## Title
Neural oscillatory and ERP indices of prediction in emotional speech



## Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

## Authors
Kikoťová, Natàlie
Chladkova, Katerina

## Publication Date
2024

Peer reviewed

# Neural oscillatory and ERP indices of prediction in emotional speech

**Natálie Kikoťová (kikotova@praha.psu.cas.cz)**
Institute of Psychology, Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 00, Praha
Department of Psychology, Faculty of Arts, Charles University, Nám Jana Palacha 2, 116 38 Praha

**Kateřina Chládková (chladkova@praha.psu.cas.cz)**
Institute of Psychology, Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 00, Praha
Department of Medical Biophysics, Faculty of Medicine in Hradec Králové, Charles University, Šimkova 870, 500 03, Hradec Králové

## Abstract

The experiment reported here investigated the neural correlates of predictive processing of angry and neutral speech. Twenty-six participants listened to recordings of angry and neutral conversation segments, as well as to speech-shaped noise, while their EEG was recorded. Oscillatory power in the gamma band (30–80 Hz) and the N400 component of event-related potentials (ERP) to sentence-final words were analyzed. In comparison to neutral words, negative emotional valence significantly reduced the amplitude of the N400 elicited by sentence-final words. Furthermore, there was larger gamma power during exposure to angry speech in comparison to neutral speech. The results generally suggest increased prediction and facilitated semantic integration in negative as compared to neutral speech. To date, the predictability effects on gamma power have been reported with relation to the semantic-lexical content of words. The present findings demonstrate that gamma power is also modulated by the emotional content of speech.

**Keywords:** emotional speech; neural oscillations; gamma power; N400; predictive processing

## Introduction

Emotional speech plays an important role in everyday social interactions. The rapid decoding and successful comprehension of emotional speech is necessary for detection of threats or life-sustaining opportunities, as well as maintaining interpersonal relationships and achieving social goals. Therefore, from an evolutionary perspective, it is highly relevant that processing resources are oriented towards an emotionally salient speech signal, and that mechanisms facilitating speech comprehension are employed.

In spoken interactions, emotional significance can be expressed via two channels: the verbal channel that conveys explicit information about the semantic content, and the vocal channel that carries the acoustic information about the rhythm, pitch, intensity, and voice quality. While these two channels are typically integrated in natural interactions, studies show that each single channel on its own does elicit an emotional percept in the listener (Ben-David et al., 2016; Pell et al, 2009; Pell & Kotz, 2011; Van Bezooijen et al, 1983).

Findings from previous neuroimaging studies which used either written or spoken stimuli have revealed enhanced perceptual processing (Johnstone et al., 2006; Sander et al., 2005) and motivational evaluation (Fields & Kuperberg, 2016) of emotional words, as well as automatic attentional capture by emotional relative to neutral words (Wang & Bastiaansen, 2014). For example, color naming is slowed in an emotional Stroop task when the presented word has an emotional meaning (Williams et al., 1996), indicating that the emotional connotation leads to attentional disengagement from the task. Using an Affective Lexical Decision Task, Carretié et al. (2008) demonstrated that highly arousing negative taboo words lead to poorest performance, indicating an interference with the task due to attentional capture. While causing poorer performance on associated tasks, it is very likely that the arousal and/or the attentional capture associated with emotional speech facilitates the processing of the (emotional) speech itself.

The literature indeed shows that the emotional salience of words extends the semantic context of a sentence and has the ability to serve as an additional contextual resource which might serve to support the integration of incoming words into sentence context.

Previous studies investigating the effect of high-arousing context on the processing of neutral target words suggest that arousing words affect the attentional allocation for subsequent lexical processing (Hinojosa et al., 2012; Ding et al., 2015) and that emotional context influences the prediction of upcoming words (Ding et al, 2020).

At the level of neural language processing, semantic prediction, semantic integration, and contextual analysis of a word is relatively robustly indexed by the N400 component of event-related potentials (Berkum et al., 1999; Kutas & Hillyard, 1980). Using electroencephalography, the N400 is recorded as a negative-peaking component occurring approximately 400 ms after stimulus onset. Leaving the emotional dimension aside, the less predictable a word is given the lexical (or visual) context, the stronger N400 response it elicits. With respect to emotional stimuli, Zhang et al. (2021) demonstrated the effect of congruent emotional context on sentence comprehension. They used visually presented two-sentence discourses where negative emotion was conveyed either by both sentences or by neither sentence; followed by a target word, emotionally congruent with the preceding sentential context. In the former scenario, i.e. in congruent negative discourse, the target word elicited a



In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* 

smaller amplitude of the N400 than did a neutral target word in neutral discourse (i.e. the latter scenario). This suggests that emotional context might facilitate the semantic integration of an emotionally congruent word in a sentence, and that it might be easier to predict an emotional word in an emotionally congruent sentence context than a neutral word in a neutral sentence context.

A bit more recently, the neural processing of (not only semantic) information has been studied through analyzing the neural oscillatory activity during speech processing. Neural oscillations reflect the synchronized spiking of neuronal populations in cortical and subcortical areas of the brain. They are crucial for communication across different brain networks and support information transfer in different brain regions through the coordinated alternation of excitatory and inhibitory phases of firing neurons (Buzsaki & Draguhn, 2004).

With regard to speech processing, the oscillatory activity in the low, delta and theta bands (0.5–4 Hz, and 4–8 Hz, respectively) is predominantly responsible for speech segmentation along the word and syllable rates (Tune & Obleser, 2022). Higher, gamma band oscillations (at frequencies above 30 Hz) are related to sentence-level semantic processing and as such they reflect the predictability of the incoming words based on the preceding sentence context (Bastiaansen & Hagoort, 2015, Mai et al., 2016; but note that gamma band activity, and particularly gamma phase coupling to the amplitude of lower-rate oscillations seems to be linked also to phonetic/phonological, segmental or subsegmental level of processing, Attaheri et al. 2022). Hald et al. (2006) observed an increase in the gamma band power during the processing of semantically correct sentences, which might suggest that gamma oscillations support semantic unification operations, and that larger gamma power could index more integrated semantic processing. That is, more predictable context seems to be reflected in greater gamma power. According to Wang et al. (2012) gamma power increase could be related to the agreement between the pre-activation of the neural representations of the predicted word and the actually incoming word.

However, it is important to note that oscillatory activity in the delta, theta, and gamma band has also been related to other cognitive, domain-general, processes. For instance, theta, delta, and gamma activity are considered to be involved in lexical memory retrieval (Osipova et al., 2006), attention (Pulvermüller et al., 1997), and concentration (Harmony, 2013).

While the N400 has been researched both in the context of semantic language processing and emotional speech, the gamma oscillatory activity has been studied only with respect to the processing of linguistic information as such. The processing of emotional speech is underresearched in the brain-rhythm literature in general; only a few studies have investigated the oscillatory dynamics associated with the processing of individual emotional words (Chen et al., 2013; Ku et al., 2022; Wang & Bastiaansen, 2014).

Here we aim to investigate whether both the N400 and the gamma activity index the integration or predictability of emotional content in conversational speech. Based on prior studies on the N400, we predict that in emotionally *congruent* sentential context, emotional (negative) words will elicit a smaller N400 than non-emotional (neutral) words. This is because a negative word is more predictable in negative context than a neutral word is in neutral context. In line with the literature on neural speech tracking, which suggests that gamma oscillations are stronger for semantically predictable contexts than for nonpredictable contexts, we predict in emotionally *congruent* sentential context, gamma power will be greater while listening to the emotional (negative) speech than while listening to non-emotional (neutral) speech.

As the emotional condition, we chose anger. This is because across emotion recognition studies, anger is the least misclassified emotion based on a speaker's voice (Fenster et al., 1977; Scherer et al., 2001), and several studies have shown that the emotional modulations are more pronounced in negative compared to positive stimuli (Fields & Kuperberg, 2012; Huang & Luo, 2006; Ito et al., 1998; Scott et al., 2009).

# Method

## Participants

Twenty-six native speakers of Czech participated in the experiment (18 females, 8 males, mean age = 22.12 years, age range = 19–27 years). All subjects were right-handed and reported normal hearing. None of the subjects suffered from any psychiatric disease nor had any neurological impairment. The experiment was approved by the ethics committee of the Institute of Psychology, Czech Academy of Sciences.

## Stimuli

**Speech material** For the speech material we created a list of 100 conversation segments (50 in the angry and 50 in the neutral condition). Each segment consisted of two sentences that were formulated in a way that the speaker is talking to another person. The last word of the second sentence was either a negative or a neutral word, according to the condition, emotionally congruent with the preceding context. An example of a neutral segment was *Nemusíš nic nastavovat manuálně. Tohle čidlo to všechno* ***kontroluje***. "You don't have to set anything manually. This sensor controls everything," and an example of an angry segment was *Přestaň na mě takhle blbě čumět. Ten tvůj přiblblej ksicht mě* ***irituje***. "Stop staring at me like that. This dumb face of yours is annoying me." The sentences, as well as the segment-final words (in bold font in the examples above), in the angry and neutral conditions were cross-matched for syllable count. The conversation segments were recorded by a female speaker with acting experience. The average intensity of each of the 100 individual conversation segments was equalized across segments using Praat (Boersma & Weenink, 1992–2024).

**Pilot rating tasks** The materials were rated on several scales in two pilot experiments, administered online (using Psychtoolkit, Stoet, 2010, 2017) with native speakers of Czech (different individuals than in the subsequent EEG experiment). The first pilot was a judgement task with written representation of the conversation segments in which participants rated valence, arousal, plausibility, and categorized the conversation segments selecting from the six basic emotion categories, namely, anger, disgust, fear, sadness, happiness, surprisal. After the first pilot, 12 out of original 112 segments were excluded from the materials (as outliers in valence or plausibility). The second pilot was a judgment task with the actual audio recordings in which a different group of pilot participants judged the segments on valence (on a three-point scale -1, 0, +1) and naturalness (on a 7-point scale between 1 and 7). The 50 neutral and the 50 negative conversation segments were perfectly separated on valence, and were comparable in naturalness, as shown in Figure 1.

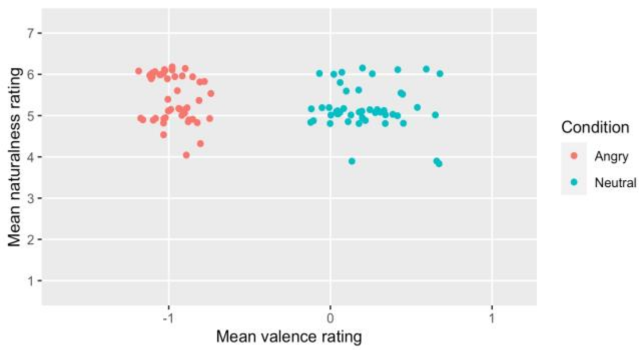


Figure 1: Valence and naturalness ratings per conversation segment and condition. Dots represent mean ratings pooled across 32 participants.

**Speech-shaped noise** The recorded conversation segments were resynthesized into speech-shaped noise using Praat and its native functions (Boersma & Weenink, 1992–2024, using the materials provided by ListenLab**,** 2023**)**. First, we calculated a long-term average spectrum for the two corresponding speech segments from the neutral and the angry condition (i.e. the two sentences from the neutral segment and the two sentence from the corresponding angry segment) and filtered a white noise signal with that spectral object. Then we converted each speech segment (this time, for the neutral and the angry condition separately) to an intensity and subsequently to an amplitude tier. And finally, we multiplied the filtered noise (common for the neutral and the respective angry segment) with the amplitude tier (specific to the neutral or to the angry segment). With this procedure, for each speech segment in each condition we obtained a speech-shaped noise that had the same envelope as the corresponding angry or neutral speech stimulus but spectral content that was identical between the angry and the neutral speech stimulus.

A block of speech-shaped noise segments always preceded the speech block. The speech-shaped noise stimuli served as a baseline condition allowing to compare the semantically rich speech condition to rhythmically similar but non-speech like stimuli. (The noise condition, crucially, allows to assess neural tracking of emotional prosody in particular, which will be analyzed and presented elsewhere).

## Procedure

The order of the conversation segments was randomized, with each conversation segment being repeated twice (not immediately after itself) within the presentation block. This resulted in a total of 100 trials per block. The trials were presented with an inter-trial interval randomly jittering between 390 and 410 ms.

Before the experiment, participants signed an informed consent form and filled out the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) questionnaire measuring their current affective state (the data from which is not part of the present analysis). Participants were tested in a quiet room, seated in a comfortable chair approximately 1 m away from a frontally located computer screen. Two loudspeakers were placed in 30° angles next to the monitor, and stimuli were presented at 65 dB SPL measured at the location of the participant's head.

Participants were presented with the four blocks of stimuli, with short intervening breaks between blocks for relaxation and refreshment. A noise block was always presented first and was followed by the speech block from the same condition, all participants heard blocks from both the neutral and the angry condition, and the order of the two conditions was counterbalanced (through random assignment) across the participants. For the speech-shaped noise blocks participants were asked to passively listen to the recordings. For the speech blocks, participants were instructed to listen attentively as in some of the trials they were to answer a question about what the speaker said; these comprehension checks occurred in ~1/10 of trials and were included to ensure the participants were paying attention to the semantic content of the recordings.

## EEG Recording & Preprocessing

EEG was recorded from 19 scalp electrodes placed according to the international 10/20 system, with an additional FCz electrode serving as an online reference. Two external sensors were placed at left and right mastoid, one at the outer canthi of the right eye, one below the right eye, and one on the nose. The signal was recorded at a 200-Hz sampling frequency, impedances were kept below 10 kΩ.

The EEG data were preprocessed using the EEGLAB toolbox (Delorme & Makeig, 2004) in Matlab (The Mathworks Inc., 2022). The signal was bandpass filtered at 0.1 and 80 Hz and re-referenced to the nose. The epoching and artifact rejection steps for the ERP and power analysis are described below.

**ERP data analysis** For the ERP analysis, the filtered data were epoched from -0.1 s to 1 s relative to target word onset (i.e., the final word in each conversation segment). The epochs were baseline-corrected to the 100-ms pre-stimulus interval. Epochs in which the absolute amplitude exceeded 100 μV were automatically rejected as artifacts. Subjects that had more than 70 % of rejected epochs were excluded from further analyses (n = 2).

For each subject, an average ERP waveform was computed for the angry and the neutral condition separately. A grand-average negative peak per condition (i.e. across all participants) was determined between 200 ms and 500 ms after word onset. In the per-participant average waveform, the N400 amplitude was then quantified as the mean amplitude in a 100-ms window centered at a grand-average negative peak, for the Cz and the Fz electrode separately.

**Time-frequency analysis** For the analysis of total power, the EEG data were segmented into 10-s epochs (leaving out cases in which the 10-s interval would have been interrupted by a comprehension question). This segmenting procedure yielded a total of 74 epochs in each speech condition and 98 epochs in each speech-shaped noise condition, per participant (the 10-s epochs allowed us to analyse the low, *delta* frequency range). Epochs in which the absolute amplitude exceeded 210 μV were automatically rejected as artifacts. Participants who had more than 60 % of rejected epochs would have been excluded from further analyses (n = 0). The epoched data were decomposed using a Morlet wavelet transform across 200 sliding windows (with the *newtimef* function in EEGLAB; which gradually adapts the number of wavelet cycles across the analyzed frequency range). The transformation was calculated in 0.1-Hz steps between 0.2 and 80 Hz, with 1 cycle at the lowest frequency and increasing by a factor of 0.5 for the higher frequency bins.

For each subject, total power in the gamma band (30–80 Hz) was computed across all epochs for each condition. The average total power in the noise epochs was then subtracted from the average total power in the corresponding speech epochs.

## Statistical analysis

The data were analyzed in R (R Core Team, 2024) with linear mixed-effects models (packages lme4, Bates et al., 2015, lmerTest, Kuznetsova et al., 2017). One model was fitted for the N400 amplitude and one model for the gamma power. Each model estimated the effect of condition, with a sum-to-zero contrast -negative vs. +neutral, and random intercepts for channel and for participant. Marginal means were estimated using the package ggeffects (Lüdecke, 2018).

# Results

## N400

The grand-average ERPs, averaged across Fz and Cz channels, are shown in Figure 2.

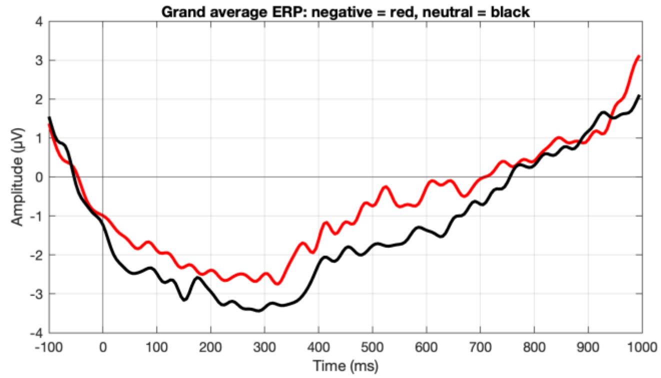


Figure 2: Grand average ERPs for negative (red) vs. neutral (black) words. The obligatory auditory ERP components such as N1 or P2 are not clearly discernible because the waveform is computed from target words that were non-repeating, 50 unique lexical tokens, embedded in a continuous stream of auditory speech.

The N400 model detected a significant intercept (estimate = -2.702, SE = 0.539, df = 4.124, t = -5.015, $p$ = 0.007) indicating that there was an overall negative ERP response. There was also a significant effect of condition (ß = -0.484, SE = 0.180, df = 70, t = -2.686, $p$ = .009) showing that neutral words evoked a larger negative response (N400) than the emotional negative words. The estimated means and confidence intervals per condition are shown in Figure 3.

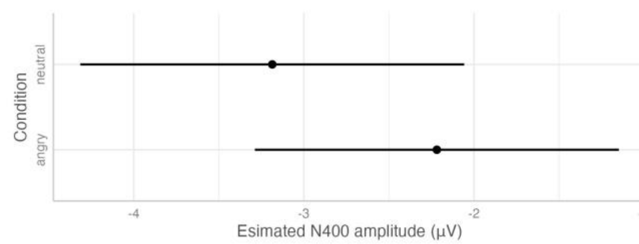


Figure 3: The estimated N400 amplitude for negative and neutral words (means and 95% confidence intervals).

## Gamma power

Figure 4 shows the grand-averaged total power in the range between 0 and 80 Hz in the angry and neutral condition.

The model for the gamma power revealed a main effect of condition (ß = -0.722, SE = 0.248, df = 25, t = -2.908, $p$ = .008) showing that gamma power was larger during angry compared to neutral speech; Figure 5 plots the estimated means and confidence intervals.

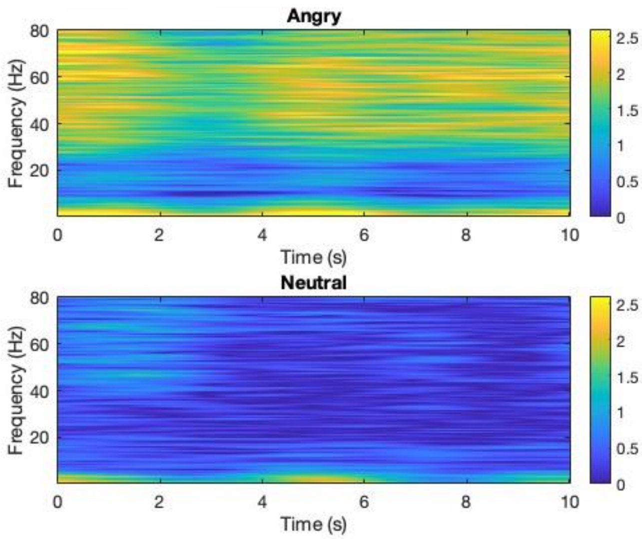


Figure 4: Grand-average total power in the angry (top) and the neutral (bottom) speech condition, as subtracted from the respective noise condition.

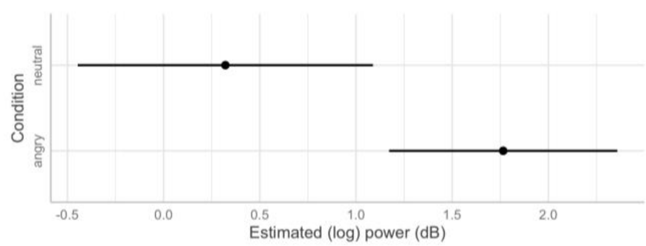


Figure 5: Estimated gamma power in the angry and neutral speech condition (referenced to the respective speech-shaped noise condition).

## Discussion

The present study focused on the perception of angry and neutral speech, measuring the oscillatory power in the gamma band and the ERP component N400, two indicators of prediction in speech comprehension. Participants listened to conversational two-sentence segments with either a negative or neutral valence and ending in words emotionally congruent with the preceding sentence context. Before each speech block, participants were exposed to speech-shaped noise.

The emotional valence of the heard conversational speech impacted the total power in the gamma band; gamma power (referenced to the respective speech-shaped noise conditions) was significantly larger during exposure to angry speech than during exposure to neutral speech. Furthermore, the emotional valence influenced the N400 response to the segment-final words, resulting in a decreased amplitude in response to negative words compared to neutral words.

### N400 as an index of semantic-emotional access

The present results contribute to the understanding of how N400 may index the processing the emotional content of words. The findings from previous studies investigating the N400 to emotional words are difficult to interrelate and interpret, as experimental paradigms and task demands seem to heavily affect the observed results. For example, some studies report an increased amplitude of the N400 when negative words are presented in a neutral context (De Pascalis et al., 2009; Grass et al., 2016; Herbert et al., 2008; Holt et al., 2009). This would suggest that the surprising valence of an emotional word after a preceding neutral context (where the context can be established by a preceding sentence or by the experimental task) might result in a deeper semantic evaluation or a more demanding semantic access, and thus also a larger N400. Other studies found a decrease in N400 to emotional words that are presented in an emotionally congruent context. In an attentional cueing paradigm, Kanske et al. (2011) reported that cues correctly predicting the emotional category of an upcoming word facilitate semantic integration such that the target emotional words that were preceded by a valid emotional cue caused a decreased amplitude of the N400. To make the picture even more complex, studies investigating emotional words presented in isolation report a reduced N400 to emotional words compared to neutral words, for example in a lexical decision task (Kanske & Kotz, 2007; Wang et al., 2019). Such reduced N400 to emotional words presented in lexical decision tasks has been interpreted as an index of facilitated lexical access for emotional compared to neutral words (Wang et al., 2019). The studies reviewed above tell us about how contextual predictability (or a complete lack of it) influences the N400 to an emotionally congruent or incongruent target word.

But does emotional valence of the target words alone affect the N400 response in cases where the preceding context is always emotionally congruent with the target word? This is what we can answer with our results. We find that, in emotionally congruent conversational context, emotional (namely, negative) words systematically elicit a smaller N400 than neutral words. The present findings of smaller N400 to negative words in the angry conversation, compared to neutral words in the neutral conversation are in line with studies reporting a reduced amplitude of the N400 in response to emotional words preceded by an emotionally congruent context than to neutral words in a neutral context. With the present auditory conversational stimuli, we replicated the findings of Zhang et al. (2021) on a reduced N400 to orthographically presented negative words in congruent negative context compared to neutral words in neutral context.

### Gamma power as an index of facilitated prediction

Besides the effects of emotional valence on the N400, we detected larger gamma power during the perception of angry speech compared to neutral speech. The literature indicates that the oscillatory activity in the gamma band is associated with the predictability of the incoming words based on the preceding sentence context (Bastiaansen & Hagoort, 2015; Hald et al., 2006; Mai et al., 2016). To date, the predictability effects on gamma power have been researched with respect to the semantic-lexical content of words. To what extent the emotional content influences gamma power has been unknown. The present findings show that the emotional content of conversation, too, modulates gamma power. Our study finds larger gamma power in angry conversation than in neutral conversation. Considering the literature on greater gamma activity in semantically predictable contexts, one can conclude that the context of angry speech is generally more predictable, and that the additional dimension of emotional valence facilitates predictive and semantic unification processes.

However, as noted in the Introduction, many studies have linked gamma activity to sustained attention (Pulvermüller et al, 1997), feature binding (Basar-Eroglu et al., 1996), as well as memory processes (Osipova et al., 2006). Therefore, another possible interpretation for the larger gamma power during exposure to angry speech is that listening to emotional speech results in a greater employment of attentional perceptual mechanisms.

## Concluding remarks

Thus, our joint findings suggest that at both the sentence and word level, prediction and semantic integration is facilitated during angry speech perception. Prediction is one of the crucial mechanisms enabling speech comprehension in spoken interactions, and through continuous updating based on preceding context allows for the anticipation of upcoming information. Situationally, and from the evolutionary perspective, too, it is often highly relevant to correctly decode and comprehend emotional speech, because it carries important information that might have potential consequences for achieving or obstructing one's goals. Emotional valence provides additional contextual information that facilitates comprehension and communication in general. Thereby, it might allow the individual to by-pass deeper semantic processing and re-allocate processing resources towards motivational evaluation or decision-making processes.

## Future directions

The data reported here are part of a larger experiment that investigates the effect of negative emotion on the cortical tracking of speech. It explores the oscillatory dynamics of emotional speech processing (at the sentence level), focusing on the lower frequency bands (namely theta and delta, associated with syllabic and prosodic word processing), as well as on oscillations in higher frequencies (>30 Hz, gamma). Comparing the accuracy and strength of neural speech tracking at those various times scales, in speech and rhythmically similar noise segments, will allow us to tease apart the contribution of the vocal (acoustic, prosodic) and verbal (lexical, semantic) channels to the processing of emotion in speech. In future analyses, we plan to include the data about participants' affective state (measured by the PANAS questionnaire), as some previous studies indicate that mood might influence semantic processing and language comprehension (Chwilla et al., 2011).

## Acknowledgments

This work was supported by the Czech Science Foundation, grant no. 21-09797S, by the Czech Academy of Sciences, project no. LQ300252401, and by the European Regional Development Fund, project Brain Dynamics, reg.no. CZ.02.01.01/00/22_008/0004643.

## References

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., ... & Goswami, U. (2022). Delta- and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698.

Başar-Eroglu, C., Strüber, D., Schürmann, M., Stadler, M., & Başar, E. (1996). Gamma-band responses in the brain: a short review of psychophysiological correlates and functional significance. *International Journal of Psychophysiology*, 24(1-2), 101-112.

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *Journal of cognitive neuroscience*, *27*(11), 2095-2107.

Bates D, Mächler M, Bolker B, Walker S (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable channels in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech, Language, and Hearing Research*, 59(1), 72-89.

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

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

Boersma, Paul & Weenink, David (2024). Praat: doing phonetics by computer [Computer program]. Version 6.4.05, retrieved 27 January 2024 from http://www.praat.org/

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926-1929.

Carretié, L., Hinojosa, J. A., Albert, J., López-Martín, S., De La Gándara, B. S., Igoa, J. M., & Sotillo, M. (2008). Modulation of ongoing cognitive processes by emotionally intense words. *Psychophysiology*, *45*(2), 188-196.

Chen, X., Yuan, J., Guo, J., & You, Y. (2013). Neural oscillatory evidence of the difference between emotional and conceptual processing in language comprehension. *Neuroscience Letters*, *553*, 159-164.

Chwilla, D. J., Virgillito, D., & Vissers, C. T. W. (2011). The relationship of language and emotion: N400 support for an embodied view of language comprehension. *Journal of Cognitive Neuroscience*, *23*(9), 2400-2414.

De Pascalis, V., Arwari, B., D'Antuono, L., & Cacace, I. (2009). Impulsivity and semantic/emotional processing: An examination of the N400 wave. *Clinical Neurophysiology*, *120*(1), 85-92.

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

Ding, J., Wang, L., & Yang, Y. (2015). The dynamic influence of emotional words on sentence processing. *Cognitive, Affective, & Behavioral Neuroscience*, *15*, 55-68.

Ding, J., Wang, L., & Yang, Y. (2020). The influence of emotional words on predictive processing during sentence comprehension. Language, *Cognition and Neuroscience*, *35*(2), 151-162.

Fenster, C. A., Blake, L. K., & Goldstein, A. M. (1977). Accuracy of vocal emotional communications among children and adults and the power of negative emotions. *Journal of Communication Disorders*, *10*(4), 301-314.

Fields, E. C., & Kuperberg, G. R. (2012). It's all about you: An ERP study of emotion and self-relevance in discourse. *NeuroImage*, *62*(1), 562-574.

Fields, E. C., & Kuperberg, G. R. (2016). Dynamic effects of self-relevance and task on the neural processing of emotional words in context. *Frontiers in Psychology*, 6, 2003.

Grass, A., Bayer, M., & Schacht, A. (2016). Electrophysiological correlates of emotional content and volume level in spoken word processing. *Frontiers in Human Neuroscience*, *10*, 326.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90-105.

Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, 7, 83.

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, *45*(3), 487-498.

Hinojosa, J. A., Méndez-Bértolo, C., & Pozo, M. A. (2012). High arousal words influence subsequent processing of neutral information: Evidence from event-related potentials. International *Journal of Psychophysiology*, *86*(2), 143-151.

Holt, D. J., Lynn, S. K., & Kuperberg, G. R. (2009). Neurophysiological correlates of comprehending emotional meaning in context. *Journal of cognitive neuroscience*, *21*(11), 2245-2262.

Huang, Y. X., & Luo, Y. J. (2006). Temporal course of emotional negativity bias: an ERP study. *Neuroscience Letters*, 398(1-2), 91-96.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75(4), 887.

Johnstone, T., Van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, *1*(3), 242-249.

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain research*, *1148*, 138-148.

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, *49*(11), 3121-3129.

Ku, L. C., Allen, J. J., & Lai, V. T. (2022). Attention and regulation during emotional word comprehension in older adults: Evidence from event-related potentials and brain oscillations. *Brain and Language*, *227*, 105086.

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

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, 82(13), 1–26.

Lin, Y., Fan, X., Chen, Y., Zhang, H., Chen, F., Zhang, H., ... & Zhang, Y. (2022). Neurocognitive dynamics of prosodic salience over semantics during explicit and implicit processing of basic emotions in spoken words. *Brain Sciences*, 12(12), 1706.

ListenLab (2023). Make speech shaped noise from sound selection [Software]. Retrieved November 10, 2023 from https://raw.githubusercontent.com/ListenLab/Praat/master/Make_Speech_shaped_noise_from_sound_selection.txt

Lüdecke D (2018). "ggeffects: Tidy Data Frames of Marginal Effects from Regression Models." *Journal of Open Source Software*, *3*(26), 772.

Mai, G., Minett, J. W., & Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *Neuroimage*, *133*, 516-528.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience*, *26*(28), 7523-7531.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS One*, *6*(11), e27256.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417-435.

Pulvermüller, F., Birbaumer, N., Lutzenberger, W., & Mohr, B. (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Progress in Neurobiology*, 52(5), 427-445.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage*, *28*(4), 848-858.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural Psychology*, *32*(1), 76-92.

Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, *80*(1), 95-104.

Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24-31

The MathWorks Inc. (2022). MATLAB version: 9.13.0 (R2022b), Natick, Massachusetts: The MathWorks Inc. https://www.mathworks.com

Tune, S., & Obleser, J. (2022). A parsimonious look at neural oscillations in speech perception. In *Speech Perception* (pp. 81-111). Cham: Springer International Publishing.

Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, *14*(4), 387-406.

Wang, L., & Bastiaansen, M. (2014). Oscillatory brain dynamics associated with the automatic processing of emotion in words. *Brain and Language*, *137*, 120-129.

Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, *3*, 187.

Wang, X., Shangguan, C., & Lu, J. (2019). Time course of emotion effects during emotion-label and emotion-laden word processing. *Neuroscience Letters*, *699*, 1-7.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, *120*(1), 3.

Zhang, Q., Ding, J., Zhang, Z., Yang, X., & Yang, Y. (2021). The effect of congruent emotional context in emotional word processing during discourse comprehension. *Journal of Neurolinguistics*, *59*, 100989.