:-vĩ:-ŋ ‘squash-PL-GEN.F.SG’
si:-vã: ‘I make wet’	si:-vã:-ŋ ‘I will make wet’
ra:-vã: ‘bean-PL’	ra:-vã:-ŋ ‘bean-PL-GEN.F.SG’
li:-vĩ: ‘we mud’	li:-vĩ:-ŋ ‘we will mud’
pi:-jã: ‘I sit’	pi:-jã:-ŋ ‘I will sit’
pãdʒi-jõ: ‘toe-PL’	pãdʒi-jõ:-ŋ ‘toe-PL-GEN.F.SG’

The primary focus of the experiment was to explore whether coarticulatory nasalization in the neutralized oral /V:-N/ condition is phonetically distinct from the coarticulatory nasalance observed in the two conditions with underlying nasal vowels (nasal /Ṽ:/ and neutralized nasal /Ṽ:-N/). Section 4.3.2.1 replicates findings from experiment 1, showing that the neutralization of nasalance on the vowels themselves is phonetically complete. Section 4.3.2.2 then shows that the coarticulatory nasalization on segments preceding neutralized oral /V:-N/ vowels is identical to that observed on segments preceding underlying nasal vowels.

4.3.2 Results

Of 2,432 possible tokens, 2,331 were included in the analysis; 101 tokens were excluded due to poor recording quality. Table 4.5 shows the breakdown in token

count by condition.

Condition	# Words	Expected # tokens	Actual # tokens
oral /V:/	11	704	684
nasal / $\tilde{V}$ /	8	512	478
neutralized oral /V:-N/	11	704	695
neutralized nasal / $\tilde{V}$:-N/	8	512	474
Total	38	2,432	2,331

Table 4.5: Words, expected tokens (16 speakers x 38 words x 4 repetitions), and actual tokens used in experiment 2 separated by condition.

In Figure 4.10, GAM curves fit to 33 timesteps in the /V:GV:/ sequence show the mean nasalance over time for each condition.⁵ As expected, nasalance across the entire /V:GV:/ sequence in the oral /V:/ condition remains steady at about .2. By contrast, the two nasal conditions and the neutralized oral /V:-N/ condition exhibit strikingly similar trajectories.

⁵Because nasalance measurements were taken at 11 equidistant timesteps for each segment, Figure 4.10 involves some distortion of time. For instance, glides were typically much shorter than either of the vowels in a given token, and so the timesteps during [G] in Figure 4.10 are not strictly comparable to those for the adjacent vowels [V:].

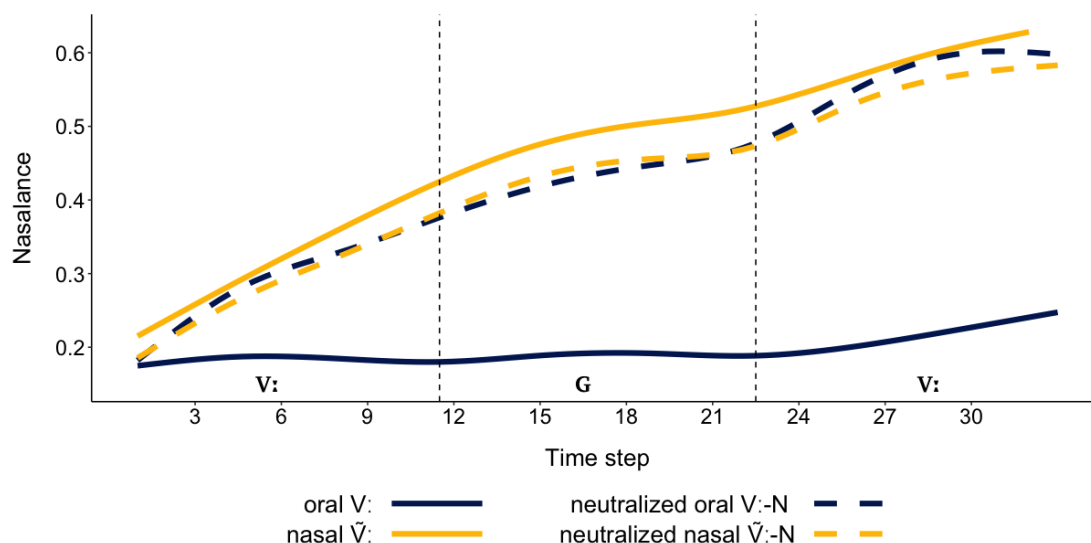


Figure 4.10: GAM curves separated by condition showing mean vowel nasalance at 33 normalized timesteps across the /V:GV:/ sequence for all tokens in experiment 2 produced by 16 speakers. Vertical dashed lines represent segment boundaries.

4.3.2.1 Neutralization on the final vowel

Table 4.6 summarizes the results of a linear mixed-effects model fit to the nasalance data of the final vowel of the underlying nasal / $\tilde{V}:/$, neutralized nasal / $\tilde{V}:-N/$ and neutralized oral / $V:-N/$ conditions. The interaction terms in the final three rows show a significant, but practically minimal (at most .017 difference in nasalance), effect at all three Timesteps of the final vowel. This means that nasal / $\tilde{V}:/$ and neutralized oral / $V:-N/$ vowels exhibit distinct nasalance trajectories across the vowel, albeit at relatively modest levels.

Examining the main effect of vowel condition, while neutralized nasal / $\tilde{V}:-N/$ approached significance in its nasalance difference with underlying nasal / $\tilde{V}:/$, nasalance of neutralized oral / $V:-N/$ was significantly different. These decreased nasalance values in the two nasalance conditions may be due to differing post-vowel contexts. That is, the final nasal / $\tilde{V}:/$ vowels always occurred in an open

syllable, with the immediate context being a pause. Speakers may have initiated oral occlusion toward the end of the vowel while velopharyngeal port closure lagged behind due to the independent timing of velic and lingual gestures (Byrd et al., 2009; Krakow, 1989, 1993). This would increase nasalance values relative to pre-N contexts at the end of phrase-final vowels, as observed in the present study.

Predictor	β	95% CI	t	p
Intercept (nasal / $\tilde{V}:/$)	.636	[.581, .692]	22.51	< .001 ***
neutralized nasal / $\tilde{V}:-N/$	-.043	[-.087, .001]	-1.93	.06
neutralized oral / $V:-N/$	-.066	[-.109, -.023]	-3.04	.006 **
Beg	-.040	[-.043, -.036]	-21.83	< .001 ***
Mid	.006	[.002, .01]	3.06	.002 **
Fin	.034	[.030, .038]	15.63	< .001 ***
u:	.096	[.053, .138]	4.4	< .001 ***
i:	-.045	[-.073, -.016]	-3.11	.005 **
o:	-.005	[-.052, .042]	-.22	.83
ɑ:	-.046	[-.071, -.02]	-3.52	.002 **
neutralized oral / $V:-N/$.Beg	-.012	[-.017, -.008]	-5.32	< .001 ***
neutralized oral / $V:-N/$.Mid	.007	[.002, .011]	2.68	.007 **
neutralized oral / $V:-N/$.Fin	.006	[0, .011]	2.09	.04 *

Table 4.6: LMER model summarizing the effects of CONDITION (3 levels), TIMESTEP (3 levels), VOWEL QUALITY (4 levels), and the interaction between CONDITION and TIMESTEP on mean nasalance of the final stem vowel for all tokens uttered by 16 speakers in experiment 2: LMER(nasalance~CONDITION*TIMESTEP + VOWEL QUALITY + (1 + CONDITION|speaker) + (1|word)).

To assess practical equivalence in nasalance for the final vowel between the two neutralized conditions (oral / $V:-N/$ and nasal / $\tilde{V}:-N/$), I defined a Region of Practical Equivalence (*RoPE*) following the procedure used in previous sections, using the final stem vowel in the baseline nasal / $\tilde{V}:/$ condition; this yielded *RoPE* = $\pm.119$.

I then extracted a pairwise comparison between the two neutralized vowel conditions using the emmeans() package in R (Lenth, 2024), as in the previous sec-

tion. Table 4.7 summarizes the pairwise comparison between the nasalance of neutralized oral /V:-N/ vowels and neutralized nasal / $\tilde{V}$:-N/ vowels, showing no significant difference in nasalance between the two vowel types.

Comparison	β	95% CI	t	p
(neut. oral V:-N) – (neut. nasal $\tilde{V}$:-N)	-.023	[-.076, .030]	-1.03	.91

Table 4.7: Pairwise comparison showing the effect of vowel condition on nasalance. Results were extracted using `emmeans()`, which averages over other predictors in the model: `emmeans(lmer-exp2, ~ condition)`.

Comparing the 95% confidence interval from this pairwise comparison to the *RoPE*, the interval lies well within $\pm.119$, thus replicating the experiment 1 results that no practically meaningful difference in nasalance between the stem-final neutralized oral and neutralized nasal vowels exists.

4.3.2.2 Neutralization of anticipatory nasalization

To examine potential differences in anticipatory coarticulation between the neutralized oral /V:-N/ and neutralized nasal / $\tilde{V}$:-N/ conditions, I fit an additional LMER model to nasalance across the V:G sequence. The fixed- and random-effects structure matched that of the model summarized in Table 4.6, except that timestep was coded as a six-level factor (three bins for both the penultimate vowel and glide). The results – summarized in Table 4.8 – show no evidence for a difference in anticipatory coarticulation between the two neutralized vowel conditions.

Predictor	β	95% CI	t	p
Intercept (Neut. Oral /V:-N/)	.359	[.278, .44]	8.7	< .001 ***
Neut. Nas. / $\tilde{V}$:-N/	.027	[-.045, .100]	.73	.47
penult Beg 1/3	-.083	[-.093, -.072]	-14.81	< .001 ***
penult Mid 1/3	-.043	[-.053, -.032]	-8.22	< .001 ***
penult Fin 1/3	-.002	[-.012, .008]	-.41	.68
glide Beg 1/3	.013	[.003, .023]	2.57	.01 *
glide Mid 1/3	.046	[.036, .056]	8.91	< .001 ***
glide Fin 1/3	.068	[.058, .078]	13.55	< .001 ***
u:	-.004	[-.013, .005]	-.96	.33
i:	.010	[-.003, .022]	1.48	.14
e:	-.060	[-.073, -.046]	-8.65	< .001 ***
ə	.087	[.073, .101]	12.25	< .001 ***
ɑ:	-.039	[-.045, -.032]	-11.37	< .001 ***
j	-.051	[-.075, -.028]	-4.24	< .001 ***
v	-.028	[-.052, -.004]	-2.30	.02 *
Neut. Nas. / $\tilde{V}$:-N/:penult Beg 1/3	-.008	[-.019, .004]	-1.33	.18
Neut. Nas. / $\tilde{V}$:-N/:penult Mid 1/3	-.008	[-.017, .001]	-1.82	.07
Neut. Nas. / $\tilde{V}$:-N/:penult Fin 1/3	-.003	[-.011, .004]	-.84	.4
Neut. Nas. / $\tilde{V}$:-N/:glide Beg 1/3	.009	[.002, .017]	2.42	.015 *
Neut. Nas. / $\tilde{V}$:-N/:glide Mid 1/3	.010	[.002, .018]	2.38	.017 *
Neut. Nas. / $\tilde{V}$:-N/:glide Fin 1/3	-.0004	[-.008, .007]	-.11	.91

Table 4.8: LMER model summarizing the effects of CONDITION (2 levels), TIMESTEP (6 levels), SEGMENT QUALITY (7 levels), and the interaction between CONDITION and TIMESTEP on mean nasalance of the V:G sequence for all tokens uttered by 16 speakers in experiment 2: LMER(nasalance~CONDITION*TIMESTEP + SEGMENT QUALITY + (1 + CONDITION|speaker) + (1|word)).

To test for equivalence, I established six *RoPE*s using data from the underlying nasal $\tilde{V}$: condition, one for each timestep included in the LMER analysis. For each *RoPE*, the speaker with the narrowest standard deviation for that given time bin set the threshold, yielding a conservative criterion for equivalence between conditions. Table 4.9 displays six pairwise comparisons between the two neutralized conditions — one per time bin — along with columns reporting the corresponding *RoPE* and whether the 95% confidence interval (CI) of the difference lies entirely

within it. No comparison shows a significant nasalance difference between conditions. Moreover, the 95% CIs for four of the six comparisons fall within their respective *RoPE*s; only the first two time bins fail to meet the equivalence criterion. This likely reflects onset-related differences that affect the early trajectory of nasalance: in the neutralized nasal / $\tilde{V}$:-N/ condition, 3 of the 8 onsets were liquids, whereas in the neutralized oral /V:-N/ condition, all onsets were obstruents (cf. Pouplier et al. 2024b).

/V:-N/ - / $\tilde{V}$:-N/	β	95% CI	t	p	<i>RoPE</i>	<i>Equivalent?</i>
penult Beg 1/3	-.020	[-.093, .054]	-.52	.60	$\pm$.082	No
penult Mid 1/3	-.019	[-.092, .054]	-.51	.61	$\pm$.083	No
penult Fin 1/3	-.024	[-.097, .049]	-.64	.52	$\pm$.196	Yes
glide Beg 1/3	-.036	[-.109, .036]	-.98	.33	$\pm$.173	Yes
glide Mid 1/3	-.037	[-.110, .036]	-1.00	.32	$\pm$.159	Yes
glide Fin 1/3	-.027	[-.099, .046]	-.72	.47	$\pm$.145	Yes

Table 4.9: Pairwise comparisons show the effect of vowel condition on nasalance. Results were extracted from the fitted LMER using `emmeans()`, averaging over the other predictors in the model.

Regardless, the evidence presented here overwhelmingly suggests that the neutralization of vowel nasality on oral vowels is, by all measurable accounts, phonetically complete, both on the vowel itself and in relation to anticipatory nasalization.

4.4 Discussion

The experiments presented above provide strong evidence for a case of phonetically complete neutralization. While other comparable examples of phonetically complete neutralization have been reported in the literature, such cases remain relatively rare, in part because the evidential bar is higher. That is, it is not enough to show the absence of an effect, but rather statistical equivalence — es-

established with the use of a *RoPE* — between all potential phonetic correlates of two underlyingly contrastive categories must also be established. This was the aim of the present experiments. Only a handful of other studies have reported complete neutralization, including the loss of a manner distinction between stops and fricatives in Korean (Kim and Jongman, 1996), a voicing contrast neutralization between word-initial stops in the Southern Bavarian variety of Tyrolean (Vietti et al., 2018), the neutralization of vowel length in open syllables in Dutch (Lahiri et al., 1987), pre-stop nasal consonant place assimilation in English (Ellis and Hardcastle, 2002), and pre-stop place assimilation of coronal stops in Korean (Kochetov and Pouplier, 2008).

By contrast, cases of incomplete neutralization are much more widely attested, since the evidential burden is lighter. Namely, it suffices to detect a residual difference between two underlyingly contrastive forms in any phonetic property. Well-known examples include word-final obstruent devoicing in German (Port and O’Dell, 1985), Catalan (Dinnsen and Charles-Luce, 1984), Dutch (Warner et al., 2004), and Russian (Matsui, 2015); Japanese vowel lengthening in monomoraic forms (Braver, 2013, 2019; Braver and Kawahara, 2016); flapping in American English (Braver, 2014); morphological tone in Cantonese (Yu, 2007); tone sandhi in Standard Mandarin (Zeng et al., 2025) and Huai’an Mandarin (Du and Durvasula, 2022, 2025); and coda aspiration in Eastern Andalusian (Gerfen, 2002). For at least some of these examples, certain methodological task effects (e.g., hints from orthography or the inclusion of minimal pairs creating priming effects) could artificially induce incomplete neutralization (Du and Durvasula, 2025, p.21-23), but there are clear cases of incomplete neutralization that have been replicated, as in the case of word-final obstruent devoicing in German (Roettger et al., 2014).

The existence of both complete and incomplete neutralization raises a broader question: which phonological theories can effectively account for both types of neutralization? In recent years, incomplete neutralization has often been treated as a poster child for usage-based approaches of linguistic representation. A few features of the theory make this true. First, under usage-based approaches, speech representations are conceived as networks of phonetically rich episodic memory traces (or exemplars) formed through language use, meaning that detailed traces of prior productions are always available at varying degrees of strength as influencers to future speech utterances (Bybee, 2023, p.14–15). Second, exemplars that are semantically and/or phonetically similar influence one another’s representations and productions more strongly, leading to subtle carry-over effects across items in memory (Bybee, 2001, p.23). In line with this, grammar itself is said to *emerge* with repeated use across tokens and is not separate from the network of lexical representations (Bybee, 2001, p.3, 26–27). In other words, massive amounts of practiced repetition of patterns across words and individual speech creates an *entrenchment* effect, in which a pattern that is used repeatedly is reinforced. Over time, this reinforcement creates the effect of a systematic, rule-like, pattern without a formal grammar needed to enforce it. Finally, token frequency plays a central role: more frequent exemplars become strengthened in memory and therefore exert greater influence on production patterns than less frequent tokens (Bybee, 1985, pp. 119–123).

Taken together, these assumptions predict that phonetic traces of underlying contrasts will survive in the lexicon and resurface in production, yielding gradient cases of incomplete neutralization. The neutralization of word-final obstruent voicing in German provides a clear example. In German, voiced obstruents un-

dergo devoicing in word-final position, resulting in the phonological neutralization of voiced and voiceless obstruents (e.g., both /bad/ ‘bathroom’ and /bat/ ‘ask’ surface as [bat]). However, Port and O’Dell (1985) demonstrated that this devoicing is phonetically incomplete: there is a small but reliable acoustic difference in the duration of preceding vowels depending on the underlying voicing of the obstruent. Specifically, the vowel in /bad/ → [bat] ‘bathroom’ is consistently longer than the vowel in /bat/ → [bat] ‘ask’.

A usage-based approach explains this asymmetry through the highly associative structure of the lexicon. Semantically and phonetically similar forms, such as [bat] ‘bathroom’ and [badə] ‘bathrooms’, are closely linked in memory, allowing stored productions of the plural form to influence productions of the singular form. This association leaves subtle but systematic phonetic traces of the underlying voicing contrast in [bat] ‘bathroom.’ By contrast, [bat] ‘ask’ lacks semantically related forms that maintain voicing on the stop, so no comparable effects arise in its exemplars. The end result is incomplete neutralization: the voicing cues in [bat] ‘bathroom’ differ consistently, though subtly, from those in [bat] ‘ask’.

In contrast to usage-based theories, incomplete neutralization has been viewed as a thorn in the side of generative theories that maintain static symbolic representations and a modular separation between phonology and phonetics.. In classic generative approaches (Chomsky and Halle, 1968), symbolic representations serve as the input to the phonological grammar, and if two distinct representations are neutralized by the grammar, their output from the phonological grammar will be identical. Under a strictly modular interpretation of this architecture, the phonetics component operates over these neutralized outputs rather than directly over lexical representations themselves. As a result, the persistence of system-

atic phonetic differences between neutralized forms is surprising. In the German word-final devoicing example, then, when the phonological grammar neutralizes the obstruent voicing contrast between /bat/ ‘ask’ and /bad/ ‘bathroom’, both forms are mapped onto the same surface representation [bat]. The existence of consistent phonetic differences between these outputs therefore raises the question of how information about the underlying contrast continues to influence phonetic realization despite phonological neutralization.

In sum, incomplete neutralization is relatively straightforward to explain in usage-based models and has engendered questions about the classic generative model of speech production.

In comparison to the rich literature on the role of incomplete neutralization in understanding the speech production process, phonetically *complete* neutralization has received little attention. In what follows, I evaluate potential explanations for both types of neutralization under usage-based and generative frameworks.

4.4.1 Accounting for neutralization in generative frameworks

The classic generative view of speech production holds that the phonological and phonetic components of the grammar are organized in a feed-forward relationship. Underlying forms (URs) are first filtered through a phonological grammar, where phonological processes apply in every relevant environment. The output of this system is a surface representation (SR), which then serves as the input to the phonetic component of the grammar. In this stage, the features of the SR are implemented in time and space, yielding the phonetic representation (PR) of the word.

While the basic architecture just described remains uncontroversial, views on what additional *performance* factors beyond the output of the phonological grammar can serve as inputs to the phonetics module, thus impacting the productions of lexical items. Additional factors that have been proposed in the literature include phonological planning (Du and Durvasula, 2025), contrastive intent (Nelson and Heinz, 2025), and cascading activation (Zeng et al., 2025), among others.

Given the findings from extensive work in the usage-based tradition, it is likely that the number of influences on the phonetic realization of lexical items is large – and likely includes phonetically detailed traces of actual exemplars – each of which must be carefully specified. For neutralization in particular, phonetically complete neutralization follows naturally in a classic generative framework if the effects of such additional factors are negligible. That is, if the output of the phonological grammar (the SR) is the only input to the phonetic component with sufficient strength to impact its shape, we should expect to observe complete phonetic neutralization. For example, the contrasting vowels in the two Mankiyali forms, /dɪ:-ŋ/ ‘giant-GEN.F.SG’ and /dĩ:-ŋ/ ‘pine firewood-GEN.F.SG’ undergo phonological neutralization of vowel nasality when filtered through the phonological grammar. Though they originate from distinct URs, a phonological process requiring vowels to be [+NAS] before a nasal consonant yields identical SRs: [dĩ:-ŋ]. Assuming no other factors play a significant role, these SRs should be indistinguishable to the phonetic system and should therefore produce indistinguishable productions (i.e., phonetically complete neutralization) precisely as observed in the experiments reported here.

In cases of incomplete neutralization, however, generative approaches must assume some other factor is impacting the production of otherwise phonologically

neutralized forms. Specifically, the lexical identity of neutralized forms appears to introduce an additional factor that serves as another input to the phonetic component. The precise nature of this factor remains a matter of theoretical debate, but the assumption that it exists is fully compatible with classic generative phonology. In fact, at least three recent accounts have proposed explicit mechanisms for how lexical identity functions in this way, which I summarize below.

4.4.1.1 The Blueprint Model of Production

The Blueprint Model of Production (BMP), proposed in Nelson and Heinz (2025), offers a modified model of the speech production process within the generative framework that has the potential to capture both complete and incomplete neutralization. While preserving the view that phonology operates over discrete representations that are fed into the phonetics, the BMP reconceptualizes the two modules as distinct higher-order functions, which can either take functions as inputs or return functions as outputs. Under this view, the phonetic component constructs the phonetic realization of a word not only from the output of the phonological component, but also potentially from additional inputs.

In cases of incomplete neutralization, Nelson and Heinz (2025) identify three relevant inputs to the phonetics module: (1) the SR, (2) the set of all URs in the lexicon that can serve as inputs to the phonological component, and (3) the *contrastive intent* (I) of the speaker to maintain a phonetic distinction between two neutralized URs. In this way, the underlying makeup of a lexical item can influence its PR.

In practice, the Intent of Contrast (I) operates as a scalar weight, modulating the relative influence of the UR or SR on the outcome of the PR. The equation in (4)

from Nelson and Heinz (2025) illustrates the inverse relationship of I between the influence of the UR and SR on the PR.

(4) Blueprint Model of Production and Neutralization

$$PR = UR \times I^2 + SR \times (1 - I^2)$$

To determine the phonetic realization of a word in any given context, the UR and SR are scaled by the appropriate contrastive intent (I) value in that context. As I approaches zero, indicating that the speaker does not intend to highlight the underlying contrast of the neutralized words, the effects of the UR on PR diminish exponentially, thereby yielding phonetically complete neutralization. Conversely, as I approaches 1, the UR exerts increasing influence on the PR, producing varying degrees of incomplete neutralization. Such gradient effects in the degree of incomplete neutralization are precisely what Port and Crawford (1989) find in their experiment on word-final obstruent devoicing in German, in which the relative presence of voicing contrast cues positively correlated with a speaker’s conscious intent to maintain an underlying obstruent voicing contrast.

In sum, under the assumptions of the BMP, a generative framework operating over discrete symbolic representations can successfully model both incomplete and complete neutralization.

4.4.1.2 Incremental Phonological Planning Effects

A different explanation for incomplete neutralization is proposed by Du and Durvasula (2025), who attribute the effect to incremental phonological planning. The key idea is that in multi-morphemic words, speakers construct surface phonological representations (SRs) incrementally, revising them as new information becomes available.

For instance, in the case of incomplete neutralization of German obstruent voicing word-finally, a speaker may begin processing the initial morpheme in the UR /bad/ ‘bathroom’ without yet knowing the following context, and thus plans [bad] as the appropriate SR. Later, when it becomes apparent that the obstruent will surface in word-final position, devoicing applies and a revised SR [baɖ] is planned. As a result, a speaker now holds multiple ‘antagonistic’ SRs for the same production. Because more recently planned representations exercise a stronger influence on the eventual PR (Rundus, 1970), the outcome is skewed toward the most recent SR, [baɖ]. The result is incomplete neutralization with a small but consistent effect size on voicing cues.

Clearly the explanation of incomplete neutralization under this view is plausible. However, it is less clear how the approach accounts for cases of *complete* neutralization, like that of vowel nasality neutralization in Mankiyali. If incremental planning effects are indeed a universal property of human cognition responsible for incomplete neutralization, it is not readily apparent why such effects would emerge in the German devoicing case but not in Mankiyali or in other reported cases of complete neutralization.

One possibility is that the absence of planning effects in the Mankiyali experiment owes to the modality of the experimental task. Previous research suggests that written-word processing does not unfold strictly left to right (or right to left in Perso-Arabic scripts), but instead involves parallel character evaluation or whole-word processing (Coltheart et al., 2001; Rayner, 1998). Yet, given that many studies reporting incomplete neutralization also rely on written stimuli, it remains unclear why Mankiyali should diverge.⁶

⁶It is possible that infrequent reading of Mankiyali could correlate with part vs. whole-word reading strategies. However, all Mankiyali speakers that participated in these experiments are

In sum, the attribution of incomplete neutralization in cases like word-final obstruent devoicing in German to incremental phonological planning effects provides a clear explanation firmly rooted in the generative framework. But why these same planning effects do not preclude complete neutralization in Mankiyali remains an open question.

4.4.1.3 Cascading activation

Zeng et al. (2025) propose another performance explanation for deriving various degrees of phonetic neutralization within a generative framework. Planning during speech production moves through four stages: (1) conceptualization of the meaning one wants to communicate, (2) selection of the appropriate lexical item among alternatives (3) phonological preparation of the selected lexical item, and (4) articulation. Importantly, under a cascading-activation model, preparation of a lexical item’s phonological form can begin before the selection process is complete. Consequently, alternative representations, such as a lexical item’s UR, may remain active and influence articulation even after a different representation has been selected as the target output.

Applied to phonological neutralization, this architecture allows a pre-neutralization representation to leave phonetic traces in the production of an otherwise neutralized surface form. Zeng et al. (2025) illustrate this mechanism using Mandarin third-tone sandhi, in which an underlying tonal sequence /T3+T3/ surfaces as [T2+T3]. Although the first syllable is targeted as T2, activation associated with its underlying T3 representation may cascade into articulatory planning. The resulting derived T2 may therefore retain measurable phonetic properties of T3 and

adept at reading Urdu, which uses the same general script.

differ subtly from an underlying lexical T2. Under this account, the phonological grammar can neutralize the contrast categorically while cascading activation produces incomplete neutralization at the phonetic level.

Crucially, Zeng et al. (2025) propose that the strength of this residual influence depends on the processing resources available to the speaker. In their experiment, Mandarin speakers produced derived T2 more similarly (though not equivalent) to underlying T2 when simultaneously performing a verbal working-memory task. They interpret this result as evidence that the additional processing demand degraded the activation of non-target representations, thus reducing their impact on articulation. Cascading activation, therefore, predicts that stronger cascading activation will produce greater phonetic residue, whereas weaker cascading activation will shift productions toward more complete neutralization.

While cascading activation provides an avenue to explain the occurrence of both types of neutralization, it is unclear whether its predictions can account for the case of phonetically complete neutralization of vowel nasality in Mankiyali observed here. Specifically, phonetically complete neutralization was achieved despite no obvious presence of increased task load for Mankiyali speakers. All that was required was for speakers to read a word as it appears on the screen, a relatively simple task which, under the cascading activation account suggests that incomplete neutralization should have been observed in this study.

In sum, cascading activation provides a possible mechanism for deriving both complete and incomplete neutralization, but its application to the Mankiyali results remains uncertain. The present Mankiyali experiments did not manipulate working-memory load, so it is unclear why the experiment did not yield incomplete neutralization. The Mankiyali findings thus reveal an important explanatory

gap in the cascading activation account: phonetically complete neutralization occurred during a simple word-reading task, without any experimentally imposed increased task load. If cascading activation allows upstream representations to influence the phonetic realization of Mandarin tone sandhi under comparable production conditions, it remains unclear why activation of the underlying oral vowel representation in a form like /dɪːŋ/ → [dĩːŋ] ‘giant-GEN.F.SG’ leaves no detectable trace in Mankiyali.

The cascading activation account can, in principle, accommodate the Mankiyali findings by assuming that activation of the underlying vowel representation is sufficiently weak by the time articulation occurs. Nevertheless, without an independently motivated explanation for why cascading activation persists in some neutralization processes but not others, it does not yet predict the cross-linguistic distribution of complete and incomplete neutralization. A beneficial test would be to compare Mankiyali productions of neutralized forms under low-load and high-load conditions to determine whether processing demands modulate the degree of neutralization.

4.4.2 Accounting for neutralization in eliminativist models

One of the core arguments levied by proponents of usage-based theories against the discrete lexical representations of generative theories is the purported inability of discrete representations to account for phenomena like incomplete neutralization (Braver, 2019; Goldrick and Cole, 2023). However, as just discussed, this critique is only valid if one holds to a conception of generative theory in which the output of the phonological grammar is the *only* permissible input to the phonetics module. Conversely, as outlined above, eliminativist usage-based models have been argued

to account for incomplete neutralization naturally.

Much less attention, however, has been given to whether usage-based models can also account for phonetically complete neutralization. Crucially, the very mechanisms in usage-based models that predict incomplete neutralization make it difficult to capture complete neutralization. In Mankiyali, for instance, the present study found that the vowel nasality contrast is fully neutralized before nasal suffixes, with no residual differences between underlying oral and nasal vowels. Given the semantic and phonetic ties between forms such as [d̪ɪ] ‘giant’ and [d̪ɪ̃ŋ] ‘giant-GEN.F.SG’, an eliminativist usage-based approach may incorrectly predict carry-over influences from the oral form, making the observed complete neutralization of vowel nasality in the genitive form unexpected.

One direct lexicon-internal mechanism available to usage-based models is to appeal to token frequency effects. By positing that oral base forms like [d̪ɪ] ‘giant’ occur with sufficiently low frequency, one could weaken their influence on paradigmatically related forms like [d̪ɪ̃ŋ] ‘giant-GEN.F.SG’. In the next chapter, I develop a computational model simulating exemplar accumulation over time to evaluate whether frequency manipulation can, in fact, provide a viable usage-based account of phonetically complete neutralization. The simulations demonstrate that, without some form of linguistic knowledge external to the lexicon of representations, pure usage-based models fail to predict both phonetically complete and phonetically incomplete neutralization.

More precisely, eliminativist usage-based models lack a mechanism external to the lexicon (e.g., a grammar) that is able to regulate the relative strength of influence from various factors on the eventual production of a word.

Appendix

NOM CV:	NOM C $\tilde{V}$:	GEN CV:-N	GEN C $\tilde{V}$:-N	CV:N
i: 3P.SG.ERG	ĩ: Eid.NOM			pi:ŋg swing (cf. pan)
pí: fear.NOM				si:n Indus River
si:d straight.NOM	sĩ: needle.NOM			
zi:b tongue.NOM	zĩ: louse.NOM		zĩ:-ŋ louse-GEN.F.SG	
ki: some	kĩ: insect.NOM		kĩ:-ŋ insect-GEN.F.SG	
pí:q food.NOM	bĩ: arms.NOM		bĩ:-ŋ arms-GEN.F.SG	pĩ:ŋ sister
qĩ: giant.NOM	qĩ: pine firewood.NOM			qĩ:m water dam
ʒi: then	ʒĩ: which		ʒĩ:-ŋ which-GEN.F.SG	
ki: that	kĩ: that which			kĩ:ŋ blind women
pu: drink.IMP	pũ: flatulence.NOM			pu:ŋ to filter
du: two.NOM	tũ: smoke.NOM		tũ:-ŋ smoke-GEN.F.SG	
lu: heatwave.NOM				lu:ŋ salt
gũr brown	gũ: feces.NOM		gũ:-ŋ feces-GEN.F.SG	
po: soil.NOM	põ:ts to approach	po:-ŋ soil-GEN.F.SG		
	jõ: 3.PL.ERG		jõ:-ŋ 3.PL-GEN.F.SG	
ko: who	kõ: why			ko:m nation (cf. urd)
kjó: cooking oil.NOM	gjø: wheat.NOM	kjó:-ŋ cooking oil-GEN.F.SG		
bɑ: joint disease.NOM	bã: arm.F.SG	bɑ:-ŋ joint disease-GEN.F.SG		bɑ:nd clear.F.SG
bɑ:y garden.NOM	bã:y call to prayer.NOM			
tɑ: fever.NOM	tã: thread.NOM			tɑ:ŋ tautness
sɑ: purlin.NOM		sɑ:-ŋ purlin-GEN.F.SG		
gɑ: cow.NOM	gã: village.NOM			gɑ:m long gate

Table 4.10: Mankiyali experiment 1 tokens, consisting of 62 monosyllabic words separated into five groups: oral (CV:), nasal (C $\tilde{V}$:), monomorphemic pre-nasal (CV:N), genitive oral (CV:-N), and genitive nasal (C $\tilde{V}$:-N).

Chapter 5

Simulating Complete and Incomplete Neutralization in an Exemplar-Based Learning System

In this chapter, I outline a usage-based computational model that simulates episodic speech memory accumulations for an individual speaker over time. Using this model, I show that, barring empirically unrealistic lexical frequency effects, eliminativist usage-based theories can effectively model *either* phonetically incomplete neutralization *or* phonetically complete neutralization, but not both. Specifically, I argue that variation in the stable end state of a phonologically neutralized segment's phonetic realization necessarily arises from differences in the strength of systematic production biases (Pierrehumbert, 2001, p.146-148) rather than from the substance contained within episodic exemplar memories themselves. These biases may originate from articulatory, perceptual, or other systematic pressures on speech production, but crucially, the *target* (or stable state) of a production bias is

not reducible to the phonetic content within the distribution of stored exemplars. Eliminativist usage-based frameworks assert that no additional mechanisms beyond the network of exemplar memories is needed to account for cross-linguistic phonetic and phonological patterns. Instead, linguistic structure is *emergent* from the characteristics of past exemplars and their relationships to and influence on one another. Consequently, the results of the simulations presented in this chapter suggest that, contrary to usage-based claims, the speech production process must include a global phonological mechanism capable of exerting planned speaker control over the intended phonetic output of neutralized forms (i.e., a bias toward a specific realization).

5.1 Central Tenets of Eliminativist Usage-Based Theories

Six core principles relevant for the current discussion form the basis of eliminativist usage-based frameworks of linguistic cognition: redundant representations, similarity-based association, frequency effects, emergent structure, articulatory bias, and decay. Each of these principles is discussed in more detail below. Usage-based theories are not unified into a single theory, but rather a cluster of theories with the shared assumption that language use — as opposed to an innate formal grammar — is the main driver of linguistic cognition. The usage-based model outlined here is based primarily on exemplar-theoretic models proposed in Bybee (2001), Pierrehumbert (2001), Wedel (2004), and Kaplan (2015), but these ideas are held by most proponents of eliminativist frameworks.

5.1.1 Redundant Representations

Whereas classic generative theories treat mental representations of morphemes as abstract symbolic categories with a single representation that omits predictable phonetic content, exemplar theories maintain that speakers store memories of individual sounds, whole words, and sometimes phrases (Wedel, 2004, p.33). Moreover, the phonetic detail associated with an individual exemplar memory is a central component of these representations, rather than being discarded as an irrelevant redundancy of phonetic implementation. For instance, the contextual realization of vowels before nasal consonants is a predictable process in American English, occurring in words like /pæn/→[p^hæ̃n] ‘pan’ (Cohn, 1993; Solé, 1992, 1995). Rather than stripping token-specific details present in an individual experience of [p^hæ̃n] from the representation, exemplar theories treat the fine-grained details of its nasality and context — such as subtle changes in the degree of nasality over the time course of the vowel and the social characteristics of the speaker who uttered the token — as the core substance of the representation.

Moreover, storage of these fine-grained details does not occur just a handful of times, but with every instance of a word heard or produced. This means that representations of words and phrases do not consist of a single form representing an entire category, but rather an interconnected ‘cloud’ of richly detailed exemplars that collectively contribute to a speaker’s stored knowledge of a linguistic form. For example, Figure 5.1 depicts a simplified exemplar cloud for the English word ‘pan’. I set aside higher dimensions of the exemplar network and the inclusion of social information for the sake of exposition. Each waveform in the cloud is an individual memory trace, and collectively, the traces as a whole constitute the language user’s linguistic knowledge of the word ‘pan’.

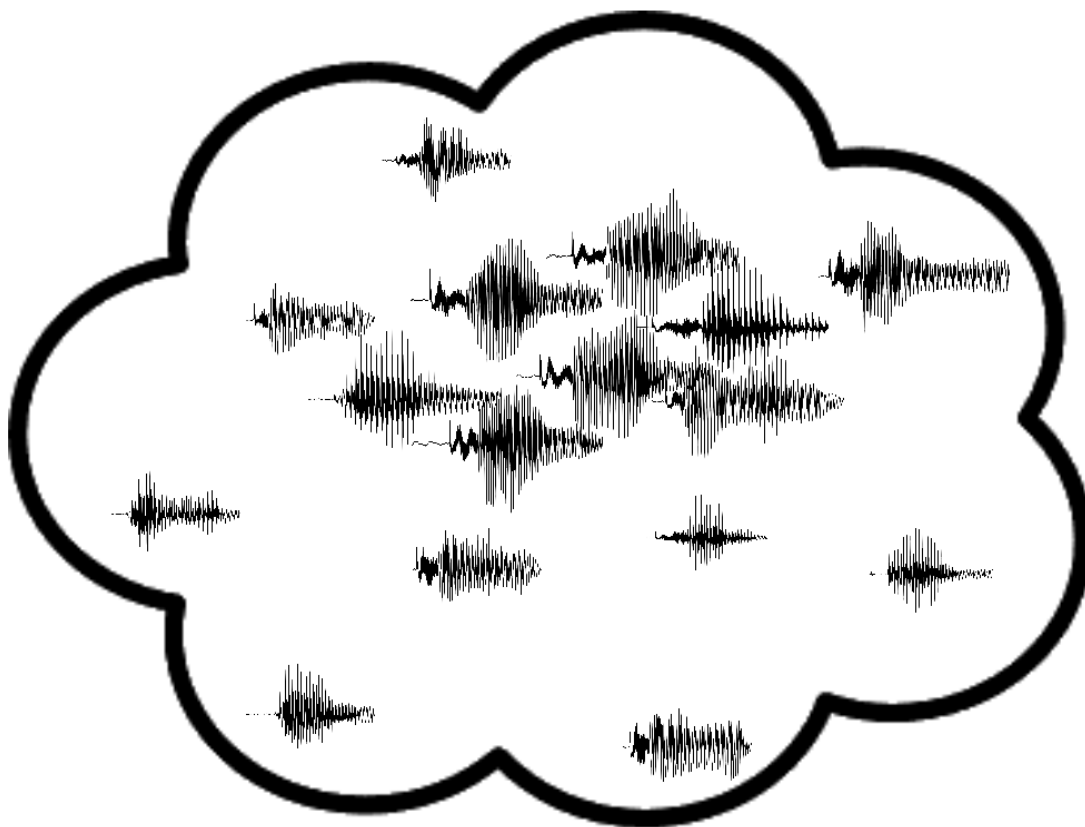


Figure 5.1: Hypothetical exemplar cloud for the English word ‘pan’.

5.1.2 Similarity-Based Association

Individual exemplar memories that are semantically and/or phonetically similar are strongly linked in memory (Bybee, 2001, p.22-23). The stronger these ties, the greater the influence an exemplar imposes on the future production patterns of its related forms. For example, the collective structure of exemplar memories of ‘pan’ contribute to shaping future pronunciation patterns of related words like ‘pans’, ‘pancake’, and ‘panhandler’.

However, Kaplan (2015) demonstrates that the influence exerted by paradigmatically related forms cannot be fully symmetric. In a series of exemplar-based simulations modeling positional neutralization through word-final devoicing, Ka-

plan implemented both (i) a production bias favoring devoicing in word-final position and (ii) a paradigmatic similarity bias that caused morphologically related forms to influence one another equally. Under this symmetric similarity bias, non-neutralizing forms containing voiced obstruents in non-final position (e.g., [gibi]) were gradually pulled toward the devoiced variants found in paradigmatically related neutralizing forms (e.g., [gip]), eventually losing the voicing contrast altogether. As a result, devoicing spread throughout the entire paradigm rather than remaining restricted to word-final position. In other words, the model converged on paradigms in which stops were devoiced in all forms whenever they surfaced word-finally in any member of the paradigm. However, when Kaplan modified the simulations so that paradigmatic influence was asymmetric — such that neutralizing forms were influenced by non-neutralizing forms, but not vice versa — the simulations successfully produced stable positional neutralization patterns in which voicing contrasts were preserved in non-final contexts while neutralization remained restricted to word-final position.

If non-neutralized forms of a paradigm were influenced by neutralized forms, phonological contrasts would, as Kaplan (2015) shows, attenuate over time. If, for instance, a vowel nasality contrast is neutralized before nasal consonants, as we observed in the Mankiyali word /d̪rːŋ/ → [d̪ĩːŋ] ‘giant-GEN.F.SG’ in chapter 4, under a symmetric influence model, the vowels in these neutralized forms should increase the degree of nasality in future productions of paradigmatically related non-neutralized forms, given their close ties within the exemplar network. This would result in a gradual erosion of the vowel nasality contrast in Mankiyali until all paradigms with at least one form containing a vowel in a neutralizing condition were fully nasalized throughout the paradigm.

However, if the influence between neutralized and non-neutralized forms is asymmetric, as argued by Kaplan (2015), we should only expect to see the influence of non-neutralized forms on related neutralized forms, so no such contrast attenuation is predicted. Under this view, lexical associations between neutralizing and non-neutralizing forms give rise to incomplete neutralization, as we observed for non-neutralized voiced stop in German words like [badə] ‘bathrooms’ exerting influence on the neutralized final devoiced stop in [bat] ‘bathroom’, yielding measurable phonetic traces of voicing in the neutralized form but no traces of devoicing in [badə].

5.1.3 Frequency Effects

Because exemplars accumulate over time, speech patterns within a given lexical item are primarily shaped by the highest concentration of exemplars — the center of gravity — within that lexical item. As more exemplars accumulate around this center of gravity, their specific phonetic details collude to direct production patterns of future exemplars, resulting in lexical *entrenchment* (Pierrehumbert, 2001, p.149-152), in which a specific phonological pattern is reinforced and eventually cemented within an exemplar cloud, barring external pressures from related processes or exemplar clouds. Returning to the hypothetical exemplar cloud for ‘pan’ in figure 5.1, a dense cluster of exemplars with similar phonetic properties has formed, while exemplars with less typical phonetic properties remain dispersed around the periphery. As a result, the dense cluster of exemplars exerts a disproportionate influence on future productions of the word ‘pan’, reinforcing the phonetic properties of the cluster over time. Eventually this cluster becomes so dense that the phonetic characteristics of ‘pan’ for this individual are entrenched

in this region of the cloud.

Entrenchment also takes effect at a more general ‘phonological’ level across related forms in the lexicon. Because word forms that are used more frequently accumulate exemplars more rapidly than less frequent forms (i.e., a higher token frequency), the sound patterns present in frequent forms direct the evolution of the exemplar network as a whole, to the effect that the patterns present in high-frequency forms permeate all relevant forms in the lexicon over time (Wedel, 2004). Exemplar theories use these predicted frequency effects to explain why phonological patterns of high-frequency lexical items exhibit an outsized influence on related forms and are simultaneously more resistant to undergoing phonological changes themselves (e.g., Bybee 2001).

5.1.4 Emergent Structure

Exemplar frameworks contend that the richly detailed and interconnected nature of the lexicon is sufficient to explain the presence of regular phonological patterns in a lexicon. That is, instead of positing the existence of a formal grammar through which representations are filtered to generate systematic phonological patterns across the lexicon, phonological patterns that are repeated with high degrees of frequency across exemplars and forms generate the appearance of a formal rule, when in reality the pattern is an emergent property of the exemplar network itself (Bybee, 2001, p.3-4). In this view, then, grammar and representations are inseparable; there is no independent level of abstract linguistic knowledge beyond individual exemplar memories and their relations.

For example, within the generative framework, the representation for English /pæn/ is filtered through a grammar that applies a pre-N assimilation rule, re-

sulting in the surface realization, [p^hæn]. This pre-N assimilation process applies systematically as a consequence of a categorical rule mapping underlying representations to surface forms.

Within the exemplar-theoretic framework, however, no such grammar exists. Instead, a speaker has accumulated many exemplars of forms with a pre-N vowel that is nasalized in a specific way, and these exemplars are stored with detailed phonetic information and linked to phonetically and semantically similar forms in the lexicon. The next time the speaker produces a word like ‘pan’, then, production is biased by the activation of these related exemplars, particularly those occupying dense regions of the exemplar cloud, such that forms with pre-N vowels shape the target production of the vowel in ‘pan’. The resulting production reflects the weighted influence of these stored exemplars rather than the application of a discrete rule. In this way, the appearance of a systematic process like pre-N assimilation emerges from the influence of related exemplars on future productions.

5.1.5 Articulatory Biases

According to Pierrehumbert (2001, p.147), historical sound change occurs when articulatory biases are introduced into a language. These production biases can take the form of lenition of a certain sound or, particularly relevant for the present discussion, nasal coarticulation on a preceding vowel. When such changes are permitted, they induce gradual shifts in the distribution of exemplar clouds over time. Ultimately, regardless of how long a category takes to arrive at a new stabilized production point, historical sound changes converge on a specific equilibrium state at which the articulatory bias is optimally satisfied.

Importantly, Pierrehumbert explicitly acknowledges that it is unclear how the initial change in the linguistic system that permits such biases arises within an exemplar account (i.e., the actuation problem, Yu 2023). As I will argue in more detail in section 5.3 based on the results of the model simulations, because the eliminativist exemplar framework rejects linguistic knowledge beyond the network of stored exemplar representations, the theory not only lacks a mechanism to explain how and why such biases arise (a problem shared by all theories), but it also lacks a clear way to implement the fact that patterns across languages and phenomena stabilize at distinct levels of phonetic neutralization.

5.1.6 Decay

Memories are not permanent if not strengthened through repetition, they fade and decay over time (Pierrehumbert, 2001, p.140-141). Exemplar memories residing in the outskirts of an exemplar cloud distribution are especially susceptible to decay due the lack of nearby exemplars to reinforce their activation. Decay is clearly affected by recency as well, in which exemplars heard or produced in the more distant past are more likely to decay.

5.2 Modeling sound change through exemplar accumulation

The main function of the present model is the simulation of the contextual neutralization of a vowel nasality contrast through the gradual accumulation of exemplars over time.¹ Constructed from the six principles of exemplar theory outlined above,

¹Morley (2023) develops a similar class of models to the one outlined here, and her results implicitly entail the need for linguistic mechanisms external to exemplar representations in order

the model demonstrates that changes in a speaker’s nasality patterns converge toward a stable state of nasality that maximally satisfies the competing pressures shaping the exemplar space. Although different starting points of nasality may alter the *rate* of convergence, the model consistently settles on the same nasality state unless the global parameters that determine the relative strength of competing pressures are intentionally manipulated by the speaker.

The model takes as an initial input words from four inflectional categories with pre-specified frequency distributions. Exemplars accumulate within each category over time based on these frequencies through an iterative process of exemplar creation, optimization, and storage. The output of the model is a plot showing the evolution of nasality within these four categories over time and the stable state that each category arrives at. The Python code for each model presented in this chapter can be found at https://github.com/joncparamore/Exemplar_Theory_Modeling_Neutralization.

5.2.1 Model input

The four input categories are comprised of two ‘non-neutralizing’ forms with contrastive oral (ORAL) and nasal (NASAL) vowels respectively, and two ‘neutralizing’ forms inflected with a nasal suffix (ORAL-N and NASAL-N). As summarized in Table 5.1, each of the four categories in the model contains two lexical items, yielding eight distinct word-level exemplar clouds. To simplify the exposition, I focus on a single phonetic property of these word categories, nasality. In reality, exemplar categories for words exist in a high-dimensional space, with a large number of phonetic, contextual, and social factors influencing the shape of the cloud. I set

to achieve a specific stable state in an exemplar system.

these complexities aside in what follows.

Oral	Nasal	Oral-N	Nasal-N
tɑ:	tã:	tɑ:-n	tã:-n
ki:	kĩ:	ki:-n	kĩ:-n

Table 5.1: The model input consists of four categories with two lexical items each, for a total of eight exemplar clouds.

At initialization, each of the eight clouds is seeded with a single exemplar, the substance of which is a single scalar value bounded between 0 and 1 that corresponds to the mean nasality of the exemplar’s vowel. Nasality values near 0 correspond to minimal or absent nasality, whereas values near 1 indicate a completely nasal segment. Unless otherwise noted, the initial exemplars in the two oral categories are randomly drawn from a normal distribution with $\mu = .2$ and $\sigma = .025$, while the initial exemplars in the two nasal categories are randomly drawn from a normal distribution with $\mu = .6$ and $\sigma = .025$. These distributions approximate the baseline nasalance values of oral and nasal vowels from the Punjabi and Mankiyali data in chapters 2 and 4. Thus, the initial state of nasality in the model lacks any neutralization of nasality in the neutralizing ORAL-N and NASAL-N forms.

Finally, the relative frequency of the lexical items is also specified at model initialization. These frequencies determine the probability that a given lexical item is selected to generate a new exemplar during each iteration of the accumulation process. Unless otherwise noted, selection probabilities were uniform across the eight lexical items, with each item having a .125 probability of being chosen to generate a new exemplar in a given iteration.

5.2.2 Exemplar creation, optimization, and storage

To accurately model the core principles of exemplar theory outlined in section 5.1, the exemplar accumulation process proceeds in six ordered steps: (1) selection, (2) entrenchment, (3) optimization, (4) noise, (5) storage, (6) and decay. These steps are repeated 1,000 times in a single simulation to model potential changes in the distribution of exemplar clouds over time. To clearly demonstrate the implementation of each component in the model, I begin with a minimalist model and slowly build up to a full-fledged model that incorporates all the necessary exemplar-theoretic principles.

5.2.2.1 Model 1 - Selection, Assignment, Storage

At minimum, the exemplar accumulation process requires the selection of a lexical item for the creation of a new exemplar, the assignment of a nasality value to the new exemplar, and the storage of that exemplar in the appropriate cloud.

For each iteration, selection occurred by randomly selecting one of the four categories (ORAL, ORAL-N, NASAL, or NASAL-N) and then randomly selecting one of the two lexical items within that category for classifying the newly generated exemplar. When frequency is equivalent across categories and lexical items, each of the lexical items has a .125 probability of being selected. After the specific lexical item is selected for classification of the new exemplar, the nasality value of one of the current exemplars in that cloud is randomly chosen as the value for the new exemplar. The new exemplar is then stored in its associated exemplar cloud, and the next iteration commences. Figure 5.2 shows the output of a model with these specifications after 1,000 iterations.

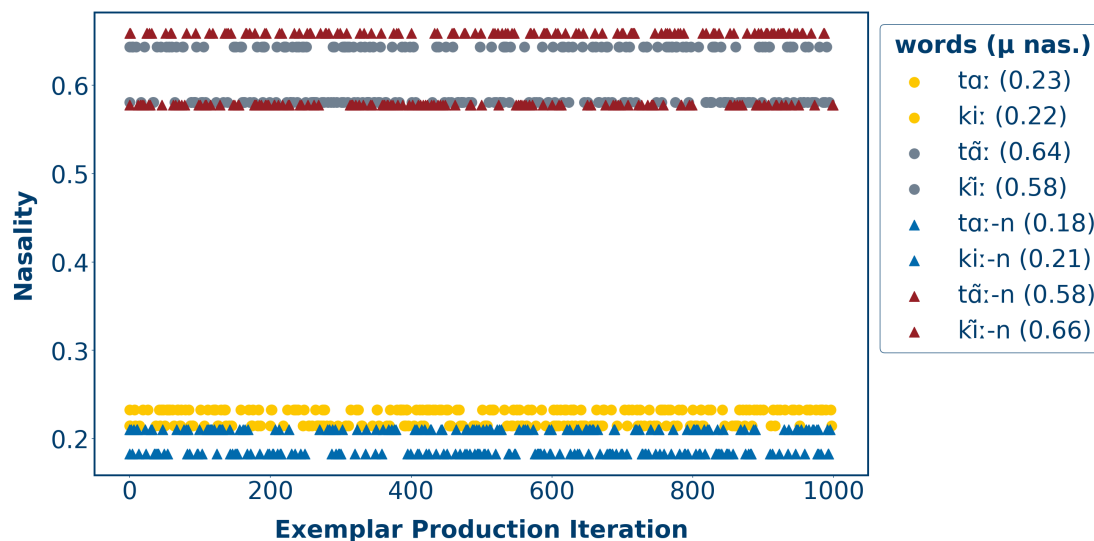


Figure 5.2: Model 1: selection, assignment, storage. The plot shows nasality values of individual exemplars over 1,000 iterations. Non-neutralizing ORAL exemplars are in yellow, non-neutralizing NASAL exemplars in grey, neutralizing ORAL-N exemplars in blue, and neutralizing NASAL-N exemplars in red. Values in the key show mean nasality of each lexical item’s cloud at the end of the simulation.

While the model shows the accumulation of exemplars across the eight lexical clouds, the nasality of all exemplars within each cloud are equivalent. This arises because each cloud is initialized with a single exemplar, and all subsequent exemplars are assigned the same value without deviation. As a result, each word’s exemplars exhibit a static nasality identical to the value of the initial exemplar.

5.2.2.2 Model 2 - Selection, Assignment, Noise, Storage

To account for inherent imprecision and variability in the production of a nasality target, random noise ($\sigma = .025$) was applied to the nasality value of each exemplar after being assigned its initial value. The results of this addition to the model are shown in Figure 5.3 and provide a more accurate illustration of the inherent variability in precision that occurs during the production of a phonetic target.

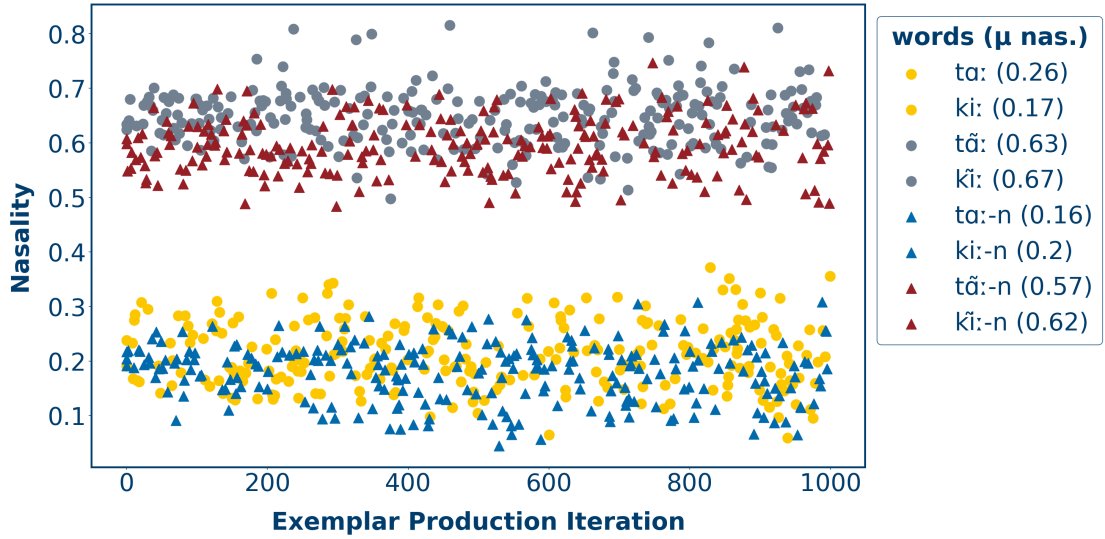


Figure 5.3: Model 2: selection, assignment, noise, storage.

5.2.2.3 Model 3 - Selection, Assignment, Pre-N Bias, Noise, Storage

While the addition of noise in model 2 more closely simulates realistic motor control variation from token to token within exemplar clouds, no effects have been included to instigate the gradual shift in nasality of the ORAL-N forms over time toward the neutralization of nasality. To achieve a pressure toward neutralization, an articulatory bias pressuring pre-N vowels to be more nasalized is needed.² As such, after a newly created exemplar's lexical identity is selected and its initial nasality copied from one of the current exemplars in that cloud, a pre-N bias favors an earlier onset of velar lowering during the pre-N vowel in anticipation of the following nasal consonant. This bias reflects a production ease pressure: initiating velum lowering during the vowel reduces the need for a rapid articulatory shift at the consonant boundary. In the model, this is encoded as a penalty on

²Morley (2023) develops another approach to implementing production biases. In her models, the bias and the target stable state are distinct mechanisms. Here, however, my implementation combines the production bias and the target into a single component. Specifically, the bias is implemented as a penalty for deviating from a certain phonetic value, and the target is the point at which the bias penalty is minimized.

insufficient nasality in pre-N vowels.

I implement this bias through the minimization of an objective function by gradient descent using `scipy.optimize` in Python. In the current model, the objective function value is computed solely by determining the deviation of an exemplar’s nasality from a typical nasal vowel in the NASAL category. As shown by the equation in (1), the exact penalty is calculated by squaring the difference between the nasality of the newly created exemplar and the mean nasality of exemplars in the NASAL category (i.e., the mean of the NASAL category). In effect, the penalization of the newly created exemplar will be exponentially larger as its nasality value diverges from that of a typical vowel in the NASAL category.³ Crucially, because the pre-N articulation bias only applies to forms with a pre-N vowel, the objective function value will be zero for exemplars selected from either the non-neutralizing ORAL or NASAL categories. That is, there is no articulatory pressure to lower the velum in anticipation of an upcoming nasal consonant in these forms, so the bias does not apply. For both ORAL-N and NASAL-N exemplars, on the other hand, the bias is applicable, albeit vacuously for NASAL-N exemplars.

³There are several advantages to defining production biases as squared deviations. First, squared deviations are continuously differentiable, which makes them well-suited for optimization via gradient descent. By contrast, an absolute-deviation formulation — $|NASAL\mu_{nas} - exemplar_{nas}|$ — yields a piecewise linear objective function whose derivative is undefined at the optimum and constant elsewhere, which often leads to less stable convergence behavior. Second, squared deviations introduce curvature into the objective function, which guarantees a unique interior minimum when multiple biases are in conflict. This property is crucial for modeling incomplete neutralization (section 5.2.2.6), as it allows the optimal value (i.e., the stabilization point in the system) to fall between competing targets rather than necessarily collapsing to the bias with the highest weight. In contrast, absolute-deviation penalties tend to produce boundary solutions (i.e., selection of one target over another) unless the competing weights are exactly equal, in which case the objective is flat between the targets. Finally, squared deviations penalize larger departures from the target more strongly than smaller ones, reflecting the theoretically grounded intuition that exemplars that are substantially misaligned with a bias should incur disproportionately greater cost than those that deviate only minimally.

- (1) Pre-N Nasality Bias
 $(NASAL\mu_{nas}-exemplar_{nas})^2$

An additional restriction on the optimization process involves the limitation of changes in an exemplar’s nasality to $\pm .075$ for a single iteration. Without such a restriction, the optimization function — which at this point is solely composed of the bias to maximize nasality on pre-N vowels — would instigate a dramatic shift in nasality at the outset of the simulation to maximally satisfy the pre-N nasality bias in (1), resulting in instant changes in both pre-N category’s nasality values toward the mean of the NASAL category’s nasality. In reality, sound change is gradual, and effects on individual exemplars are only slight. Capping changes in nasality for an individual exemplar to $\pm .075$ prevents one-fell-swoop changes and more closely imitates the gradual shift in categories over time.

Figure 5.4 shows the results of incorporating the pre-N bias into the model. Clearly, the pressure for pre-N forms to be more nasalized results in a gradual shift in nasality of ORAL-N exemplars towards a fully nasal vowel over time. Nevertheless, notice that ORAL-N exemplars with low nasality still appear throughout the simulation. These non-conforming exemplars arise more frequently early in the simulation because low-nasality exemplars constitute a larger proportion of the exemplar cloud at earlier stages. When generating a new exemplar, the model randomly selects an existing exemplar from the lexical item’s cloud to provide the initial nasality target prior to optimization, and all exemplars are equally likely to be selected. Consequently, because low-nasality exemplars remain present in the distribution, they can continue to seed newly created exemplars at any point in the simulation.

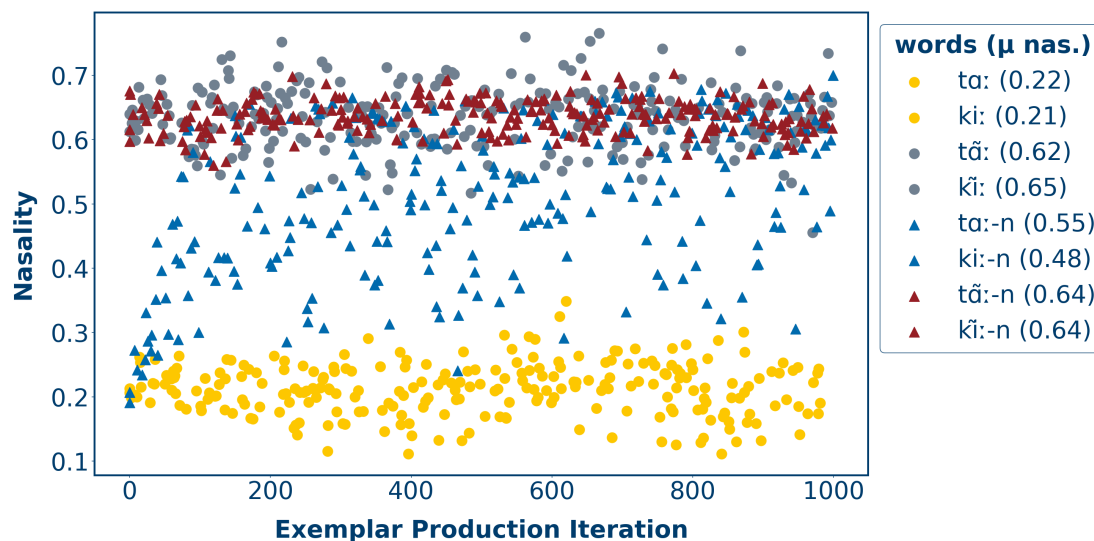


Figure 5.4: Model 3: selection, assignment, pre-N bias, noise, storage.

To demonstrate that the outcome presented in Figure 5.4 is consistent across trials, I conducted 100 independent simulations using the parameters of *Model 3* and recorded the final mean nasality of each lexical category at the end of each simulation. Thus, each category is represented by 100 mean nasality values, one from the endpoint of each simulation. Figure 5.5 summarizes the distribution of these endpoint means by plotting the mean nasality and 95% confidence intervals for each of the four word categories' final nasality means. As the figure shows, the ORAL-N category consistently shifts toward the nasality levels of the NASAL and NASAL-N categories. However, the possibility of selecting early, low-nasality exemplars as baseline forms introduces a persistent source of lower nasality forms. This interaction between the pre-N bias and inheritance of low nasality from early forms results in a stable mean nasality of approximately .51 for the ORAL-N category across simulations.

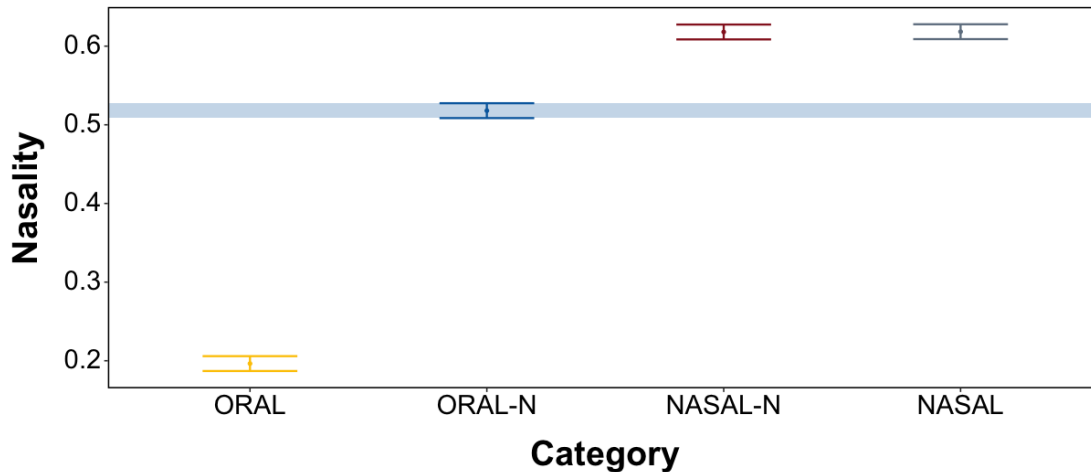


Figure 5.5: Model 3: selection, assignment, pre-N bias, noise, storage. The plot shows the average of the final mean nasality of the four word categories in the model over 100 simulations. Tails represent 95% confidence intervals around the means.

While the mean nasality of the two ORAL-N word forms did not reach the level of the two nasal categories when Model 3 was run over 1,000 iterations, this difference disappeared when the model was extended to 50,000 iterations, as shown in Figure 5.6. This convergence at higher iteration counts indicates that repeated production gradually reshapes the ORAL-N exemplar distributions toward higher nasality values. Over time, high-nasality exemplars become sufficiently dominant within the distributions that lower-nasality exemplars become increasingly less likely to seed new exemplars.

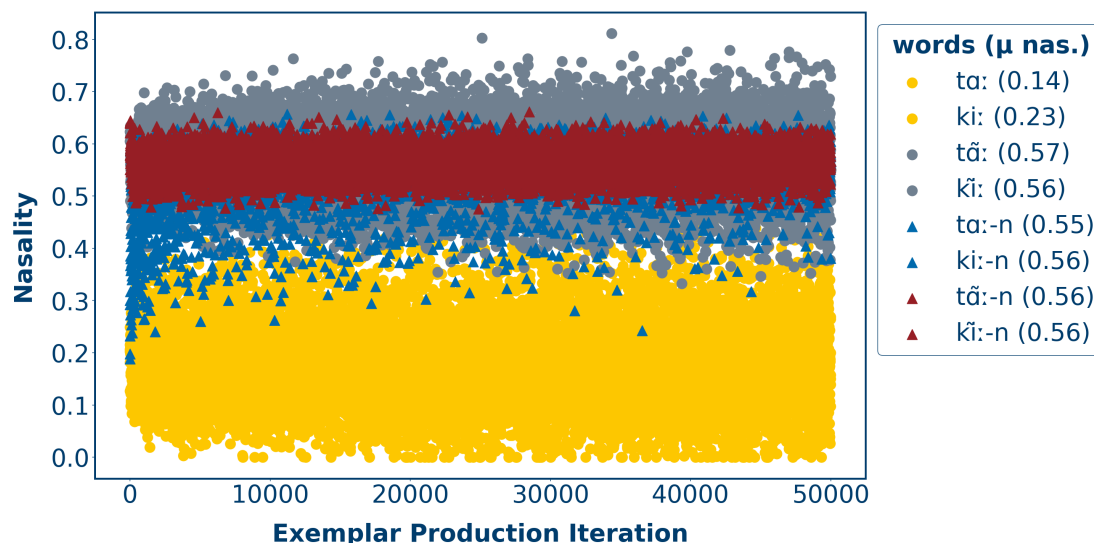


Figure 5.6: Model 3 but with 50,000 exemplars rather than 1,000.

Allowing newly created exemplars to inherit their nasality values from a single stochastically selected exemplar, rather than from the aggregate influence of the entire exemplar cloud, successfully models the entrenchment effects described in Section 5.1.3 at high iteration counts. However, visualizations of models with tens of thousands of iterations become increasingly difficult to interpret, as demonstrated by the visualization in Figure 5.6. To remedy this issue, I introduce a different method to model entrenchment in the next section: defining the initial nasality of newly created exemplars as the center of gravity (i.e., mean nasality) of an exemplar cloud.⁴

⁴An alternative approach to modeling entrenchment could combine the stochastic selection of an exemplar from the model's current distribution, which determines the base nasality of each new exemplar, with the continuous decay of older exemplars. Once decay is incorporated into the model (Section 5.2.2.5), entrenchment would emerge through the gradual accumulation of high-nasality exemplars alongside the loss of low-nasality exemplars produced primarily during earlier iterations.

5.2.2.4 Model 4 - Selection, Entrenchment, Pre-N Bias, Noise, Storage

Model 4 introduces an entrenchment mechanism to capture the tendency for speakers to progressively concentrate their neuromotor productions on the most practiced, frequently repeated patterns within a category’s distribution (Bybee, 2001, p.54). This type of entrenchment effect is evident in the production patterns of children, in which their articulations are initially highly variable, but as their vocabulary grows, children silo the use of articulations into a manageable set of highly practiced patterns (i.e., phonemes) that can be reused to facilitate the acquisition of new words (Bybee, 2001; Studdert-Kennedy, 1987). To model this, rather than randomly selecting an existing exemplar from anywhere in the cloud to serve as the template for a new token, this model uses the mean nasality of the lexical item’s exemplar cloud as the baseline for generating each new exemplar. After this baseline is assigned, the optimization process proceeds in the same manner as *Model 3*.

Using a category’s mean as the initial production baseline for new exemplars is one way to implement entrenchment. The mean provides a good approximation of a category’s most typical nasality value, and as additional exemplars accumulate, this mean becomes increasingly stable in the face of extreme productions. This approach also effectively models why higher token frequency increases a lexical item’s resistance to large-scale language changes sweeping across the lexicon (Bybee, 2001, p.11-12). In effect, the center of gravity of each category becomes progressively ‘cemented’ as the simulation proceeds, and newly generated exemplars reinforce this central value. Only a systematic articulatory bias or overwhelming frequency changes in related forms is capable of shifting a category away from its entrenched mean. This resistance to change in high-frequency categories explains

why morphological irregularities always occur in forms with a high usage-rate (e.g., English high-frequency ‘keep’~‘kept’~*‘keept’ resists analogical leveling compared to lower frequency ‘weep’~‘wept’~‘weept’ and ‘leap’~‘lept’~‘leaped’; Bybee 2001, p.12).

Because no external bias affects the original shape of the ORAL, NASAL, or NASAL-N categories, these categories will remain tightly clustered around their initial mean values throughout the simulation, with subtle variation arising solely from random noise. In contrast, the initial values of the ORAL-N category are antagonistic to the pre-N nasality bias introduced in *Model 3*. Consequently, the bias incrementally pushes the ORAL-N category mean upward, bounded by the .075 limit for each iteration. Because each new exemplar is produced from that evolving mean, the upward shift is reflected in the baseline values of new exemplars. Once the ORAL-N mean reaches nasality levels typical of a nasal vowel, the category stabilizes, and the stabilized nasality pattern becomes increasingly entrenched around the optimized mean.

Figure 5.7 depicts a representative simulation using this entrenchment-based baseline. As predicted, the three categories not affected by the pre-N bias show more tightly clustered nasality values with only small fluctuations due to random noise, illustrating the stabilizing force of entrenchment. Meanwhile, the ORAL-N exemplars — pressured by the pre-N bias — exhibit a clear upward trajectory toward the nasality space occupied by the NASAL and NASAL-N categories. Notably, because the baseline for new productions is the category mean rather than a randomly chosen exemplar, the ORAL-N categories converge to the nasality of the NASAL and NASAL-N forms more quickly. Modeling entrenchment by initializing new exemplars at the value of a category’s mean ensures that once the category

mean begins to rise, new exemplars follow suit.

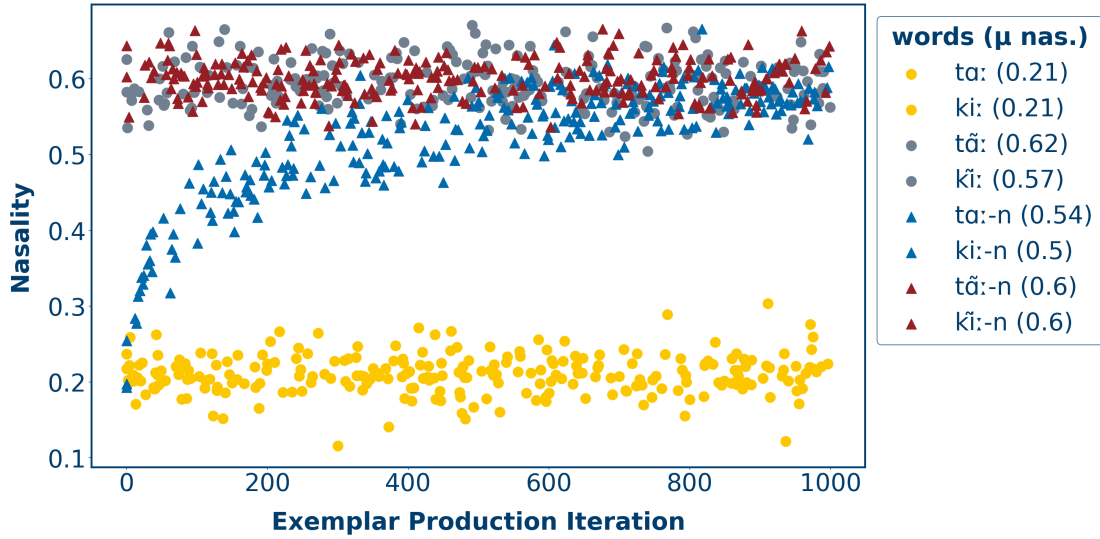


Figure 5.7: Model 4: selection, entrenchment, pre-N bias, noise, storage.

The dot-and-whisker plot in Figure 5.8 shows the estimated stabilized mean nasality and 95% CIs across 100 simulations for each of the four categories. The estimated mean of stabilized nasality for the ORAL-N category is shifted up around .53 due to implementation of entrenchment effects. However, because early exemplars with low nasality still contribute equally to the final nasality mean of the ORAL-N cloud, the nasality mean is not completely neutralized.

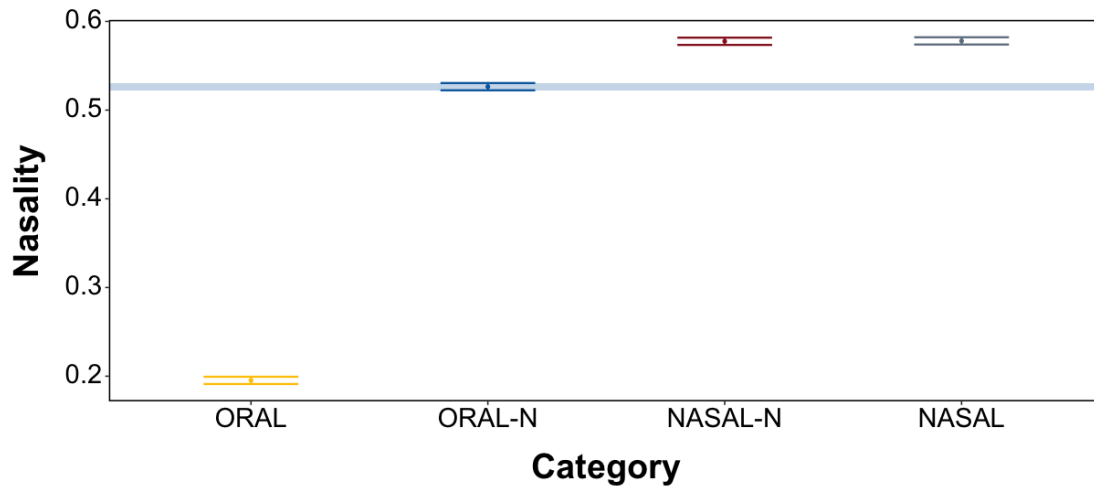


Figure 5.8: Model 4: selection, entrenchment, pre-N bias, noise, storage.

If the simulations were run with more iterations (e.g., 10,000) this difference in means would eventually be washed out by the increasing number of high-nasality exemplars, as shown by the visualization in Figure 5.9.

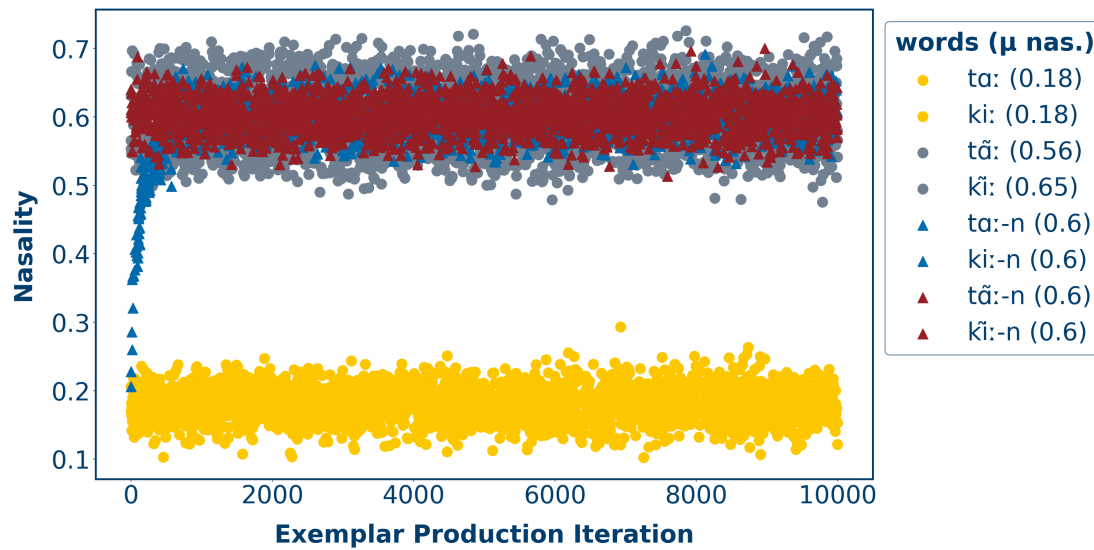


Figure 5.9: Model 4, run over 10,000 iterations.

5.2.2.5 Model 5 - Selection, Entrenchment, Pre-N Bias, Noise, Storage, Decay

To account for the fact that memories are not permanent (see section 5.1.6 above), *Model 5* incorporates an additional mechanism that approximates the decay of exemplars over time. This is implemented in the model by imposing a memory capacity limit of fifty exemplars per lexical form. Once a lexical form accumulates 50 exemplars, the generation of newly created exemplars in that lexical form's cloud triggers the removal of the oldest exemplar in the cloud, ensuring that it no longer contributes to the lexical form mean estimations, either during baseline nasality assignment within an iteration or in the final mean calculation at the end of a simulation.

Under normal model conditions, in which all lexical forms have an equal probability of being selected for production, each lexical cloud accumulates approximately 125 exemplars over the course of a single simulation. This means that the 75 earliest exemplars for each form decay by the end of the simulation. This decay mechanism also serves to further reinforce entrenchment of high-frequency patterns in the lexicon: early low-nasality exemplars in a shifting category like ORAL-N lose their influence in shaping the mean as more entrenched exemplars take their place in the cloud.

Figure 5.10 illustrates this effect. When operating under the *Model 5* parameters that include the additional decay mechanism, ORAL-N forms that are pressured to be nasal by the pre-N nasality bias converge to the nasality of underlying nasal vowels faster than in *Model 4*. By the end of the simulation, the final mean nasality of the two ORAL-N forms show complete neutralization, reflecting the combined influence of entrenchment and decay: Early ORAL-N exemplars with lower levels of

nasality slowly decay over the timecourse of the simulation, allowing the center of gravity to shift more rapidly toward an increasingly nasalized state. Because new exemplars are molded from this center of gravity, they too become increasingly nasalized over time. By the end of the simulation, all low-nasality tokens have decayed, allowing the newer high-nasality exemplars to define the category mean.

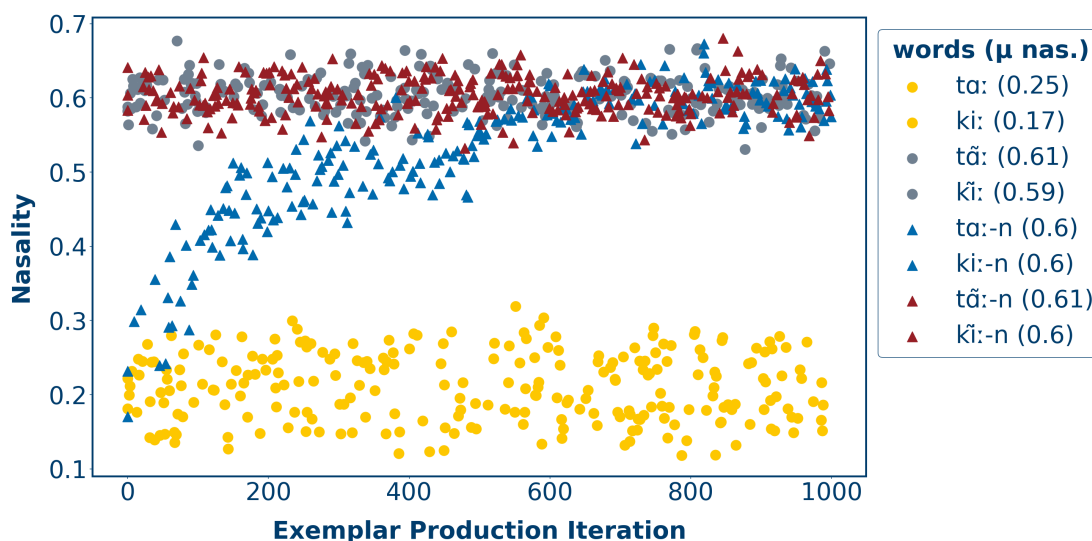


Figure 5.10: Model 5: selection, entrenchment, pre-N bias, noise, storage, decay.

The decay effect described in *Model 5* emerges not only within individual simulations, but also as a robust cross-simulation pattern, as shown in Figure 5.11. Across 100 simulations, the estimated mean nasality and 95% confidence intervals for the ORAL-N category forms converge fully with those of the NASAL and NASAL-N categories. This confirms that exemplar decay, when combined with entrenchment and the pre-N bias, reliably produces complete neutralization of pre-N vowel nasality. However, as the similarity-based association effects described in section 5.1.2 have not been included in the model, this is not a surprising result. No bias currently exists to counteract the steadily increasing nasality of

the ORAL-N category. In the next model, we will see that the introduction of a similarity bias — in which non-neutralizing forms attract their paradigmatically related neutralizing forms — works to prevent complete neutralization of vowel nasality in the ORAL-N category.

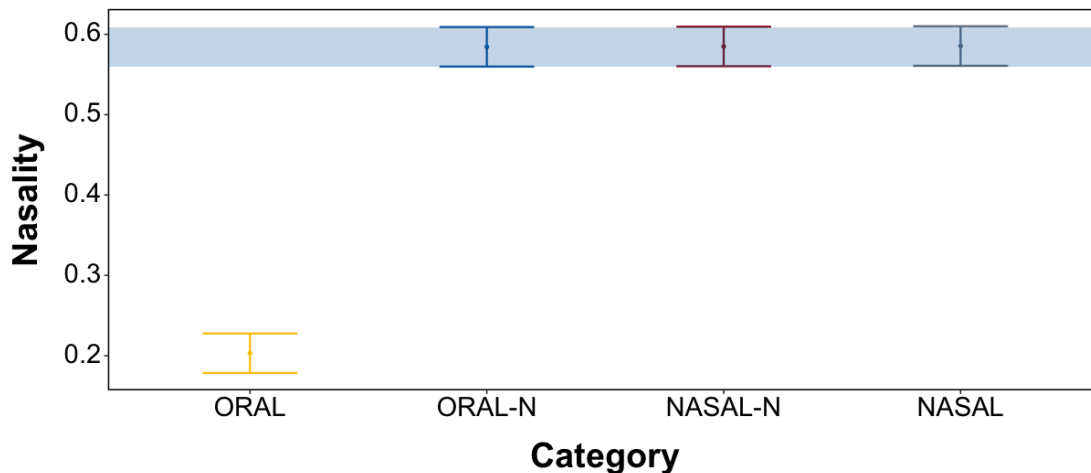


Figure 5.11: Model 5: selection, entrenchment, pre-N bias, noise, storage, decay.

5.2.2.6 Model 6 - Selection, Entrenchment, Pre-N/Similarity Biases, Noise, Storage, Decay

A fundamental predictor of sound change, neutralization patterns, and many other related processes under exemplar-theoretic frameworks is the interconnectedness of phonetically and semantically similar exemplars *across* lexical categories (see section 5.1.2). In these approaches, the production pattern of a new exemplar is shaped not only by the tokens belonging to its own lexical category, but also by paradigmatically-related exemplars. Crucially, this paradigmatic similarity effect is asymmetric, in that exemplars of non-neutralizing forms in a paradigm unidirectionally attract exemplars of neutralizing forms (Kaplan, 2015).

To incorporate these similarity effects into the simulations of *Model 6*, a morpho-

logical similarity bias was added to the objective function introduced in *Model 3* (section 5.2.2.3). As formalized in (2), the similarity bias penalizes exemplars of neutralizing forms (i.e., ORAL-N and NASAL-N) that deviate from the mean nasality of their related non-neutralizing form, α (i.e., ORAL and NASAL). Like the pre-N nasality bias in (1), the similarity bias penalty increases quadratically as discrepancy between a new exemplar and the non-neutralizing forms (α) in its paradigm grows (See footnote 3 for the advantages of constructing biases in this way).

(2) Morphological Similarity Bias

$$(\alpha\mu_{nas}-exemplar_{nas})^2$$

The objective function that integrates both biases is provided in (3). In *Model 6*, the pre-N bias is modulated by a scalar of five. This scalar amplifies the influence of the pre-N bias relative to the morphological similarity bias. Without it, the two pressures would be roughly equal in magnitude but point in opposite directions for ORAL-N exemplars, producing an equilibrium midway between the ORAL and NASAL categories. By introducing this scalar, the pre-N bias exercises an outsized influence in the optimization process, yielding higher degrees of nasality neutralization as the scalar increases.

(3) Objective Function of Model 6

$$\mathcal{O}_{Model6} = \underbrace{(\alpha\mu_{nas}-exemplar_{nas})^2}_{\text{Morphological Similarity Bias}} + \underbrace{5 \times (NASAL\mu_{nas}-exemplar_{nas})^2}_{\text{pre-N Nasality Bias}}$$

The plot in Figure 5.12 demonstrates the effects of removing the scalar weight from the pre-N nasality bias. Because both biases exert equal pressure on the

overall objective function penalty, the stabilized state for the ORAL-N category falls directly between the non-neutralizing ORAL and underlying NASAL categories.

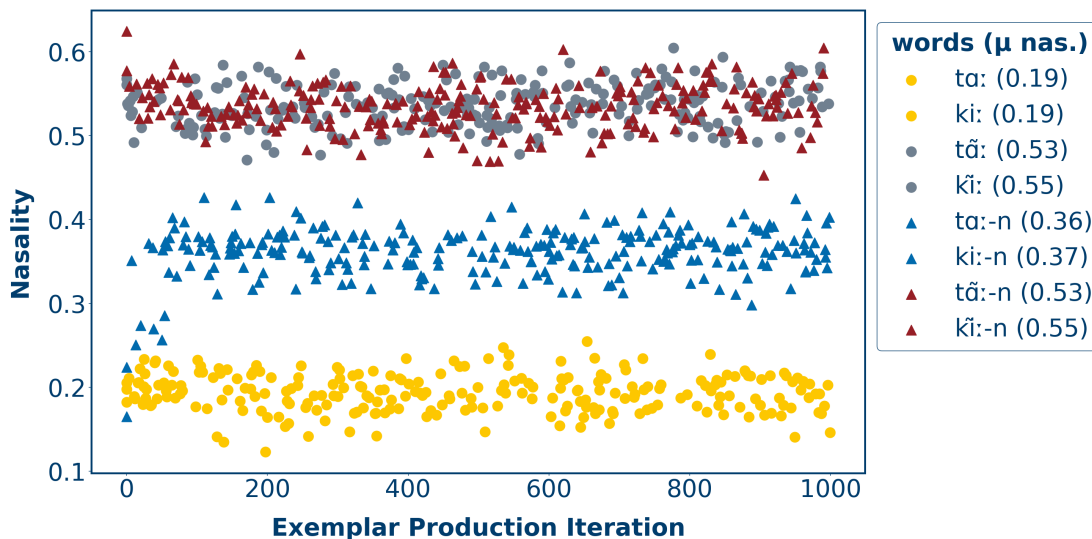


Figure 5.12: Model 6 without a scalar weight increasing the pull of the pre-N nasality bias.

Although a scalar of five is arbitrary, the utility of the scalar in general is important. Exemplar-theoretic frameworks explicitly specify that linguistic knowledge is wholly contained within the relational exemplar network. As such, no higher-order grammatical mechanism that could modulate the relative strength of individual biases across speakers or languages — beyond physiological variation such as speed of velum lowering — is available. Adjusting the weight of a single bias globally implicitly presupposes the existence of a higher-order layer of linguistic knowledge beyond the exemplar network. Since eliminativist models reject such structure, *the scalar is, in principle, universally fixed under these theories*. Here the scalar is set to five as a convenient illustrative value. In Section 5.2.2.9, I revisit this scalar when asking whether eliminativist exemplar theories can account for both complete and incomplete neutralization.

During the optimization stage of the simulation, the nasality value of a newly created exemplar is adjusted to minimize the objective function in (3). At early stages of the accumulation process, ORAL-N exemplars differ only slightly from the ORAL forms in their paradigm, making the similarity bias term very small. In contrast, the difference between early ORAL-N tokens and underlying nasal vowels is large, so the pre-N term dominates the objective function. Consequently, early ORAL-N exemplars undergo substantial upward shifts in nasality while the model seeks to reduce the overall penalty. This is illustrated for a hypothetical ORAL-N exemplar in (4).

(4) Effects of the Objective on an early ORAL-N exemplar (x) in Model 6

Objective Function: $\mathcal{O}(x) = (x - 0.2)^2 + 5(x - 0.6)^2$

Derivative: $\frac{d\mathcal{O}}{dx} = 2(x - 0.2) + 10(x - 0.6)$

At $x = 0.21$: $\frac{d\mathcal{O}}{dx} \big|_{x=0.21} = -3.88$

Interpretation: large negative derivative $\rightarrow$ increase x to reduce $\mathcal{O}$

Updated value: $x \rightarrow 0.29$

Over many iterations, these increases gradually shift the mean of the ORAL-N distribution upward toward the nasality level of the NASAL and NASAL-N categories. As this occurs, the pre-N bias term decreases, while the similarity bias term penalty, reflecting the growing distance from the non-neutralizing ORAL form, increases. The interaction between these two terms behaves like children on a seesaw: as one term falls the other rises. However, because the pre-N bias is multiplied by a scalar, the equilibrium is not reached where the two squared distances are equal, but instead at a point where the weighted pull toward the nasal category balances the pull back toward the non-neutralizing form. Just as a heavier child on a seesaw rests closer to the ground, the stable point for ORAL-N tokens settles closer to the mean of the underlying nasal categories. An example of a smaller increase at a later stage in the simulation is given in (5). Here, because

the mean of the ORAL-N category has approached the nasality level of NASAL vowels, the pre-N bias has shrunk such that it now exerts only a slight upward pull toward the NASAL category.

- (5) Effects of the Objective on a late ORAL-N exemplar (x) in Model 6

Objective Function: $\mathcal{O}(x) = (x - 0.2)^2 + 5(x - 0.6)^2$

Derivative: $\frac{d\mathcal{O}}{dx} = 2(x - 0.2) + 10(x - 0.6)$

At $x = 0.53$: $\frac{d\mathcal{O}}{dx} \big|_{x=0.53} = -0.04$

Interpretation: small neg. derivative $\rightarrow$ slightly increase x to reduce $\mathcal{O}$

Updated value: $x \rightarrow 0.55$

The optimal nasality value of ORAL-N forms under the weighted objective function in (3) is illustrated in Figure 5.13. The green solid and pink dotted curves plot the individual biases that make up the objective function, and the thick navy curve shows their summed penalties. The objective function is maximally satisfied at the x value that minimizes the entire function, as delineated by the blue vertical dashed line labeled ‘optimal ORAL-N’.

Increasing or decreasing the scalar of the pre-N bias shifts the optimal point of the objective function accordingly. Notably, because the similarity bias attracts exemplars toward the mean of the ORAL category (yellow dashed line) and the pre-N bias attracts them toward the NASAL mean (gray dashed line), the stabilization point is necessarily bounded between these two values.

This provides an interesting prediction: ‘over-neutralization’ — in which a neutralizing segment like an ORAL-N vowel *surpasses* the phonetic value of underlying category that attracts it — is not expected to occur under the implementation proposed here. As Du and Durvasula (2025, p.35) observe, over-neutralization appears unattested, and they cite this as a weakness of models like exemplar

theory that permit gradient representations.⁵ The present model shows that over-neutralization is not, in fact, a necessary prediction of exemplar-based systems.

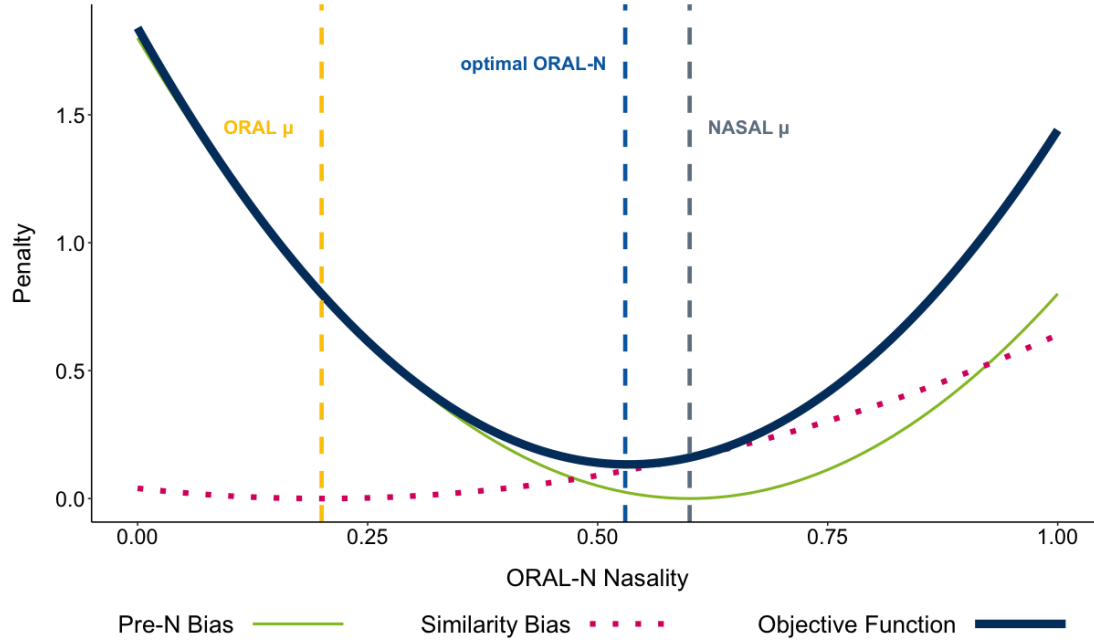


Figure 5.13: Model 6 objective function: penalty landscape for ORAL-N forms as determined by the two bias components that make up the objective function. The solid green curve shows the pre-N bias when scaled by 5, which favors nasality similar to the NASAL category; the dotted pink curve shows the similarity bias, which favors values closer to the non-neutralizing ORAL mean. The thick navy curve plots their combined objective function, with its minimum indicating the optimal nasality target for ORAL-N tokens. Vertical dashed lines mark the typical ORAL mean (yellow) at 0.2, the typical NASAL mean (gray) at 0.6, and the resulting optimal ORAL-N value (blue). The figure illustrates how competing pressures jointly determine an intermediate nasality target for ORAL-N forms.

A representative simulation of *Model 6* is shown in Figure 5.14. As is evident, exemplars in the ORAL-N category begin with nasality similar to ORAL vowels but over time increase in nasality until reaching a plateau and stabilizing just below the NASAL and NASAL-N categories, achieving incomplete neutralization.

⁵In the analysis of neutralization of vowel nasality before nasal consonants in Mankiyali in the previous chapter, the results could lend themselves to a case of over-neutralization, but future experiments are needed to determine this.

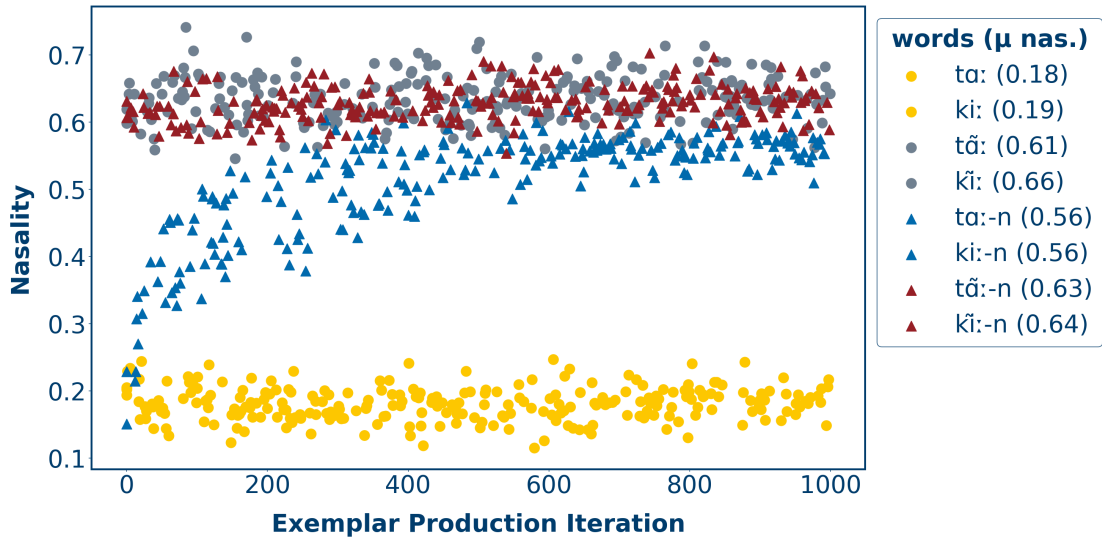


Figure 5.14: Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

The summary of 100 simulations of *Model 6* in Figure 5.15 confirms the robustness of this result. Across simulations, the ORAL-N category consistently approaches, but remains phonetically distinct from, the underlying nasal categories, with non-overlapping confidence intervals. *Under the parameters of the current model defined in the objective function*, then, exemplar theory predicts incomplete neutralization.

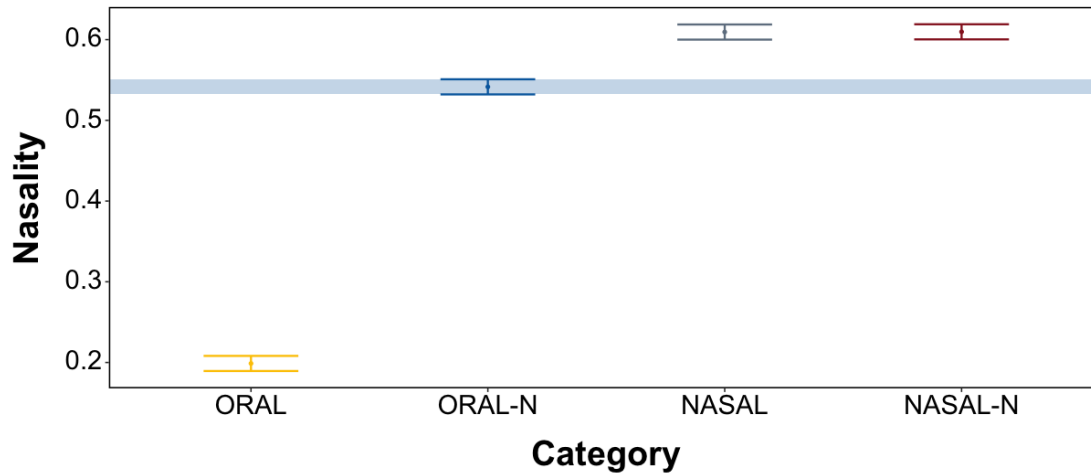


Figure 5.15: Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

The simulation in Figure 5.14 demonstrates why the morphological similarity bias must operate asymmetrically, such that non-neutralizing forms influence their neutralizing counterparts but not vice versa. When this restriction is removed, the neutralizing ORAL-N forms exert a corresponding influence on the non-neutralizing ORAL forms. As a result, the ORAL forms become progressively more nasal over the course of the simulation until all four categories converge and the nasality contrast disappears even in non-neutralizing contexts. This outcome is consistent with the demonstration in Kaplan (2017) that symmetrical similarity-based influence will lead to the eventual loss of contrast.

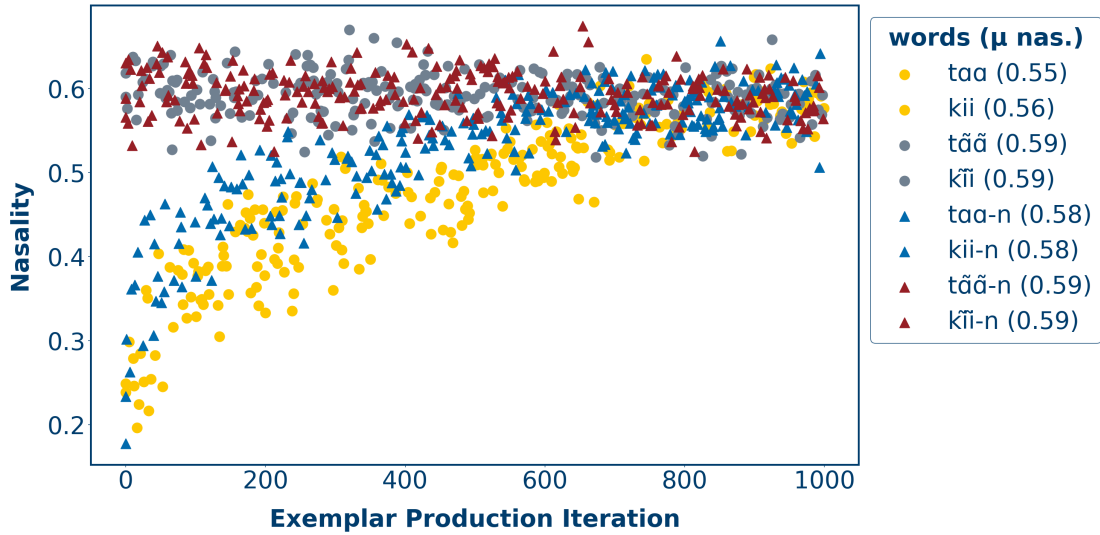


Figure 5.16: Model 6: selection, entrenchment, pre-N/similarity biases, noise, storage, decay. Here the similarity bias is symmetrical, such that neutralizing forms are able to influence non-neutralizing forms.

5.2.2.7 Model 7 - Complete Neutralization via Frequency Effects

With the core mechanisms of exemplar theory implemented in *Model 6*, it becomes clear that eliminativist frameworks naturally predict phonetically incomplete neutralization. *Model 7* explores how such a framework might be extended to account for phonetically *complete* neutralization, as observed in the neutralization of the vowel nasality contrast in Mankiyali in chapter 4.⁶

Two strategies are available to simulate complete neutralization. One is to increase the scalar of the pre-N nasality bias, thereby further penalizing low-nasality ORAL-N vowels and shifting the optimum of the objective function closer toward the NASAL category. However, as argued above, eliminativist exemplar models lack

⁶Complete neutralization is achieved using model 5 (Section 5.2.2.5), but the model there does not represent a full-fledged usage-based framework, primarily because it lacks a similarity bias. Consequently, it is empirically inadequate: without a morphological similarity bias, there is no mechanism preventing the ORAL-N category from collapsing entirely into the nasal vowel categories. Model 5 therefore *only* predicts complete neutralization, and cannot account for systems exhibiting stable, phonetically incomplete neutralization.

a mechanism for directly adjusting such bias weights. Because these approaches locate linguistic knowledge entirely within episodic memories and their associative structure, no independent locus of knowledge exists at which these scalars can be modified. Even if tokens exhibiting strong nasal coarticulation in NASAL-N forms are introduced into the system (section 5.2.2.8), the fixed relative weighting of the biases in the usage-based model will still yield stabilization at a phonetically incomplete equilibrium.

With that said, an alternative strategy available within an eliminativist model involves modulating influence indirectly through token frequency effects. Specifically, reducing the frequency of non-neutralizing ORAL forms reduces the strength of the similarity bias that penalizes high-nasality ORAL-N exemplars. When fewer non-neutralizing ORAL exemplars exist in the network, their pull on the ORAL-N category weakens. This follows from the basic assumption that high-frequency categories exert disproportionate influence in an exemplar network, while low-frequency categories exert comparatively little influence due to the sparse episodic evidence they contribute. In this way, frequency functions as an implicit scalar in a usage-based framework. As the token frequency of a category decreases, the attraction strength of the associated bias term effectively decreases as well.

To implement this in the simulation, I initialize Model 7 identically to Model 6, but with the probability of selecting non-neutralizing ORAL forms for exemplar generation reduced to one-tenth the probability of selecting any of the other three categories. As the simulation proceeds, the influence of the ORAL category diminishes relative to the growing exemplar populations of the other categories. The accumulation trajectory in Figure 5.17 demonstrates that, under these frequency conditions, eliminativist exemplar models successfully produce phoneti-

cally complete neutralization, with ORAL-N forms converging on the nasality of the underlying NASAL categories.

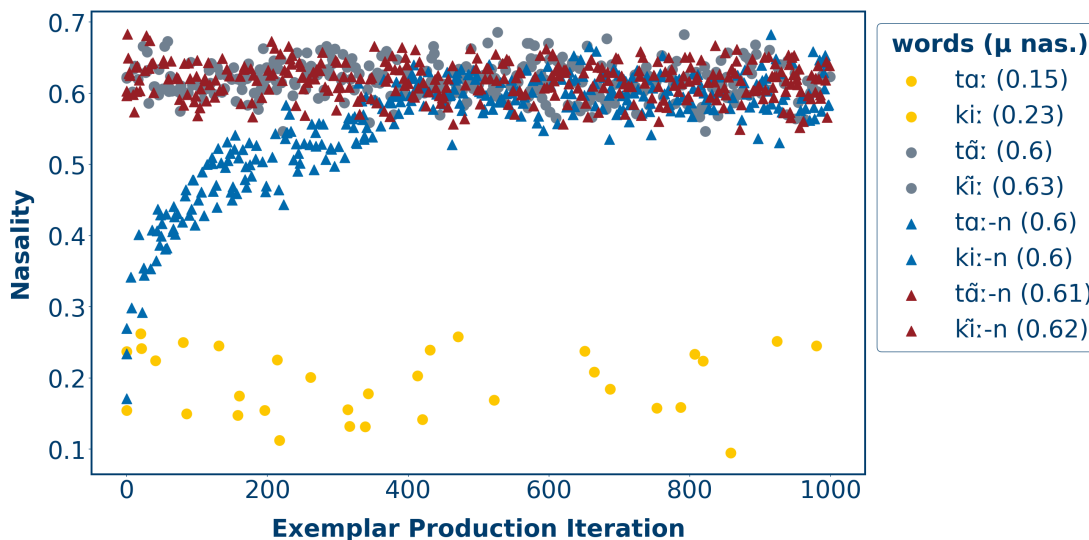


Figure 5.17: Model 7: selection (low ORAL freq.), entrenchment, pre-N/similarity biases, noise, storage, decay.

Figure 5.18 confirms that reducing the token frequency of non-neutralizing ORAL forms substantially weakens the similarity bias opposing neutralization. Across 100 simulations, the ORAL-N category converges with the NASAL and NASAL-N categories, producing a level of nasality consistent with phonetically complete neutralization.⁷

⁷Given the objective function in (3), true complete neutralization is mathematically impossible. Because the function is a weighted sum of squared distances to two distinct attraction points (the non-neutralizing ORAL mean and the NASAL mean), its minimum must always fall strictly between these points. A quadratic attractor function of this form can only reach one endpoint if the weight on the opposing attraction point is exactly zero. However, as the ORAL-N category approaches the NASAL mean, the two distributions may eventually become statistically indistinguishable, which may underlie empirical reports of complete neutralization. Alternatively, once two categories become sufficiently perceptually similar, they may collapse into a single category altogether. This process could be modeled by gradient attraction dynamics such as those proposed in Burzio (2000).

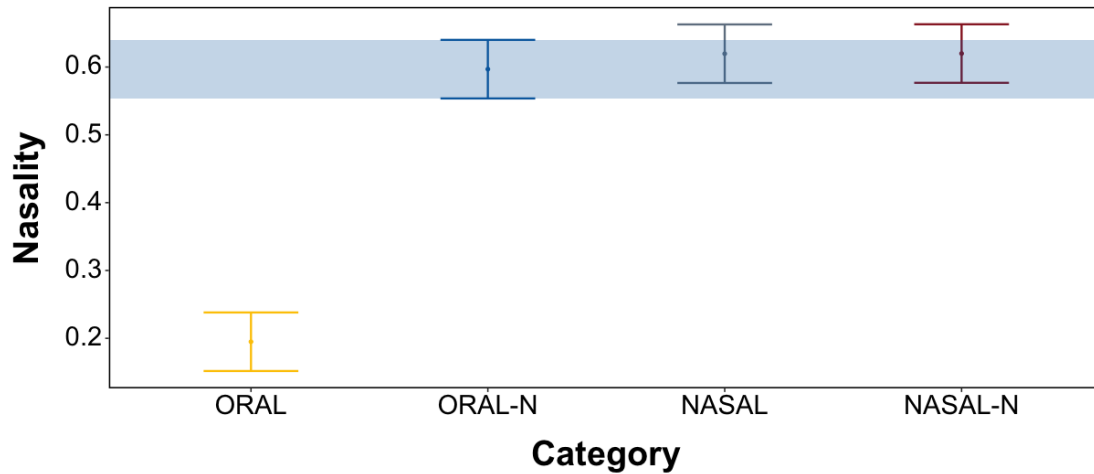


Figure 5.18: Model 7: selection (low ORAL freq.), entrenchment, pre-N/similarity biases, noise, storage, decay.

While manipulating the token frequency of non-neutralizing ORAL forms successfully produces complete neutralization in the model, this mechanism is unlikely to be a natural or general explanation for complete neutralization in real languages. There are two reasons for this. First, this approach predicts that in any language exhibiting complete neutralization of a feature (e.g., vowel nasality in Mankiyali), non-neutralizing forms across all paradigms that participate in the neutralizing alternation must be substantially less frequent than their corresponding derived forms. Whether such a distributional asymmetry is attested remains an empirical question, but it seems improbable that it would consistently hold across languages and morphological paradigms that exhibit complete neutralization.

To take Mankiyali as one concrete example, no formal corpus exists from which to extract exact token frequencies, but it is unlikely that all non-neutralizing ORAL nouns such as /qɪ:/ ‘giant.NOM’ are markedly rare. On the contrary, the nominative form occurs in a wide range of syntactic environments: as the agent of transitive verbs and the subject of intransitive verbs in the imperfect aspect, and

as the patient of transitive verbs and the subject of intransitive verbs in the perfect aspect (Englert, 2022, p.8). By contrast, the primary inflection involving a nasal suffix is the genitive case marker in /d̪ɪːn/ ‘giant-GEN.F.SG’. It is therefore unlikely that the non-neutralizing forms participating in the neutralizing alternation are consistently low-frequency in the way required by the model.

A second potentially undesirable prediction arising from a frequency-based account of neutralization is that it implies a direct correlation between non-neutralizing-form frequency and the type of neutralization observed. Specifically, if complete neutralization results from low-frequency non-neutralizing forms, then languages with phonetically complete neutralization of a particular feature should systematically exhibit unusually low-frequency non-neutralizing forms in relevant paradigms, whereas languages with phonetically incomplete neutralization should exhibit comparatively higher non-neutralizing-form frequencies. These are indeed clear and testable predictions, but it is far from obvious that real languages validate them. Whether such frequency asymmetries reliably align with the typology of neutralization remains an open empirical challenge.

5.2.2.8 Model 8 - Inclusion of higher nasality does not shift the optimization endpoint

If manipulating token frequency is not a plausible pathway for achieving both complete and incomplete neutralization, one might instead suggest that introducing high-nasality ORAL-N exemplars into the system could drive complete neutralization. Under this view — which is fully compatible with eliminativist assumptions — a speaker might encounter strongly nasalized ORAL-N tokens from another speaker, and these tokens could function as influential attractors within

the ORAL-N cloud, perhaps due to sociolinguistic salience leading to increased activation strength of these ‘high-nasality’ exemplars.

To test this possibility, *Model 8* retains all parameters of *Model 6*, where incomplete neutralization was consistently observed for ORAL-N clouds, except that the current model initiates the two ORAL-N lexical category exemplars at $\mu = .8$ and $\sigma = .025$ rather than at $\mu = .2$ and $\sigma = .025$ in all previous models. This new starting point for the ORAL-N category distributions simulates the introduction of an unusually high-nasality exemplar into the system, thus simulating a system in which speakers begin with, or are exposed to, exemplar evidence that is already strongly biased toward the nasal end of the spectrum.

As shown in Figure 5.19, however, injecting high-nasality ORAL-N exemplars into the network — while simultaneously removing all low-nasality ORAL-N exemplars — does not alter the ultimate stabilization point of the ORAL-N category. While the trajectory and time course of change differ, the ORAL-N category still settles at the same incomplete neutralization endpoint identified in *Model 6*.

This result demonstrates that although the phonetic substance of individual exemplars can influence the speed and shape at which a category approaches its stable state, the *location* of that stable state is determined entirely by the relative bias weights encoded in the objective function. In other words, the interaction of articulatory biases dictates the endpoints of neutralization without regard to individual exemplar memories.

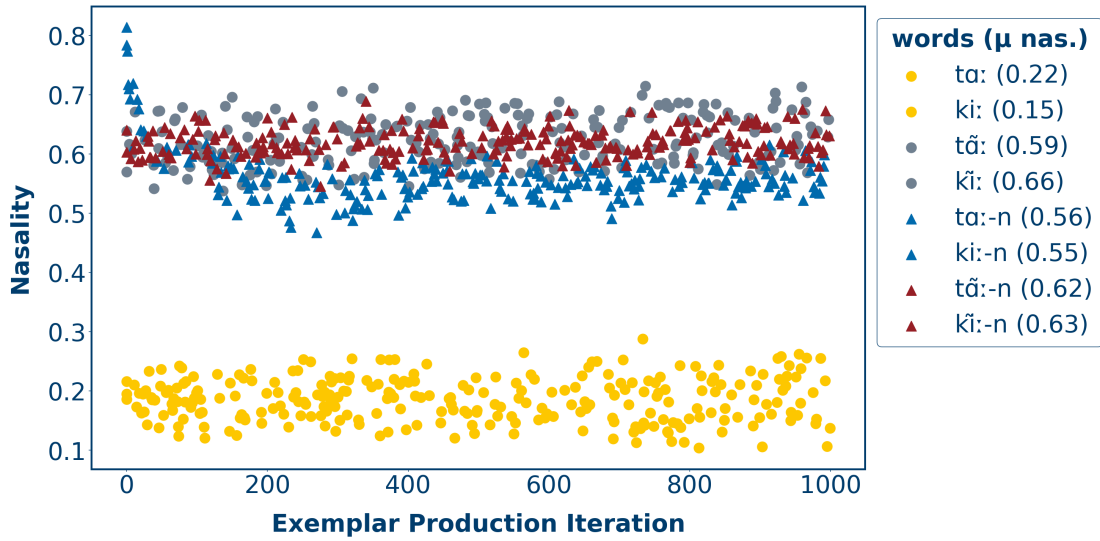


Figure 5.19: Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

The results of 100 simulations shown in Figure 5.20 corroborate this conclusion. Even when exclusively high-nasality ORAL-N exemplars are introduced at the outset, the final mean nasality of the ORAL-N category consistently stabilizes below the underlying NASAL and NASAL-N categories. While the initial conditions influence the rate and trajectory of change, the endpoint of ORAL-N neutralization remains relatively stable across simulations. This pattern demonstrates that altering the phonetic substance of early exemplars cannot shift the model's ultimate equilibrium. The relative strengths of the pre-N and similarity biases are the sole determiners of the final stabilization point.

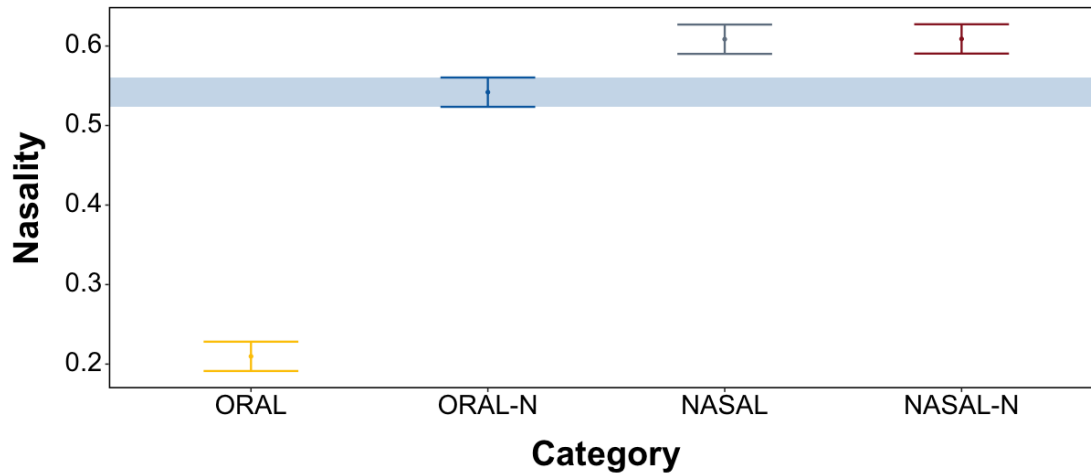


Figure 5.20: Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

The determination of the system’s stable state by the relative strengths of these biases is further supported by the model shown in Figure 5.21. Here, the nasality of vowels in the ORAL-N category is initialized at .2 and rises normally during the first 200 iterations due to the pre-N production bias. Between iterations 200 and 400, however, all newly produced ORAL-N exemplars were artificially initialized with a nasality value of .85. This simulates the temporary introduction of highly nasalized exemplars into the ORAL-N clouds, perhaps through the imitation of a socially prestigious speaker whose productions exhibit especially high degrees of nasality. As Figure 5.21 shows, this intervention temporarily shifts individual ORAL-N exemplars toward the nasal end of the continuum but does not durably alter the final mean nasality of either of the two ORAL-N lexical clouds. Once the intervention ends, the category shifts back toward the same incomplete-neutralization endpoint observed in the preceding simulations. Within the present model, then, even a sustained influx of highly nasalized exemplars can alter the category’s trajectory without changing its eventual stabilization point. A last-

ing shift in that point appears to require a corresponding change in the relative weights of the pre-N and similarity biases.

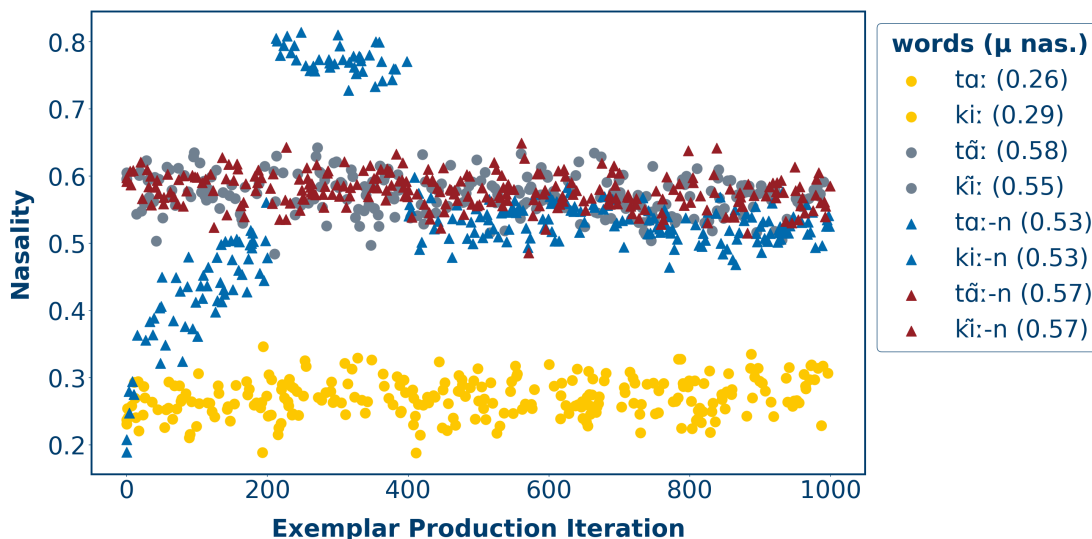


Figure 5.21: Model 8: selection, entrenchment, pre-N/similarity biases, noise, storage, decay. From iterations 200–400, all exemplars generated for lexical items in the ORAL-N category were initialized with .85 nasality.

5.2.2.9 Model 9 - change in production bias strength

Since neither frequency manipulations nor the introduction of high-nasality tokens are viable options for achieving complete neutralization, an alternative explanation is that the strength of the relevant articulatory biases themselves varies across languages and/or speakers. At the individual level, such variation may arise from physical factors, such as differences in vocal tract physiology or the timing of velum lowering, and thus need not be represented as a core component of linguistic knowledge. However, systematic, stable differences in the strength of these biases across entire speech communities require a different explanation. When such differences yield distinct phonetic outcomes across languages, they must be encoded as part of the speaker’s linguistic knowledge, rather than emerging solely

from the distribution of exemplars. While this type of higher-order manipulation of an intentionally implemented phonetic target is not available under eliminativist frameworks, it nonetheless appears necessary for any model that aims to capture both complete and incomplete neutralization in a unified and natural way.

Model 9 implements this higher-order knowledge explicitly by increasing the scalar on the pre-N nasality bias. Strengthening this bias increases its contribution to the overall objective-function penalty, shifting the optimal stabilization point closer to the mean of the NASAL category. As shown in Figure 5.22, when the scalar is set to 25, the ORAL-N forms converge to nasality values that are nearly identical to those of the NASAL and NASAL-N categories, yielding complete neutralization.

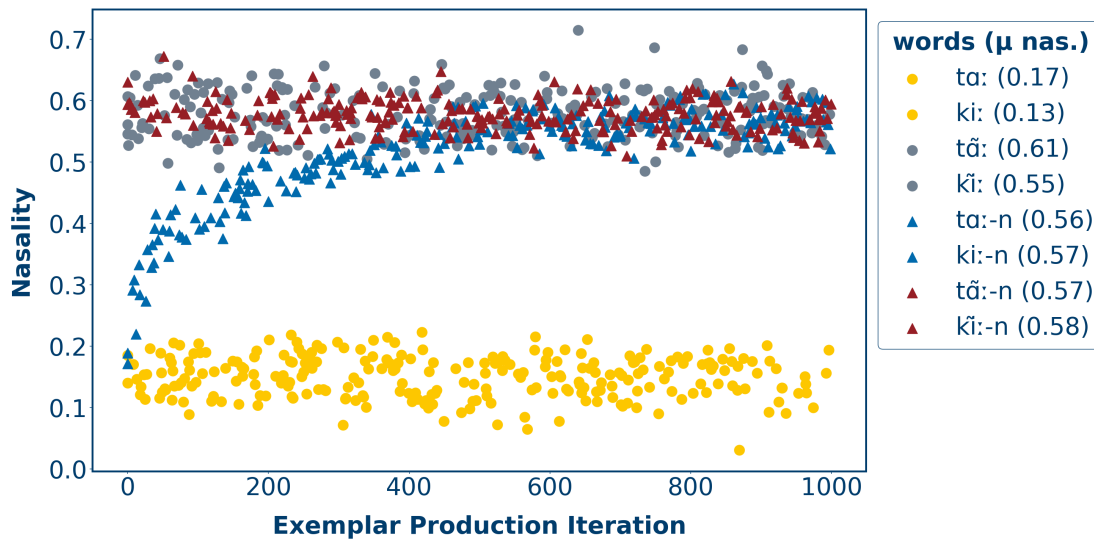


Figure 5.22: Model 9: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

The summary across 100 simulations in Figure 5.23 confirms that this pattern is uniform across samples: increasing the scalar on the pre-N bias consistently produces phonetically complete neutralization. These results indicate that, barring unnatural frequency manipulations, only by modifying the underlying articulatory

bias strengths can the model generate both incomplete and complete neutralization outcomes.

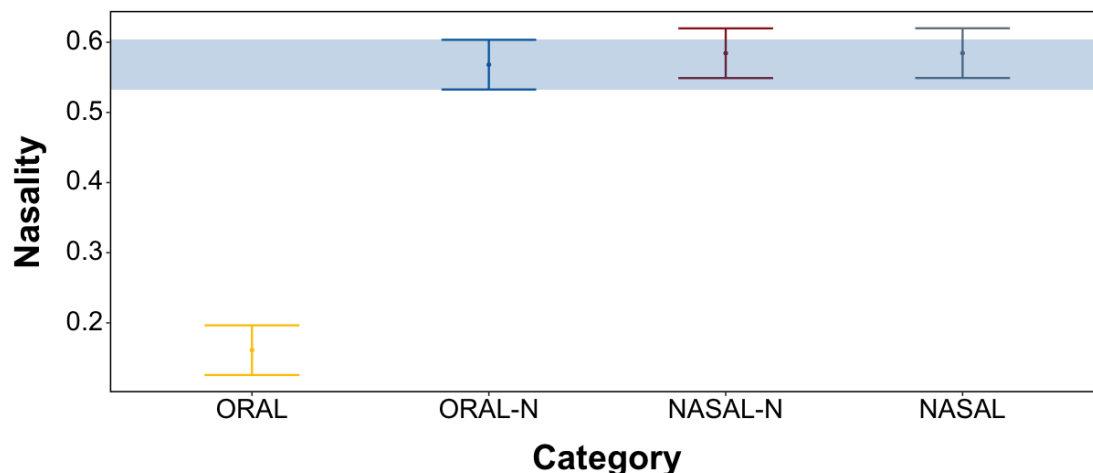


Figure 5.23: Model 9: selection, entrenchment, pre-N/similarity biases, noise, storage, decay.

5.3 Discussion

The model outlined in this chapter was constructed from the foundational commitments of eliminativist usage-based theories. Relying only on phonetically detailed exemplars, similarity-based association, frequency effects, emergent structure, articulatory biases, and decay, the model reflects an exemplar-only architecture. The central empirical question I have wrestled with is whether a system that lacks grammatical structure external to the exemplar network itself can account for both phonetically complete and phonetically incomplete neutralization.

When the model is initialized with fixed articulatory bias weights, it reliably produces incomplete neutralization across simulations (*Model 6*). This occurs because the morphological similarity bias and the pre-N nasality bias interact in a stable and predictable way: each new exemplar is pulled upward by the pre-N bias and

downward by the similarity bias, and the point at which these two pressures balance is fixed by the relative bias weights. Under the conditions initialized in *Model 6*, ORAL-N forms consistently converge on an intermediate nasality value that is phonetically distinct from both related non-neutralizing ORAL forms and from the underlying nasal categories. This behavior matches the patterns observed for languages like San Martín Peras Mixtec (SMPM) (Bennett and Eischens, 2026). In SMPM, vowel nasality is contrastive following an oral consonant (CV vs. C $\tilde{V}$), but vowel nasality is neutralized following a nasal consonant (N $\tilde{V}$ but *NV). Crucially, the realization of nasality on post-N vowels is weaker than the nasality on contrastive C $\tilde{V}$ vowels, such that a measurable and perceptually salient phonetic distinction exists between predictably nasal N $\tilde{V}$ vowels on the one hand, and both contrastive oral CV vowels and contrastive nasal C $\tilde{V}$ vowels on the other hand.

A crucial finding, however, is that the stabilizing scalar that determines the strength of the pre-N bias in *Model 6* is, by the very logic of eliminativist theories, static (or at best variable based on physiology alone). Exemplar-only frameworks prohibit mechanisms for modulating the strength of articulatory biases separate from the substance of exemplars. Because all linguistic structure is predicted to be emergent from stored exemplars and their associative architecture, there is no grammatical structure through which a speaker could globally strengthen or weaken a particular bias. As a result, eliminativist models inherently predict only a single equilibrium outcome for any given alternation. They can generate a Mixtec-type system with incomplete neutralization, or they can generate a Mankiyali-like system in which neutralization is phonetically complete, but they have difficulty generating both.

Two potential explanations within an eliminativist architecture were explored.

The first manipulated the token frequency of non-neutralizing forms, effectively weakening the similarity bias opposing neutralization. While this manipulation does yield fully neutralized outputs, it requires extreme and typologically dubious frequency asymmetries. There is no evidence that languages with complete neutralization, including Mankiyali, systematically exhibit unusually low-frequency non-neutralizing forms relative to neutralized forms. Of course, further typological work is needed to determine whether languages exhibiting complete neutralization conform to or contradict this prediction.

The second explanation involved injecting strongly nasalized ORAL-N exemplars into the system, simulating exposure to high-nasality tokens. Although this manipulation changes the trajectory and speed of category movement, it does not change the endpoint. The system still stabilizes at a level of incomplete neutralization. Both strategies therefore fail to provide a general mechanism for deriving cross-linguistic variation in neutralization.

Given these results, it seems that, within the production-based exemplar system examined here, deriving both incomplete and complete neutralization requires allowing the strength of articulatory biases themselves to vary across languages, speech communities, and/or individual speakers. In other words, the contrast between incomplete neutralization in Mixtec and complete neutralization in Mankiyali appears to require some additional level of representation that governs the relative weighting of competing phonetic pressures across these languages. Crucially, in languages like Mixtec, where paradigmatically related oral forms are absent, this competition cannot be attributed to a morphological similarity bias. Instead, incomplete neutralization would need to arise from the interaction between the pre-N nasality bias and another independent pressure disfavoring

complete nasality, yielding a stable intermediate outcome. Ultimately, then, the present results suggest that the relative strengths of production biases responsible for different types of neutralization are not uniform across languages.

One possibility not examined here concerns a listener discarding perceptually ambiguous tokens that fall between two similar categories. In Wedel (2004, 2006), for instance, tokens that strongly activate multiple lexical categories are more likely to fail to be categorized and, consequently, not to be stored as new exemplars. Such perceptual filtering could contribute to the maintenance of lexical contrast and might therefore help sustain phonetically incomplete neutralization in a case like Mixtec.

This solution, however, does not account for the production biases proposed in the present model. During production, the intended category is already known by the speaker, and the biases under investigation affect how a token from that category is produced. It therefore remains unclear whether perceptual filtering of ambiguous tokens could, by itself, account for neutralization caused by production biases. While further research incorporating the production-perception feedback loop into the model is needed to examine this, it seems that speakers possess some level of higher-order control over the strength of relevant production biases, impacting the end state of neutralization.

Another possible solution within the eliminativist framework is that different patterns of phonetic neutralization arise from differences in the phonetic properties of the languages themselves. For example, it is possible that languages with differing neutralization patterns also differ in vowel length: shorter vowels could be biased to have a higher proportion of the vowel be nasalized, which would predict that languages with shorter vowels are more likely to undergo complete neutralization

of vowel nasality in neutralizing contexts. Nothing about this analysis necessarily requires ‘knowledge’ outside of the exemplar representations themselves, only convergence on (or emergence/development of) a language which happens to have relatively shorter vowels. I leave it to future research to determine whether a consistent correlation between type of phonetic neutralization and related phonetic factors that facilitate that neutralization align with this proposal. An interesting place to start could be a comparison between vowel durations in Mankiyali and SMPM. It is worth noting, however, that Mankiyali exhibits complete neutralization in phonologically long vowels, whereas SMPM, which exhibits incomplete neutralization, does not appear to have a phonological vowel-length contrast (Eischens and Hedding, 2024). While this observation does not rule out the durational account outlined above, it provides some preliminary reason to question its validity.

What might constitute such higher-order knowledge? Several possibilities exist, none of which are compatible with a strictly eliminativist architecture. One possibility, suggested by Beddor (2009), is that speakers come to reinterpret anticipatory nasal coarticulation as a crucial property of pre-N vowels. For instance, a language user who produces ORAL-N forms with incomplete neutralization may, over time, begin to treat the nasality on ORAL-N vowels as a reliable perceptual cue to the following suffix. Under this reanalysis, the increased importance of pre-N coarticulation leads the language user to increase the strength of the pre-N bias scalar to reinforce this perceptual cue for listeners. In this way, the stabilization point of ORAL-N forms begins to shift toward complete neutralization.

A second possibility, from Ohala’s (1993) perception-based account of sound change, is that listeners sometimes misattribute the source of nasal coarticulation

to the vowel rather than the nasal consonant. This analysis is possible, though it runs against experimental evidence suggesting that listeners tend to attribute nasal coarticulation in the context of a nasal consonant as belonging to the nasal consonant (Beddor and Krakow, 1999; Rysling, 2017). Nevertheless, if the nasality of a following consonant is misperceived as belonging to the vowel, a listener may adjust the pre-N nasality bias scalar upward, again yielding a Mankiyali-type outcome.

A third possibility attributes the shift in the pre-N nasality bias scalar to sociolinguistic factors. Specifically, speakers may recalibrate their bias weights to more closely match the productions of high-prestige interlocutors who use strongly nasalized pre-N vowels. Across generations, such shifts could produce stable differences in neutralization patterns across speech communities.

What unites all three potential explanations outlined above is that they require speakers to exert control over mechanisms that operate outside of a network of associated exemplars. In each case, the relevant adjustment cannot be derived solely from some change (drastic or otherwise) in the distribution of stored exemplars, but instead reflects a form of higher-order linguistic knowledge that intentionally shapes and directs the evolution of category distributions over time.

This type of knowledge is precisely what eliminativist usage-based theories reject. The simulation results from this chapter, therefore, lead to a principled conclusion: if accounting for both phonetically incomplete and phonetically complete neutralization requires control over higher-order grammatical levers, eliminativist usage-based systems are insufficient in principle. Importantly, this conclusion is reinforced by the empirical landscape. Languages like Mixtec exhibit stable incomplete neutralization, while languages like Mankiyali exhibit complete neutral-

ization, and any adequate theory must capture both patterns. To do so, speakers must possess control over how competing phonetic biases interact to influence production. While the formal implementation of this knowledge could take many forms — whether it be a traditional generative grammar or a global weighting over biases in an optimization framework — this higher-order knowledge is necessary.

Chapter 6

Evidence for Episodic Speech Representations

In chapter 2, I provided evidence from Pakistani Punjabi that speech representations exist at a highly abstract level, such that speakers sometimes store a lexical item in a distinct form from any of its surface realizations. Nevertheless, as mentioned in the introduction, a bevy of experimental work over the past few decades has provided convergent evidence that a language user’s competence in speech production is not easily whittled down to a single symbolic representation for each lexical item, filtered through a formal grammar to generate a phonetic production target. To cite just a few examples (see Goldrick and Cole 2023 for a recent overview), rates of reduction and syncope are often lexically idiosyncratic and strongly correlated with token frequency (e.g., schwa deletion is more prevalent in high-frequency words like ‘memory’ than in low-frequency words like ‘mammary’ (Hooper, 1976); listeners process speech more efficiently when the voice characteristics of a talker are judged to be more stereotypically male or female

(Johnson, 2006); the speech of a listener-turned-talker is directly impacted by their recent linguistic exposure, such that they adopt fine-grained phonetic properties from their recent linguistic experiences (Goldinger, 1998); Long-term exposure to different languages/dialects results in perceptually salient acoustic changes to a speaker’s speech characteristics (Sancier and Fowler, 1997); and the perceived social identity of a voice can impact the degree to which a speaker adopts the voice’s phonetic properties in future productions (Babel, 2012).

The unifying thread connecting all of these studies is that the minute phonetic and social details associated with individual linguistic episodes are not filtered out in the storage/access of speech representations. To the contrary, they are latched onto and leveraged by language users in such a way that shapes subsequent production behavior. In short, the claim that usage shapes representation is difficult to dispute. As an illustration of these facts, in what follows I present data from a pilot study which adds support to the claim that episodic exemplar representations characterized by fine-grained phonetic detail are retained and exploited to communicate extralinguistic social information. A result from this study is that the same words I argue in chapter 2 require an abstract representation to account for nasal harmony patterns in Punjabi seem to be stored at a more granular level of representation as well.

6.1 Evidence for Detailed Representations from Punjabi

At present a single native Pakistani Punjabi speaker (male, aged 41 from Mirpurkhas, Pakistan) has participated in the following experiment. As such, no firm

conclusions can yet be drawn; rather, the results presented here are intended to illustrate patterns observed in previous work using Punjabi data that representations retaining high levels of detail form a core component of linguistic competence. The pilot study investigates phonetic convergence in pitch, exploring whether changes to a speaker's pitch properties through interactions with other speakers might be influenced by the perceived social status of other speakers. This question is motivated by findings like those in Labov (1963), which suggest that social pressures are a driving force in language change.

The stimuli in the study consisted of the 61 words used in Experiment 1 (Chapter 2), along with an additional five Punjabi words that served as filler items in that experiment. There were four phases to the experiment, during which the participant both produced and was repeatedly exposed to the 66 stimuli. The goal of Phase 1 was to establish baseline pitch values for the participant, which were later used for comparison following exposure to two voices of differing social status.

In Phase 1, the participant engaged in a game-like task in which a word appeared on the screen and he was given two seconds to read it clearly aloud before clicking “Next” to earn points. At scheduled intervals, the participant received artificial feedback indicating that his production lacked clarity, resulting in a loss of points. This feedback was included to encourage clarity of speech while maintaining a rapid speaking rate to earn more points. An illustration of the interface from Phase 1 is shown in Figure 6.1, where the word [pī:g] ‘swing’ was successfully produced, earning the participant thirty points.

The game structure was designed to promote a consistent and relatively rapid speech rate, both in this phase and in the later post-production phase (Phase 4).

In all phases involving speech production, recordings were made using a head-worn Audio-Technica BP894X cardioid condenser microphone connected to a Zoom H5 Handy Recorder. The recordings from Phase 1 served as the baseline pitch measurements for the participant.

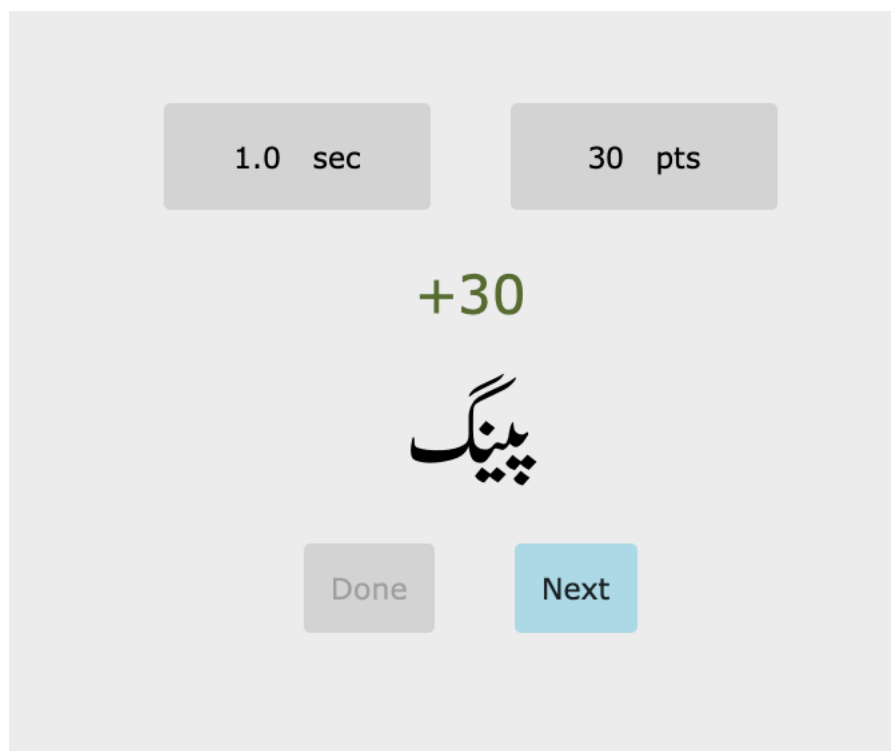


Figure 6.1: Phase 1 baseline productions interface for the word [pī:g] ‘swing’.

In Phase 2, the participant was exposed to words produced by two Punjabi voices with perceptually distinct pitch properties. Visual and written information framed one voice as a ‘low-prestige’ brick kiln worker (left image in Figure 6.2) and the other voice as a ‘high-prestige’ scholar of Punjabi (right panel in Figure 6.2).

The two voices were selected from Experiment 1 (Chapter 2) on the basis of exhibiting the most distinct pitch characteristics. This selection ensured a strong phonetic contrast between the low- and high-prestige voices, maximizing the like-

likelihood that participants would perceive and encode the difference.



Figure 6.2: Images presented for the two Punjabi voices. The left image corresponds to the ‘low-prestige’ brick-kiln laborer and the image on the right to the ‘high-prestige’ Punjabi scholar.

Figure 6.3 shows the average pitch (f_0) trajectories for vowels extracted from all words produced by both voices in the experiment. As is evident, the high-prestige voice exhibits substantially higher f_0 values than the low-prestige voice. In addition, the f_0 trajectory of the high-prestige voice rises over the course of the vowel, whereas that of the low-prestige voice falls. These sharply contrasting pitch profiles provide a strong testing ground to examine whether the participant selectively stores phonetic information from experienced speech based on a talker’s perceived social status, as evidenced by both his immediate imitations (Phase 3) and his delayed post-exposure productions (Phase 4).

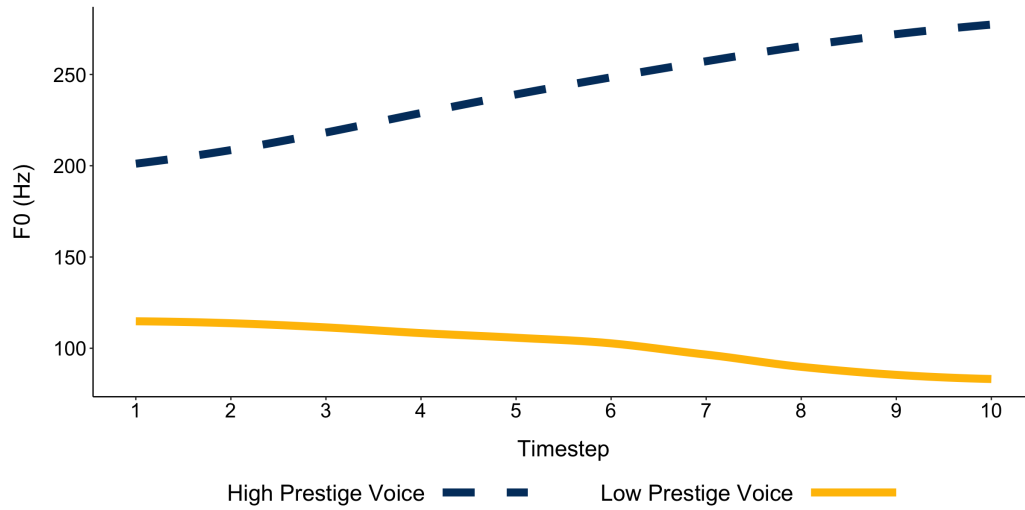


Figure 6.3: Average f0 trajectories across 66 Punjabi words produced in isolation. For each vowel token, f0 was measured at 10 equally spaced time points. Generalized additive model (GAM) curves were then fit to the f0 trajectories and averaged across tokens, grouped by voice.

Phase 2 consisted of a game in which the participant heard a word produced by one of the two voices and was given five seconds to identify the correct word from a matrix of 16 options presented in the Shahmukhi script. The participant heard 25 words from each voice, with each word repeated three times over the course of the phase (75 recordings per voice). The remaining 16 words from the 66-word list were not presented during this phase.

Stimuli were organized into 15 blocks of five words per voice, alternating between the low- and high-prestige voices. At the beginning of each block, the participant was shown an image and a brief description associated with the upcoming voice. The low-prestige voice was introduced as a brick-kiln laborer learning to read in order to obtain a position as an office helper, and the participant was instructed to select the word corresponding to what he said to help him learn to read. The high-prestige voice was introduced as a Punjabi scholar compiling a dictionary to

promote the Punjabi language, and the participant was instructed to help him collect words for the dictionary. The display presented at the beginning of each low-prestige block is shown in Figure 6.4.



Figure 6.4: Description and imagery for the low-prestige voice that participants were exposed to at the beginning of each ‘low-prestige’ block in Phase 2.

In total, the participant was exposed to 150 tokens during Phase 2, 75 words from each voice drawn from a set of 25 items repeated three times by that voice. An example of the interface following an incorrect selection for the high-prestige voice is shown in Figure 6.5. During each trial, an image associated with the current voice was displayed in the top-left corner of the screen, serving as a continuous reminder to the voice’s social status.

The participant received feedback after each selection indicating whether the correct or incorrect spelling had been learned (for the low-prestige voice) or added to the dictionary (for the high-prestige voice). These feedback messages were designed to reinforce the contrast in perceived social status between the two voices.

Importantly, the participant did not produce any speech during Phase 2, but was only exposed to the two voices through repeated lexical items.

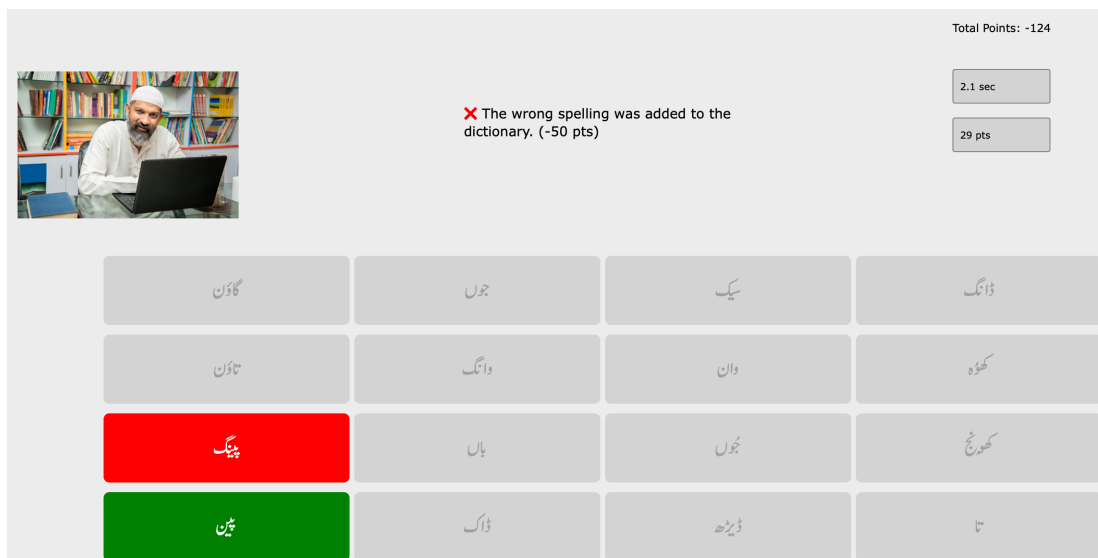


Figure 6.5: Example interface from a high-prestige voice block for Phase 2 of the experiment. The high-prestige voice uttered the word [pĩ:g] ‘swing’ (highlighted in green), but the participant incorrectly selected [pin] ‘drink.IMP’ (highlighted in red).

Phase 3 aimed to examine whether the differing presentation of the social statuses of the two voices impacted the participant’s imitation patterns (Babel, 2012; Goldinger, 1998). In this phase, the participant heard a word with the picture of the associated voice on the screen and had 2 seconds to repeat the word. The faster he repeated the word and clicked next, the more points he received. Again, the game-like structure here was instituted in an effort to obtain stable speech rates from the participant across the baseline production (phase 1), imitation (phase 3), and post-production (phase 4) phases. As in the previous phase, each of the 25 words from each voice were repeated and imitated three times during phase 3, resulting in 75 imitations of each voice. Words spoken by each voice were spoken in blocks of five, with one of the voices randomly selected as the first voice.

Figure 6.6 shows the participant’s imitations of the high-prestige voice across the three shadowing iterations in Phase 3. The solid pink curve shows the participant’s baseline f0 trajectory, averaged across vowels from the 25 relevant words produced in Phase 1. This trajectory begins around 140 Hz and falls over the course of the vowel, reaching approximately 120 Hz by the final timepoint. The blue curve shows the high-prestige voice’s f0 trajectory for the same set of words. The remaining three curves show the participant’s f0 trajectories after the first, second, and third shadowing iterations of the high-prestige voice.

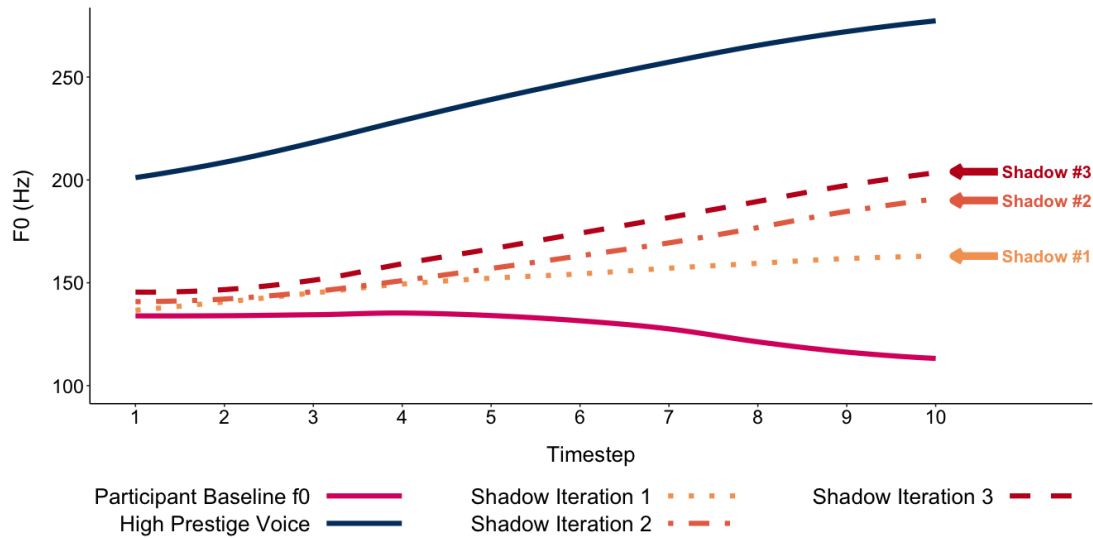


Figure 6.6: Results from Phase 3 imitations of the high-prestige voice by a single participant. The solid pink curve shows the participant’s baseline f0 trajectory, averaged across vowel tokens from 25 words produced in Phase 1. The blue curve shows the high-prestige voice’s f0 trajectory for the same set of words. The remaining curves show the participant’s productions grouped by shadowing iteration (first, second, and third repetitions of each word), with f0 trajectories averaged across the 25 words within each iteration. For each token, f0 was measured at 10 equally spaced timepoints, and values are averaged at each time point.

From a qualitative examination of the plot, the participant’s productions shift upward across the three shadowing iterations, moving progressively closer to the high-prestige voice’s f0 trajectory. This pattern is consistent with a usage-based

view of speech representations: as the participant is repeatedly exposed to exemplars from the high-prestige voice, his own productions appear to be increasingly shaped by the high and rising pitch profile of that voice. These preliminary results suggest that, at least in the short term, speakers may retain and reproduce fine-grained phonetic detail associated with an individual speaker's voice.

Crucially, the results for imitations of the low-prestige voice in Figure 6.7 do not show the same pattern. Whereas f_0 trajectories in the high-prestige condition crept progressively toward the high-prestige voice with increased exposure, the three shadowing iterations in the low-prestige condition lack a clear pattern of progressive convergence toward the low-prestige voice.¹

¹One possible explanation for this dichotomy is that the participant may have had limited ability to further lower his f_0 , which could attenuate observable imitation effects. To account for this potential confound, a full-scale experiment should incorporate a matched-guise design in which prestige labels are counterbalanced across participant groups.

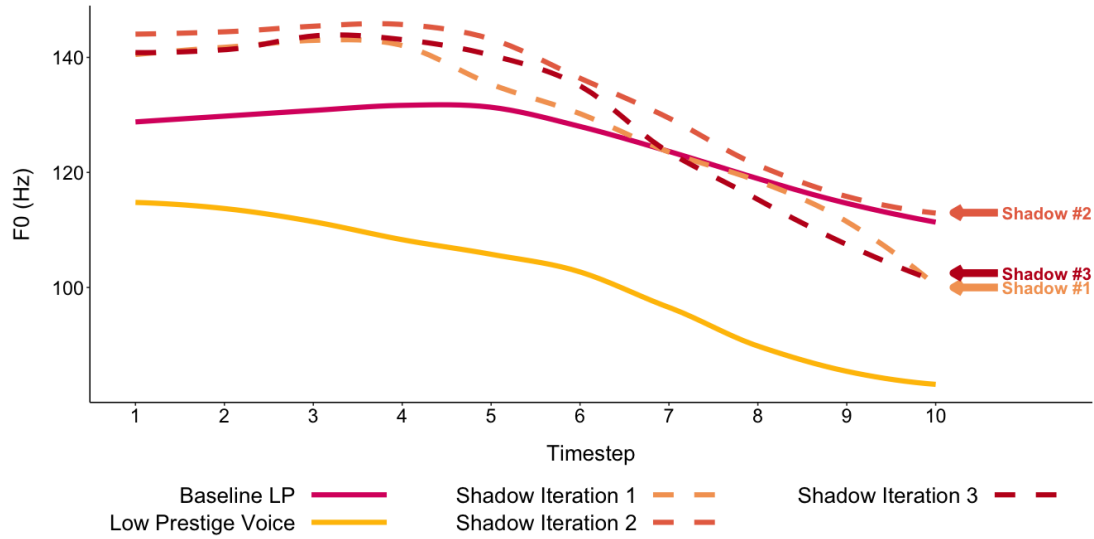


Figure 6.7: Results from Phase 3 imitations of the low-prestige voice by a single participant. The solid pink curve shows the participant’s baseline f0 trajectory, averaged across vowel tokens from 25 words produced in Phase 1. The yellow curve shows the low-prestige voice’s f0 trajectory for the same set of words. The remaining curves show the participant’s productions grouped by shadowing iteration (first, second, and third repetitions of each word), with f0 trajectories averaged across the 25 words within each iteration. For each token, f0 was measured at 10 equally spaced timepoints, and values are averaged at each time point.

This asymmetry in imitation is consistent with the hypothesis that, while fine-grained phonetic detail is encoded in memory, social context may bias attention toward particular properties of the speech signal. In particular, these preliminary results suggest that representations associated with socially high-prestige voices may be more strongly encoded or more readily accessible during speech production.

The final phase of the experiment tested whether the prestige-based imitation effects observed in Phase 3 persist as long-term influences on production. It may be that the fine-grained phonetic and social information used during imitation is only stored in working memory for very short durations. If so, these effects should

degrade rapidly and disappear after even a short delay. To evaluate this, after a 5-minute break between phases 3 and 4, the participant produced the 66 words in the experiment again, following the Phase 1 procedure.

Visual results of these post-productions are shown in Figure 6.8. While the Phase 1 baseline productions and the Phase 4 post-exposure productions exhibit a similar downward f0 trajectory over the timecourse of the vowel, the post-productions show a clear increase in overall f0 at essentially every point in the vowel.

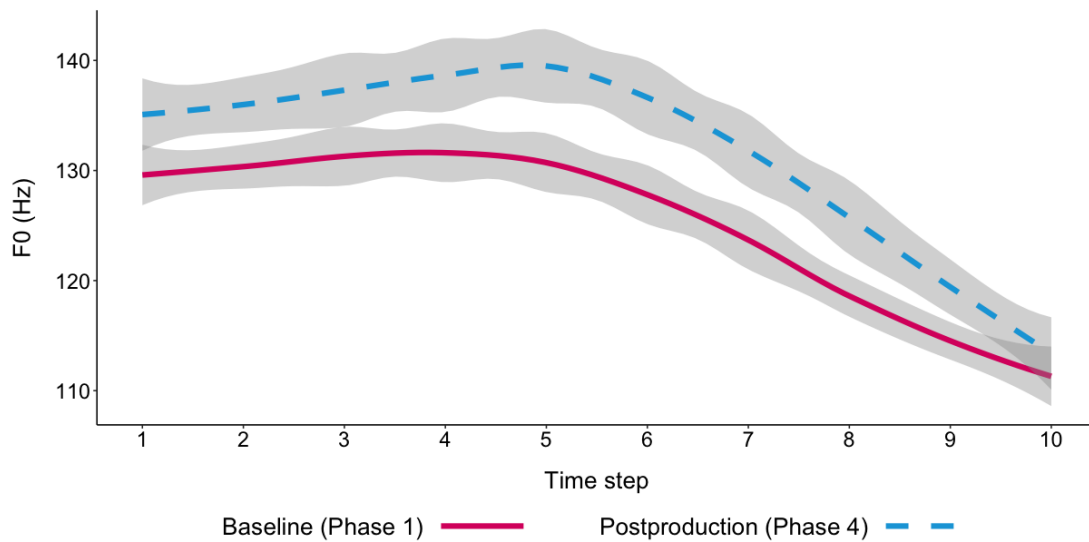


Figure 6.8: Results from Phase 4 postproductions (dashed blue line) compared to the Phase 1 baseline productions (solid pink line). Shading around each line represents the 95% confidence intervals.

To quantitatively assess this shift toward the high-prestige voice, I fit a linear mixed effects model, examining the effect of production phase (baseline vs. post-production) on f0 values. A summary of the model is provided in Table 6.1. PHASE was dummy coded, such that the intercept refers to the participant's baseline f0 (126 Hz), averaged over timestep and vowel quality. TIMESTEP (binned into thirds of the vowel), VOWELQUALITY (6 levels), and the voice imitated during

Phase 3 (3 levels) were included as effect-coded fixed effects. Random intercepts were included for word. Focusing on the key coefficients of the model, rows 1 and 2 show that the participant exhibited a statistically significant increase in f0 of approximately 7 Hz from baseline to postproduction.

Predictor	β (Hz)	95% CI	t	p
Intercept (baseline)	126	[124, 128]	107.73	< .001***
postproductions	7	[6, 8]	11.68	< .001***
voice imitated (high-prestige)	2	[-1, 4]	1.31	.20
voice imitated (low-prestige)	0	[-3, 2]	-0.12	.91
voice imitated (None)	-2	[-4, 1]	-1.05	.30
Beg	6	[5, 7]	10.19	< .001***
Mid	4	[3, 5]	7.47	< .001***
Fin	-10	[-12, -9]	-17.01	< .001***
u	5	[0, 9]	2.04	.046*
i	2	[-3, 7]	0.74	.46
e	1	[-5, 8]	0.33	.74
o	-1	[-5, 4]	-0.25	.80
ɔ	0	[-6, 6]	-0.01	.99
ɑ	-7	[-11, -4]	-4.31	< .001***
postproductions: Beg	0	[-2, 1]	-0.40	.69
postproductions: Mid	2	[1, 4]	2.68	.007**
postproductions: Fin	-2	[-3, 0]	-2.08	.04*

Table 6.1: Linear mixed-effects model results for f0. Phase (baseline vs. postproduction) was dummy coded, timestep, voice imitated during Phase 3 (high-prestige, low-prestige, or None) and vowel quality were effect-coded, and the interaction between TIMESTEP and PHASE was included as well: `lmer(f0~PHASE*TIMESTEP + VOWELQUALITY + IMITATIONVOICE + (1|word))`.

For dynamically changing synthetic speech with an f0 centered around 120 Hz, Klatt (1973) reported just-noticeable differences (JNDs) in f0 of approximately 2-2.5 Hz. The observed 7-Hz increase here is therefore substantially larger than these experimentally established discrimination thresholds. Although the perceptibility of the present difference would ultimately need to be confirmed experimentally, its magnitude suggests that the shift toward the f0 pattern of the high-prestige

voice is not only measurably distinct, but is also perceptible to listeners.

While more speakers are needed to confirm the results presented here, the data supports previous research that speakers store detailed phonetic and social information from the speakers they interact with. Moreover, this information is not simply retained in working memory for short durations, but exposure to certain phonetic properties from voices perceived as socially prestigious produces measurable changes in a speaker's later productions. In short, these findings suggest that stored speech representations must contain detailed information. The kinds of abstract symbolic representations posited by generative theories do not capture the full gamut of linguistic competence speakers possess.

Chapter 7

Conclusion

In this dissertation, I have addressed two central theoretical questions: (i) how abstract can mental representations of speech be, and (ii) is a lexicon of detailed speech memories sufficient to explain a speaker’s linguistic knowledge of their language? Drawing on empirical evidence from two Pakistani languages — Punjabi and Mankiyali — I have argued, first, that stored speech representations can be completely *covert*, containing information that is not present in any of their surface realizations (chapter 2). In chapter 3, I showed that, given an appropriate learning architecture, a learner can recover this covert analysis within the hypothesis space considered.

Chapter 6 returns to the question of representational substance from the opposite direction, presenting preliminary evidence that speakers retain highly detailed phonetic and socially indexed information alongside more abstract representations. Taken together, chapters 2 and 6 suggest that the same Punjabi lexical items are represented at multiple levels: the forms used to motivate covert URs in chapter 2 are also those for which talker-specific phonetic detail appears to be

retained and reflected in subsequent production in chapter 6.

Turning to consider the locus of linguistic knowledge in chapter 4, I presented phonetic and phonological evidence from Mankiyali demonstrating that the neutralization of a phonological contrast can be phonetically complete. Then, in chapter 5, I outlined a series of usage-based simulations of exemplar accumulation to argue for the existence of higher-order linguistic knowledge — separate from speech representations stored in the lexicon — that operates over and shapes future speech production patterns. In these simulations, once the interaction of production and similarity biases defined a stable target for a given acoustic parameter, introducing exemplars with different values did not by itself shift that target. A persistent shift of the target state instead required recalibration of the biases that shape production. Crucially, control over this recalibration constitutes a form of higher-order linguistic knowledge that cannot be located within episodic representations themselves. A lexicon of detailed speech memories is therefore insufficient on its own to account for production patterns.

Taken together, these conclusions about representational substance and the locus of linguistic knowledge challenge eliminativist usage-based conceptions of speech, in which linguistic knowledge is said to be reducible to a lexicon of richly detailed exemplars accumulated over an individual's linguistic experience. At the same time, they challenge classical generative theories, in which fine-grained phonetic detail and social information are not considered integral components of linguistic competence. This means that neither theoretical tradition, on its own, provides an adequate account of the full range of representational and production patterns examined in this dissertation.

Because multiple levels of representation appear to coexist, this synthesis raises the

question of how they interact. Do more abstract levels of representation emerge over time after sufficient distributional evidence has been gathered by the language user? If so, as in the case of Punjabi (chapter 2), what prompts symbolic representations to become decoupled from the detailed network of experienced memory traces? These questions require further investigation and point toward several promising directions for future research in speech production, sound change, and sociolinguistics. One important next step is a full-scale study examining the robustness of the prestige effects reported in chapter 6, using more speakers and counterbalanced voice–prestige pairings to avoid confounding f_0 with perceived prestige. Such a study could also investigate whether socially salient linguistic experience can induce the recalibration of production biases implicated by chapter 5. Another important direction involves testing the predictions of the exemplar simulations across languages exhibiting different degrees of phonetic neutralization. Ultimately, these lines of research must converge onto a formal model specifying how episodic memories, abstract categories, and higher-order linguistic knowledge jointly shape speech production.

The central conclusion of this dissertation is that speech production cannot be explained by any single level of representation or by a lexicon of speech representations alone. Speakers appear to retain detailed memories of experienced speech, maintain abstract categories that may diverge substantially from that experience, and possess higher-order knowledge that regulates how representations and phonetic pressures shape production. An adequate theory of speech production must account for all three of these components and, crucially, explain how they interact. Taken together, the Punjabi and Mankiyali data examined here motivate a hybrid architecture in which abstract categories and detailed episodic memories are com-

plementary components of linguistic knowledge, while a higher-order mechanism regulates their joint influence on speech production.

Bibliography

- Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2):119–161.
- Albright, A. C. (2002). *The identification of bases in morphological paradigms*. PhD thesis, UCLA.
- Allen, B. and Becker, M. (2015). Learning alternations from surface forms with sublexical phonology. Ms., University of British Columbia and Stony Brook University.
- Avelino, H., Zariquiey, R., and Pérez-Silva, J. I. (2020). Nasal coarticulation and prosody in Kakataibo. *Phonetica*, 77:29–54.
- Baart, J. L. G. (2014). *Tonal features in languages of northern Pakistan*, pages 1–13. Amsterdam and Philadelphia, PA: John Benjamins.
- Baayen, R. H., Davidson, D. J., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390–412.
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40:177–189.
- Baković, E. (2011). *Opacity and Ordering*. Blackwell Publishing Ltd.
- Baković, E. and Blumenfeld, L. (2024). A formal typology of process interactions. *Phonological Data and Analysis*, 6(3):1–43.
- Barker, M. A.-A.-R. (1964). *Klamath dictionary*. Berkeley: University of California Press.
- Bashir, E. and Conners, T. J. (2019). *A descriptive grammar of Hindko, Panjabi, and Saraiki*. Mouton-CASL Grammar Series. De Gruyter Mouton.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2025). Linear mixed-effects models using ‘eigen’ and s4. *[Computer software program]* Available from <https://cran.r-project.org/package=lme4>.

- Becker, M., Clemens, L., and Nevins, A. (2017). Generalization of French and Portuguese plural alternations and initial syllable protection. *Natural Language & Linguistic Theory*, 35(2):299–345.
- Becker, M., Ketrez, N., and Nevins, A. (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87(1):84–125.
- Becker, M., Nevins, A., and Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 88(2):231–268.
- Beddor, P. S. (2007). Nasals and nasalization: The relation between segmental and coarticulatory timing. In *Proceedings of the XVIth International Congress of Phonetic Sciences*, pages 249–254. Universität des Saarlandes, Saarbrücken, Germany.
- Beddor, P. S. and Krakow, R. A. (1999). Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *Journal of the Acoustical Society of America*, 106(5):2868–2887.
- Bell-Berti, F. (1993). *Understanding velic motor control: studies of segmental context*, volume 5 of *Phonetics and Phonology*. Academic Press.
- Bennett, R. T., Aguinda, S., Lucitante, H., and Anderbois, S. (to appear). Phonetic evidence for contextual underspecification of nasality in A’ingae. *Phonology*.
- Bennett, R. T. and Eischens, B. (2026). "La neutralización incompleta de la nasalidad en el Mixteco: El rol de la abstracción" [Incomplete neutralization of nasality in Mixtec: the role of abstraction]. Presentation at Sound Systems of Latin America (SSLA) 5.
- Benua, L. (2000). *Phonological relations between words*. Garland, New York.
- Bermúdez-Otero, R. (2011). Cyclicity. In van Oostendorp, M., Ewen, C., Hume, E., and Rice, K., editors, *The Blackwell companion to phonology*, pages 2019–2048. Wiley-Blackwell, Malden, MA.
- Bermúdez-Otero, R. (2018). In defence of underlying representations: Latin rhotacism, French liaison, Romanian palatalization. *Probus*, 30(2):171–214.
- Bhatia, T. K. (1993). *Punjabi*. Descriptive Grammars. Routledge.
- Bloomfield, L. (1933). *Language*. New York: Holst, Rinehart and Winston, Inc.
- Boersma, P. and Weenink, D. (2015). Praat: Doing phonetics by computer (version 5.4.21).

- Bonett, D. (2021). *Statistical Methods for Psychologists: Introduction to Linear Statistical Models*, volume 2. University of California, Santa Cruz. <https://dgbonett.sites.ucsc.edu/statistical-methods-for-psychologists/>.
- Bowers, D. A. (2015). *A system for morphophonological learning and its consequences for language change*. PhD thesis, University of California, Los Angeles, Los Angeles, CA.
- Braunschweiler, N. (1997). Integrated cues of voicing and vowel length in German: A production study. *Language and Speech*, 40(4):353–376.
- Braver, A. (2013). *Degrees of incompleteness in neutralization: paradigm uniformity in a phonetics with weighted constraints*. PhD thesis, Rutgers University.
- Braver, A. (2014). Imperceptible incomplete neutralization: production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, 152:24–44.
- Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology*, 36:1–36.
- Braver, A. and Kawahara, S. (2016). Incomplete neutralization in Japanese monomoraic lengthening. In Albright, A. C. and Fullwood, M. A., editors, *Proceedings of the Annual Meeting on Phonology*, volume 2.
- Brinkerhoff, M. L. (2025). *Voice Quality and Laryngeal Complexity in Santiago Laxopa Zapotec*. PhD thesis, University of California, Santa Cruz.
- Brinkerhoff, M. L. and McGuire, G. (2025). Using residual h1* for voice quality research. *JASA Express Letters*, 5(2):025501.
- Brown, V. (1990). *The social and linguistic history of a merger: /i/ and /e/ before nasals in Southern American English*. PhD thesis, Texas A&M University.
- Burzio, L. (1996). *Surface constraints versus underlying representations*, pages 118–136. Salford: European Studies Research Institute, University of Salford.
- Burzio, L. (2000). Segmental contrast meets output-to-output faithfulness. *The Linguistic Review*, 17:367–384.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia, PA: John Benjamins.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2023). *What is usage-based linguistics?* Wiley-Blackwell.
- Byrd, D., Tobin, S., Bresch, E., and Narayanan, S. (2009). Timing effects of

- syllable structure and stress on nasals: a real-time MRI examination. *Journal of Phonetics*, 37(1):97–110.
- Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics*, 63:87–105.
- Chai, Y. and Garellek, M. (2022). On H1–H2 as an acoustic measure of linguistic phonation type. *Journal of the Acoustical Society of America*, 152:1856–1870.
- Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America*, 102(4):2360–2370.
- Chomsky, N. and Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics*, 1(2):97–138.
- Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper and Row, New York.
- Cohn, A. C. (1993). Nasalization in english: Phonology or phonetics. *Phonology*, 10(1):43–81.
- Cohn, A. C. (2006). *Is there Gradient Phonology?*, chapter 2, pages 25–44. Oxford University Press.
- Cole, J. and Hualde, J. (2011). Underlying representations. In van Oostendorp, M., Ewen, C., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*, pages 1–26. Wiley-Blackwell, Malden, MA.
- Cole, J. and Kisseberth, C. (1994). An optimal domains theory of harmony. *Studies in the Linguistic Sciences*, 24(2):101–114.
- Cole, J. and Kisseberth, C. (1995). Nasal harmony in optimal domains theory. *WECOL*, 7:44–58.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual rout cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1):204–256.
- Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., and Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, 28(2):197–234.
- Dawson, M. R. W. (2013). *Mind, body, world: Foundations of cognitive science*. AU Press.
- Delvaux, V., Demolin, D., Harmegnies, B., and Soquet, A. (2008). The aerodynamics of nasalization in French. *Journal of Phonetics*, 36(4):578–606.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. In *Royal Statistical Society*.

- Desmeules-Trudel, F. and Brunelle, M. (2018). Phonotactic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Québécois French and Brazilian Portuguese. *Journal of Phonetics*, 69:43–61.
- Dhillon, R. (2007). Stress in punjabi. In *Berkeley Linguistics Society*, volume 33, pages 84–95.
- Dinnsen, D. A. and Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, 12:49–60.
- Dow, M. (2014). Hybrid opacity in Berbice Dutch Creole. In Kingston, J., Moore-Cantwell, C., Pater, J., and Staubs, R., editors, *Proceedings of the Annual Meetings on Phonology*, volume 2, Washington, DC. Linguistic Society of America.
- Du, N. and Durvasula, K. (2022). Phonetically incomplete neutralisation can be phonologically complete: evidence from Huai’an Mandarin. *Phonology*, 39(4):559–595.
- Du, N. and Durvasula, K. (2025). *Psycholinguistics and Phonology: The forgotten foundations of Generative phonology*. Cambridge Elements in Phonology. Cambridge: Cambridge University Press.
- Eischens, B. and Hedding, A. (2024). San martin peras mixtec. *Journal of the International Phonetic Association*, 54:811–852.
- Ellis, L. and Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from epg and ema data. *Journal of Phonetics*, 30(3):373–396.
- Englert, E. G. (2022). A first look at Mankiyali morphology. Master’s thesis, University of North Texas.
- Ettlinger, M. (2008). *Input-drive opacity*. PhD thesis, University of California, Berkeley.
- Ettlinger, M. (2009). Variation as a window into opacity. In Elliott, M., Kirby, J., Sawada, O., Staraki, E., and Yoon, S., editors, *Chicago Linguistics Society (CLS) 43*, Chicago. University of Chicago, Chicago Linguistic Society.
- Garellek, M. (2019). *The phonetics of voice*, chapter 4, pages 75–106. Routledge.
- Garellek, M., Ritchart, A., and Kuang, J. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics*, 59:110–121.
- Gerfen, C. (2002). Andalusian codas. *Probus*, 14:247–277.
- Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological Review*, 105(2):251–279.

- Goldrick, M. and Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*, 99:1–16.
- Gorman, K. (2013). *Generative phonotactics*. PhD thesis, University of Pennsylvania, Philadelphia, PA.
- Gouskova, M. (2012). Unexceptional segments. *Natural Language & Linguistic Theory*, 30(1):79–133.
- Gouskova, M. (2023). MSCs in positional neutralisation: the problem of gapped inventories. *Phonology*, 40(3–4):229–265.
- Gouskova, M. and Becker, M. (2013). Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language & Linguistic Theory*, 31(3):735–765.
- Grimm, J. (1822). *Deutsche Grammatik*. Gottingen: Dieterichschen Buchhandlung.
- Halle, M. (1978). Knowledge unlearned and untaught: What speakers know about the sounds of their language. In Halle, M., Bresnan, J., and Miller, G. A., editors, *Linguistic Theory and Psychological Reality*, pages 294–304. MIT Press, Cambridge, MA.
- Hamann, S. R. (2003). *The phonetics and phonology of retroflexes*. PhD thesis, Utrecht University.
- Hauser, I. and Hughto, C. (2020). Analyzing opacity with contextual faithfulness constraints. *Glossa: a journal of general linguistics*, 5(1):1–33.
- Hayes, B. (1999). Phonological restructuring in Yidin^y and its theoretical consequences. In Hermans, B. and van oostendorp, M., editors, *The derivational residue in phonological Optimality Theory*, pages 175–205. John Benjamins, Amsterdam.
- Hayes, B. (2001). *Phonological acquisition in Optimality Theory: the early stages*, pages 158–203. Cambridge: Cambridge University Press.
- Hayes, B. and White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, 44(1):45–75.
- Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.
- Henderson, J. B. (1984). *Velopharyngeal function in oral and nasal vowels: A cross-language study*. PhD thesis, University of Connecticut.
- Hooper, J. B. (1976). *Word frequency in lexical diffusion and the source of morphophonological change*. North Holland.

- Hussain, Q., Proctor, M., Harvey, M., and Demuth, K. (2019). Punjabi (Lyallpuri variety). *Journal of the International Phonetic Association*, 50(2):282–297.
- Hyman, L. (2018). Why underlying representations? *Journal of Linguistics*, 54:591–610.
- Jakubíček, M., Kilgarriř, A., Kovář, V., Rychlý, P., and Suchomel, V. (2013). The TenTen Corpus family. In *7th International Corpus Linguistics Conference CL*, pages 125–127.
- Jarosz, G. (2006a). *Rich lexicons and restrictive grammars - Maximum likelihood learning in Optimality Theory*. PhD thesis, Johns Hopkins University.
- Jarosz, G. (2006b). Richness of the base and probabilistic unsupervised learning in Optimality Theory. In *Proceedings of the Eighth Meeting of the ACL Special Interest Group on Computational Phonology*, pages 50–59.
- Jarosz, G. (2009). Restrictiveness and phonological grammar and lexicon learning. In Elliot, M., Kirby, J., Sawada, O., Staraki, E., and Yoon, S., editors, *Proceedings of the 43rd Annual Meeting of the Chicago Linguistics Society*, volume 43, pages 125–134.
- Jarosz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language. Special Issue on Computational Models of Child Language Learning*, 37(3):565–606.
- Jarosz, G. (2015). Expectation driven learning of phonology. University of Massachusetts manuscript.
- Jarosz, G. (2019). Computational modeling of phonological learning. *Annual Review of Linguistics*, 5:67–90.
- Johnson, K. (1997). *Speech Perception without Speaker Normalization*, chapter 8, pages 145–165. San Diego: Academic Press.
- Johnson, K. (2006). Resonance in exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34(4):485–499.
- Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (1994). Infants’ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33:630–645.
- Kaplan, A. (2015). Positional neutralization in an exemplar model: The role of unique inflectional bases. In *Proceedings of the 2014 Annual Meeting on Phonology*.
- Kaplan, A. (2017). Incomplete neutralization and the (a)symmetry of paradigm uniformity. In Kaplan, A., Kaplan, A., McCarvel, M. K., and Rubin, E. J.,

- editors, *West Coast Conference on Formal Linguistics (WCCFL) 34*, pages 319–328, Somerville, MA. Cascadilla Proceedings Project. Available online at <http://www.lingref.com/>, document #3326.
- Kawahara, S. (2001). Similarity among variants: Output-variant correspondence. Senior Thesis, International Christian University. Available online as ROA-497, Rutgers Optimality Archive, <http://roa.rutgers.edu/>.
- Kenstowicz, M. and Kisseberth, C. (1977). *Topics in Phonological Theory*. Academic Press.
- Kenstowicz, M. and Kisseberth, C. (1979). *Generative phonology: description and theory*. New York: Academic Press.
- Kim, H. and Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, 24:295–312.
- Kiparsky, P. (1973). *Abstractness, opacity, and global rules*, pages 57–86. Tokyo: TEC.
- Kiparsky, P. (1976). *Abstractness, opacity, and global rules*. The Hague: Mouton.
- Kiparsky, P. (1982). *How abstract is phonology?*, chapter 6, pages 119–163. Foris Publications.
- Kirby, J. and Puggaard-Rode, R. (2025). *PraatSauce. Praat-based tools for spectral processing*. Version 1.0.3. GitHub: [kirbyj/praatsauce](https://github.com/kirbyj/praatsauce).
- Kirchner, R. (2011). Preliminary thoughts on “phonologisation” within an exemplar-based speech processing system. *UCLA Working Papers in Linguistics*, 6.
- Kirchner, R., Moore, R. K., and Chen, T.-Y. (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics*, 38(4):540–547.
- Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *Journal of the Acoustical Society of America*, 53(1):8–16.
- Kochetov, A. and Pouplier, M. (2008). Phonetic variability and grammatical knowledge: an articulatory study of korean place assimilation. *Phonology*, 25(3):399–431.
- Krakow, R. A. (1989). *The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures*. PhD thesis, Yale University.
- Krakow, R. A. (1993). *Nonsegmental influences on velum movement patterns: Syllables, sentences, stress, and speaking rate*. Academic Press.

- Krämer, M. (2012). *Underlying Representations*. Cambridge: Cambridge University Press.
- Kruszewski, M. (1883). *Očrk Nauki O Jazyke (An outline of linguistic science)*. *Writings in general linguistics*, pages 43–173. John Benjamins.
- Kuo, J. (2023). Evidence for prosodic correspondence in the vowel alternations of Tgdaya Seediq. *Phonological Data and Analysis*, 5:1–31.
- Kuo, J. (2024a). Phonetic naturalness in the reanalysis of Samoan thematic consonant alternations. *Journal of Phonetics*, 107:1–18.
- Kuo, J. (2024b). Phonological reanalysis is guided by markedness: the case of Malagasy weak stems. *Phonology*, 41(3):1–35.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19(3):273–309.
- Lahiri, A., Schriefers, H., and Kuijpers, C. (1987). Contextual neutralization of vowel length: Evidence from Dutch. *Phonetica*, 44:91–102.
- Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence testing for psychological research: a tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2):259–269.
- Lapierre, M. (2023). Two types of [NT]s in Panãra: Evidence for temporally ordered subsegmental units. *Glossa*, 8(1).
- Lenth, R. V. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.3.
- Lloret, M.-R. and Pons-Moll, C. (2016). Catalan vowel epenthesis as evidence for the free ride approach to morphophonemic learning. *Linguistic Inquiry*, 47(1):147–157.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: W.H. Freeman.
- Mascaró, J. (2016). Morphological exceptions to vowel reduction in Central Catalan and the problem of the missing base. *Catalan Journal of Linguistics*, 15:27–51.
- Matsui, M. (2015). *Roshia-go ni okeru yuusei-sei no tairitsu to tairitsu no jyakka: onkyo to chikaku*. PhD thesis, Hiroshima University. Voicing contrast and contrast reduction in Russian: acoustics and perception.
- McCarthy, J. J. (1999). Sympathy and phonological opacity. *Phonology*, 16(3):331–399.

- McCarthy, J. J. (2000). Harmonic serialism and parallelism. In Hirotani, M., Coetzee, A., Hall, N. E., and Kim, J. y., editors, *Proceedings of the 30th meeting of the North East Linguistic Society*, pages 501–524.
- McCarthy, J. J. (2005a). Optimal paradigms. In Downing, L., Hall, T. A., and Raffelsiefen, R., editors, *Paradigms in phonological theory*, pages 170–210. Oxford University Press, Oxford, UK.
- McCarthy, J. J. (2005b). Taking a free ride in morphonemic learning. *Catalan Journal of Linguistics*, 4(1):19–55.
- McCarthy, J. J. (2007a). Derivations and levels of representation. In de Lacy, P., editor, *The Cambridge handbook of phonology*. Cambridge University Press, Cambridge, UK.
- McCarthy, J. J. (2007b). *Hidden Generalizations: Phonological Opacity in Optimality Theory*. Sheffield: Equinox.
- Monahan, P. J. and Idsardi, W. J. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and Cognitive Processes*, 25(6):808–839.
- Morley, R. (2023). *Sound structure and sound change: A modeling approach*. Number 4. Language Science Press, Berlin.
- Nelson, S. and Heinz, J. (2025). The blueprint model of production. *Phonology*, 42(12):1–33.
- O’Hara, C. (2017). How abstract is more abstract? learning abstract underlying representations. *Phonology*, 34(2):325–345.
- Paramore, J. C. (2021). Mankiyali phonology: description and analysis. Master’s thesis, University of North Texas, <https://digital.library.unt.edu>.
- Paramore, J. C. (2025). Learning covert URs via disparity minimization. In *Eighth Annual Meeting of the Society for Computation in Linguistics (SCiL)*, volume 8.
- Paramore, J. C. and Aurangzeb (2025). The acoustic correlates of word-level stress and focus-related prominence in Mankiyali. *Journal of the International Phonetic Association*, pages 1–42.
- Paramore, J. C. and Bennett, R. T. (2025). Covert URs: evidence from nasalization in Pakistani Punjabi. Manuscript, UC Santa Cruz.
- Pater, J., Staubs, R., Jesney, K., and Smith, B. (2012). Learning probabilities over underlying representations. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, pages 62–71.

- Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46(2-3):115–154.
- Pierrehumbert, J. (2006). The statistical basis of an unnatural alternation. In Goldstein, L., Whalen, D., and Best, C., editors, *Papers in Laboratory Phonology VIII*, pages 81–107. Mouton de Gruyter, Berlin.
- Pierrehumbert, J. B. (2001). *Exemplar dynamics: Word frequency, lenition and contrast*, pages 137–157. Amsterdam: John Benjamins.
- Pierrehumbert, J. B. (2002). Word-specific phonetics. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology*, volume 7, pages 101–140.
- Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.
- Pinker, S. (2002). Four decades of rules and associations, or whatever happened to the past tense debate? In *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*. MIT Press, Cambridge, MA.
- Plug, L., Shitaw, A., and Heselwood, B. (2019). Inter-consonantal intervals in Tripolitanian Libyan Arabic: Accounting for variable epenthesis. *Laboratory Phonology*, 10(1):5.
- Port, R. F. and Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17:257–282.
- Port, R. F. and O’Dell, M. L. (1985). Neutralization of syllable-final voicing in german. *Journal of Phonetics*, 13:455–471.
- Poupplier, M., Rodriquez, F., Alderton, R., Lo, J. J., Reinisch, E., Evans, B. G., and Carignan, C. (2023). The window of opportunity: Anticipatory nasal coarticulation in three languages. In Skarnitzl, R. and Volín, J., editors, *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, pages 2085–2089. International Phonetic Association.
- Poupplier, M., Rodriquez, F., Lo, J. J., Alderton, R., Evans, B. G., Reinisch, E., and Carignan, C. (2024a). Language-specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics*, 107:101365.
- Poupplier, M., Rodriquez, F., Lo, J. J., Alderton, R., Evans, B. G., Reinisch, E., and Carignan, C. (2024b). Language-specific and individual variation in anticipatory nasal coarticulation: A comparative study of American English, French, and German. *Journal of Phonetics*, 107.
- Prince, A. and Smolensky, P. (1993/2004). *Optimality Theory: constraint interaction in Generative grammar*. Malden, Mass: Blackwell Publishers.

- Prince, A. and Tesar, B. (2004). *Learning Phonotactic Distributions*, pages 245–291. Cambridge: Cambridge University Press.
- Rasin, E. and Katzir, R. (2016). On evaluation metrics in Optimality Theory. *Linguistic Inquiry*, 47(2):235–282.
- Rasin, E. and Katzir, R. (2018). Learning abstract underlying representations from distributional evidence. In Hucklebridge, S. and Nelson, M., editors, *North East Linguistic Society*, volume 48. MA: GLSA.
- Rasin, E. and Katzir, R. (2020). A conditional learnability argument for constraints on underlying representations. *Journal of Linguistics*, pages 745–773.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3):372–422.
- RCoreTeam (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Roettger, T. B., Winter, B., Grawunder, S., Kirby, J. P., and Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 43:11–25.
- Rundus, D. (1970). Analysis of rehearsal processes in free recall. Technical Report 159, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA.
- Rysling, A. (2017). *Preferential early attribution in segmental parsing*. PhD thesis, University of Massachusetts, Amherst.
- Sancier, M. L. and Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25(4):421–436.
- Scobbie, J. M., Coleman, J. S., and Bird, S. (1996). Key aspects of declarative phonology. In Durand, J. and Laks, B., editors, *Current trends in phonology: Models and methods*, volume 2, pages 685–709. University of Salford, European Studies Research Institute.
- Seidenberg, M. S. and Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive science*, 38(6):1190–1228.
- Shackle, C. (2003). *Panjabi*, chapter 16, pages 581–621. Routledge Language Family Series. Routledge.
- Shosted, R. K. (2015). Nasal vowels are not [+NASAL] oral vowels. In Smith, J. and Ihsane, T., editors, *Romance Languages and Linguistic Theory*, volume 16, pages 63–76.

- Shosted, R. K., Carignan, C., and Rong, P. (2012). Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. *Journal of the Acoustical Society of America*, 131(1):455–465.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138.
- Smolensky, P. (1996). The initial state and ‘Richness of the Base’ in Optimality Theory. *Rutgers Optimality Archive*, ROA-154.
- Solé, M.-J. (1992). Phonetic and phonological processes: The case of nasalization. *Language and Speech*, 35(1):29–43.
- Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1):1–23.
- Solé, M.-J. (2007). *Controlled and mechanical properties in speech*, chapter 18. Oxford University Press.
- Speelman, D., Heylen, K., and Geeraerts, D., editors (2018). *Mixed-effects regression models in Linguistics*. Springer.
- Studdert-Kennedy, M. (1987). *The phoneme as a perceptuomotor structure*. Haskins Laboratories.
- Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America*, 142(4):2469–2482.
- Sylak-Glassman, J., Farmer, S., and Michael, L. (2014). An Agreement by Correspondence analysis of Máihiki nasalization harmony. Presentation at *ABC↔* conference.
- Tesar, B. (2014). *Output-Driven Phonology*. Cambridge: Cambridge University Press.
- Tesar, B. (2016). Phonological learning with output-driven maps. *Language Acquisition*, 24(2):148–167.
- Tesar, B. and Prince, A. (2003). Using phonotactics to learn phonological alternations. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 29, pages 241–269. Chicago Linguistic Society.
- Trommer, J. (2013). Stress uniformity in Albanian: morphological arguments for cyclicity. *Linguistic Inquiry*, 44(1):109–143.
- van Heuven, V. J. and Turk, A. (2021). *Phonetic correlates of word and sentence stress*, pages 150–165. Oxford University Press.

- Vietti, A., Alber, B., and Vogt, B. (2018). Initial laryngeal neutralisation in Tyrolean. *Phonology*, 35(1):79–114.
- Walker, R. (2003). *Reinterpreting transparency in nasal harmony*, pages 37–72. Amsterdam: John Benjamins.
- Walker, R. (2005). Weak triggers in vowel harmony. *Natural Language & Linguistic Theory*, 23(4):917–989.
- Walker, R. (2014). Nonlocal trigger-target relations. *Linguistic Inquiry*, 45(3):501–523.
- Wang, Y. and Hayes, B. (2025). Learning phonological underlying representations: the role of abstractness. *Linguistic Inquiry*, pages 1–39.
- Warner, N., Jongman, A., Sereno, J., and Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics*, 32:251–276.
- Wedel, A. B. (2004). *Self-organization and categorical behavior in phonology*. PhD thesis, University of California, Santa Cruz.
- Wedel, A. B. (2006). Exemplar models, evolution and language change. *Linguistic Review*, 23(3):247–274.
- Wedel, A. B. (2007). Feedback and regularity in the lexicon. *Phonology*, 24:147–185.
- Wedel, A. B. (2012). Lexical contrast maintenance and the organization of sub-lexical contrast systems. *Language and Cognition*, 4(4):319–355.
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. Taylor & Francis Group, 2 edition.
- Whitney, W. D. (1889). *Sanskrit Grammar*. Cambridge: Harvard University Press.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.
- Wiese, R. (2000). *The Phonology of German*. Oxford University Press.
- Wilson, C. and Gallagher, G. (2018). Accidental gaps and surface-based phonotactic learning: A case study of South Bolivian Quechua. *Linguistic Inquiry*, 49(3):610–623.
- Wolf, M. (2008). *Optimal interleaving: Serial phonology-morphology interaction in a constraint-based model*. PhD thesis, University of Massachusetts, Amherst, Amherst, MA.

- Yu, A. C. L. (2007). Understanding near mergers: the case of morphological tone in Cantonese. *Phonology*, 24:187–214.
- Yu, A. C. L. (2023). The actuation problem. *Annual Review of Linguistics*, 9:215–231.
- Zahid, S. and Hussain, S. (2012). An acoustic study of vowel nasalization in Punjabi. In *Proceedings of the Conference on Language & Technology*, volume 61.
- Zellou, G. (2022). *Coarticulation in Phonology*. Cambridge Elements in Phonology. Cambridge: Cambridge University Press.
- Zeng, Y., Wang, C., and Zhang, J. (2025). Cascading activation in spoken word production drives incomplete neutralization: An internet-based study of Mandarin 3rd tone sandhi. *Journal of Phonetics*, 112.