CHARLES UNIVERSITY

Faculty of Arts

Department of Psychology

Bachelor's Thesis

Natálie Kikoťová

The Effect of Negative Emotion on Neural Speech Tracking and Lexical Processing

Vliv negativních emocí na neuralní sledování řeči a zpracování lexikální informace

Supervisor: Kateřina Chládková, Ph.D.

2024

Acknowledgements

Firstly, I would love to express my gratitude to Dr. Kateřina Chládková for supervising my thesis, for her patience and her enthusiasm. The support she has provided goes beyond what I could have ever imagined, and I feel privileged to have been able to work alongside her in her lab for the past four years.

I am grateful to the Laboratory of Behavioral and Linguistic Studies, the Institute of Psychology of the Czech Academy of Sciences, and the Faculty of Arts of Charles University for generously providing access to their equipment and enabling me to conduct the experiments.

Additionally, I would like to thank Ester Salajová for her assistance in recording the angry and neutral speech segments. I would also like to thank all the participants who took part in the pilot and final EEG experiment.

Finally, I am very grateful to my friends and family for their loving support. A special thank you goes to my colleagues from the SPEAKIN Lab research group for their helpful insights, long discussions about neural speech tracking and unwavering moral support.

Declaration

I hereby declare that I have prepared my thesis independently. All sources, references, and literature used or excerpted during the elaboration of this work are properly cited and listed in completed reference to the due source. Thesis was not used at another university or to obtain another or the same degree.

Natálie Kikoťová

Prague, 30.4. 2024

Abstract

The thesis investigates the differences in the neural processing of emotional and non-emotional speech. The theoretical part summarizes previous research findings on emotional language processing and the underpinnings of speech processing in the domain of neural oscillations. The empirical part reports the results of an EEG study that was conducted to explore the differences in neural speech tracking during exposure to 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. Neural speech tracking, which was quantified as oscillatory power and the inter-trial phase coherence, and the N400 component of event-related potentials (ERP) to sentence-final words were analyzed. The results revealed larger *delta*, *theta*, and *gamma* power during exposure to angry speech in comparison to neutral speech. Negative emotional valence also significantly reduced the amplitude of the N400 elicited by sentence-final words. The results demonstrate enhanced neural processing and facilitated prediction in angry as compared to neutral speech. The present study represents one of the first investigations of the oscillatory dynamics during continuous emotional speech processing.

Key words: emotional speech; neural speech tracking; anger; EEG

Abstrakt

Tato práce se zabývá rozdíly v neurálním zpracování emočně zabarvené a neutrální řeči. Teoretická část shrnuje předchozí výzkumné poznatky ohledně zpracování emočně zabarvené řeči a představuje základní mechanismy neurálního zpracování řeči z pohledu neurálních oscilací. Empirická část popisuje výsledky EEG studie, která zkoumala rozdíly v neurálním sledování řeči během poslechu naštvané a neutrální řeči. Experimentu se zúčastnilo dvacet šest participantů, kteří byli vystaveni nahrávkám naštvaných a neutrálních úryvků konverzací a nahrávkám rytmického šumu. Neurální sledování řeči bylo kvantifikováno jako síla neurálních oscilací a koherence oscilační fáze. Dále byly analyzovány evokované potenciály (konkrétně komponenta N400) na poslední slovo každého úryvku konverzace. Výsledky ukazují vyšší sílu *delta*, *theta* a *gamma* oscilací během poslechu naštvané řeči ve srovnání s neutrální řečí. Negativní emoční valence vedla také k významnému snížení amplitudy na komponentě N400. Výsledky naznačují zvýšení neurální aktivity a facilitaci prediktivních procesů při poslechu emočně zabarvené řeči. Tento experiment představuje jednu z prvních studií oscilační dynamiky během kontinuálního zpracování emočně zabarvené řeči.

Klíčová slova: emočně zabarvená řeč; neurální sledování řeči; hněv; EEG

Table of contents

List of abbreviations

ADS	Adult-directed speech
CaC	Cerebral-acoustic coherence
EEG	Electroencephalography
EPN	Early posterior negativity
ERP	Event-related potentials
ERSP	Event-related spectral perturbations
IDS	Infant-directed speech
ITPC	Inter-trial phase coherence
LABELS	Laboratory of Behavioral and Linguistic Studies
LPC	Late positivity component
MEG	Magnetoencephalography
MI	Mutual information
PANAS	Positive and Negative Affect Schedule
TRF	Temporal response function

Introduction

Emotional speech holds significant importance in our daily social interactions. Rapidly decoding and effectively comprehending emotional speech is essential for identifying potential threats or life-saving opportunities, fostering interpersonal relationships, and achieving social goals. Hence, considering an evolutionary standpoint, it is crucial that processing resources are directed towards the processing of emotionally significant speech and that mechanisms enhancing speech comprehension are employed.

The question of how humans process emotional words has been thoroughly studied in the past decades and has yielded a wide range of findings on how the human brain detects emotional significance in verbal material. However, there is a limited number of studies investigating the processing of *continuous* emotional speech, that integrates both emotional lexical-semantics and emotional prosody. With respect to electrophysiology, emotional speech is underresearched in the brain-rhythm literature; the most common method to study emotional language processing are the event-related potentials.

Therefore, the aim of this thesis is to contribute to a comprehensive understanding of emotional speech processing and the mechanisms distinguishing it from the processing of non-emotional speech. Our aim is to investigate the oscillatory dynamics of emotional speech processing following the procedures from the neural speech tracking literature. Using this methodology may aid answering the question of how perceptual processes are modulated to decode emotionally salient signals rapidly and effectively.

The first section of the theoretical part will review the latest electrophysiological findings on emotional language processing, focusing predominantly on lexical processing and the attentional and predictive mechanisms supporting visual and spoken emotional word perception. The second section will summarize the key principles of neural speech processing and introduce the approaches and methods which are applied in the empirical part. The empirical part of the thesis describes the methodology and the results of an EEG experiment on the neural tracking of emotional speech and discusses the results in relation to the findings presented in the theoretical part.

I. Theoretical part

1 Emotions and language

1.1 Conveying emotions in speech

1.1.1 Two main channels

Humans have the ability to express emotions through language. While conveying emotions in written discourse is restricted to the use of semantic information, the process of translating one's emotional state into spoken interactions can occur via both the verbal (semantic) and vocal (prosodic) channel. The **verbal channel** provides information about linguistic units in speech, its syntactic structure, and meaning (which is represented by the content words the speaker uses for their expression). The **vocal channel** provides the listener with non-lexical (sometimes termed paralinguistic) information represented by changes in prosody, which is modulated by the speaker's emotional state (Berkmoes & Vingerhoets, 2004).

From the perspective of acoustics, in order to successfully decode emotional meaning from continuous speech the listener must integrate a variety of parameters, such as the percept of pitch (cued by the fundamental frequency), rhythm (duration of syllables and pauses, i.e., speech rate), loudness, and voice quality (e.g., breathy voice, creaky voice). The decrease or increase in one acoustic parameter can be associated with conveying various emotional states. Therefore, one isolated acoustic cue cannot be considered a perfect predictor of a specific emotional state (Ilie & Thompson, 2006), and a flexible interpretation of a specific combination of parameters is in order. For example, the melodic contour of angry speech, compared to neutral speech, is generally more pronounced and contains greater pitch and loudness changes. With respect to speech rate, angry speech is generally characterized by faster, irregular rhythm. Conversely, the intonation of joyful speech is smooth, rounded, and slow-varying and its rhythm tends to unfold in a more regular manner (Sbattella et al., 2014).

1.1.2 Modality differences in emotion processing

Findings from studies on emotion processing consistently indicate that the processing of emotional speech is longer, deeper, and more elaborate than the processing of neutral speech (Fields & Kuperberg, 2012; Kissler et al., 2008; Sander et al., 2005). That the emotional valence of a stimulus modulates its neural processing has been thoroughly demonstrated with emotional pictures (Lane et al., 1999) and facial expressions (Batty & Taylor, 2003), and unsurprisingly, similar processing differences have been observed with emotional verbal

material as well. However, as many authors argue, the magnitude of difference when contrasting emotional and non-emotional stimuli is often much larger for pictures and facial expressions as opposed to verbal material. This may be because the detection of emotional significance is more automatic for pictorial stimuli than for verbal stimuli, as the latter requires the additional pre-requirement of extracting the semantic representation in order for the emotional valence to be encoded. Despite that and as evidenced by a vast number of studies, the emotional content of words, too, modulates cognitive processes and their underlying neural mechanisms.

Notably, Grandjean (2021) argues that an important advantage of auditory processing in general is that it represents space as a 360° sphere which allows humans (and animals) to process information that is not directly available through other senses. For example, we cannot see if someone is standing behind us, but we can hear them talking or making noises which is crucial for detection of threats and survival. From an evolutionary standpoint, this poses a functional difference between auditory and visual processing of emotional stimuli.

In social settings, we encounter emotional language in the form of a continuous auditory stream that unfolds in real time. Our ability to decode the speaker's emotional state from speech is dependent upon a swift analysis of the speech signal, which, in comparison to written emotional material, integrates emotional information not only from the verbal but also the vocal channel. Therefore, it is essential to understand the mechanisms that support such parallel processing of emotional semantic content and prosody and enable humans to detect emotional relevance in spoken interactions.

1.2 The neural time course of emotional language processing

To date, the most common method used for investigating emotional language processing have been the event-related potentials (ERP). ERP are electrical responses evoked by an external stimulus and are typically recorded from electrodes placed on the human scalp. The different peaks of the response (ERP components) are characterized by their polarity (negative or positive going) and latency.

A significant advantage of ERP is their ability to capture the precise time course of processing. In the context of emotional word processing, ERP aid the investigation of how the emotional valence of a word impacts specific cognitive processes.

1.2.1 Lexical processing

To determine the emotional significance of a word, first, its meaning has to be extracted through accessing lexical representations in the brain and relating them to the presented verbal material. With regard to the lexical processing of individual emotional words, Kissler & Herbert (2013) demonstrated that in a lexical decision, task lexical access was faster for emotional than for non-emotional words, manifesting as an earlier word-pseudoword differentiation for negative relative to neutral words (252 and 324 ms after word onset, respectively). It is, however, unclear whether this effect can be attributed specifically to the emotionality of words, or whether it is related to the differential lexical processing of various semantic classes.

Similarly, Siakaluk et al. (2016) found facilitatory effects of emotionality on reaction times in three lexical decision tasks using three sets of emotional verbal stimuli – concrete nouns, abstract nouns, and verbs. The authors concluded that the emotional dimension of a word (and the ease with which a word elicits emotions) is an integral part of its lexical-semantics and leads to facilitated lexical decision.

To some extent, the time course of lexical access in individual visual word processing is reflected in the latency of the early posterior negativity (EPN) component, peaking typically around 250 ms after word onset (with a parieto-occipital distribution). The EPN indexes implicit processing of the emotional valence of a word regardless of task demands and is associated with increased attentional capture (Citron, 2012), which is reflected by a greater amplitude of the EPN to emotional (negative, positive) words than to neutral words (Kissler & Herbert, 2013).

Formerly, the N400 component was also considered to partly reflect the extraction of a word's semantic content. However, recently, it is increasingly clearer that the N400 does not reflect the semantic analysis per se, but rather the semantic agreement of the presented word with its preceding context. Such semantic prediction and integration represent an integral component of speech comprehension, as speech processing is vastly based on the continuous formation of semantic predictions and their subsequent integration with the preceding sentence context (as described in section 1.3.1.).

Another ERP component that is robustly affected by the emotional dimension of a word is the late positivity component (LPC), which has been observed in response to both visually and

auditorily presented emotional verbal material. The LPC is a slow, positive wave occurring around 600 ms after the word onset and is considered to reflect sustained processing and evaluation of emotional stimuli based on task demands. Some authors report differences in LPC amplitudes based on valence, with negative words eliciting greater amplitudes than positive and neutral words (Fields & Kuperberg, 2012); however other studies have reported advantages for both negative and positive in contrast with neutral words (Kanske & Kotz, 2007).

While earlier components are typically related to the processing of sensory information (whether it is visual or auditory), later ERP components generally reflect higher level cognitive processes, such as sustained analyses that integrate prior knowledge and contextual information (Brandeis & Lehmann, 1986).

1.2.2 Emotional prosody processing

The differential processing of emotional and non-emotional prosody has been the subject of experimental studies for many decades. For a long time, a traditional perspective has been that the right hemisphere regulates the processing of emotional prosody and plays an important role in the detection of emotion from the voice. This view was mostly based on results of clinical studies with patients who had lesions in right frontal and temporal areas of the brain (Kotz et al., 2011). Later experimental studies (in which emotional words were presented either in the right or left hemifield), however, found little evidence supporting a strict dominance of the right hemisphere in emotional prosody processing, rather suggesting a bilateral activation during the perception of emotional prosody. For example, using functional magnetic resonance imaging (fMRI), Grandjean et al. (2005) demonstrated enhanced bilateral activation of the mid superior temporal sulcus (STS) in response to angry relative to neutral prosody.

According to Schirmer and Kotz (2006), emotional prosody processing can be divided into three stages (Figure 1A). Within the first 150 ms, incoming acoustic information is processed in a bottom-up manner which employs subcortical pathways from the ear to the brainstem, thalamus, and the core and belt areas of the auditory cortex in the temporal lobe (Figure 1B). Such encoding of the acoustic features of a stimulus is reflected in the N100 component of the ERP, which reflects the differences in the physical properties of emotional and non-emotional prosody. Subsequently, emotional acoustic cues are integrated in order to derive an emotional meaning and significance. This process is organized along the auditory pathway from the

superior temporal gyrus to the superior temporal sulcus. This implicit attentional orienting towards an emotionally significant stimulus is reflected in the increased amplitude of the P300 to emotional (angry, sad, fearful) than to non-emotional (neutral) intonation (Kotz & Paulmann, 2007). In the stage that follows, the information about emotional significance is made available for more complex cognitive processes that recruit frontal cortical areas, as such that it can modulate semantic processing or deeper emotional evaluative judgements.

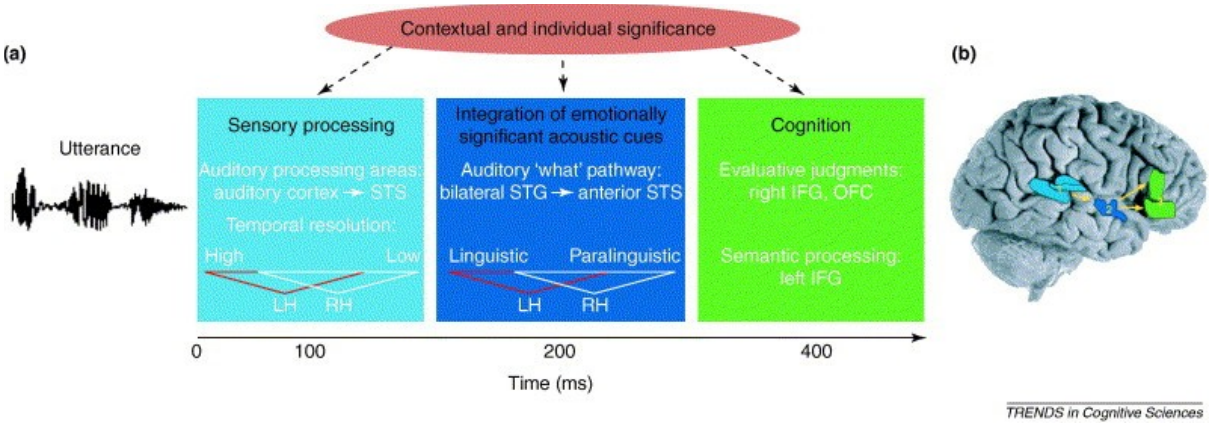

Figure 1: (A) Three-stage model of emotional prosody processing. During the first stage (up to ~150 ms), acoustic information is processed in a bottom-up manner. Around 200 ms, emotionally significant cues are integrated. At a later processing stage (around 400 ms), the extracted emotional information is available for deeper cognitive processing. (B) Areas employed in emotional prosody processing – right sagittal view (light blue: auditory cortex), dark blue: anterior part of the superior temporal sulcus, green: frontal and orbitofrontal gyrus, yellow arrows: processing directions). From *Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing* by A. Schirmer & S.A. Kotz, 2006. Copyright 2006 by Trends in Cognitive Sciences.

1.3 Cognitive mechanisms and biases associated with emotion processing

1.3.1 Predictive processing in emotional contexts

In electroencephalography, semantic prediction and semantic integration are relatively robustly indexed by the N400 component of the event-related potentials (Berkum et al., 1999; Kutas & Hillyard, 1980). The N400 is a negative peaking waveform around the latency of 400 ms after the onset of a stimulus. With respect to neural language processing in general, a less predictable word elicits a stronger N400 response than a word that is more predictable based on preceding (sentence) context. For instance, in a sentence *"I take my coffee with cream and sand*" the final word "*sand*" would elicit a strong N400 effect because it is semantically

incongruent with the preceding context and represents a violation of the prediction that was formed based on semantic context. The N400 wave is elicited by every word within a sentence and reflects the amount of processing resources directed towards the semantic integration and semantic analysis of that word even in cases of no semantic violation (Berkum et al., 1999).

With respect to emotional stimuli, the literature indeed shows that the emotional dimension of words extends the semantic context of a sentence. It can serve as an additional contextual resource influencing the prediction of upcoming words and supporting their integration into the preceding sentence context. Several studies investigating the effect of high-arousing words on the processing of subsequent neutral targets suggest that the presentation of arousing words may lead to perceptual potentiation, as their presence may affect the attentional allocation for subsequent lexical processing of the neutral target words (Hinojosa et al., 2012; Ding et al., 2015). Ding et al. (2020) showed that the emotional arousal of presented action verbs influenced the anticipatory processing of their subsequently presented agents which manifested as a larger sustained negativity for the neutral verbs (in contrast to emotional verbs) in a highly predictable context.

In emotionally congruent sentences (i.e., where the valence of the emotional target word matches the preceding context), semantic prediction and integration of the target might be facilitated due to the additional contextual information provided by the emotional valence of the sentence. In simpler terms, it might be easier to predict an emotional word in an emotionally congruent sentence context than a neutral word in a neutral sentence context. As the authors argued, this could explain the results observed by Zhang et al. (2021) who demonstrated the effect of congruent emotional context on sentence comprehension during silent reading. They compared two-sentence discourses where either both sentences or neither sentence conveyed a negative emotion. The target word, which was presented in the second sentence, was emotionally congruent with the preceding sentential context. In congruent negative discourse, the target word elicited a smaller amplitude of the N400 than did a neutral target word in neutral discourse.

However, the results of previous studies investigating the N400 to emotional words are difficult to interrelate, most likely due to differences in experimental paradigms and task demands. For example, some studies report an increased amplitude of the N400 to negative words presented in a neutral context (De Pascalis et al., 2009; Grass et al., 2016; Herbert et

al., 2008; Holt et al., 2009). These results might be interpreted by a surprisal effect caused by the unexpected emotional valence of a presented word after a preceding neutral context that had been established either by a preceding sentence or by the experimental task itself. The attentional capture by an emotional word might then result in a deeper semantic evaluation or a more demanding semantic access, and thus also a larger N400. Other studies report a decrease in the amplitude of the N400 component to emotional words that were presented in an emotionally congruent context. Kanske et al. (2011) tested the predictive value of an emotional cue on semantic integration using an attentional cueing paradigm. They observed that cues that correctly predicted the emotional category of an upcoming word facilitated semantic integration which manifested as a decrease in the N400 amplitude to the target emotional words when they were preceded by a valid emotional cue. Interestingly, some studies investigating the processing of emotional words in lexical decision tasks report a reduced N400 to emotional words compared to neutral words (Kanske & Kotz, 2007; Wang et al., 2019) which has been interpreted as a reflection of facilitated lexical access for emotional compared to neutral words (Wang et al., 2019).

Chen et al. (2013) compared the neural oscillatory processing of emotional and conceptual congruency during sentence comprehension using event-related spectral perturbations (ERSP). They presented participants with written sentences where the target word was either emotionally incongruent, conceptually incongruent, or congruent with the preceding context. They found an increase in *gamma* activity (relative to baseline) for the emotionally incongruent condition compared to the conceptually incongruent and congruent condition. This finding is particularly interesting, as studies investigating *gamma* activity in relation to semantic prediction in language comprehension report increased *gamma* activity to semantically congruent sentences in contrast to sentences containing a semantic violation (see section 2.2.). A possible explanation is that different neural mechanisms underlie the processing of emotional and conceptual incongruities; however, further investigation is needed in order to disentangle the effects of emotional valence and semantic prediction on *gamma* activity.

To summarize, contextual predictability influences the N400 (and *gamma* oscillatory activity) to an emotionally congruent or incongruent target word as evidenced by the results of the studies reviewed above.

1.3.2 Attentional mechanisms

The emotional dimension of a word automatically attracts attention, as such it engages attention irrespective of whether the stimulus is explicitly attended. One of the paradigms used to investigate the attentional processing of emotional words is the Emotional Stroop Task, where emotional (typically taboo) and neutral words are presented in various colors and participants are instructed to name the respective color as quickly as possible. McKay et al. (2004) demonstrated that color-naming times are longer for taboo words relative to neutral words, arguing that reading a taboo word leads to reallocation of attentional resources from the word's color to the presented word itself, interrupting the primary task and consequently slowing down color naming. In a different adaptation of the Stroop task, Bertels & Kolinsky (2016) replicated this effect in an auditory paradigm, highlighting that this phenomenon is not restricted to the visual modality.

Bertels et al. (2010) also found similar attentional biases towards emotional words using an auditory adaptation of the dot probe task. In their experiment, two words were simultaneously presented from a left and right loudspeaker. The words were followed by a to-be detected beep. Negative words induced attentional biases towards their location which manifested as reduced reaction times to the detection of the subsequent beep delivered to the same loudspeaker, suggesting that they served as a more efficient cue.

These studies indicate that the attentional modulations caused by emotional valence pertain to lexical processing in both the visual and auditory modality across experimental paradigms.

Such attentional effects of the emotionality of a stimulus have been replicated in electrophysiological studies as well (Kissler at al., 2009; Wang & Bastiaansen, 2014). Their results generally imply that the observed differences in the neural processing of emotional and non-emotional stimuli are independent of explicit attention. For instance, Kissler et al. (2009) investigated the processing of emotional words during silent reading. Participants were instructed to count either the number of adjectives or verbs. The authors found an enhanced amplitude of the EPN to emotional words irrespective of whether the word came from the target category (adjectives or verbs), indicating that the emotional content of a word was processed automatically. With regard to the processing of emotional prosody, an fMRI study by Sander et al. (2005) showed increased bilateral activation of the middle STS both in trials during which angry prosody was attended, as well as in trials in which angry prosody was ignored.

1.3.3 Negativity bias

Humans process emotions in an asymmetrical manner, displaying an advantage for negative emotional material early in socio-emotional development. This higher sensitivity to potentially threatening stimuli can be manifested through increased attention to negative material, better recall of negative material, and greater weight of negative material during evaluative judgments (Unkelbach et al., 2020).

Specific to the processing of verbal stimuli, several electrophysiological studies provide evidence in support of the negativity bias. ERP data from Fields & Kuperberg (2012) revealed an increased amplitude of the LPC to unpleasant as compared to pleasant words. That could be attributed to their greater motivational significance, as late positivity typically reflects higher cognitive processing and evaluative judgements. Field & Kuperberg's study aptly illustrates the negativity bias in verbal emotional processing and replicates previous studies demonstrating the negativity bias with non-verbal visual stimuli (Huang & Luo, 2006; Ito et al., 1998; Scott et al., 2009).

The processing advantage may be attributed to greater informational value of negative stimuli, as typically, the presence of negative stimuli indicates potential threat and danger and can have serious consequences for one's life and wellbeing. This creates demand for increased attentional and cognitive resources directed towards a negative emotional stimulus (Vaish et al., 2008).

2 Neural mechanisms of speech processing

2.1 Temporal and spectral modulations in speech

The speech signal is rhythmic on multiple temporal scales. The most prominent rhythm in speech is the interchanging of syllables and of prosodic words. In normal speech tempo, there is typically 5 to 6 syllables per second and approximately 1 to 2 words per second which corresponds to neural oscillatory activity in the *theta* (4–8 Hz) and the *delta* (0.5–4 Hz) band, respectively. The speech signal also contains smaller chunks than syllables, that is, speech segments or coarticulated parts of segments, which can be reflected in neural oscillatory activity in the higher frequencies (> 30 Hz) (Tune & Obleser, 2022). These temporal modulations contained in the acoustic envelope of speech, rather than the fine-grained structure of the spectrum, are critical for the decoding of the speech signal and successful speech comprehension (Tune & Obleser, 2022).

2.2 Neural oscillations in speech processing

To date, the most widely used method for the study of language processing have been the event-related potentials. The main limitation of ERP is, however, that they are unable to capture the dynamics of continuous speech processing as they can only measure the neural response locked to one specific time point in the speech signal. A bit more recently, the neural processing of speech has been studied through analyzing patterns of rhythmic neural activity in the brain – neural oscillations.

Neural oscillations reflect the synchronized spiking of neuronal populations in cortical and subcortical areas of the brain. They are crucial for communication across brain networks and support information transfer in different brain regions through the coordinated alternation of excitatory and inhibitory phases of firing neuronal populations (Buzsaki & Draguhn, 2004). Neural oscillations integrate information about the firing patterns of neurons at different frequency scales. This hierarchical structure of neural oscillations (which is similar to the structure of speech) might allow for the parallel processing of time-varying features at different frequencies contained in the acoustic envelope, i.e., sensory demultiplexing (Hyafil et al., 2015). Through the process of sensory demultiplexing, the brain can analyze the different rhythmic patterns of the speech signal and combine them hierarchically in order to create an integrated speech percept.

Cumulative evidence from experimental studies on speech processing suggests that slow oscillatory activity in the *delta* and *theta* band (0.5–4 Hz, and 4–8 Hz, respectively) is predominantly responsible for syllable and word segmentation as its frequency corresponds to the rate of syllables (~ 6 syllables per second in Czech; Weingartová & Volín, 2014) and prosodic words (~2 prosodic words per second) in continuous speech (Tune & Obleser, 2022).

However, it is important to note that besides speech-specific processes, activity in the *delta* and *theta* band has been researched in relation to more domain-general functions. *Theta* oscillations are generally associated with lexical memory retrieval, while *delta* oscillations have been related to timing and binding processes (Herweg et al., 2020; Tune & Obleser, 2022).

The precise role of *gamma* band oscillatory activity (> 30 Hz) remains debated in research on speech processing. From the perspective of neural speech tracking, some authors argue that *gamma* synchronization to the subsyllabic patterns in speech reflects higher linguistic processes – for instance, phonemic categorical perception (Giraud & Poeppel, 2012). Others have linked such *gamma* synchronization to acoustic-perceptual processes and suggested that *gamma* oscillations collaborate with low-frequency activity via phase-amplitude coupling to integrate acoustic representations in speech (Attaheri et al., 2022a; for more on phase-amplitude coupling, see section 2.3.).

Moreover, *gamma* band oscillatory activity has been linked to semantic processing at the sentence level (Tune & Obleser, 2022). It is considered to be involved in semantic unification processes, as such it reflects the predictability of incoming words based on preceding sentence context (Bastiaansen & Hagoort, 2015; Mai et al., 2016). Hald et al. (2006) observed an increase in *gamma* power during the processing of sentence-final words that were semantically congruent with the preceding sentence compared to words that produced a semantic violation. According to their results, a more coherent and predictable context might be, to some extent, reflected by greater *gamma* power. Wang et al. (2012) argue that such observed *gamma* power increases could be related to the agreement between the pre-activation of the neural representations of the predicted word and the incoming word. They propose that the *gamma* power increase is more likely to be related to the checking of the incoming linguistic representation against top-down contextual predictions, rather than reflecting a direct prediction of a word.

2.3 Neural speech tracking

Neural oscillations support the segmentation and identification of discrete linguistic units (phonemes, prosodic words, and syllables) by tracking the acoustic envelope of the incoming speech signal and by temporarily aligning their phase to its acoustic patterns (Figure 2). The oscillatory phase angle represents the relative position of peaks and throughs in the signal and can be reset in response to an external stimulus, for example speech onset (Obleser & Kayser, 2019). Phase-locking to an external input represents the process through which neurons in the auditory cortex time their high-excitatory and low-excitatory activity to match the information load in the incoming acoustic structure so that the arriving information is processed during the episodes of excitatory activity with maximal gain (Giraud & Poeppel, 2012; Peelle & Davis, 2012). This ability to track the speech envelope has been proposed as a mechanism supporting the parsing of speech into linguistic units, thus facilitating speech comprehension.

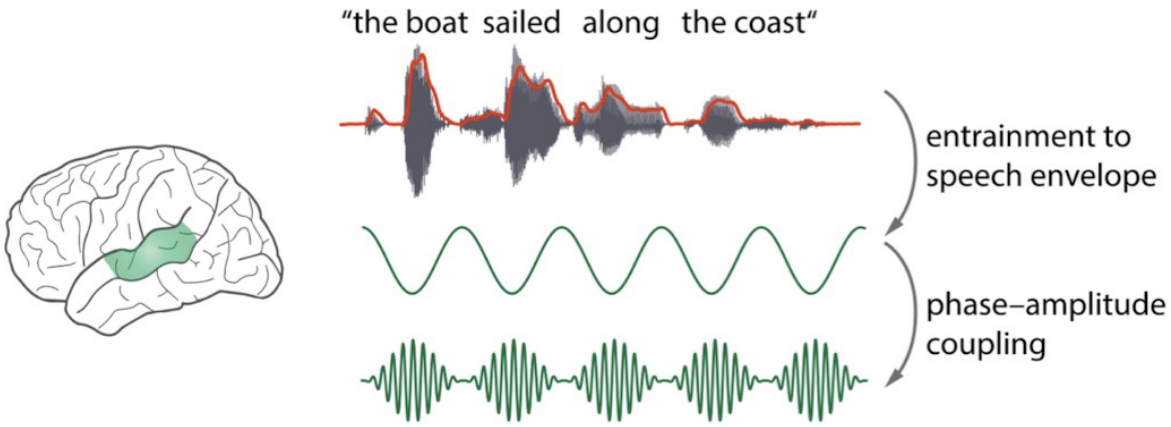

Figure 2: Neural tracking of speech. From *A parsimonious look at neural oscillations in speech perception* by S. Tune & J. Obleser, 2022. Copyright 2022 by Springer International Publishing.

In a seminal MEG study Luo & Poeppel (2007) demonstrated that the phase of oscillations at 4–8 Hz (i.e., *theta)*, tracks the acoustic structure of sentences and showed that sentences could be distinguished from one another based on information about the oscillatory phase only. Furthermore, tracking performance was correlated with speech intelligibility as less intelligible speech signal (due to higher levels of degradation) lead to worse *theta* phase tracking.

Neural tracking occurs most prominently at the rate of syllables and prosodic words (*theta* and *delta* band, as described above). Neural tracking of speech in the *theta* frequency band corresponds to the alignment of neural activity to the sustained rhythmic patterns of speech occurring at the temporal modulation rate, while neural tracking in the *delta* band reflects the processing of acoustic onsets in the speech envelope (Chalas et al., 2023). That is, increases in *delta* power tracking are related to acoustic transitions from periods of silence to speech onsets rather than sustained rhythmic activity at lower frequencies (Chalas et al., 2023).

Both *delta* and *theta* tracking are involved in the phonological analysis of speech but whether they play a role in higher-level speech processing remains unclear. Mai et al. (2016) found higher *theta* and *delta* power when participants listened to speech as opposed to non-speech sounds supporting the notion that *theta* and *delta* power reflect syllabic and supra-syllabic patterns in speech. However, the authors found no differences between normal speech containing real words and pseudo-speech. Their results indicate that *theta* and *delta* support phonological rather than semantic processes. Number of studies, however, suggests that *delta* band tracking reflets not only the acoustic analysis of the speech signal but also higher-order linguistic processes, such as the processing of syntactic structure and semantics (Coopmans et al., 2022; Molinare & Lizarazu, 2018). Etard & Reichenbach (2019) report increased *delta* band tracking when participants were exposed to their native language as opposed to a foreign language with comparable acoustics, highlighting the role of *delta* tracking in speech comprehension.

Coupling between oscillations in the lower and higher frequency range represents another mechanism underlying the parsing of continuous speech into discrete linguistic units and the integration of acoustic representations in speech (Attaheri et al., 2022a). The phase of low-frequency oscillations (mainly *theta*) modulates the amplitude of higher oscillations (*gamma*) which may represent one of the fundamental mechanisms underlying the integration of rhythmic patterns occurring at different temporal scales. Authors have hypothesized that the phase-amplitude coupling between *theta* and *gamma* oscillations subserves the syllabification of phonemes based on the temporal dynamics of the auditory signal (Hovsepyan et al., 2020; Morillon et al., 2012). In line with this hypothesis, Lizarazu et al. (2019) discovered that *theta-gamma* coupling follows alternations in speech rate (accelerated or decelerated), suggesting that the coupling mechanism does, in fact, reflect the tracking of the input's acoustics.

Furthermore, it has been suggested that *theta-gamma* coupling is related to phonological working memory, and lexical semantic processes, specifically semantic integration (Mai et al., 2016). Besides language specific functions, *theta-gamma* coupling has been associated with domain-general processes, such as memory and learning (for a review see Colgin, 2015).

2.3.1 Selected contributions of prosody

Prosody comprises of suprasegmental information about the rhythm, intonation, and stress in speech and supports the detection of linguistic boundaries, such as phonemes, syllables, and phrases. Not only does prosody aid the segmentation of the speech stream, but it also conveys important cues about the speaker's intent, such as irony, sarcasm, and affective states, which extend the message by adding further contextual information and create another layer of meaning (Myers et al., 2019). The constituent parts of prosody vastly contribute to speech comprehension, and unsurprisingly, to the neural tracking of speech.

Bachmann et al. (2021) recorded cortical and subcortical EEG activity to continuous speech and found that adding relative pitch as a predictor lead to significant improvement of the decoding model in case of the cortical (but not subcortical) responses, most likely due to clear dissociation between pitch tracking and envelope fluctuations in cortical responses. These findings are consistent with those of Teoh et al. (2019) showing that relative pitch tracking is band-limited to *delta* phase and that this measure of pitch processing is, indeed, separable from other acoustic features conveyed by the amplitude envelope. That is in line with previous studies reporting coupling between the fundamental frequency contour and cortical responses at around 0.5 Hz, corresponding to the *delta* band (Bourguignon et al., 2013).

A specific demonstration of the role of prosody in neural speech processing and language development is infant-directed speech (IDS). IDS is distinguished by exaggerated prosodic patterns, such as prolonged vowel length, slower tempo, and easier syntactic structure, relative to adult-directed speech (ADS) (Fernald & Simon, 1984). Findings from behavioral studies indicate that infants have a significant preference for IDS, such that they fixated on a visual stimulus longer when it produced IDS in contrast with ADS (Cooper & Aslin, 1990; ManyBabies Consortium, 2020) and chose social partners that used IDS rather than those using ADS (Schachner & Hannon, 2011).

Compelling evidence from neural speech tracking studies with infants suggests that IDS facilitates speech decoding by enhancing low-frequency cortical tracking (Attaheri et al.,

2022b; Kalashnikova et al., 2018; Menn et al., 2022). Menn et al. (2022) investigated the advantages for the neural tracking of IDS both at the rate of syllables and prosodic stress, finding stronger speech tracking only for the latter. The authors argue that the advantage of IDS rests predominantly in the enhanced prosodic stress which may aid word segmentation by establishing a more regular rhythm or by holding infants' attention better than less pronounced prosodic stress in adult speech. This is in line with Attaheri et al. (2022b) who report higher *delta* relative to *theta* speech tracking when infants were exposed to sung nursery rhymes. Attaheri et al. (2022a) found that data from the same experiment run with adults broadly replicated the patterns found in infants, suggesting that prosodic stress might play a key role in facilitating speech comprehension across development.

Taken together, the studies mentioned above demonstrate the contribution of pitch and stress cues to neural speech tracking.

2.3.2 Attentional modulations

Besides being determined by bottom-up acoustic information in speech, such as the alternations of syllables and word stresses, neural tracking is strongly modulated by attention. A convincing number of studies show that neural representations of speech can be enhanced or suppressed as a function of selective attention and that the coupling of neural oscillations and speech can be adjusted to converge with the listener's goals (Holtze et al., 2021; Obleser & Kayser, 2019; Vanthornhout et al., 2019). In experimental settings, this is best demonstrated by the "cocktail party" situation where the listener needs to isolate one talker among multiple competing speech streams. In an EEG study by Kerlin et al. (2010), participants listened to two simultaneously heard speech streams preceded by a visual cue that indicated whether they should attend to stream presented on the left or right side (left/right ear). Their results showed that speech representations in the 4–8 Hz range (corresponding to the *theta* band) were enhanced when speech was presented to the to-be attended ear. This was supported by hemispheric differences in *alpha* power which indicated the direction of attention (ipsilateral *alpha* enhancement and contralateral alpha suppression); furthermore, the strength of *alpha* lateralization was associated with the extent to which selective attention enhanced the cortical representations of the speech signal (Kerlin et al., 2010).

Similarly, Rimmele et al. (2015) demonstrated that neural speech tracking is modulated by selective attention when participants were simultaneously exposed to natural and vocoded speech and had to attend to only one stream while ignoring the other. However, attentional

enhancement effects were detected only in the case of natural speech. The neural tracking response for vocoded speech was similar to that of unattended natural speech, pointing to a significant contribution of higher-order linguistic processes to the more precise tracking of the attended speech stream.

Other studies show that attentional capture (Holtze et al., 2021) but also other higher-order processes, such as listening effort (Decruy et al., 2020) and predictive processing (Golumbic et al., 2013; Park et al., 2015) can be reflected in neural speech tracking.

2.3.3 Semantic contributions

As demonstrated by the studies reviewed in the section above, neural speech tracking is not exclusively driven by low-level acoustic features. Besides top-down (but non-linguistic) modulations, such as attention or listening effort, a listener's ability to extract linguistic information from speech represents another integral component of neural tracking.

The phase-locking of cortical oscillations to the rhythmic patterns in speech is stronger when linguistic information is available to the listener (Peelle et al., 2013). The authors presented participants with speech stimuli with varying levels of intelligibility while preserving the speech envelope to control the amount of non-sensory information contained in the signal. Their results (and similarly, the results of Rimmele et al., 2015, as mentioned in the section above) are consistent with the hypothesis that neural tracking is dependent on various informational sources that are not exclusively conveyed in the speech envelope.

To further investigate the lexical-semantic contributions to speech tracking, Lizarazu et al. (2021) investigated how language proficiency affects the cortical tracking of syllables and words in speech. They compared the neural speech tracking of second language (L2) speakers of Basque in the *delta* and *theta* band and found that cortical activity in these bands was related to L2 proficiency.

Taken together, results from the above-mentioned studies provide evidence that speech perception and processing, reflected in neural speech tracking, constitutes a constant interplay of low-level sensory and high-level cognitive processes.

2.3.4 Measures of neural speech tracking

Across studies the quantification of neural speech tracking varies as there is a number of measures that can be used to capture the similarity between the speech signal and the neural

response. One fundamental approach is to measure the **total oscillatory power** in frequency bands across the range that is defined with respect to the power spectrum of the acoustic envelope of the speech signal (generally *delta* and *theta* band). The second approach pertains to measuring the **oscillatory phase** which can be analyzed by computing the inter-trial phase coherence (ITPC, phase-locking value). ITPC indicates how consistently a certain phase angle occurs at a specific point in time across trials (Wöstmann et al., 2017) and is calculated from the frequency decomposed signal for a specific frequency range based on the properties of the speech signal (e.g., *delta* and *theta* band). Coherent phase across trials points to stronger speech tracking.

The neural response during exposure to speech is then compared to the neural response during exposure to a baseline condition. This allows for filtering out the neural response that is shared with the processing of acoustic features and rhythm in general and is not speech specific. The baseline condition is chosen based on study design; it is typically an unintelligible auditory signal which to some extent preserves several acoustic features contained in natural speech. For example, vocoded speech (speech distorted by noise, e.g., Rimmele et al., 2015), spectrally rotated speech (e.g., Lizarazu et al., 2020), white noise, or amplitude modulated (speech-shaped) noise (e.g., Molinaro & Lizarazu, 2018).

More computationally demanding methods quantify neural speech tracking as a direct comparison between the amplitude envelope of speech and the neural signal. The most straightforward approach is to directly cross-correlate the amplitude envelope of speech and the neural signal. Recently there has been an increase in studies that use forward modelling to predict the EEG response based on the amplitude envelope using the temporal response function (TRF). After modelling the neural signal, the predicted response is then correlated with the real neural response, indicating the extent to which input information is represented in the neural signal.

Several other methods, such as stimulus reconstruction, mutual information (MI) or cerebral-acoustic coherence (CaC), are used to quantify neural speech tracking. For further explanation of those see Wöstmann et al. (2017) or Harding et al. (2019).

II. Empirical part

3 Research aim

In this study, we aimed to investigate the differences in the neural tracking of emotional (angry) and non-emotional (neutral) speech. Our motivation was to use continuous conversational stimuli in order to explore the processing of emotional speech on the level of whole sentences as opposed to isolated words only.

Up to date, emotional speech (and specifically, emotional prosody) has been unexplored in the neural tracking literature (with the exception of recent research on the neural tracking of IDS as mentioned in section 2.3.1.). Many authors have thus inquired that novel experimental studies be carried out in order to comprehend how the human brain integrates information at various temporal scales to process emotional speech (Grandjean, 2020; Myers et al., 2019).

As the emotional speech condition, we chose anger. This is because (1) across emotion recognition studies, anger is the least misclassified emotion based on prosodic cues from the speaker's voice (Fenster et al., 1977; Scherer et al., 2001), and (2) several studies have shown that the emotion effects are more pronounced in negative compared to positive material (Fields & Kuperberg, 2012; Huang & Luo, 2006; Ito et al., 1998; Scott et al., 2009).

3.1 Hypotheses and predictions

H1: Based on prior ERP studies indicating a processing advantage for emotional stimuli, we hypothesized that the processing of angry speech would be enhanced in comparison to neutral speech and predicted observing greater neural speech tracking in the *theta* and *delta* band, indexed by increased oscillatory power in the *delta* and *theta* band and higher ITPC for the *delta* and *theta* band.

H2: Based on prior studies on the N400, we hypothesized that the additional contextual information conveyed by the emotional valence of speech would facilitate predictive processing of upcoming emotional words and predicted that in emotionally congruent sentential context, emotional (negative) words would elicit a smaller N400 than non-emotional (neutral) words.

H3: In line with the literature on neural speech tracking, which suggests an increase in *gamma* band oscillations for semantically predictable contexts than for nonpredictable contexts, we predicted that in emotionally congruent sentential context, *gamma* power would be stronger while listening to the emotional (angry) speech than while listening to non-emotional (neutral) speech.

4 Methods

4.1 Participants

We recruited twenty-six participants (18 females, 8 males, mean age = 22.12 years, SD = 2.49, age range = 19–27 years) from the participant pool of the Laboratory of behavioral and linguistic studies (LABELS). All subjects were right-handed native speakers of Czech, who reported normal hearing and did not have a history of any psychiatric disease nor neurological impairment. After data preprocessing, we excluded two participants from the ERP analysis due to a low number of artifact-free trials.

4.2 Stimuli

4.2.1 Speech material

For the speech material we created a set of 100 conversation segments (50 in the angry and 50 in the neutral condition). Each segment comprised of two sentences where a female speaker is talking to another person. The last word of the second sentence was emotionally congruent with the preceding sentence context, i.e., it was either negative (in the angry speech condition) or neutral (in the neutral speech condition). 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 full list of the conversational segments can be found in Appendix A. The sentences and 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 actress who was instructed to produce the sentences in a congruent emotional prosody. The average intensity of each of the conversation segments was equalized across segments using Praat (Boersma & Weenink, 1992–2024). The duration of the silent gap between the two sentences within a conversation segment was artificially edited to 400 ms in Praat (Boersma & Weenink, 1992–2024) in cases when the duration exceeded the mean duration by more than 1.5 standard deviation (n = 14).

4.2.2 Pilot rating task

The speech material was rated on several scales in two pilot experiments, administered online (using Psychtoolkit, Stoet, 2010, 2017) with native speakers of Czech (different participants than in the subsequent EEG experiment). In the first pilot, participants performed a judgement task with written representation of the conversation segments. Participants rated the valence,

arousal, and plausibility of the conversation segments and categorized them 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 final list (as outliers in valence or plausibility). The second pilot was a judgment task with the audio recordings in which a different group of participants judged the conversation segments on valence (on a three-point scale -1, 0, +1) and naturalness (on a 7-point scale between 1 and 7).

A linear mixed-effects model with varying slopes for each participant revealed a significant difference in valence between the angry and neutral recordings (B = 0.605, SE = 0.012, df = 32.02, t = 49.36, $p < .001$). No significant difference in naturalness between conditions was detected (B = -0.086, SE = 0.084, df = 31, t = -1.025, $p = .314$). Table 1 summarizes the linguistic characteristics of the speech stimuli. Figure 3 plots the mean valence and naturalness ratings from the second pilot with recordings.

Table 1: Linguistic characteristics of the speech material in the angry and neutral condition: segment duration (ms), duration of silent gaps (ms), syllable rate (syllables per second), valence (on a three-point scale -1, 0, +1), naturalness (on a 7-point scale between 1 and 7), mean pitch (Hz), pitch change (i.e., maximum pitch - minimum pitch, measured in Hz), minimum intensity (dB), maximum intensity (dB). The table shows the mean and standard deviation (in brackets) for each characteristic.

Characteristic	**Condition**	
	Angry speech	**Neutral speech**
Segment duration	4.357 (0.443)	3.974 (0.343)
Duration of silent gaps	0.414 (0.119)	0.437 (0.125)
Syllable rate	5.09 (0.49)	5.62 (0.39)
Valence	-0.987 (0.112)	0.223 (0.469)
Naturalness	5.12 (1.62)	4.95 (1.53)
Mean pitch	290.77 (22.36)	267.558 (13.137)
Pitch range	269.762 (46.626)	210.919 (53.258)
Maximum intensity	83.372 (0.708)	83.669 (0.545)
Minimum intensity	36.534 (4.534)	38.732 (5.232)

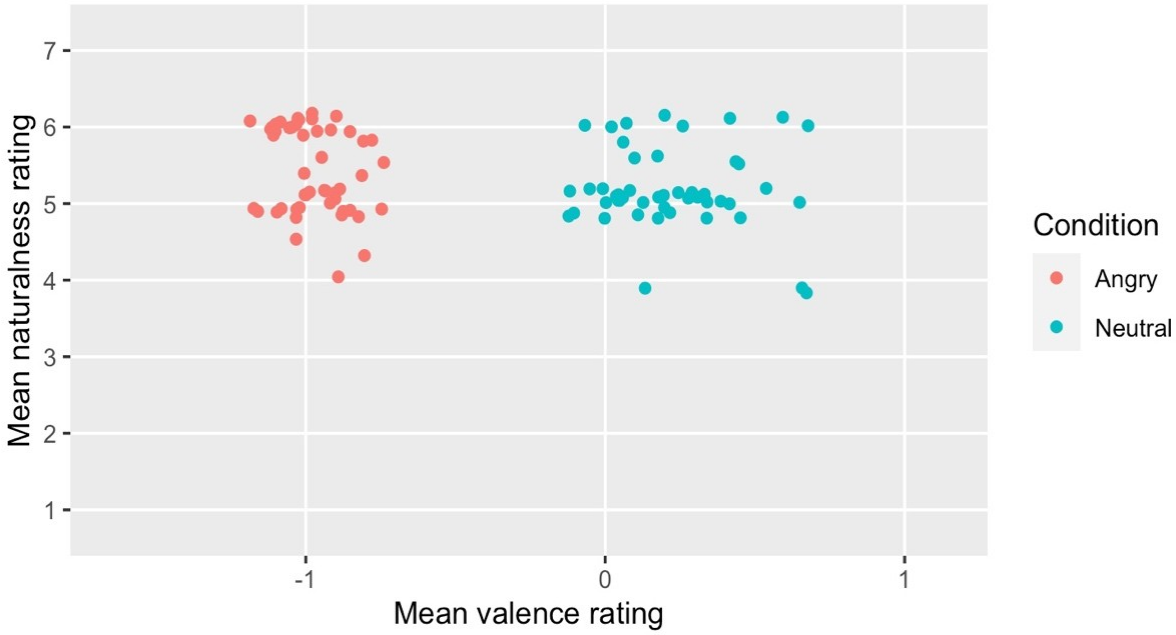

Figure 3: Mean valence and naturalness ratings per conversation segment and condition. Dots represent ratings pooled across 32 participants from the second pilot rating task with recordings.

4.2.3 Speech-shaped noise

The recordings of speech were transformed into speech-shaped noise using Praat and its native functions (Boersma & Weenink, 1992–2024, using the materials provided by ListenLab, 2023). First, we derived a long-term average spectrum for the two sets of speech segments from the neutral and the angry condition (i.e., the neutral conversation segment and the corresponding angry conversation segment) and filtered a white noise signal with that spectral object. Then we converted each speech segment (separately for the neutral and the angry condition) to an intensity and subsequently to an amplitude tier. Finally, we multiplied the filtered noise (common for the two corresponding segments) with the amplitude tier specific to the neutral or to the angry segment. This process resulted in speech-shaped noise for each segment in each condition, maintaining the envelope of the original neutral or angry speech stimulus while keeping identical spectral content between the two conditions.

Each speech block was always preceded by a block of speech-shaped noise. The speech-shaped noise stimuli established a baseline allowing for comparison of the semantically rich speech condition to rhythmically similar but non-speech like stimuli (similarly to Molinaro et al., 2018), enabling the assessment of the neural tracking of emotional prosody.

4.3 Supplementary methods

Before the experiment, participants filled out the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988). PANAS consists of 20 items which measure current positive (10 items) and negative (10 items) affective state of the participant. Each item is an adjective expressing a either a positive or negative feeling (e.g., curious, ashamed, tense) that participants rate according to how strongly it represents their current mood (on a 1–5 scale). The cumulative score, separately for the positive and negative affect, reflects the intensity of the subjectively perceived affective experience. The score ranges from 10 to 50 indicating low/high intensity of the affective state of the respective emotional valence.

After the completion of the experiment, participants filled out a form with questions about the perceived differences in the neutral and angry speech block, and the frequency of exposure to angry speech in their daily life. Specifically, we asked how often (on a scale 1–5) they use the angry tone of voice and curse words themselves, and how often they are on the receiving side of this type of speech. The final questionnaire can be found in Appendix B.

4.4 Procedure

The order of the conversation segments was randomized. Each conversation segment was repeated twice within the presentation block, resulting in a total of 100 trials per block. The inter-trial interval varied randomly between 390 and 410 ms.

Prior to the experiment, participants provided informed consent (see Appendix C) and completed the PANAS questionnaire. Participants were tested individually in a quiet room, seated comfortably about 1 meter away, frontally from a computer screen. Stimuli were presented at 65 dB SPL (measured at the location of the participant's head) from two loudspeakers placed in 30° angles next to the monitor.

We presented participants with the four blocks of stimuli, each block lasting approximately eight minutes. Participants were allowed to take short breaks between blocks for relaxation and refreshment. First, a speech-shaped noise block was presented, followed by the corresponding speech block (from the same condition). Each participant listened to blocks from both the neutral and the angry condition – the order of the two conditions was counterbalanced through random assignment across participants. Participants were instructed to listen passively during the speech-shaped noise blocks and to listen attentively during the speech blocks as in some of the trials they were to answer a comprehension question about

what the speaker said. This comprehension task occurred after ~1/10 of trials to ensure that participants were paying attention to the semantic content of the recordings.

4.5 EEG recording and preprocessing

The EEG data were acquired at a sampling rate of 200 Hz from 19 scalp electrodes placed according to the international 10/20 system. An additional FCz electrode served 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. Impedances were kept below 10 kΩ.

Preprocessing was performed using the EEGLAB toolbox (Delorme & Makeig, 2004) in Matlab (The Mathworks Inc., 2022). The data were bandpass filtered at 0.1 and 80 Hz and re-referenced to the nose. A notch filter at 50 Hz was applied to filter out the electrical line noise. Further preprocessing steps, including artifact rejection methods, for the ERP and power analysis are described below.

4.5.1 ERP data analysis

The filtered data were epoched from -0.1 s to 1 s relative to target word onset (i.e., the sentence- final word in each conversation segment). The epochs were baseline-corrected to the 100-ms pre-stimulus interval. Using an automatic artifact rejection approach, epochs in which the absolute amplitude exceeded 100 μV were marked and subsequently rejected as artifacts. Participants who had less than 30 % of remaining epochs after artifact rejection were excluded from further analyses (N = 2).

For each participant, an average ERP waveform was computed for the angry and the neutral condition separately. A grand-average negative peak per condition was determined between 200 ms and 500 ms after target word onset. In the per-participant average waveform, the N400 amplitude was measured by calculating the mean amplitude within a 100-ms window centered around the grand-average negative peak, separately for the Cz and Fz electrodes.

4.5.2 Time-frequency analysis

The EEG data were epoched into 10-s segments (leaving out cases in which the 10-s interval would have been interrupted by a comprehension question). This procedure resulted in a total of 74 epochs in each speech condition and 98 epochs in each speech-shaped noise condition, per participant. Using the automatic artifact rejection approach, epochs in which the absolute amplitude exceeded 210 μV were marked and subsequently rejected as artifacts. Participants

who had less than 40 % of remaining epochs after artifact rejection were excluded from further analyses (N = 0). The epoched data were decomposed using a Morlet wavelet transform across 200 sliding windows. 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 *delta* (defined here as 0.1–2 Hz), *theta* (defined here as 3–7.9 Hz), and *gamma* band (30–80 Hz) for the Cz electrode was computed across all epochs for each condition. As the speech-shaped noise blocks served as a baseline condition, the average total power in the noise epochs was subtracted from the average total power in the corresponding speech epochs.

Inter-trial phase coherence was used to quantify the phase synchrony across trials. Higher ITPC indicates stronger phase-locking with respect to external stimulation and has been shown to reliably reflect speech comprehension (Batterink & Paller, 2017). For each subject, ITPC for the *delta* band (corresponding to the rate of prosodic words, 0.1–2 Hz) and *theta* band (corresponding to the rate of syllables, 3–7.9 Hz) was computed at the Cz electrode for all epochs, in each condition. ITPC was calculated as circular average at time point and frequency and averaged across time points and respective frequency band, using the *newtimef* function of EEGLAB. Then, the average ITPC in the speech-shaped noise condition was subtracted from the average ITPC in the corresponding speech condition.

4.6 Statistical analysis

All statistical analyses were conducted in R (R Core Team, 2024) using linear mixed-effects models (packages *lme4*, Bates et al., 2015; *lmerTest*, Kuznetsova et al., 2017). A separate model was fitted for each of the frequency bands (*delta*, *theta*, *gamma* – after the subtraction of speech-shaped noise) and the predicted measure (total band power, ITPC). Another model was fitted for the N400 component of ERP.

For each model, the effect of condition (with a sum-to-zero contrast -negative vs. +neutral) was estimated, with random intercepts for channel and for participant. Marginal means were estimated using the package *ggeffects* (Lüdecke, 2018).

4.7 Ethical aspects

The experiment was approved by the Ethics Committee of the Institute of Psychology of the Czech Academy of Sciences. After receiving information about the experimental procedure,

participants signed an informed consent form agreeing to participation in the experiment and the use of the collected data for research purposes. All collected data were anonymized. Each participant was given a unique identifier under which they participated in the experiment. Participant names were recorded only in case of students who wanted to exchange their participation for ECTS credits and were deleted after the credits were collected.

Before the experiment, the administrator disclosed that negative words and insults were included in the to-be presented speech material. Participants were informed of their right to withdraw their participation in the study without providing explanation. After the experiment, participants were debriefed about the true purposes of the study.

5 Results

5.1 Positive and Negative Affect Schedule

Table 2 provides the mean score, standard deviation, median, minimum, and maximum for the positive and negative scales acquired from PANAS. Figure 4 plots the distribution of the scores for the positive and negative scale. There is low variability in the data; most participants scored low on the negative mood scale and around the average for the positive mood scale. Therefore, we discarded our original intention which was to control for mood variability across participants by including the positive and negative scale in the final models for ITPC and total power.

Table 2: Descriptive statistics for the positive and negative scales.

Scale	Mean (standard deviation)	Median	Min–max
Positive	27.20 (5.07)	27.5	18–36
Negative	15.85 (6.70)	14	10–41

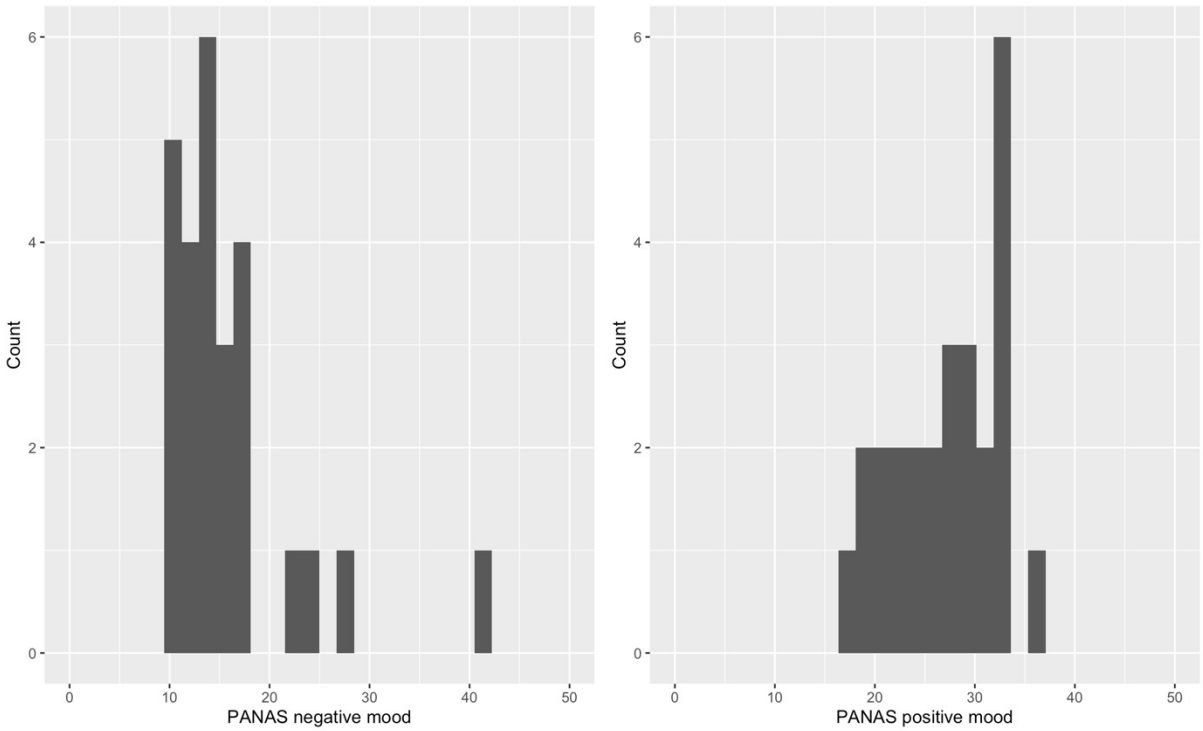

Figure 4: Distribution of the positive (right) and negative (left) scores from PANAS.

5.2 Neural speech tracking: speech vs. noise

Oscillatory power was significantly greater during the exposure to speech in comparison to speech-shaped noise in both the *delta* (ß = 0.938, SE = 0.143, df = 75, t = 6.566, $p < .001$) and *theta* (ß = 0.528, SE = 0.089, df = 75, t =5.909, $p < .001$) band. Furthermore, ITPC was significantly stronger for speech in comparison to speech-shaped noise in both the *delta* (ß = 0.017, SE = 0.002, df = 75, t = 10.037, $p < .001$) and *theta* (ß = 0.013, SE = 0.002, df = 75, t = 6.599, $p < .001$) band. Figure 5 illustrates the estimated means and confidence intervals.

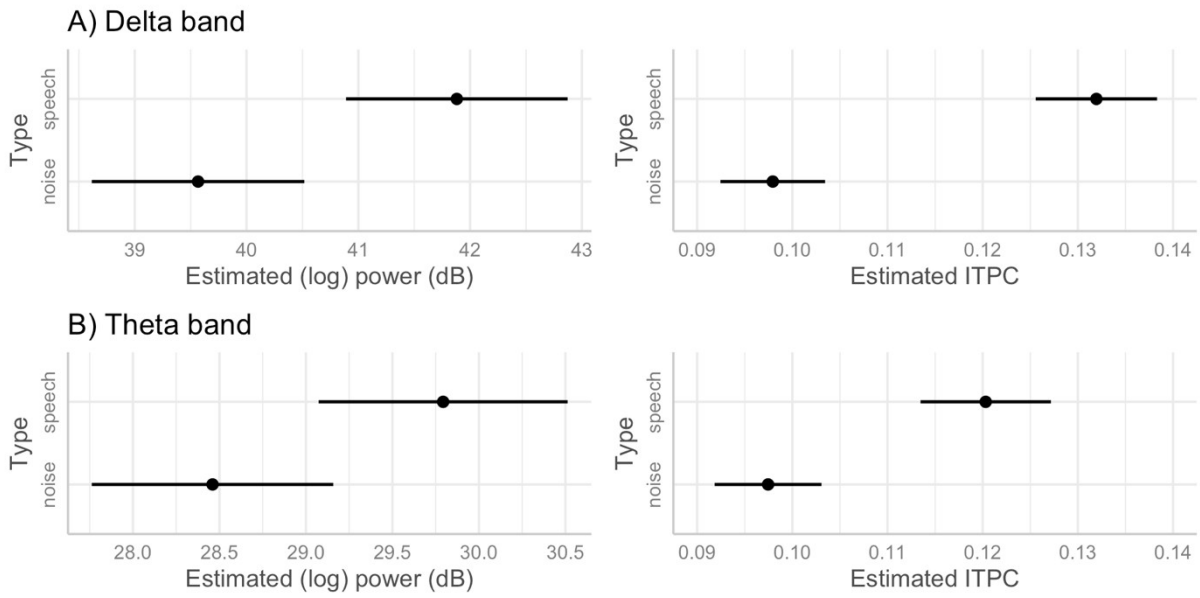

Figure 5: (A) Estimated power (left) and ITPC (right) for the speech and speech-shaped noise condition in the *delta* band. (B) Estimated power (left) and ITPC (right) for the speech and speech-shaped noise condition in the *theta* band.

5.3 Valence effects

5.3.1 Total oscillatory power

Figure 6 shows the grand-averaged total power in the lower and higher frequencies. Each condition is illustrated separately, i.e., speech condition (angry, neutral) and speech-shaped noise condition (angry, neutral).

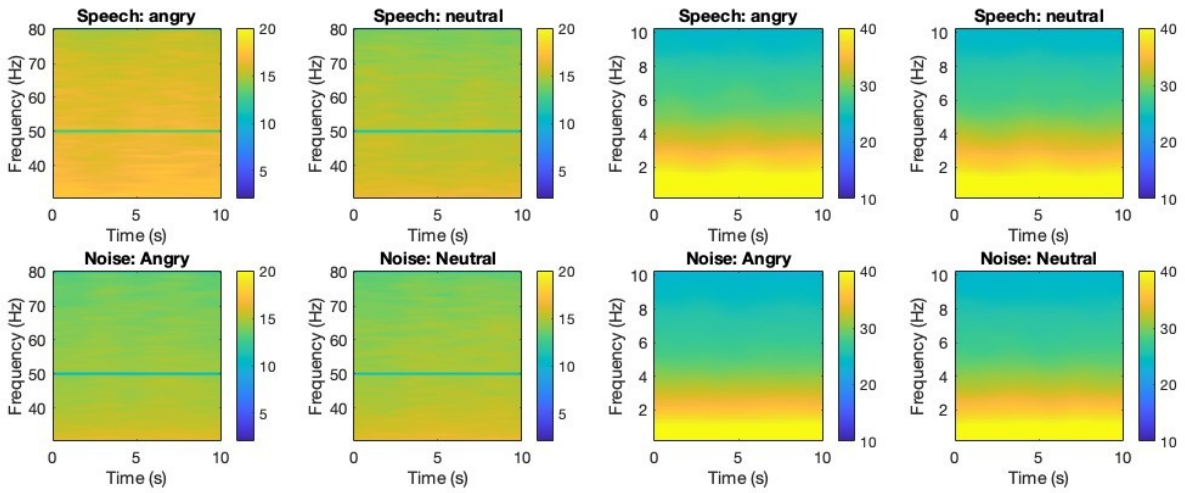

Figure 6: Grand-averaged total power in the higher frequencies (30–80 Hz) (the four graphs on the left) and lower frequencies (0.2–10 Hz) (the four graphs on the right). Yellow colors represent greater log power (dB).

The model for the *delta* power (referenced to the respective speech-shaped noise condition) revealed a significant intercept (ß = 1.876, SE = 0.298, df = 25, t = 6.286, $p < .001$) and a main effect of condition (ß = -0.439, SE = 0.209, df = 25, t = -2.101, $p = .046$) showing that *delta* power was significantly larger during angry compared to neutral speech. The model for the *theta* power (referenced to the respective speech-shaped noise condition) revealed a significant intercept (ß = 1.056, SE = 0.209, df = 25, t = 5.060, $p < .001$) and a main effect of condition (ß = -0.276, SE = 0.115, df = 25, t = -2.399, $p = .024$) indicating that *theta* power was significantly larger during angry compared to neutral speech. Figure 7 shows the grand-averaged total power in the lower frequencies. Figure 8 plots the estimated means and confidence intervals.

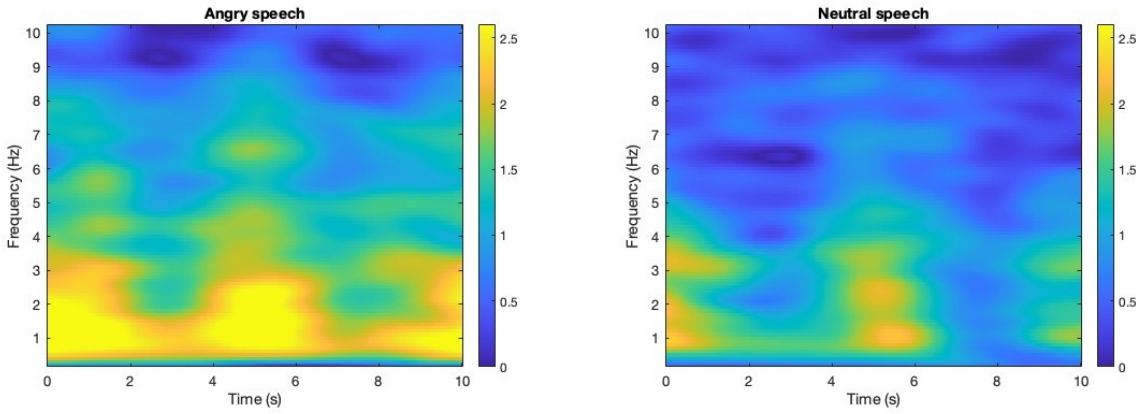

Figure 7: Grand-averaged total power in the lower frequency range (0.2–10.2 Hz) for the angry (left) and neutral (right) speech condition after the subtraction of speech-shaped noise. Yellow colors represent greater log power (dB).

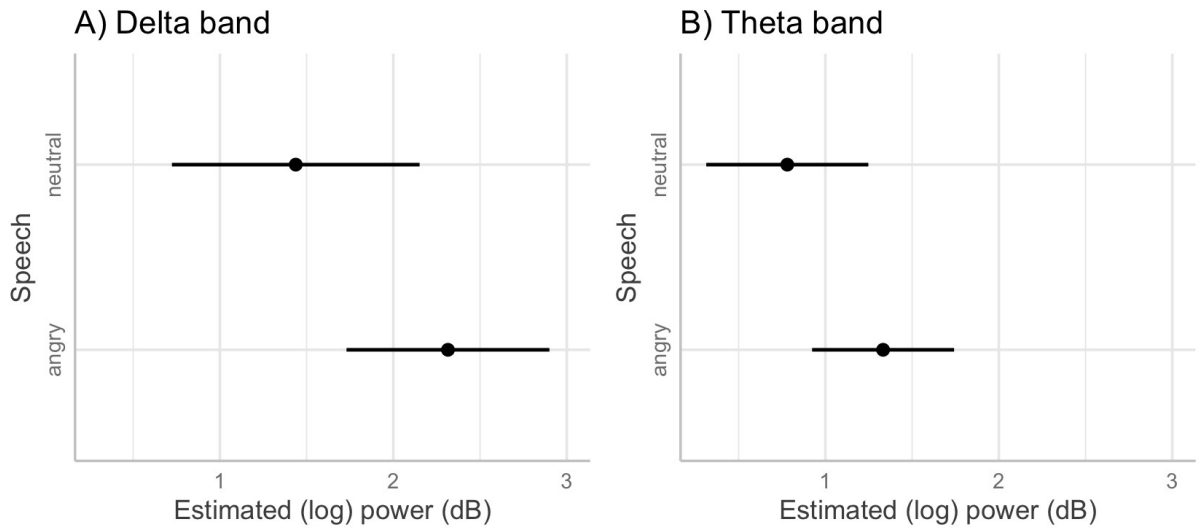

Figure 8: Estimated *delta* (left) and *theta* (right) power in the angry and neutral speech condition (referenced to the respective speech-shaped noise condition).

The model for the *gamma* power (referenced to the respective speech-shaped noise condition) revealed a significant intercept (ß = 1.043, SE = 0.303, df = 25, t = 3.446, *p* = .002) and a main effect of condition (ß = -0.772, SE = 0.248, df = 25, t = -2.902, *p* = .008) showing that *gamma* power was significantly larger during angry compared to neutral speech. Figure 9 shows the grand-averaged total *gamma* power for the angry and neutral speech condition after the subtraction of the baseline noise condition. Figure 10 plots the estimated means and confidence intervals.

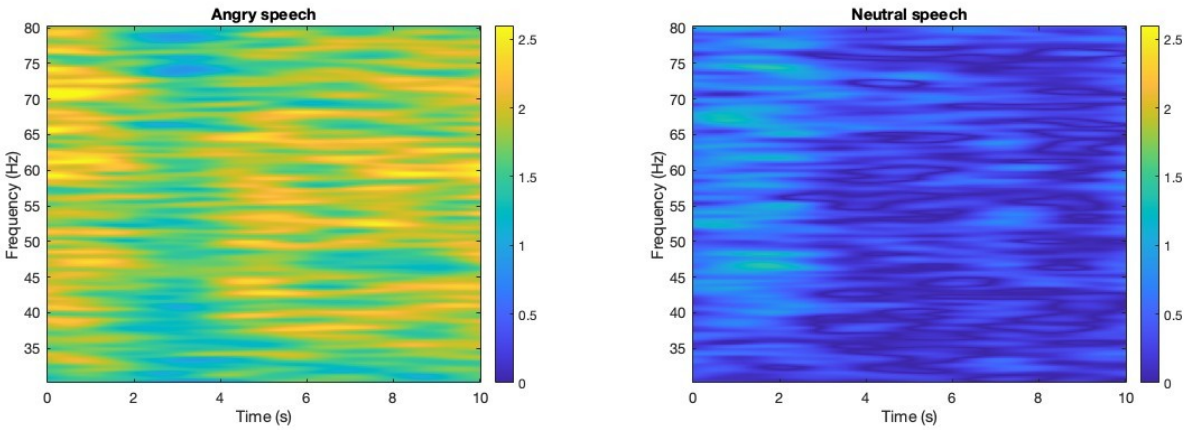

Figure 9: Grand-averaged total power in the higher frequency range (30–80 Hz) for the angry (left) and neutral (right) speech condition after the subtraction of speech-shaped noise. Yellow colors represent greater log power (dB).

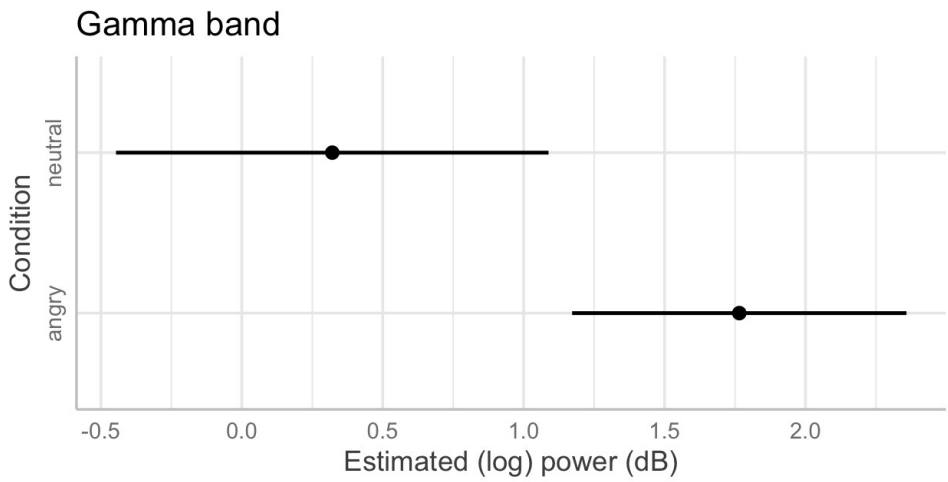

Figure 10: Estimated *gamma* power in the angry and neutral speech condition (referenced to the respective speech-shaped noise condition).

5.3.2 Inter-trial phase coherence

Figure 11 plots the grand-averaged ITPC in the frequency range 0.2–10 Hz. Each condition is illustrated separately, i.e., speech condition (angry, neutral) and speech-shaped noise condition (angry, neutral).

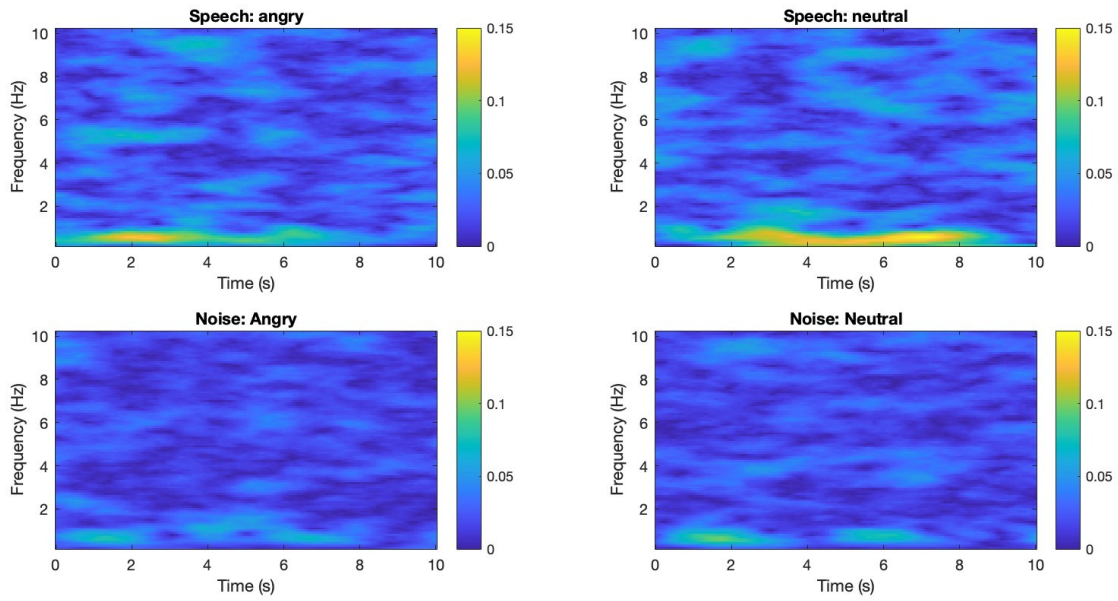

Figure 11: Grand-averaged ITPC in the angry speech, neutral speech, angry speech-shaped noise and neutral speech-shaped noise conditions. Yellow colors represent higher ITPC values.

The model for ITPC in the *theta* band (referenced to the respective speech-shaped noise condition) detected a significant intercept (ß = 0.026, SE = 0.005, df = 25, t = 5.038, $p < .001$), however no significant main effect of condition was detected (ß = 0.004, SE = 0.003, df = 25, t = 1.199, $p = .242$). The model for ITPC in the *delta* band (referenced to the respective speech-shaped noise condition) revealed a significant intercept (ß = 0.033, SE = 0.003, df = 25, t = 10.05, $p < .001$), however no significant main effect of condition was detected (ß = -0.001, SE = 0.003, df = 25, t = -0.292, $p = .772$). Figure 12 plots the grand-averaged ITPC in the frequency range 0.2–10 Hz for the angry and neutral speech condition after the subtraction of the baseline condition. Figure 13 plots the estimated means and confidence intervals for the *delta* and *theta* band.

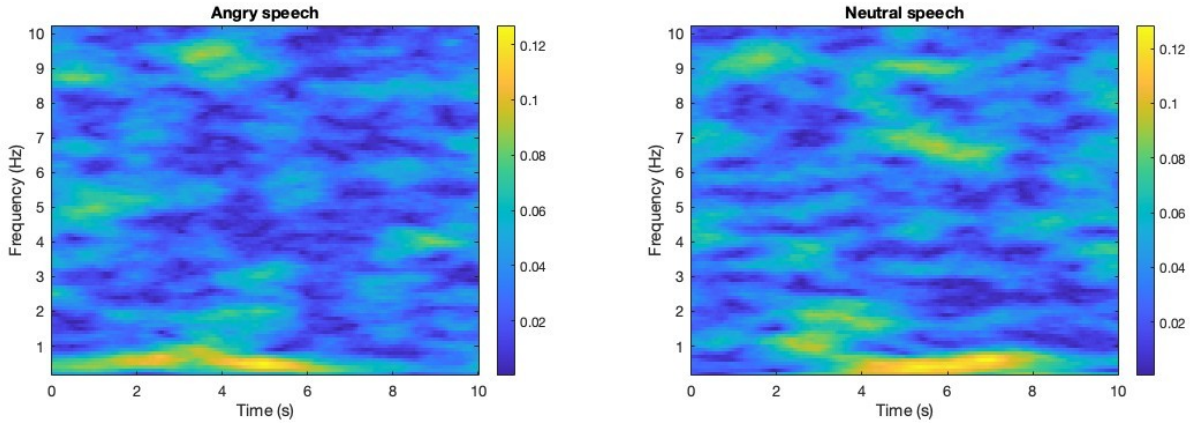

Figure 12: Grand-averaged ITPC for the angry (left) and neutral (right) speech condition after the subtraction of the speech-shaped noise. Yellow colors represent higher ITPC values.

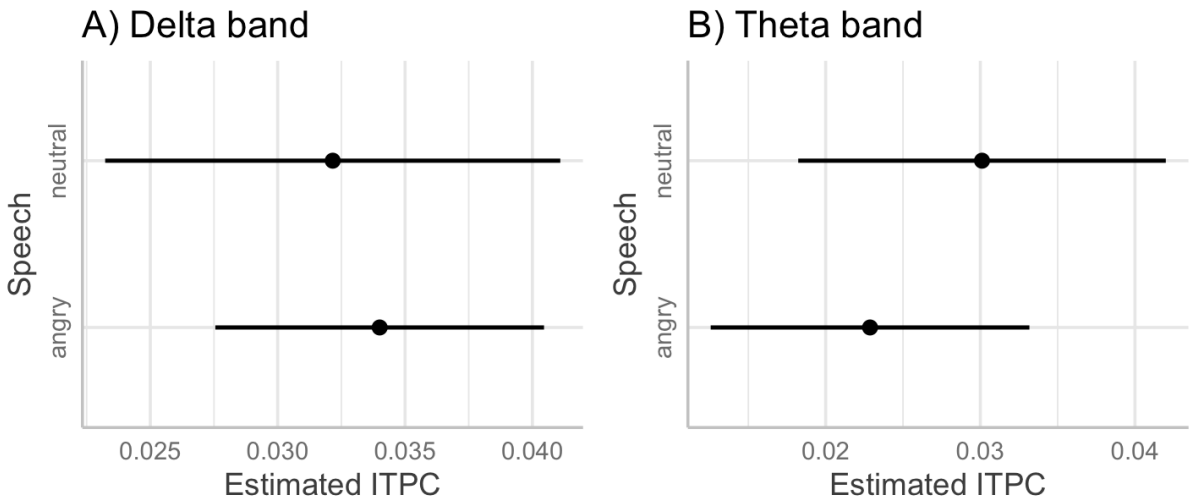

Figure 13: (A) Estimated ITPC in the *delta* and (B) *theta* band for the angry and neutral speech condition (referenced to the respective speech-shaped noise condition).

5.3.3 N400

The model for N400 revealed a significant intercept (estimate = -2.702, SE = 0.539, df = 4.124, t = -5.015, *p* = .007) indicating that the ERP response was negative overall. The model also detected a significant main effect of condition (ß = -0.484, SE = 0.180, df = 70, t = -2.686, p = .009) showing that emotional negative words evoked a significantly smaller negative response (N400) than the non-emotional neutral words. Figure 14 shows the grand average ERP averaged across Fz and Cz electrodes. Figure 15 shows the estimated marginal means and confidence intervals.

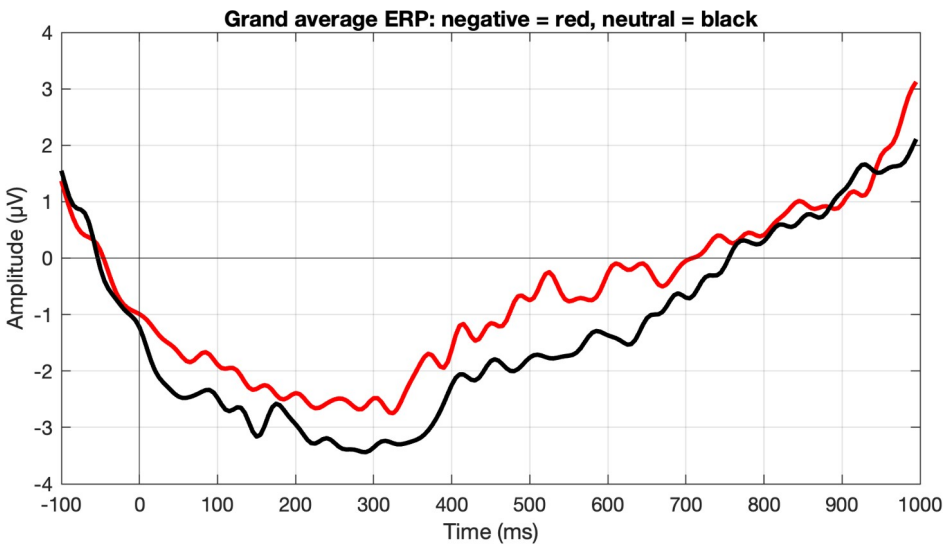

Figure 14: Grand-averaged ERP for negative (red) and neutral (black) words. Negative is plotted downward.

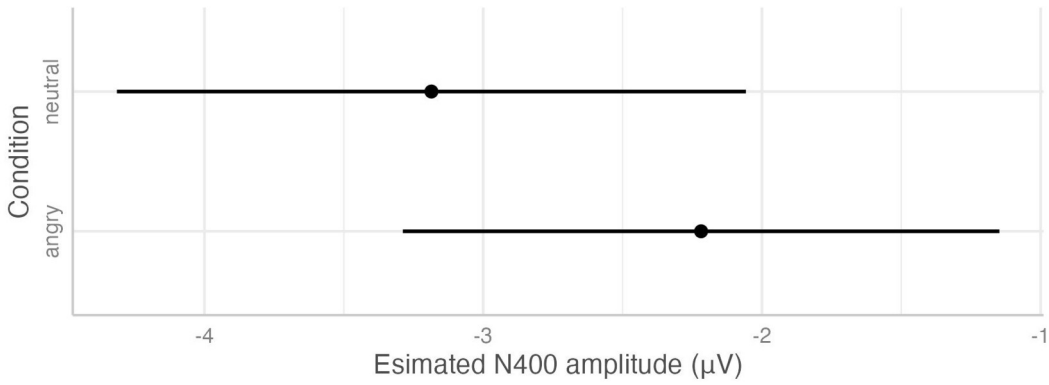

Figure 15: Estimated N400 amplitude for negative and neutral words (means and 95% confidence intervals).

6 Discussion

The present study investigated the neural processing of angry and neutral speech by measuring two indicators of neural speech tracking – oscillatory power in the *delta*, *theta*, and *gamma* band and the inter-trial phase coherence for the *delta* and *theta* band. ERP to the sentence-final words were recorded, specifically the N400 component, an index of semantic prediction and integration in speech comprehension. Participants were exposed to two-sentence segments of conversations, each carrying either a negative or neutral valence, ending with words that matched the emotional context of the preceding sentences. Prior to each block of speech, participants listened to speech-shaped noise.

Oscillatory power and inter-trial phase coherence in the *delta* and *theta* band was stronger for the speech conditions (regardless of the emotional valence) compared to speech-shaped noise. The emotional valence of heard conversation impacted the total power in the *delta*, *theta*, and *gamma* band (referenced to the respective speech-shaped noise condition which served as a baseline condition). Oscillatory power was significantly larger during angry in contrast with neutral speech exposure. No effect of emotional valence on the inter-trial phase coherence was detected in either the *delta* or *theta* band. Consequentially, the results regarding the differences in neural tracking of angry and neutral speech (H1) are ambiguous (as we observed significant differences in oscillatory power but not in ITPC) and insufficient to reject the null hypothesis that angry and neutral speech elicit similar neural tracking.

Furthermore, the emotional valence of the sentence-final word impacted the N400 component; the N400 response was significantly smaller to negative words in a preceding negative context than to neutral words in a preceding neutral context. The joint results regarding the N400 and *gamma* power support our hypotheses H2 and H3.

6.1 Positive and Negative Affect Schedule

Overall, the mean score for the positive scale suggests that most participants were in a neutral or positive mood. The distribution of the negative scores was skewed to the left, indicating that most participants did not experience negative affect before the experiment. Although some previous studies indicate that there is a relationship between affective state and the processing of emotional stimuli (Chwilla et al., 2011), we were unable to test this effect due to the low variability of our participants' affective states.

6.2 Neural speech tracking: speech vs. noise

Neural tracking was stronger for speech in comparison to speech shaped noise. That is consistent with previous studies reporting stronger tracking for normal speech compared to a non-speech baseline (Molinaro & Lizarazu, 2018; Rimmele et al., 2015). This indicates that the measured oscillatory power and ITPC in the *delta* and *theta* band during exposure to speech cannot be explained by a purely bottom-up synchronization of neural oscillations to a rhythmical sensory input, but rather reflect processing that is specific to speech comprehension.

6.3 Enhanced processing of emotional speech

Previous studies on neural speech tracking have linked *delta* and *theta* oscillations to the segmentation of the speech stream along the syllable and prosodic word rate. We wanted to compare the neural tracking of emotional (angry) and neutral speech by calculating total oscillatory power and ITPC in the *delta* and *theta* band. We expected observing differential neural tracking either due to the salient prosodic differences associated with emotional speech, or due to higher employment of cognitive resources that is related to the processing of emotionally significant stimuli in general.

The higher oscillatory power in the *delta* and *theta* band implies enhanced processing during the perception of angry in comparison to neutral speech. From the neural tracking perspective, this could mean greater tracking of the acoustic properties of the angry speech input. The difference was slightly larger for *delta* power which would support the theory that the effect was driven by the salient prosody of angry speech. Given that we used speech-shaped noise as a baseline condition (i.e., an auditory signal with the intensity contour of speech), we have solid evidence that the differences in *delta* and *theta* power are not due to different rhythmic properties of our angry and neutral speech signal. However, as we did not observe an effect for the ITPC, we cannot conclude that the differences in oscillatory power were caused by stronger or more accurate tracking of angry speech compared to neutral speech.

As mentioned in the theoretical part, besides being involved in speech processing, activity in the *delta* and *theta* band has been related to working memory. For instance, Bastiaansen et al. (2005) found event-related increases in *theta* power to open class words (e.g., verbs, nouns, and adjectives) in comparison to closed class words (e.g., conjunctions, prepositions) when subjects were reading a short story. The authors attributed the changes in *theta* power to the retrieval of lexical-semantic information from memory. Thus, another viable theory is that the

greater *theta* power we observed during the angry speech condition could to some extent reflect facilitated retrieval from lexical memory for emotional speech.

Therefore, prospectively, we intend to calculate other neural tracking indexes, such as cerebral-acoustic coherence or the temporal response function, in order to better understand whether the oscillatory differences could indeed be attributed to the differential synchronization of neural activity to angry and neutral speech.

6.4 Gamma power as an index of facilitated prediction

The literature on neural speech processing indicates that increased *gamma* activity is linked to the processing of words that are semantically congruent with their preceding context. Therefore, *gamma* power might reflect the predictability of the incoming words based on the preceding sentence (Bastiaansen & Hagoort, 2015; Hald et al., 2006; Mai et al., 2016). As previous studies on emotional priming suggest (Eder et al., 2012; Kanske et al., 2011), emotionality of a preceding cue modulates the predictability of the subsequent target which leads to facilitated integration of the target into the preceding context. According to this theory, the increased *gamma* power during the perception of emotional speech, together with the effect on the N400 component that we observed in our experiment, might indeed reflect reduced integration demands due to a facilitatory effect of the congruent emotional context. However, our results are in opposition to Chen et al. (2013) who observed *gamma* power increases during the processing of emotionally incongruent in comparison to congruent words.

Taken together with the studies reviewed above, our results lead us to conclude that neural activity in the *gamma* band is *in some way* modulated by the interaction of emotional and predictive processes.

Oscillatory activity in the *gamma* band has also been related to memory processes (Headley & Paré, 2013) and sustained attention (Jensen et al., 2007; Siegle et al., 2010). Therefore, it is important to note that the observed greater *gamma* activity during the perception of emotional speech might be attributed to other cognitive processes. To better understand the contribution of speech-specific and domain-general processes to our observed data, we intend to perform further analyses, e.g., calculate the phase-amplitude coupling of *theta* and *gamma* oscillations.

6.5 N400 as an index of semantic-emotional access

Our results regarding the effect of emotional valence on the N400 are in line with previous studies reporting a decreased amplitude of the N400 to negative in comparison to neutral words presented in an emotionally congruent context (Kanske et al., 2011; Wang et al., 2019; Zhang et al., 2021). More specifically, with the present auditory stimuli we replicated the findings of Zhang et al. (2021) on a decrease of the N400 amplitude to negative words following a negative sentence context.

The N400 effect has been related to semantic prediction and integration, with more predictable words requiring less integration efforts and thus eliciting a smaller amplitude of the N400 (Berkum et al., 1999; Kutas & Hillyard, 1980). The observed modulation of the N400 amplitude by negative valence might be explained by facilitated predictive processing of the emotional target word. The preceding sentence context (conveyed by the emotional semantic content as well as emotional prosody) provides the listener with additional contextual information which changes the predictability of upcoming words. Consequentially, the processing demands on the subsequent integration of the emotionally congruent target word might be reduced.

In the grand-average ERP, we observed differences in the amplitude of the ERP between the negative and neutral condition all the way from 0 to 800 ms after target word onset (i.e., the difference was not exclusive to the N400 component). However, it is important to highlight that we recorded auditory ERP to sentence-final words embedded in naturalistic speech. There is a high number of confounding factors that are challenging to control for when using speech as experimental stimuli (e.g., articulation, pronunciation, prosodic differences). Consequentially, this leads to higher variability in the shape of the ERP response.

In the future, calculating the predictability for each segment-final word in both conditions might be beneficial in order to determine whether the observed effect on the N400 (and possibly, the effect on *gamma* power, too) might actually be attributed to the differences in predictability due to the additional emotional context. To achieve this, surprisal values or cloze probability for segment-final words could be calculated.

6.6 Potential limitations

In our experiment, we only used one female speaker to record our stimuli. Future studies should use recordings of multiple male and female speakers to cover a wider range of possible speaker characteristics. This might increase experimental validity.

We wanted to use one set of stimuli for both the female and male participants while keeping the speech material as self-relevant as possible. Therefore, we formulated the speech stimuli to be gender neutral. That led to the exclusion of all insults that could be used only to address males or females. Despite that, we still faced the challenge of having to use formulations that would not explicitly reveal the gender of the addressed listener. In future studies, two separate speech sets (based on gender) might be more suitable in order to accentuate the participant's impression that the speech is addressed directly at them.

The emotional speech in our experiment was acted which might raise concerns as to whether the participants perceived the prosody of angry speech as emotional. Although some studies suggest that acted emotional speech might lead to exaggerated displays of affect (Shahid et al., 2009; Wilting et al., 2006), they also provide substantial evidence that emotional speech recorded by a trained actor does elicit some emotional percept in the listener. Furthermore, in the final questionnaire following the EEG experiment, participants were instructed to report the perceived differences between the two experimental blocks in the speech condition (i.e., angry and neutral speech blocks). Based on our participants' answers, there are distinct differences in perceived emotional valence of our speech blocks not only in the semantic content, but also in the emotional intonation.

It is also important to highlight the conceptual issue regarding the term *neutral word*. In most studies, researchers typically contrast emotional verbal material with relatively non-emotional, neutral material, characterizing neutral words as stimuli that do not carry an emotional charge. While this definition might be most suitable for experimental research, it is important to acknowledge its main limitation. That being the fact that the perceived emotionality of a word is heavily contingent upon individual factors, such as personal experiences and attitudes related to the object or situation in question. For example, the word "plastic" is a neutral word per se, however, for some people might have negative connotations due to environmental reasons.

6.7 Concluding remarks

Emotional speech carries important information that might have potential consequences for achieving or obstructing one's goals. Thus, the ability to detect that a speaker is angry is important to swiftly evaluate the motivational significance of the social interaction. In cases when our personal or social goals are potentially endangered, the facilitated processing of emotional stimuli can be advantageous by allowing to reallocate cognitive resources towards decision-making strategies that ultimately lead to an appropriate reaction to the stimulus.

7 Conclusion

The present study represents one of the first explorations of the oscillatory dynamics of continuous emotional speech processing. Previous EEG studies have concentrated predominantly on the investigation of evoked responses (ERP and ERSP) to written emotional words, typically presented in isolation (Wang & Bastiaansen, 2014), or embedded in sentences (Chen et al., 2013). To our knowledge, ours is the first study investigating total oscillatory power during the continuous processing of emotional auditorily presented sentences.

In this study, we attempted to challenge classical experimental paradigms for the investigation of emotional speech and word processing. Our aim was to create stimuli that would integrate congruent emotional information from the vocal and verbal channel, simulating conditions under which humans encounter emotional speech in real-life settings. For this reason, we used short conversation segments formulated in a way as if the speaker is addressing the listener (i.e., the participant). Using longer segments of speech allowed us to use processing methods from the neural speech tracking literature and explore the oscillatory processing of larger chunks of speech.

We found that negative words elicited a smaller amplitude of the N400 component of the event-related potentials, in comparison to neutral words. Listening to emotional speech led to greater *delta*, *theta*, and *gamma* power. The analyses failed to find a difference in the inter-trial phase coherence for angry and neutral speech. The results are in line with previous studies suggesting enhanced processing of emotional material; however, further investigation is necessary in order to determine whether the observed differences in oscillatory power could be attributed to stronger neural speech tracking or whether they are related to domain-general processes involved in the perception emotional speech.

Our experiment contributes to the general understanding of how the human brain is wired to process emotionally salient speech input. Prospectively, the results may aid in understanding the neural processes underlying the disrupted ability to decode emotions from speech observed in many psychiatric and neurological disorders, such as schizophrenia, Parkinson's disease or autism.

References

Attaheri, A., Choisdealbha, Á. N., Di Liberto, G. M., Rocha, S., Brusini, P., Mead, N., ... & Goswami, U. (2022b). Delta-and theta-band cortical tracking and phase-amplitude coupling to sung speech by infants. *NeuroImage*, *247*, 118698. https://doi.org/10.1016/j.neuroimage.2021.118698

Attaheri, A., Panayiotou, D., Phillips, A., Ní Choisdealbha, Á., Di Liberto, G. M., Rocha, S., ... & Goswami, U. (2022a). Cortical tracking of sung speech in adults vs infants: A developmental analysis. *Frontiers in Neuroscience*, *16*, 842447. https://doi.org/10.3389/fnins.2022.842447

Bachmann, F. L., MacDonald, E. N., & Hjortkjær, J. (2021). Neural measures of pitch processing in EEG responses to running speech. *Frontiers in Neuroscience*, *15*, 738408. https://doi.org/10.3389/fnins.2021.738408

Bastiaansen, M. C., Van Der Linden, M., Ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical—Semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, *17*(3), 530-541. https://doi.org/10.1162/0898929053279469

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. https://doi.org/10.1162/jocn_a_00829

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

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31-45. https://doi.org/10.1016/j.cortex.2017.02.004

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research*, *17*(3), 613-620. https://doi.org/10.1016/S0926-6410(03)00174-5

Berckmoes, C., & Vingerhoets, G. (2004). Neural foundations of emotional speech processing. *Current Directions in Psychological Science*, *13*(5), 182-185. https://doi.org/10.1111/j.0963-7214.2004.00303.x

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. https://doi.org/10.1162/089892999563724

Bertels, J., & Kolinsky, R. (2016). Disentangling fast and slow attentional influences of negative and taboo spoken words in the emotional Stroop paradigm. *Cognition and Emotion*, *30*(6), 1137-1148. https://doi.org/10.1080/02699931.2015.1052780

Bertels, J., Kolinsky, R., & Morais, J. (2010). Emotional valence of spoken words influences the spatial orienting of attention. *Acta Psychologica*, *134*(3), 264-278. https://doi.org/10.1016/j.actpsy.2010.02.008

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/

Bourguignon, M., De Tiege, X., De Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., ... & Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314-326. https://doi.org/10.1002/hbm.21442

Brandeis, D., & Lehmann, D. (1986). Event-related potentials of the brain and cognitive processes: approaches and applications. *Neuropsychologia*, *24*(1), 151-168. https://doi.org/10.1016/0028-3932(86)90049-7

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*(5679), 1926-1929. https://doi.org/10.1126/science.1099745

Citron, F. M. (2012). Neural correlates of written emotion word processing: a review of recent electrophysiological and hemodynamic neuroimaging studies. *Brain and Language*, *122*(3), 211-226. https://doi.org/10.1016/j.bandl.2011.12.007

Colgin, L. L. (2015). Theta–gamma coupling in the entorhinal–hippocampal system. *Current Opinion in Neurobiology*, *31*, 45-50. https://doi.org/10.1016/j.conb.2014.08.001

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant directed speech in the first month after birth. *Child Development*, *61*(5), 1584-1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Coopmans, C. W., De Hoop, H., Hagoort, P., & Martin, A. E. (2022). Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. *Neurobiology of Language*, *3*(3), 386-412. https://doi.org/10.1162/nol_a_00070

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. https://doi.org/10.1016/j.clinph.2008.10.008

Decruy, L., Lesenfants, D., Vanthornhout, J., & Francart, T. (2020). Top down modulation of neural envelope tracking: the interplay with behavioral, self report and neural measures of listening effort. *European Journal of Neuroscience*, *52*(5), 3375-3393. https://doi.org/10.1111/ejn.14753

Delorme A & Makeig S (2004) EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, *Journal of Neuroscience Methods* 134:9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Ding, J., Wang, L., & Yang, Y. (2015). The dynamic influence of emotional words on sentence processing. *Cognitive, Affective, & Behavioral Neuroscience*, *15*, 55-68. https://doi.org/10.3758/s13415-014-0315-6

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. https://doi.org/10.1080/23273798.2019.1628283

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: An ERP study. *Social Cognitive and Affective Neuroscience*, *7*(4), 436-445. https://doi.org/10.1093/scan/nsr033

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2012). Automatic response activation in sequential affective priming: An ERP study. *Social Cognitive and Affective Neuroscience*, *7*(4), 436-445. https://doi.org/10.1093/scan/nsr033

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*(29), 5750-5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104. https://psycnet.apa.org/doi/10.1037/0012-1649.20.1.104

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. https://doi.org/10.1016/j.neuroimage.2012.05.003

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511-517. https://doi.org/10.1038/nn.3063

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980-991. http://dx.doi.org/10.1016/j.neuron.2012.12.037

Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emotion Review*, *13*(1), 34-43. https://doi.org/10.1177/1754073919898522

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. https://doi.org/10.3389/fnhum.2016.00326

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. https://doi.org/10.1016/j.bandl.2005.06.007

Harding, E. E., Sammler, D., Henry, M. J., Large, E. W., & Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *NeuroImage*, *185*, 96-101. https://doi.org/10.1016/j.neuroimage.2018.10.037

Headley, D. B., & Paré, D. (2013). In sync: *gamma* oscillations and emotional memory. *Frontiers in Behavioral Neuroscience*, *7*, 170. https://doi.org/10.3389/fnbeh.2013.00170

Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, *45*(3), 487-498. https://doi.org/10.1111/j.1469-8986.2007.00638.x

Herweg, N. A., Solomon, E. A., & Kahana, M. J. (2020). Theta oscillations in human memory. *Trends in Cognitive Sciences*, *24*(3), 208-227. https://doi.org/10.1016/j.tics.2019.12.006

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. https://doi.org/10.1016/j.ijpsycho.2012.06.001

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. https://doi.org/10.1162/jocn.2008.21151

Holtze, B., Jaeger, M., Debener, S., Adiloğlu, K., & Mirkovic, B. (2021). Are they calling my name? Attention capture is reflected in the neural tracking of attended and ignored speech. *Frontiers in Neuroscience*, *15*, 643705. https://doi.org/10.3389/fnins.2021.643705

Hovsepyan, S., Olasagasti, I., & Giraud, A. L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications*, *11*(1), 3117. https://doi.org/10.1038/s41467-020-16956-5

Huang, Y. X., & Luo, Y. J. (2006). Temporal course of emotional negativity bias: an ERP study. *Neuroscience Letters*, *398*(1-2), 91-96. https://doi.org/10.1016/j.neulet.2005.12.074

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A. L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *eLife*, *4*, e06213. https://doi.org/10.7554/eLife.06213

Chalas, N., Daube, C., Kluger, D. S., Abbasi, O., Nitsch, R., & Gross, J. (2023). Speech onsets and sustained speech contribute differentially to delta and theta speech tracking in auditory cortex. *Cerebral Cortex*, *33*(10), 6273-6281. https://doi.org/10.1093/cercor/bhac502

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. https://doi.org/10.1016/j.neulet.2013.08.034

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. https://doi.org/10.1162/jocn.2010.21578

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319-330. https://doi.org/10.1525/mp.2006.23.4.319

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. https://psycnet.apa.org/doi/10.1037/0022-3514.75.4.887

Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human *gamma*-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317-324. https://doi.org/10.1016/j.tins.2007.05.001

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Scientific Reports*, *8*(1), 13745. https://doi.org/10.1038/s41598-018-32150-6

Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, *1148*, 138-148. https://doi.org/10.1016/j.brainres.2007.02.044

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, *49*(11), 3121-3129. https://doi.org/10.1016/j.neuropsychologia.2011.07.022

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *Journal of Neuroscience*, *30*(2), 620-628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Kissler, J., & Herbert, C. (2013). Emotion, Etmnooi, or Emitoon?–Faster lexical access to emotional than to neutral words during reading. *Biological Psychology*, *92*(3), 464-479. https://doi.org/10.1016/j.biopsycho.2012.09.004

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing—An ERP study. *Biological Psychology*, *80*(1), 75-83. https://doi.org/10.1016/j.biopsycho.2008.03.004

Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain research*, *1151*, 107-118. https://doi.org/10.1016/j.brainres.2007.03.015

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203-205. https://doi.org/10.1126/science.7350657

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lane, R. D., Chua, P. M., & Dolan, R. J. (1999). Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia*, *37*(9), 989-997. https://doi.org/10.1016/S0028-3932(99)00017-2

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

Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., & Molinaro, N. (2021). Language proficiency entails tuning cortical activity to second language speech. *Cerebral Cortex*, *31*(8), 3820-3831. https://doi.org/10.1093/cercor/bhab051

Lizarazu, M., Lallier, M., & Molinaro, N. (2019). Phase− amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, *1453*(1), 140-152. https://doi.org/10.1111/nyas.14099

Lüdecke D (2018). "ggeffects: Tidy Data Frames of Marginal Effects from Regression Models." *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010. https://doi.org/10.1016/j.neuron.2007.06.004

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*, 474-488. https://doi.org/10.3758/BF03195840

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. https://doi.org/10.1016/j.neuroimage.2016.02.064

ManyBabies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24-52. https://doi.org/10.1177/2515245919900809

Menn, K. H., Michel, C., Meyer, L., Hoehl, S., & Männel, C. (2022). Natural infant-directed speech facilitates neural tracking of prosody. *NeuroImage*, *251*, 118991. https://doi.org/10.1016/j.neuroimage.2022.118991

Molinaro, N., & Lizarazu, M. (2018). Delta (but not theta)□band cortical entrainment involves speech□specific processing. *European Journal of Neuroscience*, *48*(7), 2642-2650. https://doi.org/10.1111/ejn.13811

Morillon, B., Liegeois-Chauvel, C., Arnal, L. H., Bénar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Frontiers in Psychology*, *3*, 248. https://doi.org/10.3389/fpsyg.2012.00248

Myers, B. R., Lense, M. D., & Gordon, R. L. (2019). Pushing the envelope: Developments in neural entrainment to speech and the biological underpinnings of prosody perception. *Brain Sciences*, *9*(3), 70. https://doi.org/10.3390/brainsci9030070

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive sciences*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649-1653. http://dx.doi.org/10.1016/j.cub.2015.04.049

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378-1387. https://doi.org/10.1093/cercor/bhs118

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

Rimmele, J. M., Golumbic, E. Z., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, *68*, 144-154. https://doi.org/10.1016/j.cortex.2014.12.014

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. https://doi.org/10.1016/j.neuroimage.2005.06.023

Sbattella, L., Colombo, L., Rinaldi, C., Tedesco, R., Matteucci, M., & Trivilini, A. (2014, January). Extracting emotions and communication styles from prosody. In *International Conference on Physiological Computing Systems* (pp. 21-42). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45686-6_2

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. https://doi.org/10.1016/j.biopsycho.2008.03.010

Shahid, S., Krahmer, E. J., & Swerts, M. G. J. (2008). Real vs. acted emotional speech: Comparing south asian-and caucasion speakers and observers. In *Proceedings of the 4th International Conference on Speech Prosody* (pp. 669-772). Unknown Publisher.

Schachner, A., & Hannon, E. E. (2011). Infant-directed speech drives social preferences in 5-month-old infants. *Developmental Psychology*, *47*(1), 19. https://psycnet.apa.org/doi/10.1037/a0020740

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. https://doi.org/10.1177/0022022101032001009

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24-30. https://doi.org/10.1016/j.tics.2005.11.009

Siakaluk, P. D., Newcombe, P. I., Duffels, B., Li, E., Sidhu, D. M., Yap, M. J., & Pexman, P. M. (2016). Effects of emotional experience in lexical decision. *Frontiers in Psychology*, *7*, 1157. https://doi.org/10.3389/fpsyg.2016.01157

Siegle, G. J., Condray, R., Thase, M. E., Keshavan, M., & Steinhauer, S. R. (2010). Sustained *gamma*-band EEG following negative words in depression and schizophrenia. *International Journal of Psychophysiology*, *75*(2), 107-118. https://doi.org/10.1016/j.ijpsycho.2008.04.008

Stoet, G. (2010). PsyToolkit – A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24-31 https://doi.org/10.1177/0098628316677643

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta□band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*(11), 3831-3842. https://doi.org/10.1111/ejn.14510

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. https://doi.org/10.1007/978-3-030-81542-4_4

Unkelbach, C., Alves, H., & Koch, A. (2020). Negativity bias, positivity bias, and valence asymmetries: Explaining the differential processing of positive and negative information. In *Advances in Experimental Social Psychology* (Vol. 62, pp. 115-187). Academic Press. https://doi.org/10.1016/bs.aesp.2020.04.005

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383. https://psycnet.apa.org/doi/10.1037/0033-2909.134.3.383

Vanthornhout, J., Decruy, L., & Francart, T. (2019). Effect of task and attention on neural tracking of speech. *Frontiers in Neuroscience*, *13*, 474279. https://doi.org/10.3389/fnins.2019.00977

Wang, L., & Bastiaansen, M. (2014). Oscillatory brain dynamics associated with the automatic processing of emotion in words. *Brain and Language*, *137*, 120-129. https://doi.org/10.1016/j.bandl.2014.07.011

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*, 20589. https://doi.org/10.3389/fpsyg.2012.00187

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. https://doi.org/10.1016/j.neulet.2019.01.028

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. https://psycnet.apa.org/doi/10.1037/0022-3514.54.6.1063

Weingartová, L., & Volín, J. (2014). Temporální charakteristiky. Fonetická identifikace mluvčího. Praha: Filozofická fakulta, Univerzita Karlova v Praze, 95-103.

Wilting, J., Krahmer, E., & Swerts, M. (2006, September). Real vs. acted emotional speech. In Interspeech (Vol. 2006, p. 9th).

Wöstmann, M., Fiedler, L., & Obleser, J. (2017). Tracking the signal, cracking the code: Speech and speech comprehension in non-invasive human electrophysiology. Language, Cognition and Neuroscience, 32(7), 855-869. https://doi.org/10.1080/23273798.2016.1262051

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. https://doi.org/10.1016/j.jneuroling.2021.100989

List of appendices

Appendices

Appendix A: Final list of conversation segments

Negative conversation segments	**Neutral conversation segments**
Okamžitě nech těch svejch debilních vtípků! Je mi z těch narážek na **blití**.	Přines mi prosím ten novej výtisk novin. Ty starý vydání odlož na **stranu**.
Už mám tvýho chování po krk. Sbal si svoje krámy a **vypadni**!	Po těch schodech musíš jít pomalu. Přidržuj se zábradlí a **nespadni**.
Už tě mám fakt plný zuby! Celej večer mě jenom **ztrapňuješ**.	Nemusíš to nikam odnášet. Stačí když tu vázu jenom **odsuneš**.
Celou dobu mě s nima akorát pomlouváš. Je mi z tebe fakt **špatně**.	To tlačítko nestačí podržet krátce. Musíš ho zmáčknout fakt **dlouze**.
Vždyť mě kvůli tobě vyhodili ze školy! Za tohle se budeš smažit v **pekle**.	Včera jsme ti koupili novou židli. Návod na její poskládání máš v **tašce**.
Celej večer si jenom stěžuješ. Už je to tady s tebou fakt **otravný**.	Klidně si půjč tuhle učebnici. Máš to tam všechno podrobně **popsaný**.
Okamžitě jí přestaň nadávat! Tohle tvoje chování je **ostuda**.	Před chvílí tady něco zvonilo. Podle mě to byl tvůj **telefon**.
Pořád se mi cpeš do mejch osobních věcí! Tuhle vlezlost na tobě **nenávidím**.	Klidně za mnou pak doraž do studovny. Každej den tam někoho **doučuju**.
Nebaví mě furt poslouchat tvoje kecy. Pro jednou už zavři **hubu**!	Klidně můžeš zůstat u mě v kanceláři. Jen potom prosím zavři **dveře**.

Ten formulář máš zase špatně vyplněnej! Ty jsi prostě úplně **debilní**!	Nemusíš s sebou brát ty těžký talíře. Stačí vzít ty lehký **plastový**.
V práci se všema jednáš hrozně nepřátelsky. Přestaň být takhle **arogantní**!	Budeme se scházet každý odpoledne. Tak to aspoň bude víc **pravidelný**.
Nebudu se o tom s tebou hádat. Prostě se přestaň chovat jako **hovado**.	Musíš to stříhat těmahle nůžkama. Pak to budeš mít rovný jako **pravítko**!
Čekám tady na tebe už dvě hodiny. Tohle chování je fakt **nespolehlivý**.	U metra jsem vás vůbec neviděla. Byli jste úplně **neviditelný**.
Tobě se v ničem nedá vůbec věřit. Celou dobu mě akorát **pomlouváš**.	Ve dvanáct hodin mám ještě přednášku. To ty v tu dobu už dávno **obědváš**.
Tyhle kalhoty ti teda vůbec nesluší. Vypadáš v nich fakt úplně **strašně**.	Nevidím na tobě vůbec žádnej rozdíl. Takhle vypadáš úplně **běžně**.
Dej tomu dítěti aspoň najíst. Tohle je od tebe strašně **nezodpovědný**.	Nemusíš tam docházet každej tejden. Naše schůzky budou celkem **nepravidelný.**
Neodhazuj ty plesnivý šlupky na stůl. Bydlet s tebou je teda fakt **odporný**.	Vyber si jeden z těch článků ode mě. Tamty jsou pro tebe zbytečně **odborný**.
Během tý schůzky musíš být zticha. Tak už se přestaň smát jako **kretén**!	Tenhle bílej plášť máš určitě novej. Vypadáš v něm úplně jako **doktor**.
Kvůli tobě je úplně prázdná lednička. Ty seš opravdu nenažraný **prase**.	Mirek jede příští týden na dovolenou. Nezapomeň pak nakrmit to jeho **zvíře**.

Radši se nad sebou trochu zamysli. Nemusíš furt reagovat **agresivně**!	Není třeba si na to hledat výpomoc. Klidně můžeš pracovat **samostatně**.
Hned se převlíkni do něčeho jinýho! Tvoje roztrhaný hadry mě **pohoršujou**.	Všichni pracujeme podle tvýho návodu. Všechna pravidla se tady **dodržujou**.
Nemusíš ho přece takhle ponižovat. Přestaň se chovat jako **sadista**!	To tvoje kafe má zajímavou strukturu. Děláš ho úplně jako **barista**.
Přestaň se před všema takhle předvádět. Tohle chování mě začíná **iritovat**.	K bakalářce si najdi dost literatury. Při psaní pak musíš všechno **citovat**.
Zase tě z tý školy vyhodili! Rodiče kvůli tobě budou **zuřit**.	To připomenutí jim nevadí. Díky tobě na to aspoň budou **myslet**.
Na mě z těch peněz vůbec nic nezbylo! Ty seš opravdu takovej **sobec**!	Tu smlouvu můžeš rovnou podepsat. Jsi totiž jedinej **kupec**.
Vždyť on ti přece vůbec nic neudělal! Přestaň se k němu chovat tak **bezcitně**!	Nechci to tu říkat úplně před všema. Řeknu ti to potom **soukromě**.
Vždyť je to jenom malinký miminko! Nesmíš s ním zacházet takhle **surově**.	Vůbec to nemusíš předělávat. Klidně to můžeš namalovat **barevně**.
Na schůzi se kvůli tobě všichni rozhádali. Do práce teď choděj fakt **zpruzený**.	Na dnešní schůzi nám řekneš svoje nápady. Všichni jsou na tvoje návrhy **zvědavý.**
Nemůžu uvěřit, co to na mě vytahuješ. Teď bych tě zlostí fakt **zmlátila**.	Potřebovala bych s tebou něco probrat. Včera jsem ti kvůli tomu **volala**.
Ty zas vypadáš jako bezdomovec. A navíc z tebe táhne strašnej **smrad**.	To tvoje tričko vypadá zajímavě. Má takovej neobvyklej **tvar**.

Kvůli tobě jsou na nás všichni naštvaný. Příště se nechovej jako **pitomec**.	Brzo se s nima mnohem víc poznáš. Budeš pro ně pracovat jako **konzultant.**
Vždyť je to všechno zase úplně špatně. Ty seš opravdu takovej **blbec!**	Někdo by s tebou měl jet jako spolujezdec. Jsi přece docela novej **řidič**.
Nikomu z nich se ani trochu nelíbíš. Určitě proto, jak furt mluvíš **sprostě**!	Klidně tu prezentaci odříkej svým tempem. Nevadí mi, když mluvíš **rychle**.
Vždyť vždycky vypadáš jak nějakej vandrák! Aspoň pro jednou se neoblíkni **vulgárně**!	Večer si vem společenský oblečení. Přes den se ale neoblíkej tak **formálně**.
Pořád se jenom válíš na gauči. Život s tebou je tak **nudnej**!	Klidně bych si s tebou vyměnila polštář. Ten tvůj je oproti mýmu **měkkej**.
Vždycky myslíš jenom na sebe. Jsi fakt neuvěřitelnej **egoista**.	Tady máš svůj sportovní dres. V něm budeš úplnej **fotbalista**.
Pořád si jen na něco stěžuješ. Přestaň být takovej **ufňukánek**.	Ty o světě nemáš žádný iluze. Jsi vlastně docela **realista**.
Nikdo se s tebou nechce kamarádit. Ta tvoje dotěrnost všechny **obtěžuje**.	Zpracováváš to zajímavým způsobem. To tě od ostatních vždycky **odlišuje**.
Přestaň strkat nos do cizích věcí. Tvoje neustálý otázky jsou dost **vlezlý**.	Tvoje výtvory jsou takový specifický. Vždycky jsou oproti ostatním dost **jiný**.
Přestaň z toho dělat takovou tragédii. Hlavně nereaguj takhle **hystericky**.	Večer se sejde celá naše rodina. Nezapomeň se oblíknout **společensky** .

Appendix B: Questionnaire

Dotazník k výzkumu Vnímání emočně zabarvené řeči a slov

ID participanta:

Věk:

Pohlaví:

V experimentu jste slyšel/a dva bloky, které obsahovaly úryvky z konverzací. Na škále 1-5 ohodnoťte, jak moc Vám tyto dva bloky přišly odlišné.

vůbec				velmi
1	2	3	4	5

V čem Vám tyto bloky přišly odlišné? *(odpovězte volně)*

Během experimentu jste slyšel/a komunikaci vyznačující se naštvaným tónem hlasu, nadávkami a sprostými slovy. Pomocí škály 1-5 ohodnoťte, jak často jste takovému stylu komunikace vystavený/á.

vůbec				denně
1	2	3	4	5

Jak často tímto způsobem komunikujete Vy?

vůbec				denně
1	2	3	4	5

Appendix C: Informed consent form

Informace o výzkumu
Vnímání emočně zabarvené řeči a slov

Milá účastnice, milý účastníku,

pro výzkum o vnímání řeči a slov hledáme 20-30 dospělých účastníků, jejichž mateřským jazykem je čeština, splňují další požadavky uvedené v tomto dokumentu a kteří by se mohli dostavit do laboratoře na Psychologickém ústavu AV ČR v Praze. Tento výzkum se zabývá vnímáním emočně zabarvené řeči a slov.

Měření budeme provádět pomocí EEG. Jedná se o neinvazivní a zcela bezpečnou metodu, díky které můžeme pozorovat mozkovou aktivitu. Další informace o tomto výzkumu a o používané metodě najdete v tomto dokumentu.

Výzkumnice
Natálie Kikoťová, členka výzkumné skupiny SPEAKIN lab

Dr. Kateřina Chládková, působí na Ústavu českého jazyka a teorie komunikace FF UK a Psychologickém ústavu AV ČR a vede výzkumnou skupinu SPEAKIN lab

Cíle výzkumu
V našem výzkumu se zaměřujeme na vnímání řeči u dospělých osob. Tento experiment nám pomůže pochopit, jak vnímáme emočně zabarvenou řeč a slova. Primárním cílem je tedy přispět k popisu mechanismů, které v mozku při poslechu emočně zabarvené řeči probíhají.

Instrukce a průběh experimentu

1. Výzkum začne krátkým rozhovorem, při kterém vám mimo jiné vysvětlíme, o čem výzkum je a zodpovíme vaše případné dotazy.
2. Potom vám na hlavu nasadíme tenkou čepici, která k sobě má připevněny EEG sensory, které snímají mozkové signály. Aplikujeme trochu gelu na vodní bázi na povrch elektrod dotýkající se pokožky vaší hlavy, což poslouží ke zlepšení vodivosti signálu. Toto měření je pro vás zcela bezpečné a bezbolestné. Používaný gel je hypoalergenní.
3. Během celého experimentu budete sledovat obrazovku a my vám budeme z reproduktorů pouštět úryvky dialogů. Po každém úryvku uslyšíte slovo a pomocí dvou tlačítek se rozhodnete, jestli dané slovo v předcházejícím úryvku zaznělo, nebo ne. Je důležité, abyste během celého experimentu seděli co nejklidněji. Hlasitost zvuku nebude více než 70 decibelů, což je hlasitost srovnatelná s běžnou řečí.
4. Na konci vám sundáme čepici a vám zůstanou ve vlasech malé zbytky gelu. Budete mít možnost si přímo na stejném patře gel otřít, případně vlasy umýt a vysušit (vše potřebné poskytneme).
5. Celý experiment trvá maximálně hodinu a půl, a to včetně uvítacího rozhovoru, přípravy čepice, přestávek a případného mytí vlasů. Výzkum samotný (tj. vlastní poslech nahrávek řeči) trvá 45 minut.

Dobrovolnost

Pokud budete souhlasit s účastí na tomto experimentu, požádáme vás o podepsání prohlášení, že souhlasíte s účastí na tomto výzkumu a s použitím získaných dat pro naše výzkumné účely. Pokud se kdykoliv během experimentu rozhodnete z jakéhokoliv důvodu svoji účast ukončit, řeknete nám to, a experiment bude ukončen.

Rizika a pojištění

Z předchozí zkušenosti s výzkumem tohoto typu nejsou známa žádná rizika. Protože z tohoto výzkumu neplynou žádná zdravotní ani bezpečnostní rizika, není uzavřeno žádné speciální pojištění.

Důvěryhodnost a data získaná ve výzkumu

V rámci výzkumu budeme sbírat údaje o věku, pohlaví a jazykovém pozadí. U studentů, kteří za účast na výzkumu dostanou kredity, bude navíc sbírán údaj o jméně, a to právě kvůli zapsání kreditů: po zapsání kreditů bude údaj o jméně smazán. U účastníků, kteří obdrží poukázku na nákup, bude na formuláři o převzetí poukazu na nákup uveden údaj o jméně a datu narození účastníka: tyto osobní údaje budou sloužit pouze pro případnou kontrolu čerpání grantových prostředků. U všech účastníků platí, že soubor s experimentálními daty bude od osobních údajů zcela oddělen a nebude je možné zpětně propojit.
Získaná data budou použita jen pro analýzu výsledků a případnou budoucí publikaci ve vědeckých časopisech. Data mohou být v budoucnu sdílena v rámci vědecké komunity. V těchto případech nebudou nikdy použity vaše osobní údaje a bude vždy zachována vaše anonymita.

Místo a datum konání

Výzkum bude probíhat v laboratoři LABELS Psychologického ústavu AV ČR a Filozofické fakulty UK na adrese Voršilská 1, Praha 1, od listopadu 2023 do jara 2024, vždy mezi 8:00 a 21:00.

Odměna

Za účast na našem výzkumu dostanete poukaz na nákup za 300 Kč nebo ECTS kredity v rámci příslušného kurzu na FF UK.

Požadavky

Jako účastnice nebo účastník v našem výzkumu byste měl(a) splňovat následující požadavky:

- jste věku mezi 18 a 45 lety
- vaším jediným mateřským jazykem je čeština
- nemáte sluchové postižení a netrpíte akutním onemocněním ucha
- netrpíte psychiatrickým nebo neurologickým onemocněním
- 24 hodin před účastí na experimentu nepožijete návykové látky (včetně většího množství alkoholu)

Informovaný souhlas – Výzkum o vnímání emočně zabarvené řeči a slov

Tímto prohlašuji, že jsem byl/a náležitě seznámen/a s podstatou a metodou výzkumu, která je popsána v dokumentu Informace o výzkumu Vnímání emočně zabarvené řeči a slov. Všechny dotazy mi byly dostatečně zodpovězeny.

Ve výzkumu budou sbírána data o věku, pohlaví a jazykovém pozadí, a nahráván záznam mozkové aktivity během EEG. U studentů, kteří za účast na výzkumu dostanou kredity, bude navíc sbírán údaj o jméně, a to právě kvůli zapsání kreditů: po zapsání kreditů bude údaj o jméně smazán. Pokud jsem studentem, souhlasím s poskytnutím svého jména.

U účastníků, kteří obdrží poukázku na nákup, bude na formuláři o převzetí poukazu na nákup uveden údaj o jméně a datu narození účastníka: tyto osobní údaje budou sloužit pouze pro případnou kontrolu čerpání grantových prostředků. Pokud za účast na výzkumu obdržím tuto poukázku, souhlasím s poskytnutím svého jména a data narození.

U všech participantů platí, že soubor s experimentálními daty bude od osobních údajů zcela oddělen a nebude možné zpětně spojit experimentální data s osobními údaji účastníka; všechna data tedy budou plně anonymizovaná.

Moje účast na tomto výzkumu je zcela dobrovolná. Ponechávám si právo zrušit svůj souhlas s účastí bez nutnosti udání důvodu. Kdykoli během experimentu mohu proceduru ukončit a odejít. Pokud budou moje naměřená data použita ve vědeckých publikacích, pak jedině zcela anonymně. Souhlasím s tím, že anonymizovaná data mohou být v budoucnu sdílena v rámci vědecké komunity.

Pokud budu chtít další informace o výzkumu, teď nebo v budoucnosti, mohu se obrátit na Natálii Kikoťovou (natalie.kikotova@ff.cuni.cz). S případnými stížnostmi ohledně tohoto výzkumu se mohu obrátit na dr. Kateřinu Chládkovou (chladkova@praha.psu.cas.cz).

Podepsáno ve dvojím vyhotovení:

……	……
Jméno a příjmení účastníka/účastnice	Podpis

Podala jsem informace a vysvětlila průběh výzkumu. Prohlašuji, že zodpovím všechny případné dotazy o výzkumu dle svého nejlepšího vědomí.

……	……
Jméno a příjmení	Podpis