 doi:10.20944/preprints202210.0424.v1

Article

Neurocognitive dynamics of prosodic salience over semantics during explicit and implicit processing of basic emotions in spoken words

Yi Lin [1], Xinran Fan [1], Yueqi Chen [1], Hao Zhang [2], Fei Chen [3], Hui Zhang [1], Hongwei Ding [1,*], Yang Zhang [4,*]

[1] Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
[2] School of Foreign Languages and Literature, Shandong University, Jinan, China
[3] School of Foreign Languages, Hunan University, Changsha, China
[4] Department of Speech-Language-Hearing Science & Center for Neurobehavioral Development, University of Minnesota, US
* Correspondence: hwding@sjtu.edu.cn (H.D.); zhanglab@umn.edu (Y.Z.); Tel.: +86-213420-5664 (H.D.); +1-612-624-7818 (Y.Z.)

Abstract: How language mediates emotional perception and experience is poorly understood. The present event-related potential (ERP) study examined the explicit and implicit processing of emotional speech to differentiate the relative influences of communication channel, emotion category and task type in the prosodic salience effect. Thirty participants (15 women) were presented with spoken words denoting happiness, sadness and neutrality in either the prosodic or semantic channel. They were asked to judge the emotional content (explicit task) and speakers' gender (implicit task) of the stimuli. Results indicated that emotional prosody (relative to semantics) triggered larger N100 and P200 amplitudes with greater delta, theta and alpha inter-trial phase coherence (ITPC) values in the corresponding early time windows, and continued to produce larger LPC amplitudes and faster responses during late stages of higher-order cognitive processing. The relative salience of prosodic and semantics was modulated by emotion and task, though such modulatory effects varied across different processing stages. The prosodic salience effect was reduced for sadness processing and in the implicit task during early auditory processing and decision-making but reduced for happiness processing in the explicit task during conscious emotion processing. Additionally, across-trial synchronization of delta, theta and alpha bands predicted the ERP components with higher ITPC values significantly associated with stronger N100, P200 and LPC enhancement. These findings reveal the neurocognitive dynamics of emotional speech processing with prosodic salience tied to stage-dependent emotion- and task-specific effects, which can reveal insights to research reconciling language and emotion processing from cross-linguistic/cultural and clinical perspectives.

Keywords: emotional speech processing; communication channel; emotion category; task type

1. Introduction

1.1 Sensory dominance effects: theoretical importance and methodological concerns

Emotion plays an essential role in successful interpersonal communication. Humans show how they feel through what they say (i.e., linguistic content) and how they say it (i.e., paralinguistic information). One important theoretical contention centering around multisensory emotional speech processing is whether a certain sensory channel is more perceptually dominant over others, which is referred to as the channel (sensory) dominance effect [1,2]. A focus on channel dominance, especially the role of prosody, in emotional speech processing is crucial for understanding the developmental trajectory and functional impairments of speech, language and hearing abilities. Existing research showed that in early development, infants are more sensitive to the prosodic aspects of

 doi:10.20944/preprints202210.0424.v1

speech which serves as the early language input providing socioaffective foundation for language acquisition [3,4]. For individuals with typical language skills, prosody is a salient part of multisensory speech communication [1,2]. In aging, emotional prosody is also difficult for individuals with hearing loss and cognitive decline [5-7]. Various clinical populations struggle with emotional speech processing, including patients with schizophrenia and autism [8,9].

While some studies observed the predominance of auditory prosodic cues over verbal content in emotional speech perception [10,11], there is also evidence pointing to a perceptual bias towards semantic content [2,12]. These empirical discrepancies in behavioral investigations may be related to differences in language and cultural background across studies. Given the cross-linguistic differences and socio-cultural nature of decoding and encoding emotions, what is considered a normal pitch or rhythm in a tonal language (e.g., Mandarin Chinese) may be considered excessive in a non-tonal language (e.g., Italian) and vice versa [13]. Notably, those studies supporting a semantic dominance effect are largely based on data collected in western countries (e.g., Germany and Canada) with a non-tonal language background and a low-context culture [14], in which interlocutors tend to rely heavily on verbal messages during speech communication. It remains to be tested to what extent the existing findings can be generalized to a different socio-contextual background, such as a high-context culture, where nonverbal information and interpersonal relationships are more important [15]. For studies investigating the neural underpinnings of emotional semantics and prosody processing, extensive efforts have been made to specify the related brain structures using functional neuroimaging [16-20]. Relatively fewer studies have explored the underlying time course using neurophysiological techniques with fine temporal resolution (e.g., electroencephalogram) [21].

Though conventional ERP waveform analysis can shed light on the event-locked regularities of brain dynamics based on time-domain information averaged across trials, it may underestimate trial-by-trial response variability in the time-frequency domain [22-24]. A line of studies have applied time-frequency analyses to explore the time-locked neural substrates of auditory processing [23,25-28], though these investigations were often conducted with non-emotional stimuli. In these studies, evoked neural synchrony can be evaluated through inter-trial phase coherence (ITPC) in five frequency bands, including delta (1–4 Hz), theta (4–8 Hz), alpha (8– 12 Hz), beta (12–30 Hz) and gamma (over 30 Hz). Higher ITPC values suggest better phase alignment of cortical oscillations, while smaller values indicate poorer consistency or larger neural "jittering" across trials [29]. Results suggested that stimulus-evoked phase alignment of EEG oscillations, especially delta, theta and alpha ITPC, forms a crucial basis for the neural generation of auditory ERP [23,28,30]. By contrast, time-frequency analyses of vocal emotion processing are sparse with even less attention on the relationship between ERP waveforms and neural oscillations [31,32]. The combination of ERP waveform and time-frequency analyses in the current study may provide meaningful insights into the underlying neural mechanisms of emotional speech processing.

In light of the theoretical and methodological issues, the primary focus of this work is to examine the temporal dynamics of emotional speech processing using the event-related potential (ERP) measure with waveform and time-frequency analyses. Importantly, we strived to characterize the neurobehavioral representations of channel dominance effects with consideration of emotional category and task type, which can contribute to the understanding of existing discrepancies in previous literature. Since we based our study on a Mandarin Chinese context, the tonal language background enabled us to investigate how pitch variations denoting lexical meaning alone are processed differently from those communicating emotional and linguistic meaning simultaneously at early and late stages. The high-context East Asian setting also allowed for a new cultural perspective on the neurobehavioral distinctions of verbal and nonverbal processing.

1.2 Effects of communication channel on multi-stage processing of emotional speech

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2022 doi:10.20944/preprints202210.0424.v1

Decoding emotional information in speech occurs rapidly, involving a multilayered process that contains temporally and functionally distinct processing stages [16,33,34]. According to Schirmer and Kotz [35], there are three stages for emotional speech processing: (1) analyzing the acoustic features in vocalizations, (2) deriving the emotional salience from a set of acoustic signals, and (3) integrating emotional significance to higher-order cognitive processes. The first two stages have largely been studied with the N100 and P200 components using the ERP technique, and the third stage can be probed with the late positive component (LPC) as well as behavioral measures [21,36-40]. However, it remains unclear how the relative salience of semantic versus prosodic channels unfolds across the different emotional speech processing stages.

There has been divided attention in the literature on prosodic and semantic aspects of emotional speech processing. For instance, the 3-stage model by Schirmer and Kotz [35] characterizes the prosodic aspect of processing for emotional speech, and many studies supporting the model focused on emotional prosody by employing non-linguistic affective vocalizations or pseudo-words/sentences [34,36,41-43]. Some studies applied a cross-splicing paradigm to temporally control when prosodic cues became available to the listener by artificially introducing discrepancies between verbal and nonverbal messages [34,38,44]. Likewise, ERP studies on semantic processing of emotional words often chose the visual modality for stimulus presentation without considering emotional prosody in speech [45-49].

Some limitations in the existing research may have prevented us from gaining a comprehensive understanding of the relationship between the two speech channels. One previous ERP investigation substantiated the predominance of semantics over prosody during deviance detection in emotional contexts [34]. However, since the effect was observed based on sentence-level stimuli, its generalizability to other linguistic representations (e.g., word) warrants further examination. It also remains to be tested whether the effect occurs based on semantic mismatch alone or depends on integrative semantic and prosodic processing. In addition, the speech stimuli especially those with unintelligible semantic content are somewhat disassociated from what we are usually faced with in daily communication. Recent behavioral studies attempted to address the joint multi-sensory multi-channel processing of emotional speech, but the behavioral data (including accuracy and reaction time) could not easily separate the final decision-making stage from the earlier processing stages [2,50,51].

1.3 Effects of emotion category on emotional speech processing

In addition to the relative salience of the communication channels, emotional speech processing is subject to a number of influential factors. One key issue is whether emotional and non-emotional signals can be distinguished from each other automatically at an early stage and if so, exactly when they start to be differentiated. There is cumulative evidence that emotional stimuli elicited larger auditory ERP responses and greater neural synchronization (esp. in the delta and theta band) than neutral stimuli [34,35,52,53]. This can be explained by the evolutionary significance of affective signals, which leads to increased automatic attentional capture and prioritized processing strategies relative to neutral stimuli [34,54]. However, findings are mixed concerning how early the significant differentiation occurs. The processing of emotional speech is generally thought to diverge from that of neutral speech around 200 milliseconds (ms) post stimulus presentation [21,35,55,56], but there is also evidence indicating the distinction as early as 100 ms [41].

A second issue is how different categories of emotion in speech are distinguished from one another. According to the differential emotion theory, a set of emotions (e.g., joy, interest, sadness, anger, fear, disgust) are distinguishable in neurochemical processes, expressive behaviors and subjective experiences [57]. These discrete emotions can also be described in a two-dimensional space with regard to their valence and arousal. Empirical evidence has shown how the two dimensions can influence emotion perception at different processing stages. For example, Paulmann, Bleichner and Kotz [43] found that valence-relevant information can be reliably deciphered at both early and late processing

 doi:10.20944/preprints202210.0424.v1

stages, while arousal is more robustly decoded during the late processing stage. Although there tends to be perceptual bias towards positive and high-arousing stimuli, these valence- and arousal-dependent processing patterns have not been conclusively established [58,59]. Some studies have shown valence- and arousal- independent emotion processing [34,60]. Notably, neurophysiological studies on emotional speech processing have generally taken valence attributes into account in stimulus design while disregarding the possible role of arousal. One example is that happiness and anger are often chosen as the two contrasting emotions [36,37], but both of them are high arousing emotions despite a distinction in valence. Thus, the relative influences of valence and arousal on emotional speech processing need to be further investigated with the inclusion of more emotional categories.

1.4 Effects of task type on emotional speech processing

A third factor is the experimental task. Task focuses can be changed under different types of tasks. In explicit emotion processing tasks, participants are required to evaluate the emotional content (e.g., valence and arousal attributes) of the stimuli. By contrast, attention in implicit tasks is diverted from the emotional attributes of the stimuli and focused on other informational dimensions [61]. Differentiated effects of attention have been found on several ERP components, with increased attention evoking enhanced N100 but diminished P200 amplitudes [62-64]. Early and late processing of emotional speech can also be modulated by task difficulty/cognitive efforts. Increased task complexity leads to enhanced early auditory ERP responses (e.g., more negative N100, more positive P200) and neural synchrony [37,65-67] but reduced brain responses and poorer behavioral performances in the post-perceptual processing stage [36,68,69]. Though some studies indicated that task types can modulate modality- (e.g., visual vs. auditory) or category-specific emotion processing [2,70-72], this is not always the case probably due to varying task requirements [43]. To what extent the observed effects of channel and emotion in speech processing can be generalized across different task types warrants further examination.

1.5 The present study

The present study aimed to examine the neurobehavioral effects of communication channel, emotional category and task type as emotional speech processing unfolded in time. Two basic emotions (i.e., happiness and sadness) and neutrality [73] were tested, and these emotional categories can be distinguished from one another on both valence and arousal scales. Emotional information was conveyed through either the prosodic or semantic channel, which constituted two types of experimental stimuli, namely semantically neutral words spoken in emotional intonations and emotional words spoken in neutral prosody. Participants were asked to identify these emotional stimuli in explicit (i.e., emotion identification tasks) and implicit (i.e., gender identification tasks) conditions. We measured N100, P200, LPC and their associated cortical oscillatory activities to characterize sensory processing of acoustic signals, initial decoding of emotional significance, and early stages of cognitive evaluation. Delta, theta and alpha ITPC were selected for evaluation as these frequency band oscillations could reflect salience detection, emotional significance and attentional modulation [53], and could better predict auditory ERP responses [23,28,30]. We also recorded accuracy and reaction time data from stimulus offset to show emotional speech processing in the decision-making stage.

Based on previous studies revealing the effects of channel, emotion and task on emotional speech processing and the relationships among different neurological and behavioral measures, we developed the following hypotheses:

- First, we expected to find ERP and behavioral differentiation of emotional prosody and semantics given the channel (prosodic) dominance effects observed in our recent studies based on a tonal language and high-context culture [2,51].
- Second, we predicted that emotional stimuli would be distinguished from the neutral ones [34,54], and differences would also be found between specific emotion types (i.e., happy and sad) [43].

 doi:10.20944/preprints202210.0424.v1

- Third, task types would modulate brain and behavioral responses during emotional speech processing, since our task instructions would lead to differences in task focuses and difficulty [68,71].
- Finally, we hypothesized that ITPC data could be potential indicators of auditory ERP responses [23,28,30]. However, processing patterns were likely to vary across the neurophysiological and behavioral indices since the adopted measures were not conceptually equivalent [22,23].

Findings from the present study will contribute new data to the multi-stage model of emotional speech processing and reveal insights to research on emotion cognition from cross-linguistic/cultural and clinical perspectives.

2. Materials and Methods

2.1 Participants

The present study was conducted with approval from the institutional review board (IRB) of School of Foreign Languages at Shanghai Jiao Tong University (SJTU). Thirty volunteers (15 females and 15 males) were recruited to take part in this experiment through an online campus advertisement. Participants averaged 23.1 (SD = 2.2) years in age and had received an average of 16.6 (SD = 2.2) years of formal school education. All participants were native speakers of Mandarin Chinese with no medical history of speech, language and hearing disorders or neurological problems. All had normal or corrected-to-normal vision and normal hearing in standard audiometric assessment (≤ 20 dB HL for 0.25-, 0.5-, 1-, 2-, 4-, and 8-kHz pure tones) [74]. All were studying at SJTU as undergraduate or graduate students at the time of testing and were non-musicians without formal musical training in the past five years and less than two years of musical training prior to that [75]. Written informed consent was obtained from all participants, who were paid for their time and involvement.

2.2 Stimuli

The stimuli contained two sets of disyllabic words in Mandarin Chinese spoken by a female and a male professional speaker. Each auditory stimulus conveyed one of the two basic emotions (i.e., happiness and sadness) [73] or neutrality in either the prosodic or semantic channel. There were altogether 180 spoken words in each stimulus set/communication channel, in which the number of words was balanced between the two speakers (i.e.,90 words for each speaker), and among the three emotional categories (i.e.,60 words for each emotion). Specifically, for *the prosodic set*, 60 semantically neutral concrete nouns were spoken in happy, neutral and sad prosody respectively. For *the semantic set*, words were spoken in a neutral tone of voice and conveyed emotional information in verbal content, including 60 adjectives with happy semantics, 60 with sad semantics, and 60 with neutral semantics. Most words and their frequencies were taken from *A Dictionary of the Frequency of Commonly Used Modern Chinese Words (Alphabetical sequence section)* [76]. The semantic word set had higher word frequency than the prosodic set ($t(394) = -3.67$, $p < .001$). See Supplemental Tables S1 and S2 for the list of included words for prosodic and semantic stimuli, respectively. All auditory stimuli were normalized in intensity (at 70 dB) using Praat (version 6.1.41) [77]. The duration and mean f0 measures of the prosodic and semantic stimuli are summarized in Tables S3 and S4 in *Supplemental Materials* respectively. The spectral images of the auditory stimuli are illustrated in Figure S1 in *Supplemental Materials.*

The stimuli were uttered by two native speakers (one woman and one man) of Mandarin Chinese in a quiet laboratory setting, and digitized onto a Macbook Pro computer with AVID Mbox Mini at a sampling rate of 44,100 kHz with a 16-bit resolution. Each word was portrayed three times by the two speakers, and the best ones were selected according to the results of a norming study. In the norming test, forty adult native speakers of Mandarin Chinese (20 women and 20 men, Mean $_{age}$ = 23.0 , SD = 3.4) who did not participate in the current research were invited to perceptually validate the experimental

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2022 doi:10.20944/preprints202210.0424.v1

stimuli using Praat [77]. These raters were randomly assigned to one of the two gender-balanced groups (20 raters, 10 women in each group). One group of subjects were asked to rate the word familiarity on a 7-point Likert scale (1 = not familiar, 7 = very familiar) and identify the emotional category of each prosodic and semantic stimulus. The other group of subjects were asked to rate the emotional arousal of each stimulus on a 7-point Likert scale (1 = low, 7 = high). Only words with an average rating of >5 for familiarity and over 85% identification accuracy for emotional categories were included in the present experiment. The mean familiarity rating, identification accuracy and emotional arousal of the finally included word stimuli are shown in Tables S5 and S6 in *Supplemental Materials*. The familiarity rating did not differ between the prosodic and semantic word sets and no significant difference was found in accuracy and arousal for words in the same emotion category between the two channels (all $p > .05$).

2.3 Procedure

During the electroencephalograph (EEG) recording session, participants were seated comfortably at a distance of 1.15 m from a 19-inch LCD computer monitor in a soundproof booth. The raw EEG was recorded with 64 Ag-AgCl electrodes attached to an elastic cap at the sampling rate of 1000 Hz by the NeuroScan system (Compumetics NeuroScan®, Victoria, Australia). All electrodes were placed according to the International 10-20 electrode placement standard with a ground electrode located at the AFz electrode, and the recording reference placed between Cz and CPz. Four bipolar facial electrodes were placed above and below the left eye and outer canthi of both the eyes to monitor vertical and horizontal eye movements (EOG channels) and two electrodes were placed on two mastoids to be used offline for re-referencing. Electrode impedances were kept at or below 8 kΩ throughout the recording.

The EEG experiment was divided into two sessions (explicit or implicit). Each session contained two blocks (prosodic or semantic). In each block, 180 spoken words of different emotional prosody or semantics (60 happy, 60 sad, 60 neutral) were presented binaurally through E-A-R TONE™ 3A Insert Earphone at 70 dB SPL. For explicit emotion perception, participants were instructed to attend to the emotional information of the stimuli. They indicated whether a word was spoken with a happy, neutral or sad tone of voice (prosodic block), and whether a word conveyed happy, neutral or sad semantic content by pressing one of the three buttons (semantic block). For implicit emotion perception, participants were instructed to attend to the gender of the speaker while ignoring the emotional information of the words. They indicated whether the word was spoken by a male or female speaker by pressing one of the two buttons in both prosodic and semantic blocks. E-prime (version 2.0.10) was used for stimulus presentation [78]. The presentation order of the session, block and button press was counterbalanced across participants.

Before each experimental block, participants were given a 10-trial training session and entered the experiment with at least 80% identification accuracy. There were 180 trials in each block. Each trial started with a fixation cross presented centrally on the screen for 1000 ms. The words were then presented auditorily, during which the fixation cross remained on the screen to minimize eye movements. Afterwards, a question mark was presented, which signaled the beginning of response. The words were presented in a pseudo-randomized manner. To reduce baseline artifacts, a variable inter-trial interval of 800-1000 ms occurred before the next trial began. A short pause of 10 seconds was provided after every 20 trials. There was a 2-minute break between the two blocks in each session, and there was a 5-minute break between the two sessions. The total duration of the experiment was approximately 60 minutes. During the experiment, behavioral (i.e., accuracy, reaction time) and electrophysiological data were recorded. The schematic illustration of the experimental protocol is presented in Figure 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2022 doi:10.20944/preprints202210.0424.v1

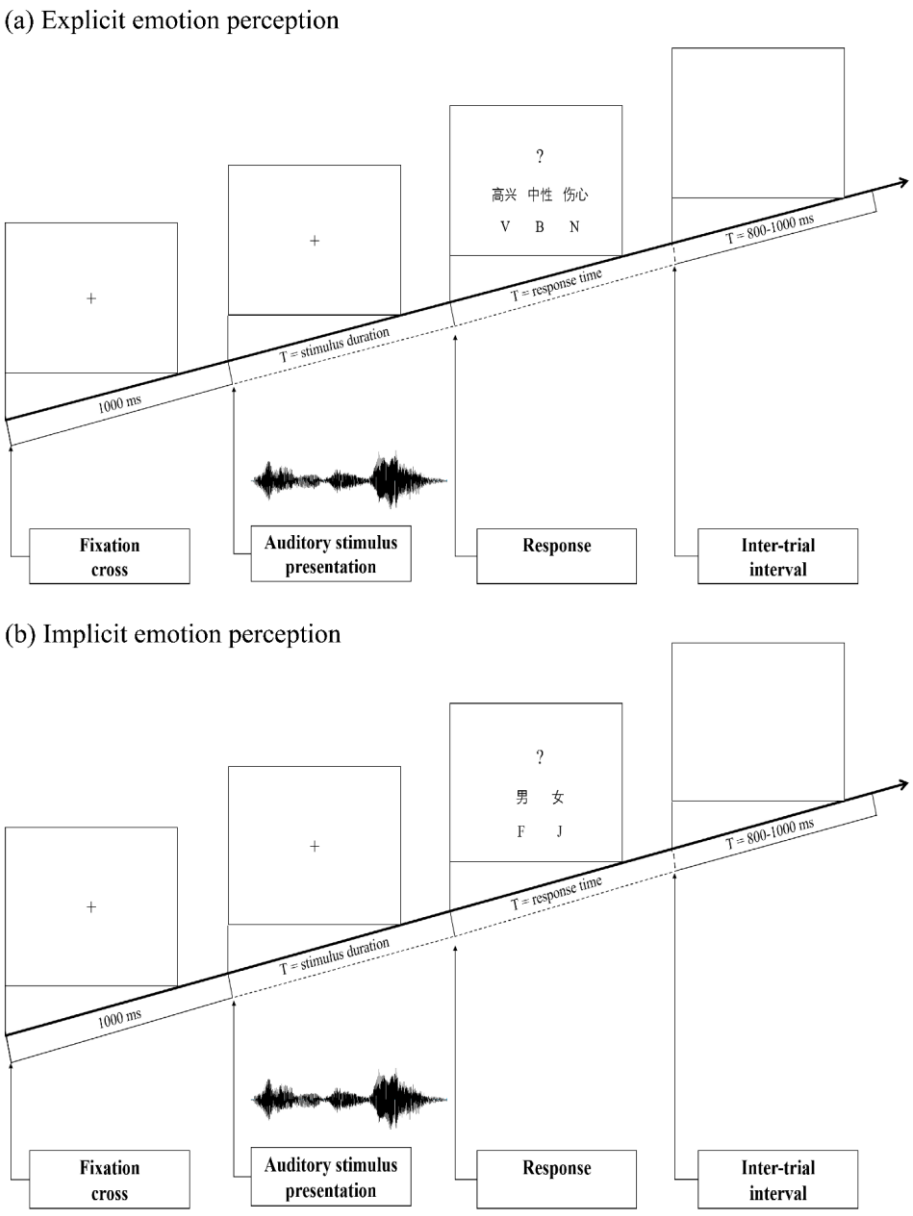

Figure 1. Schematic illustration of the experimental protocol for (a) explicit and (b) implicit emotion perception tasks.

2.4 Data analysis

ERP data analysis. EEG data processing was performed with Matlab-based (Version: R2016a) EEGLAB (Version: 14.1.2) and ERPLAB (Version: 7.0) toolboxes. Only trials with correct behavioral responses were included in the ERP waveform and time-frequency (TF) analysis. The raw EEG data were down-sampled to 250 Hz. Eye blinks and muscle movements were identified and removed using Independent Component Analysis (ICA) algorithm following the guidelines by Chaumon, *et al.* [79]. Artifact detection was performed according to the following criteria: (i) the maximally allowed amplitude difference for all EEG channels within a moving window (width: 200 ms; step: 50 ms) should not exceed ± 30 μV; (ii) the maximally allowed absolute amplitude for all EEG channels throughout the

 doi:10.20944/preprints202210.0424.v1

whole epoch should not exceed ± 100 μV. After excluding trials with incorrect responses and rejecting artifact-contaminated trials, the overall data retention rate was 95.1%. The data were re-referenced to the algebraic average of the two mastoid electrodes.

For the auditory ERP analysis, the EEG data were band-passed at 0.1-40 Hz, and were segmented into time-based epochs of 1200 ms, which consisted of a 200 ms pre-stimulus interval for baseline correction and a 1000 ms post-stimulus interval. Grand average ERP waveforms (Figure 2) were computed for each emotion (happy, neutral and sad) in each channel (semantic vs. prosodic) under each task (explicit vs. implicit). Four time windows were chosen for analyses based on previous literature and visual inspection of the grand mean auditory ERP data (i.e., N100: 65-170 ms; P200: 150-300 ms; LPC:500-900 ms) [36-38,40,41,70]. Since maximal effects were observed at the fronto-central and central sites, we selected six electrodes (FC3, FCz, FC4, C3, Cz, C4) for statistical analyses, which was consistent with previous reports [36,37,41,80]. The amplitude data were quantified by averaging data points within the time window of 40 ms around the peak of the components for each condition.

For the TF analysis, inter-trial phase coherence (ITPC) in delta (1-3.9 Hz), theta (4-7.9 Hz) and alpha (8-11.9 Hz) frequency bands at electrode Cz was computed using the "newtimef" function with the open-source EEGLAB package [81]. ITPC estimates the trial-by-trial synchronization as a function of time and frequency, the value of which in a given frequency band can range from 0 to 1. Larger ITPC values indicate better trial-by-trial synchronization, and smaller values suggest lower consistency or larger neural "jittering" across trials. A modified short-term Fourier Transform (STFT) with Hanning window tapering was implemented to extract the ITPC values for the delta, theta, and alpha frequency bands, which is recommended for the analysis of low-frequency activities. Zero-padding was applied to short epochs that did not have sufficient number of sample points with a padratio of 16 for Fourier transform. Frequencies for ITPC calculation ranged from 0.5 to 50 Hz with a step interval of 0.5 Hz. An epoch window of 1800 ms with an 800 ms pre-stimulus baseline was used. The maximum ITPC values in the designated time windows of N100 (65-170 ms), P200 (150-300 ms) and LPC (500-900 ms) were identified per participant for each emotion category in each channel under each task for statistical analyses.

Statistical analyses of the event-related potential and TF data were conducted using linear mixed-effect (LME) models in R (version 4.0.3) [82]. For the waveform analysis, N100, P200 and LPC amplitudes were analyzed as dependent variables respectively. For the TF analysis, the delta, theta and alpha ITPC in the corresponding time windows of the two components were entered as dependent variables respectively. Within-subject factors included communication channel (semantic and prosodic), emotion category (happy, neutral and sad), and task type (explicit and implicit). The semantic channel, the sad emotion, and the implicit task were set as the baseline level for communication channel, emotion category, and task type respectively. When happy stimuli were compared with the neutral ones, neutrality was set as the baseline. Subject was included as a random factor for intercepts. In case of significant main effects or interactions, Tukey's post hoc tests were carried out with the emmeans package [83]. Additionally, to examine the relationship between the auditory ERP and TF measures, LME models with ITPC values as predictor variables were fit for N100, P200 and LPC amplitudes. Delta, theta and alpha ITPC were as entered as fixed effects, and subject was entered as a random effect for intercept. Two-tailed significance level with $\alpha = .05$ was used for all statistical analyses throughout the study. The full model with intercepts, coefficients, and error terms for the analysis of each neurophysiological index is shown in *Supplemental Materials*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 October 2022 doi:10.20944/preprints202210.0424.v1

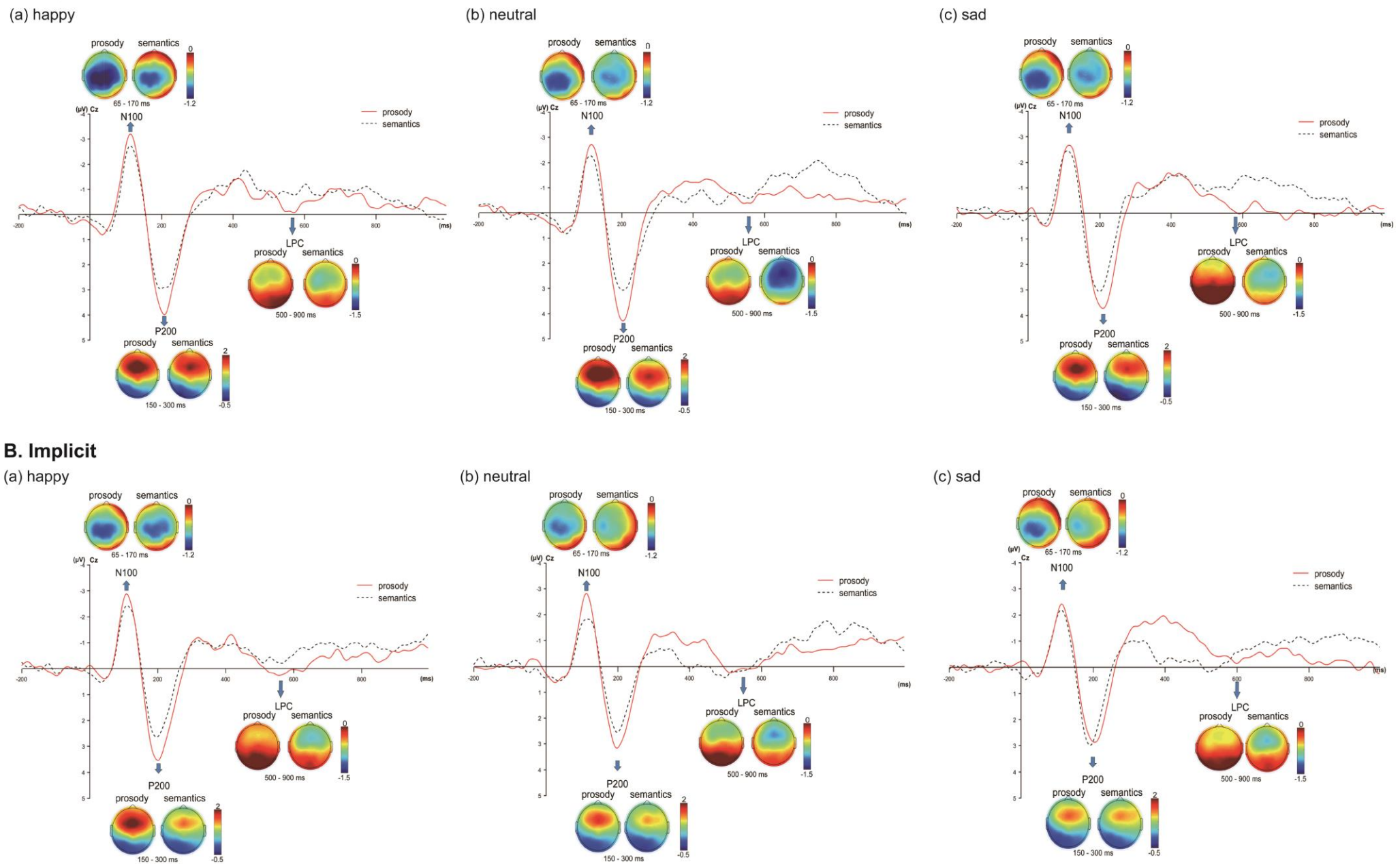

Figure 2. Grand averaged waveforms at Cz and topographical maps of mean amplitude in the N100, P200 and LPC windows for (a) happy, (b) neutral and (c) sad stimuli in prosodic and semantic channels across (A) explicit and (B) implicit tasks.

 doi:10.20944/preprints202210.0424.v1

Behavioral data analysis. A three-way multivariate analysis of variance (MANOVA) was conducted in R (version 4.0.3) [82] to investigate the statistical significance of communication channel (prosodic or semantic), emotion category (happy, neutral or sad) and task type (explicit or implicit) on identification accuracy and reaction time. The semantic channel, the sad emotion, and the implicit task were set as the baseline level for communication channel, emotion category, and task type respectively. When happy stimuli were compared with the neutral ones, neutrality was set as the baseline. To test the MANOVA assumption, we first carried out a Pearson correlation test, which suggested that the two outcome variables (i.e., accuracy and reaction time) were correlated (r = -.25, p < .001). Then the two behavioral measures were entered as dependent variables in MANOVA with Pillai's trace statistics reported. For any significant differences in the MANOVA results, we followed up the analysis with univariate analyses of variance (ANOVA). Pairwise comparisons with Tukey adjustment on p value were conducted with the emmeans package [83] in case of a significant main effect or interaction in the univariate analyses of each individual outcome measure.

3. Results

3.1 Auditory event-related potential measures

The mean and standard deviation of N100, P200 and LPC amplitudes (μV) elicited by happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks are demonstrated in Table 1 and illustrated in Figure 3.

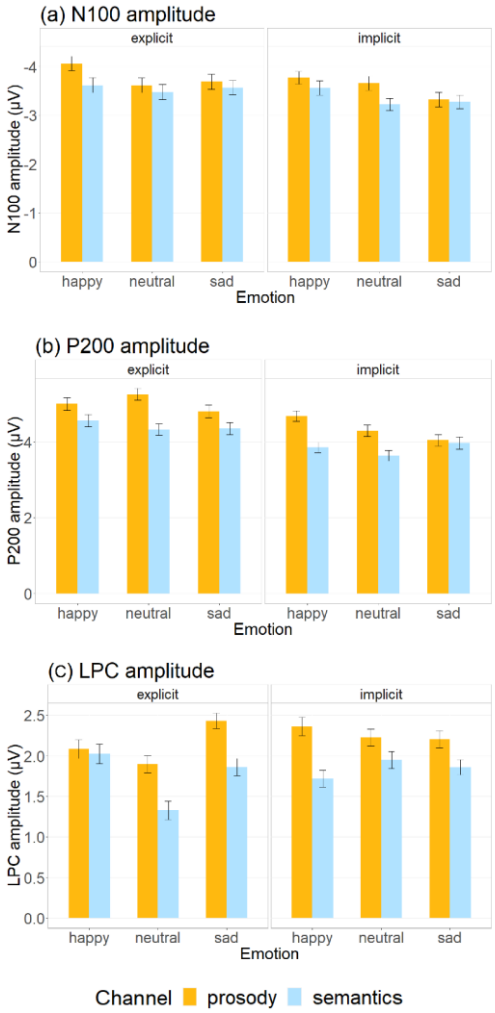

Figure 3. Bar plots of auditory ERP amplitude of (a) N100, (b) P200 and (c) LPC for happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks. Mean amplitude is displayed in the bar charts with error bars showing 95% confidence intervals.

N100. LME analyses on N100 amplitudes revealed main effects of channel (χ^2 (1) = 58.58, $p < .001$), emotion (χ^2 (2) = 72.23, $p < .001$), and task (χ^2 (1) = 43.63, $p < .001$). Post-hoc multiple-comparison tests suggested that larger N100 amplitudes were observed for emotional prosody than emotional semantics ($\hat{\beta}$ = -.23, *SE* = .03, *t* = -7.67, $p < .001$, *d* = -.18), and for explicit tasks than the implicit ones ($\hat{\beta}$ = -.20, *SE* = .03, *t* = -6.62, $p < .001$, *d* = -.16). N100 was also increased for happy stimuli relative to the neutral ($\hat{\beta}$ = -.26, *SE* = .04, *t* = -6.95, $p < .001$, *d* = -.20) and sad ($\hat{\beta}$ = -.29, *SE* = .04, *t* = -7.74, $p < .001$, *d* = -.22) ones, while there was no significant difference between neutral and sad stimuli (p =.711). Significant interactions between channel and emotion (χ^2 (2) = 12.65, p = .002) and between emotion and task (χ^2 (2) = 9.33, p = .009) were found. More importantly, there was a three-way interaction among channel, emotion and task (χ^2 (2) = 13.05, p = .001). Increased N100 was found in emotional prosody compared with emotional semantics in all three emotional categories regardless of task types, but this trend was significant or marginally significant only for happy (explicit tasks: $\hat{\beta}$ = -.44, *SE* = .07, *t* = -5.99, $p < .001$, *d* = -.35; implicit tasks: $\hat{\beta}$ = -.21, *SE* = .07, *t* = -2.87, p = .004, *d* = -.17) and neutral (explicit tasks: $\hat{\beta}$ = -.14, *SE* = .07, *t* = -1.91, p = .056, *d* = -.11; implicit tasks: $\hat{\beta}$ = -.43, *SE* = .07, *t* = -5.83, $p < .001$, *d* = -.34) stimuli but not for the sad (explicit: p = .106; implicit: p = .550) ones.

P200. LME analyses on P200 amplitudes showed main effects of channel (χ^2 (1) = 267.71, $p < .001$), emotion (χ^2 (2) = 29.81, $p < .001$), and task (χ^2 (1) = 324.60, $p < .001$). Post-hoc multiple-comparison tests suggested that larger P200 amplitudes were observed for emotional prosody than emotional semantics ($\hat{\beta}$ = .57, *SE* = .03, *t* = 16.51, $p < .001$, *d* = .39), and for explicit tasks than the implicit ones ($\hat{\beta}$ = .64, *SE* = .04, *t* = 18.22, $p < .001$, *d* = .43). P200 was also increased for happy stimuli relative to the neutral ($\hat{\beta}$ = .15, *SE* = .04, *t* = 3.43, p = .002, *d* = .10) and sad ($\hat{\beta}$ = .23, *SE* = .04, *t* = 5.4, $p < .001$, *d* = .16) ones, while there was no significant difference between neutral and sad stimuli (p =.119). Significant interactions between channel and emotion (χ^2 (2) = 42.86, $p < .001$) and between emotion and task (χ^2 (2) = 15.49, $p < .001$) were found. More importantly, we observed a three-way interaction among channel, emotion and task (χ^2 (2) = 24.45, $p < .001$). Increased P200 was found in emotional prosody compared with emotional semantics in all three emotional categories regardless of task types, but this trend was significant for happy (explicit tasks: $\hat{\beta}$ = .44, *SE* = .08, *t* = 5.30, $p < .001$, *d* = .31; implicit tasks: $\hat{\beta}$ = .83, *SE* = .08, *t* = 9.96, $p < .001$, *d* = .58) and neutral (explicit tasks: $\hat{\beta}$ = .93, *SE* = .08, *t* = 11.17 $p < .001$, *d* = .65; implicit tasks: $\hat{\beta}$ = .66, *SE* = .08, *t* = 7.92, $p < .001$, *d* = .46) stimuli. For sad stimuli, P200 amplitudes were significantly larger in the prosodic channel (relative to the semantic one) in explicit tasks, and displayed a non-significant increasing trend in implicit tasks (explicit tasks: $\hat{\beta}$ = .45, *SE* = .08, *t* = 5.37, $p < .001$, *d* = .31; implicit tasks: p = .347).

LPC. LME analyses on LPC amplitudes showed main effects of channel (χ^2 (1) = 242.33, $p < .001$), emotion (χ^2 (2) = 61.53, $p < .001$), and task (χ^2 (1) = 18.60, $p < .001$). Post-hoc analyses showed that larger LPC amplitudes were observed for emotional prosody

than emotional semantics ($\hat{\beta}$ = .41, *SE* = .03, *t* = 15.70, $p < .001$, *d* = .37), and for the implicit task than the explicit one ($\hat{\beta}$ = -.12, *SE* = .03, *t* = -4.32, $p < .001$, *d* = -.10). LPC was more positive for happy ($\hat{\beta}$ = .20, *SE* = .03, *t* = 6.09, $p < .001$, *d* = .18) and sad ($\hat{\beta}$ = -.24, *SE* = .03, *t* = -7.35, $p < .001$, *d* = -.21) relative to neutral stimuli, while no significant difference was found for happy and sad stimuli (p = .421). There was a significant interaction between task and emotion (χ^2 (2) = 97.46, $p < .001$), and more importantly, a significant interaction among channel, emotion and task (χ^2 (2) = 58.30, $p < .001$). Prosody elicited more positive LPC amplitudes than semantics for all emotional categories in both tasks, but the effect was non-significant for happy stimuli in explicit tasks (p = .331).

Table 1. Mean amplitude (μV) of N100, P200 and LPC elicited by happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks.

Measure	Task / Channel / Emotion	Explicit				Implicit			
		Prosody (Mean/SD)		Semantics (Mean/SD)		Prosody (Mean/SD)		Semantics (Mean/SD)	
N100	Happy	-4.06	1.79	-3.62	1.93	-3.77	1.63	-3.56	1.80
	Neutral	-3.62	1.89	-3.47	1.91	-3.66	1.74	-3.22	1.57
	Sad	-3.69	1.92	-3.57	1.82	-3.32	1.90	-3.28	1.74
P200	Happy	5.00	2.06	4.56	2.02	4.68	1.79	3.85	1.77
	Neutral	5.25	1.99	4.32	1.88	4.29	1.91	3.63	1.76
	Sad	4.80	2.07	4.35	1.95	4.04	1.94	3.96	2.03
LPC	Happy	2.09	1.45	2.02	1.53	2.36	1.41	1.72	1.29
	Neutral	1.90	1.32	1.33	1.40	2.23	1.30	1.95	1.30
	Sad	2.43	1.18	1.86	1.31	2.20	1.33	1.86	1.16

3.2 Inter-trial phase coherence measures

Figure 4 shows the time-frequency representations of trial-to-trial phase-locking measured by ITPC for happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks. The mean and standard deviation of delta, theta, and alpha ITPC values associated with N100, P200 and LPC amplitudes are summarized in Table 2 and illustrated in Figure 5.

N100. In the N100 window, LME analyses on delta and theta ITPC showed main effects of channel (delta: χ^2 (1) = 22.07, p < .001; theta: χ^2 (1) = 24.67, p < .001), emotion (delta: χ^2 (2) = 9.64, p = .008; theta: χ^2 (2) = 10.65, p = .005) and task (delta: χ^2 (1) = 20.05, p < .001; theta: χ^2 (1) = 19.87, p < .001). Delta and theta ITPC values were larger in the explicit task than the implicit one (delta: $\hat{\beta}$ = .03, SE = .007, t = 4.54, p < .001, d = .48; theta: $\hat{\beta}$ = .03, SE = .006, t = 4.52, p < .001, d = .48), and in the prosodic channel than the semantic one (delta: $\hat{\beta}$ = .03, SE = .007, t = 4.75, p < .001, d = .50; theta: $\hat{\beta}$ = .02, SE = .006, t = 5.03, p < .001, d = .53). Happy stimuli produced greater delta and theta ITPC than the neutral (delta: $\hat{\beta}$ = .02, SE = .008, t = 2.85, p = .013, d = .37; theta: $\hat{\beta}$ = .02, SE = .007, t = 3.14, p = .005, d = .40) and sad (delta: $\hat{\beta}$ = .02, SE = .008, t = 2.51, p = .034, d = .32; theta: $\hat{\beta}$ = .02, SE = .007, t = 2.39, p = .046, d = .31) ones, while no significant difference was found for neutral and sad stimuli (p = .736). Analyses on alpha ITPC suggested main effects of channel (χ^2 (1) = 4.00, p = .046) and task (χ^2 (1) = 10.00, p = .001). Alpha ITPC values were greater in the prosodic than the semantic channel ($\hat{\beta}$ = .01, SE = .005, t = 1.99, p = .047, d = .21), and in the explicit than the implicit task ($\hat{\beta}$ = .02, SE = .005, t = 3.18, p = .002, d = .34). *P200.* In the P200 window, LME analyses on delta ITPC exhibited main effects of channel (χ^2 (1) = 45.06, p < .001), emotion (χ^2 (2) = 7.86, p = .020) and task (χ^2 (1) = 13.17, p < .001). There was increased delta ITPC in the prosodic than the semantic channel ($\hat{\beta}$ = .04, SE = .006, t = 6.91, p < .001, d = .73), and in the explicit than the implicit task ($\hat{\beta}$ = .03, SE = .007, t = 3.66, p < .001, d = .39). Happy stimuli produced greater delta ITPC values than the neutral one ($\hat{\beta}$ = .02, SE = .008, t = 2.69, p = .021, d = .35), while no significant difference was found between happy and sad stimuli and between neutral and sad ones (p > .05). Analyses on theta and alpha ITPC indicated main effects of channel (theta: χ^2 (1) = 29.41, p < .001; alpha: χ^2 (1) = 13.31, p < .001) and task (theta: χ^2 (1) = 16.74, p < .001; alpha: χ^2 (1) = 6.45, p = .011). Larger ITPC values were found in the prosodic than the semantic channel (theta: $\hat{\beta}$ = .02, SE = .006, t = 5.51, p < .001, d = .58; alpha: $\hat{\beta}$ = .02, SE = .005, t = 3.66, p < .001, d = .39) and in the explicit than the implicit task (theta: $\hat{\beta}$ = .02, SE = .006, t = 4.14, p < .001, d = .44; alpha: $\hat{\beta}$ = .01, SE = .005, t = 2.55, p = .011, d = .27).

LPC. In the LPC window, no significant main effect or interaction was found for delta, theta or alpha ITPC (all p > .05).

Table 2. Delta, theta, and alpha ITPC measures in the windows of N100, P200 and LPC elicited by happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks.

Auditory ERP **measure**	**Frequency band**	**Task** / **Channel** / **Emotion**	**Explicit**				**Implicit**			
			Prosody (Mean/SD)		**Semantics** (Mean/SD)		**Prosody** (Mean/SD)		**Semantics** (Mean/SD)	
N100 (Mean/SD)	Delta	Happy	.37	.11	.32	.12	.34	.11	.28	.09
		Neutral	.33	.11	.31	.11	.30	.11	.28	.09
		Sad	.33	.11	.32	.10	.31	.10	.28	.08
	Theta	Happy	.33	.09	.29	.10	.31	.10	.27	.09
		Neutral	.31	.10	.28	.11	.28	.10	.25	.09
		Sad	.30	.11	.29	.10	.28	.08	.27	.09
	Alpha	Happy	.23	.09	.22	.08	.22	.11	.20	.07
		Neutral	.23	.08	.21	.10	.22	.08	.20	.08
		Sad	.23	.08	.21	.08	.19	.09	.21	.09
P200 (Mean/SD)	Delta	Happy	.41	.15	.38	.13	.40	.11	.34	.11
		Neutral	.40	.12	.35	.12	.37	.13	.31	.11
		Sad	.40	.12	.37	.12	.39	.12	.35	.12
	Theta	Happy	.34	.10	.33	.10	.33	.10	.29	.09
		Neutral	.33	.10	.30	.10	.31	.11	.27	.09
		Sad	.33	.10	.31	.11	.31	.10	.29	.09
	Alpha	Happy	.25	.09	.22	.08	.23	.10	.21	.07
		Neutral	.24	.09	.21	.09	.23	.08	.20	.09
		Sad	.24	.08	.21	.08	.20	.08	.21	.08
LPC (Mean/SD)	Delta	Happy	.19	.06	.19	.05	.18	.05	.17	.05
		Neutral	.19	.05	.18	.06	.21	.06	.19	.05
		Sad	.17	.04	.19	.05	.18	.04	.19	.04
	Theta	Happy	.17	.05	.17	.05	.16	.04	.15	.04
		Neutral	.17	.03	.15	.05	.17	.05	.17	.04
		Sad	.16	.03	.17	.04	.15	.04	.16	.03
	Alpha	Happy	.15	.04	.16	.05	.14	.03	.14	.03
		Neutral	.14	.04	.14	.03	.14	.03	.15	.04
		Sad	.14	.03	.15	.04	.15	.03	.14	.04

Table 3. Summary of LME models indicating the relationships between auditory ERP amplitude and ITPC measures.

Auditory ERP measure	**Frequency band**	**Chi-square**	**Parameter estimate**	**Standard error**	***t* value**	***p* value**
N100	Delta	48.01	-4.10	1.00	-4.09	< .001
	Theta	1.30	.51	1.37	.38	.025
	Alpha	6.88	-2.58	.97	-2.65	< .001
P200	Delta	133.15	5.27	1.02	5.16	< .001
	Theta	17.17	3.48	1.51	2.30	< .001
	Alpha	3.69	2.17	1.13	1.92	.055
LPC	Delta	13.49	3.04	.82	3.71	<.001
	Theta	6.58	3.47	1.35	2.58	.010
	Alpha	.44	-.84	1.28	-.66	.513

 doi:10.20944/preprints202210.0424.v1

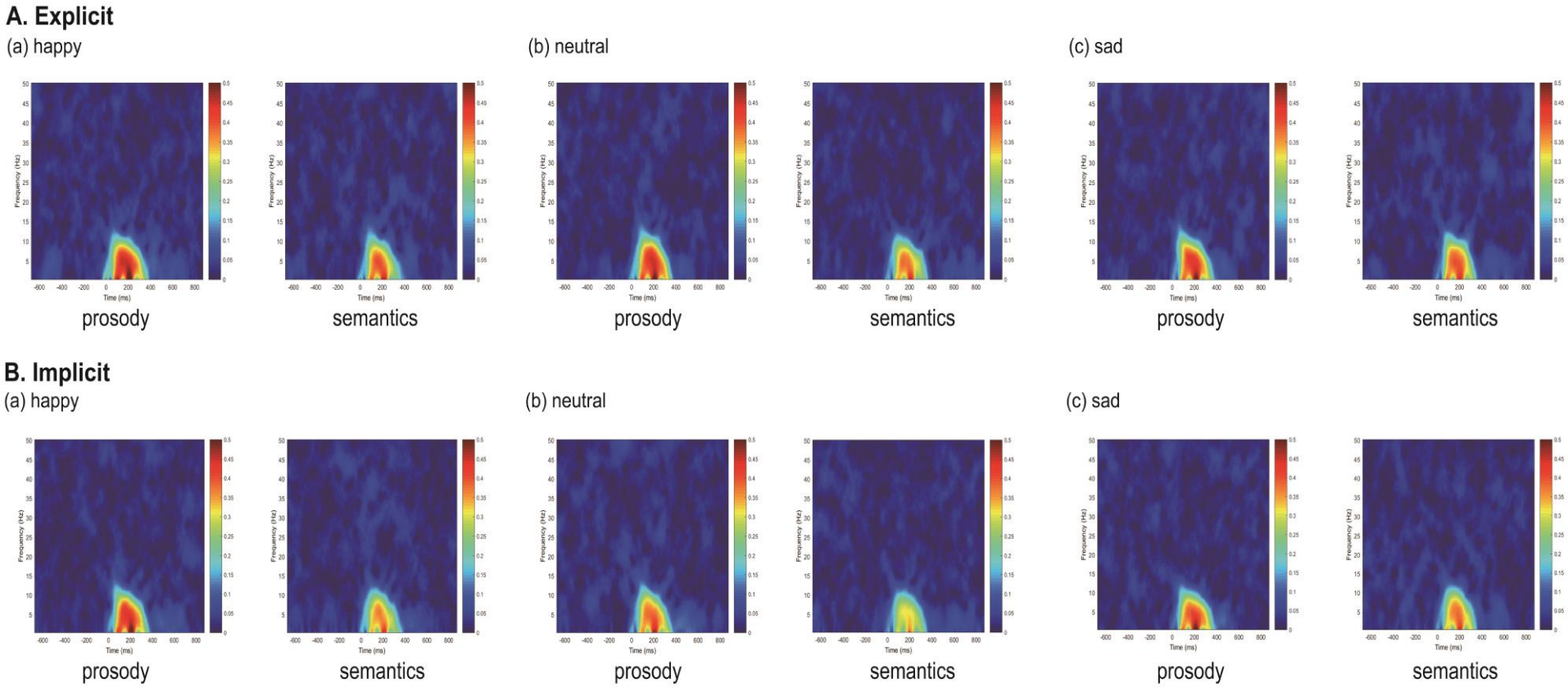

Figure 4. Time-frequency representations showing trial-to-trial phase-locking measured by ITPC for (a) happy, (b) neutral and (c) sad stimuli in prosodic and semantic channels across (A) explicit and (B) implicit tasks.

 doi:10.20944/preprints202210.0424.v1

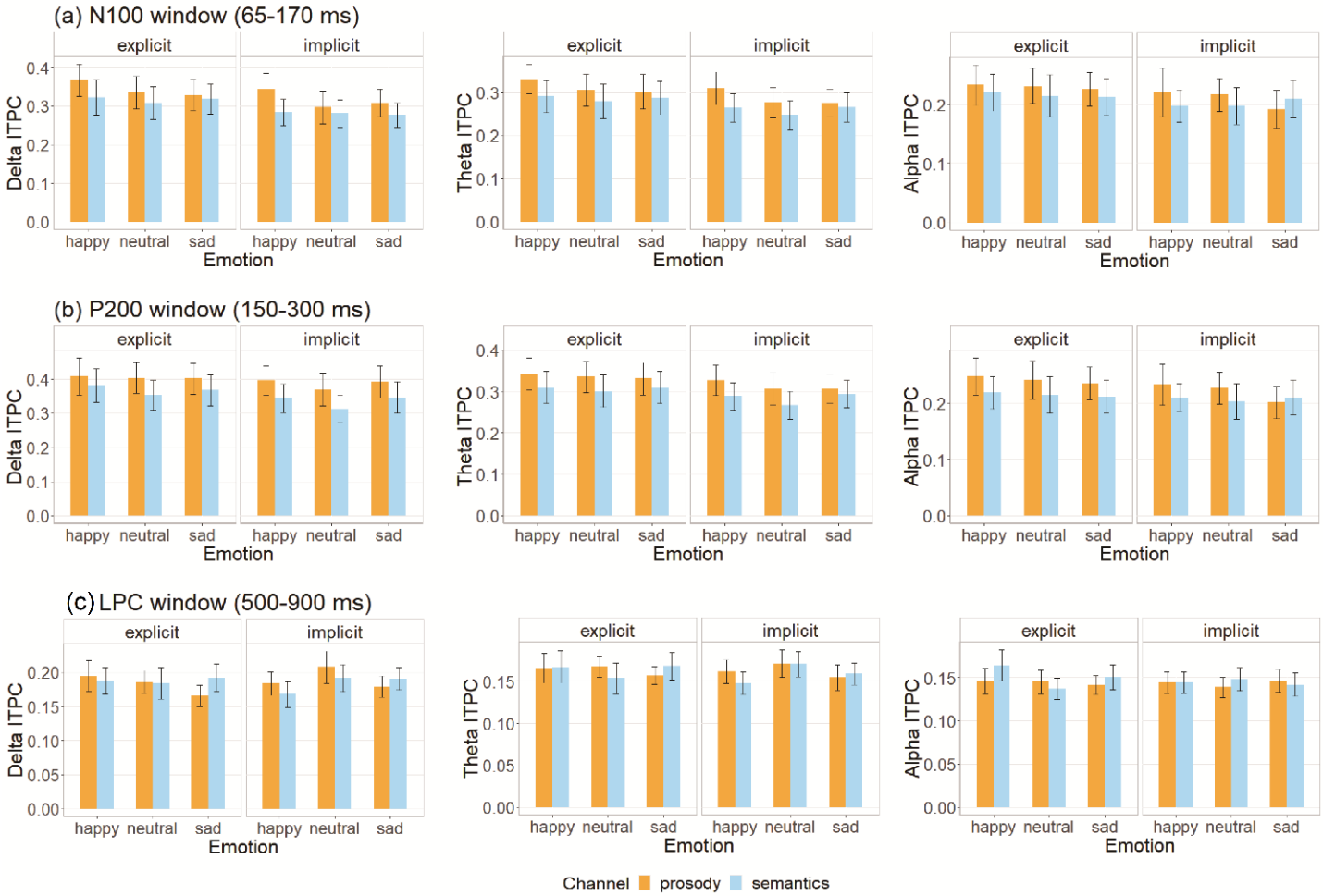

Figure 5. Bar plots of delta, theta and alpha ITPC associated with (a) N100, (b) P200 and (c) LPC for happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks. Mean phase locking value is displayed in the bar charts with error bars showing 95% confidence intervals.

3.3 Relationships between auditory ERP and ITPC

LME analyses revealed that delta (χ^2 (1) = 48.01, p < .001) and alpha (χ^2 (1) = 6.88, p = .009) ITPC were significant predictors of N100 amplitudes. In addition, all three ITPC measures were significant or marginally significant predictors of P200 amplitudes (delta: χ^2 (1) = 133.15, p < .001; theta: χ^2 (1) = 17.17, p < .001; alpha: χ^2 (1) = 3.69, p = .055). LPC amplitudes were predicted by delta (χ^2 (1) = 13.49, p < .001) and theta ITPC (χ^2 (1) = 6.58, p = .010). For these significant effects, higher ITPC values were significantly associated with stronger N100, P200 and LPC enhancement (Table 3).

3.4 Behavioral results

Identification accuracy and reaction time data of happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks are summarized in Table 4 and visualized in Figure 6.

When analyzing the behavioral data, we excluded responses over two standard deviations from the mean reaction time (3.4%) [84]. Results of MANOVA indicated main effects of channel (Pillai's trace = .03, F (2, 347) = 5.53, p = .004), emotion (Pillai's trace = .10, F (4, 696) = 9.50, p < .001), and task (Pillai's trace = .41, F (2, 347) = 122.04, p < .001), and significant interactions between emotion and task (Pillai's trace = .07, F (4, 696) = 6.64, p < .001) and between channel and task (Pillai's trace = .02, F (2, 347) = 4.44, p = .013) on accuracy and reaction time.

 doi:10.20944/preprints202210.0424.v1

Separate univariate ANOVAs on accuracy data revealed a main effect of emotion (F (2, 348) = 15.79, p < .001). Post-hoc multiple-comparison tests indicated that happy ($\hat{\beta}$ = .02, *standard error (SE)* = .005, t = 3.66, p < .001, *Cohen's d* = .47) and neutral ($\hat{\beta}$ = .03, SE = .005, t = 5.51, p < .001, d = .71) stimuli triggered more accurate responses than the sad ones. There was no significant difference between happy and neutral stimuli (p = .157). In addition, there was a main effect of task (F (1, 348) = 61.32, p < .001). Explicit tasks produced less accurate responses than the implicit ones ($\hat{\beta}$ = -.03, SE = .004, t = -7.81, p < .001, d = -.82). More importantly, significant interactions between emotion and task (F (2, 348) = 11.75, p < .001) and between channel and task (F (1, 348) = 8.55, p = .004) were found. In explicit tasks, happy ($\hat{\beta}$ = .04, SE = .007, t = 5.20, p < .001, d = .95) and neutral ($\hat{\beta}$ = .05, SE = .007, t = 7.11, p < .001, d = 1.30) stimuli elicited more accurate responses than the sad ones, and there was no significant difference between happy and neutral stimuli (p = .138). In addition, emotional prosody yielded more accurate responses than semantics when attention was focused on the emotional aspect of the stimuli ($\hat{\beta}$ = .02, SE = .006, t = 3.00, p = .003, d = .45). In implicit tasks, however, there was no significant difference between any of the two emotional stimuli nor between the prosodic and semantic channels (all p > .05).

Separate univariate ANOVAs on reaction time revealed a main effect of channel (F (1, 348) = 9.54, p = .002). Emotional prosody elicited faster responses than semantics ($\hat{\beta}$ = -44.0, SE = 14.2, t = -3.09, p = .002, d = -.33). There was also a main effect of emotion (F (2, 348) = 3.67, p = .026). Happy stimuli triggered faster responses than the neutral ($\hat{\beta}$ = -41.41, SE = 17.4, t = -2.37, p = .048, d = -.31) ones and marginally significantly faster responses than sad ($\hat{\beta}$ = -40.39, SE = 17.4, t = -2.32, p = .055, d = -.30) ones. There was no significant difference between neutral and sad stimuli (p = .998). Furthermore, a main effect of task was found (F (1, 348) = 188.88, p < .001). Explicit tasks produced slower responses than the implicit ones ($\hat{\beta}$ = 196, SE = 14.2, t = 13.73, p < .001, d = 1.45).

Table 5 summarizes the effects of channel, emotion and task and their interactions in each neurophysiological and behavioral measure.

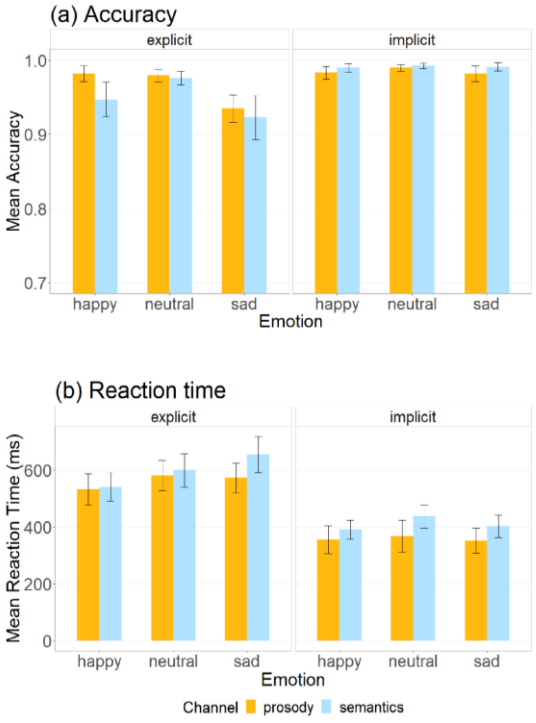

Figure 6. Identification (a) accuracy and (b) reaction time of happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks. Mean accuracy is displayed in the bar charts with error bars showing 95% confidence intervals.

Table 4. Mean identification accuracy and reaction time of happy, neutral and sad stimuli in prosodic and semantic channels across explicit and implicit tasks.

Measure	Task / Channel / Emotion	Explicit				Implicit			
		Prosody (Mean/SD)		Semantics (Mean/SD)		Prosody (Mean/SD)		Semantics (Mean/SD)	
ACC	Happy	98.21%	2.87%	94.70%	6.29%	98.30%	2.34%	98.96%	1.47%
	Neutral	97.93%	2.34%	97.59%	2.48%	98.96%	1.19%	99.25%	0.98%
	Sad	93.48%	4.99%	92.30%	7.99%	98.17%	2.77%	99.11%	1.49%
RT	Happy	532.90	147.14	541.48	134.41	355.92	131.45	390.82	89.99
	Neutral	580.87	143.07	599.90	157.93	368.54	151.08	437.44	107.08
	Sad	572.34	140.49	654.54	169.75	352.81	120.02	402.99	104.76

Note. "ACC" stands for accuracy, and "RT" stands for reaction time.

Table 5. Summary of the effects of channel, emotion and task and their interactions in each neurophysiological and behavioral measure.

Stages	**Early stages: basic auditory processing**		**Late stages: higher-order cognitive processing**		
Indices and functions	**Sensory processing of acoustic signals (N100)**	**Initial derivation of emotional meaning (P200)**	**Conscious construction of emotional meaning (LPC)**	**Behavioral identification of emotional stimuli**	
				Accuracy	**Reaction time**
Channel	Pro > Sem *(amplitudes & all ITPC)*	Pro > Sem *(amplitudes & all ITPC)*	Pro > Sem *(amplitudes)*	Pro ≈ Sem	Pro < Sem
Emotion	Hap > Neu ≈ Sad *(amplitudes, delta and theta ITPC)*	Hap > Neu ≈ Sad *(amplitudes)* Hap > Neu *(delta ITPC)*	Hap ≈ Sad > Neu *(amplitudes)*	Neu ≈ Hap > Sad	Hap < Neu ≈ Sad
Task	Exp > Imp *(amplitudes and all ITPC)*	Exp > Imp *(amplitudes & all ITPC)*	Exp < Imp *(amplitudes)*	Exp < Imp	Exp > Imp
Interaction among factors	Pro > Sem not for sadness *(amplitudes)*	Pro > Sem not for sadness in implicit tasks *(amplitudes)*	Pro > Sem not for happiness in explicit tasks *(amplitudes)*	Pro > Sem not for implicit tasks	Pro ≈ Sem

Notes. Pro = prosody; Sem = semantics; Hap = happy; Neu= neutral; Exp = explicit; Imp = implicit; "≈" indicates no significant differences.

4. Discussion

The present study investigated how communication channels, emotion categories and task types affected different stages of auditory emotional speech perception. We examined the auditory ERP responses, their corresponding oscillatory activities and the behavioral performances elicited by spoken words expressing happiness, neutrality and

sadness in either the prosodic or semantic channel under explicit and implicit emotion perception tasks. Overall, our neurophysiological and behavioral data revealed the modulatory role of channels, emotions, tasks and their reciprocal interactions in auditory emotion perception. Specifically, emotional prosody (relative to semantics) and happiness (relative to neutrality and sadness) are more perceptually dominant with greater neural activities during the sensory processing of acoustic signals and initial derivation of emotional significance, and better behavioral performance during cognitive evaluation of the stimuli. While explicit tasks also trigger greater neural responses than the implicit ones during early auditory processing, they produce reduced brain responses and poorer processing performance in the later stages. Interestingly, the prosodic dominance effect is meditated by emotional categories and task focuses, but the extent of modulation is specific to different processing stages. In addition, our study indicated that oscillation synchrony plays an important role in the neural generation of auditory event-related responses by showing increased ITPC significantly correlated with enhanced auditory ERP amplitudes. These major findings will be discussed in detail in the following subsections.

4.1 Effects of communication channels on emotional speech perception

Early auditory evoked potentials (i.e., N100 and P200) were identified for semantic and prosodic stimuli across participants, which indicates that both linguistic and paralinguistic emotion processing occurs before making judgments about the spoken stimuli [2,80,85,86]. These two types of information processing share some similarities in the time courses, which concurs with the three-stage model of emotion processing proposed by Schirmer and Kotz [35]. However, as predicted in Hypothesis 1, we observed important differences in the perceptual salience of the two communication channels: emotional prosody is consistently more perceptually salient than the semantic channel throughout emotional speech perception. It is generally assumed that early neurophysiological measures (e.g., N100, P200) primarily reflect sensory perception and late neurobehavioral measures (e.g., LPC, accuracy, reaction time) demonstrate high-order cognitive processing. Our study shows that there was a general increase in all ERP amplitudes and ITPC values, as well as shorter reaction time for emotional prosody relative to semantics. This suggests that prosody dominates over semantics not only during low-level sensory perception but also during high-level cognitive evaluation even when semantic processing is given more weight later on.

To the best of our knowledge, this is the first study to provide neurophysiological evidence showing larger auditory evoked responses with smaller neural jittering for the prosodic dominance effect during early and late emotional speech processing. The present study was also able to isolate the emotion processing in the response-making stage from the earlier perceptual and cognitive stages by measuring reaction time from the offset of auditory stimuli. The response time data demonstrated that prosody continues to dominate over semantics in the later decision-making stage, which replicates previous behavioral research on unisensory and multisensory emotion perception in our lab [2,50,51,87]. The predominance of prosody over semantics can be related to differences in stimulus characteristics of the two channels. As shown in Tables S3, S4 and S6 in *Supplemental Materials*, prosodic stimuli showed greater variations in acoustic properties, including mean duration and f0, and emotional arousal among different emotional categories compared with the semantic ones, thus enjoying greater perceptual salience throughout the three stages of emotion word processing. In addition, since our participants all spoke a tonal language (i.e., Mandarin-Chinese) as their mother tongue and lived in an East-Asian country with a high-context culture, they were likely to develop greater sensitivity to pitch-related cues that are important for prosody processing and rely heavily on contextual messages during social communication [15,88].

Interestingly, the processing dominance of prosody over semantics are modulated by emotion categories and task types, though such modulatory effects are differentially represented at the three processing stages. The prosodic dominance effect was attenuated

for sadness processing and in the implicit task during early auditory processing and decision-making. However, the effect was reduced for happiness processing in the explicit task during conscious emotion processing in the brain. Specifically, compared with emotional semantics, prosody elicited larger N100 amplitudes for happy and neutral stimuli but not for the sad ones in both explicit and implicit tasks. Larger P200 amplitudes were found in the prosodic channel for happy and neutral stimuli regardless of task focuses, but for sad stimuli in the explicit task only. Larger LPC amplitudes were also observed for emotional prosody except for happiness processing in the explicit task, though there was a general increase in accuracy irrespective of emotion category when participants were guided to focus on the emotionality of prosody than that of semantics. Though prosody elicited higher accuracy during explicit emotion identification, this channel dominance effect was somewhat reduced during earlier stages of cognitive processing, as indexed by non-significantly different LPC amplitudes between the two sensory channels when participants encountered happy stimuli in the explicit task.

The differential representations of emotional and task modulation as time unfolds may be related to the distinct functions of each processing stage. In the context of early emotional speech processing, N100 reflects the physical features of the auditory stimuli, and P200 serves as an index of the emotional salience of a vocal stimulus [21,33,86]. In this perspective, sad stimuli in the present study were characterized by longer mean duration and lower mean f0 compared with the happy and neutral ones (Tables S3 and S4 in *Supplemental Materials)*, which makes it difficult to differentiate the two communication channels for sadness processing irrespective of task requirements in the N100 window. In the P200 window, the prosodic dominance effect reached significance in explicit emotion identification tasks, while it only displayed a non-significant trend for the processing of sadness in implicit tasks. This implies that attention directed towards the emotional meaning of the stimuli plays a facilitatory role in the derivation of emotional significance from prosodic cues. Higher identification accuracy of prosodic stimuli in the explicit tasks but not in the implicit ones further suggests that task focuses not only shape early emotional speech perception but continue to interact with the channel dominance effect in the response-making stage of emotion processing. This finding is not surprising as in the implicit task, participants relied on similar vocal cues (esp. f0) for the perception of speaker's gender in both channels [89]. By contrast, while they counted on various acoustic features (e.g., f0, duration, voice quality) to determine the emotional information of prosodic stimuli, they conducted higher-order semantic analyses to determine that of verbal content, which made the two channels more distinguishable in the explicit task. Moreover, LPC is more sensitive to lexico-semantic processing than earlier sensory components [90] and to emotional stimuli especially those with high arousal [33], which may explain why no significant amplitude difference was found between the processing of happy semantics and that of happy prosody in explicit tasks.

4.2 Effects of emotion categories on emotional speech perception

One important question centering around the effect of emotion is whether emotional signals can be differentiated from the neutral ones in speech processing [34,35,52]. Some differences were identified between the emotional and non-emotional signals in the present study, but the strength of the emotionality effect tends to be valence-dependent. Consistent with previous neurophysiological and behavioral observations [80,91-93], happy stimuli were consistently more perceptually salient than the neutral ones, as reflected by significantly larger N100, P200 and LPC amplitudes, greater delta and theta ITPC values in the N100 window, and greater delta ITPC values in the P200 window. However, sadness did not differ from neutrality in the N100 and P200 windows, but elicited significantly larger LPC amplitudes later on. This is understandable as LPC reflects a more elaborate building-up of emotional meaning [33]. Such results underline the idea that the emotional salience of happiness emerges from early sensory stages, whereas sadness does not

manifest its emotional significance until high-order cognitive processing of the spoken stimuli.

During the response-making stage, in line with previous behavioral results [68], the identification accuracy of neutral stimuli was significantly higher than that of the sad stimuli, and even slightly (but not significantly) higher than that of the happy stimuli, though these differences only occurred in explicit tasks. It is likely that while both emotional stimuli contained semantics-prosody incongruency (e.g., happy/sad semantics spoken in a neutral prosody or semantically neutral words spoken in a happy/sad prosody), neutral stimuli were always congruent in prosody and semantics, thus producing more accurate identification when participants focused their attention on the emotional content of the stimuli. Interestingly, the reaction time of neutral stimuli lay between that of the happy and sad ones, which may be due to the fact that neutrality stands in the middle both for acoustic properties (e.g., f0 and duration) and emotional attributes (e.g., valence and arousal) [94]. However, the generalizability of the differences between accuracy and reaction time measures warrants investigations in future studies.

Another important finding consistent with our prediction in Hypothesis 2 was that there were significant neurobehavioral differences between specific emotion types. Compared with sadness, happiness tended to be more perceptually salient as it triggered larger N100 and P200 amplitudes, greater delta and theta ITPC values in the N100 window, higher accuracy and shorter reaction time compared with the sad ones. Our electrophysiological data suggest that the differentiation between emotional categories can start as early as around 100 ms, which might be attributable to differential acoustic and arousal characteristics of the two emotions [35,43,58,95]. For example, happiness is often characterized by a faster speech rate (shorter duration), higher intensity and mean f0, and higher emotional arousal compared to sadness, thereby triggering larger auditory ERP responses during the initial sensory and emotional decoding of the stimulus. As delta oscillations depend on the activity of motivational systems and reflect salience detection, and theta oscillations are involved in emotional regulation [53,96], better phase alignment of cortical oscillations in happiness processing implicates that happiness tends to be more motivationally and emotionally significant than sadness, which might also contribute to its sensory dominance. In addition, happiness continued to produce better identification performances compared with sadness during behavioral evaluation of the auditory stimulus, which supports the claims of a positive outlook and prosocial benevolent strategies in social communication [59].

4.3 Effects of task types on emotional speech perception

In the present study, participants intentionally directed their attention to the emotional aspect of the stimuli in explicit tasks, while they paid attention to the non-emotional property (speaker's gender) of the stimuli in implicit tasks. Our electrophysiological, time-frequency, and behavioral data confirmed the third hypothesis that explicit tasks triggered larger neural responses during earlier stages of auditory emotion perception but produced reduced brain activities and poorer behavioral performance during later cognitive processing. Previous studies demonstrated distinctive effects of attention on N100 and P200, with increased attention producing more negative N100 but less positive P200 amplitudes [62-64]. While we observed enhanced N100 as an indication of increased attentiveness in explicit tasks, there was also an increase in P200 amplitudes when attention was guided towards the emotional characteristics of the stimulus in our study. The P200 following the N100 is often referred to as part of the N1-P2 complex in auditory processing and shares many characteristics with the preceding component [97]. Another plausible account is that N100 and P200 are sensitive to cognitive efforts as increased processing demands lead to enhanced auditory ERP amplitudes [37,66]. Given the differential roles of required attentiveness and cognitive efforts in shaping the auditory ERP components, we speculate that the two effects may exert an additive effect on the more negative-going

N100 component in explicit tasks; by contrast, they may counteract in affecting the P200 amplitude with task demands exerting a more decisive influence.

The nature and difficulty of different task types can also explain the neural oscillatory patterns and late cognitive processing performances observed in the current study [67,68,92,98]. All ITPC indices in the N100 and P200 time windows showed a significant enhancement in explicit emotion recognition tasks relative to the implicit condition. ITPC differences could reflect task-induced changes in the power of oscillations or concurrent evoked responses instead of actual changes in the phase of the ongoing activity [99]. According to Weiss and Mueller [67], higher inter-trial phase coherence is often found during increased task complexity, which requires a higher level of neuronal cooperation or synchronization. In this regard, our ITPC data suggest increased synchrony of neuronal oscillations across trials in the explicit task requiring top-down control of attention on the emotional aspect of the stimuli, which is more cognitively demanding than the gender discrimination task. However, we remain cautious when drawing conclusions concerning the oscillation results since our time-frequency representations contained power all the way down to 0Hz, which may reflect transient brain responses [100].

As expected, the differences between task types continue to influence the cognitive processing of the auditory stimuli. The implicit task elicited more positive LPC than the explicit one. Since LPC is often considered as a possible variant of P300, a decline in amplitudes may indicate greater task difficulty in explicit emotional identification [69]. Similarly, we found significantly better identification performances in both accuracy and reaction time measures in the implicit relative to the explicit task. It is conceivable that while the gender discrimination task was a binary (i.e., female vs. male) alternative forced-choice (AFC) task, the emotion recognition task involved differentiation among the three emotional categories (i.e., happy, neutral and sad), which automatically required more cognitive resources in memory retrieval and introduced more judgmental confounds in the response-making stage.

4.4 Neurophysiological and behavioral measures of emotional speech perception

One noteworthy finding is that ITPC values were significant predictors of auditory ERP amplitudes across experimental conditions, which supports our final hypothesis. Specifically, increased delta and alpha ITPC were correlated with more negative N100, increased delta, theta and alpha were related to more positive P200, and increased delta and theta was predictive of more positive LPC. These patterns are consistent with findings from healthy [22,28] and clinical [23,25,27,74] populations. Although previous studies have examined whether ITPC is able to predict variations in the obligatory N1-P2 complex response to speech sounds [101], very few studies have investigated whether measures of trial-by-trial neural synchrony are potential indicators of auditory ERP responses (especially late components) using emotional speech stimuli. Therefore, our novel finding has added to the extant literature in showing that stimulus-evoked phase alignment of cortical oscillations contributes to the neural generation of auditory ERPs in early and late emotional speech processing [24,30].

It is noteworthy that different types of neurological activities and their subsequent behavioral performances did not always exhibit the same profile in characterizing emotional speech processing. For instance, while interaction effects among channels, emotions and tasks were observed for all auditory ERP components, no significant interplay was found among the three factors in the ITPC measures. Moreover, there remained some distinctions even among the results from different indices belonging to the same type of experimental measure (e.g., waveform amplitudes in different time windows, ITPC data of different frequency bands, or accuracy and reaction time as behavioral data). These differences in findings may be related to differential sensitivities to various measurement indices and processing stages [98,102]. Future work can further investigate in what measures, contexts, and processing stages the observed effects of channels, emotions and tasks can be generalized and in what conditions they may or may not be replicated, which

will offer more refined ways to interpret the underlying mechanisms of emotional speech processing [51].

4.5 Implications, limitations and future studies

The present study elucidates how the channel dominance effect, emotionality effect and task effect converge in shaping emotional speech processing, which sheds new light on the theoretical debates and underlying neural substrates and behavioral mechanisms of emotion cognition. Our findings contributed tonal language data from a high-context culture to the three-stage model of emotion cognition by delineating the temporal dynamics, neural oscillation characteristics and behavioral performances of emotional prosody and semantics processing in explicit and implicit emotion perception tasks. Apart from the three contextual factors explored in the current study, individual differences have also been repeatedly reported to influence emotion processing [35]. Future work can specify how the individual variables, including personality [103], age [104] and gender [71], can modulate emotional speech processing at different stages. Since we involved participants from a tonal language background and a high-context Chinese culture, the current work can also inspire new efforts to unravel the cross-linguistic and cross-cultural differences in emotion processing [105]. Furthermore, the current experimental protocol can be applied to testing clinical populations who reportedly display dysfunctions in auditory processing and emotion perception, such as cochlear implant users [106], individuals with schizophrenia [37,42], autism [107] and Parkinson disease [108], which can promote insightful understanding of the behavioral symptomology and underlying neural basis of the diseases.

Limitations of the current study need to be acknowledged. First, emotional information was conveyed through either the prosodic or semantic channel in our experiment. Though it is possible to communicate affective messages through a single channel (e.g., talking on the telephone or listening to news broadcast) in real-life settings [109], it is more often the case that emotions are expressed concurrently through auditory (e.g., prosody and semantics) and visual (e.g., facial expressions) channels in which congruent and incongruent information can be transmitted. Therefore, it is worthwhile to delve into the neural correlates of multisensory integration of emotions and investigate how different channels interact with one another in online emotion processing [110]. Second, findings might also be limited as we focused on two of the basic emotions (i.e., happiness, sadness) and neutrality in our study. Though such selection of emotions allowed us to compare voluntary and involuntary prosodic and semantic processing using emotional and non-emotional stimuli, it has led to some asymmetries in task difficulty between the explicit (three AFC) and implicit (two AFC) tasks as discussed earlier in the third subsection of *Discussion*. Future studies are encouraged to employ an experimental design with comparable complexity between tasks and explore whether the current findings can be extended to other categories of basic (e.g., anger, disgust, surprise, fear) and complex (e.g., embarrassment, guilt) emotions and required focuses of attention (e.g., emotional arousal of the stimuli or decoders) [43]. Third, we observed significant differences in brain responses between neutral prosody and semantics, which may be related to some intrinsic differences between the prosodic and semantic stimulus sets, such as the word frequency, word types (i.e., noun vs. adjectives), word number (i.e., 60 different words for the prosodic set vs. 180 words for the semantic set), speech features (e.g., duration, f0, voice quality), prosodic contours (e.g., tonal combination) of the disyllabic stimuli. It also seems difficult to make sure whether comparable amounts of valence were presented in each channel type. As such, it is possible that the larger ERP effects in the prosodic channel were due to more valenced stimuli used in that channel. Future studies are recommended to isolate the emotional aspect alone by controlling the potential confounds such as removing all the speech elements and presenting sound contours that differ in the same way between conditions, or using the exact same words (with or without emotional connotations) for testing different conditions. Fourth, we observed N400 amplitude differences in some conditions

(e.g., implicit neutral and sad), which may affect the subsequent measure of LPC amplitudes. This is likely due to the design of our experiment, in which we divided our EEG session into two tasks (i.e., explicit or implicit) and each task contained two blocks (i.e., prosodic or semantic). Although the order of task and block was counterbalanced across participants, whether different orders led to differential amounts of repetition effect warrants further investigation. Moreover, we can see from the topographic maps in Figure 2 that the LPC effect was partially driven by some frontal negative responses to semantic conditions, so whether these are indeed LPC effects requires closer examination. The ERP methodology is limited in spatial resolution that is important for localizing the brain regions involved in generating scalp-recorded potentials [111]. Therefore, future studies combining ERP and functional magnetic resonance imaging techniques are needed to specify the engagement of brain structures involved in the time course of emotional speech processing.

5. Conclusions

The current work studied the interplay of channel, emotion and task effects on emotional speech processing using electrophysiological and behavioral measures. The results showed that prosody (relative to semantics) and happy stimuli (relative to the neutral and sad ones) gain more perceptual salience during the sensory processing of acoustic signals, initial derivation of emotional significance, and cognitive evaluation of the stimuli. Although the explicit emotion identification task tends to trigger greater neural responses compared to the implicit gender discrimination task during early processing stages, there is evidence for greater difficulty in task completion in the later decision-making stage. The channel salience effect over semantics tends to be emotion- and task-specific at different processing stages. In addition, stimulus-evoked phase alignment of oscillatory activity at different frequency bands plays a crucial role in generating the auditory event-related responses. Taken together, communication channel, emotion category and task focus interact to shape the time course, neural oscillations and behavioral activities of emotional speech processing, which enriches theoretical understanding of auditory emotion processing and provides the basis for further investigation on individual differences in emotion cognition from cross-cultural and clinical perspectives.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: Spectral images of the (A) prosodic and (B) semantic stimuli for the (a) happy, (b) neutral and (c) sad emotions; Table S1: Words for the prosodic stimulus set; Table S2: Words for the semantic stimulus set; Table S3: Duration (milliseconds) of the experimental stimuli; Table S4: Mean f0 (Hertz) of the experimental stimuli; Table S5: Familiarity rating for the spoken words used in prosodic and semantic tasks; Table S6: Identification accuracy of emotional category and rating of emotional arousal for the experimental stimuli.

Author Contributions: Conceptualization, Y.L., H.D. and Y.Z.; methodology, Y.L., H.D. and Y.Z.; validation, Y.L. and Q.C.; formal analysis, Y.L., Hao Z., F.C., H.D. and Y.Z.; investigation, Y.L., X.F., Hao Z. and Hui Z.; resources, H.D. and Y.Z.; data curation, Y.L.; writing—original draft preparation: Y.L.; writing—review and editing, H.D. and Y.Z.; visualization, Y.L..; supervision, H.D. and Y.Z.; project administration, H.D. and Y.Z.; funding acquisition, H.D. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by National Social Science Foundation of China (18ZDA293).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of School of Foreign Languages at Shanghai Jiao Tong University (protocol code: 1903S11016; date of approval: March 26, 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated during and analyzed in the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Filippi, P.; Ocklenburg, S.; Bowling, D.L.; Heege, L.; Güntürkün, O.; Newen, A.; de Boer, B. More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cogn. Emot.* **2017**, *31*, 879-891.
2. Lin, Y.; Ding, H.; Zhang, Y. Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal Stroop effects. *J. Speech. Lang. Hear. Res.* **2020**, *63*, 896-912.
3. Blasi, A.; Mercure, E.; Lloyd-Fox, S.; Thomson, A.; Brammer, M.; Sauter, D.; Deeley, Q.; Barker, G.J.; Renvall, V.; Deoni, S.; et al. Early specialization for voice and emotion processing in the infant brain. *Curr. Biol.* **2011**, *21*, 1220-1224.
4. Graf Estes, K.; Bowen, S. Learning about sounds contributes to learning about words: effects of prosody and phonotactics on infant word learning. *J. Exp. Child Psychol.* **2013**, *114*, 405-417.
5. Lima, C.F.; Alves, T.; Scott, S.K.; Castro, S.L. In the ear of the beholder: How age shapes emotion processing in nonverbal vocalizations. *Emotion* **2014**, *14*, 145-160.
6. Dupuis, K.L.; Pichora-Fuller, M.K. Use of affective prosody by young and older adults. *Psychol. Aging* **2010**, *25*, 16–29.
7. Picou, E.M. How hearing loss and age affect emotional responses to nonspeech sounds. *J. Speech. Lang. Hear. Res.* **2016**, *59*, 1233-1246.
8. Zhang, M.; Xu, S.; Chen, Y.; Lin, Y.; Ding, H.; Zhang, Y. Recognition of affective prosody in autism spectrum conditions: A systematic review and meta-analysis. *Autism* **2021**, 1362361321995725.
9. Lin, Y.; Ding, H.; Zhang, Y. Emotional prosody processing in schizophrenic patients: A selective review and meta-analysis. *J. Clin. Med.* **2018**, *7*, 363.
10. Kitayama, S.; Ishii, K. Word and voice: Spontaneous attention to emotional utterances in two languages. *Cogn. Emot.* **2002**, *16*, 29-59.
11. Ishii, K.; Reyes, J.A.; Kitayama, S. Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychol. Sci.* **2003**, *14*, 39-46.
12. Tanaka, A.; Koizumi, A.; Imai, H.; Hiramatsu, S.; Hiramoto, E.; de Gelder, B. I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychol. Sci.* **2010**, *21*, 1259-1262.
13. Anolli, L.; Wang, L.; Mantovani, F.; De Toni, A. The Voice of Emotion in Chinese and Italian Young Adults. *J. Cross Cult. Psychol.* **2008**, *39*, 565-598.
14. Pell, M.D.; Jaywant, A.; Monetta, L.; Kotz, S.A. Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cogn. Emot.* **2011**, *25*, 834-853.
15. Hall, E.T. *Beyond culture*; Anchor, 1989.
16. Wildgruber, D.; Ackermann, H.; Kreifelts, B.; Ethofer, T. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.* **2006**, *156*, 249-268.
17. Castelluccio, B.C.; Myers, E.B.; Schuh, J.M.; Eigsti, I.M. Neural substrates of processing anger in language: Contributions of prosody and semantics. *J. Psycholinguist. Res.* **2016**, *45*, 1359-1367.
18. Adolphs, R. Neural systems for recognizing emotion. *Curr. Opin. Neurobiol.* **2002**, *12*, 169-177.
19. Buchanan, T.W.; Lutz, K.; Mirzazade, S.; Specht, K.; Shah, N.J.; Zilles, K.; Jäncke, L. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research* **2000**, *9*, 227-238.
20. Wegrzyn, M.; Herbert, C.; Ethofer, T.; Flaisch, T.; Kissler, J. Auditory attention enhances processing of positive and negative words in inferior and superior prefrontal cortex. *Cortex* **2017**, *96*, 31-45.
21. Paulmann, S.; Kotz, S.A. Early emotional prosody perception based on different speaker voices. *Cognitive Neuroscience and Neuropsychology* **2008**, *19*, 209-213.
22. Fuentemilla, L.; Marco-Pallares, J.; Grau, C. Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *Neuroimage* **2006**, *30*, 909-916.
23. Yu, L.; Wang, S.; Huang, D.; Wu, X.; Zhang, Y. Role of inter-trial phase coherence in atypical auditory evoked potentials to speech and nonspeech stimuli in children with autism. *Clin. Neurophysiol.* **2018**, *129*, 1374-1382.
24. Makeig, S.; Debener, S.; Onton, J.; Delorme, A. Mining event-related brain dynamics. *Trends Cogn. Sci.* **2004**, *8*, 204-210.
25. Bishop, D.V.; Anderson, M.; Reid, C.; Fox, A.M. Auditory development between 7 and 11 years: an event-related potential (ERP) study. *PLoS One* **2011**, *6*, e18993.
26. Chen, F.; Zhang, H.; Ding, H.; Wang, S.; Peng, G.; Zhang, Y. Neural coding of formant-exaggerated speech and nonspeech in children with and without autism spectrum disorders. *Autism Res.* **2021**, *14*, 1357-1374.
27. Edwards, E.; Soltani, M.; Kim, W.; Dalal, S.S.; Nagarajan, S.S.; Berger, M.S.; Knight, R.T. Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol.* **2009**, *102*, 377-386.
28. Koerner, T.K.; Zhang, Y. Effects of background noise on inter-trial phase coherence and auditory N1-P2 responses to speech stimuli. *Hear. Res.* **2015**, *328*, 113-119.
29. Cohen, M.X. *Analyzing neural time series data: Theory and practice*; MIT Press: London, 2014.

30. Klimesch, W.; Sauseng, P.; Hanslmayr, S.; Gruber, W.; Freunberger, R. Event-related phase reorganization may explain evoked neural dynamics. *Neurosci. Biobehav. Rev.* **2007**, *31*, 1003-1016.
31. Chen, X.; Pan, Z.; Wang, P.; Zhang, L.; Yuan, J. EEG oscillations reflect task effects for the change detection in vocal emotion. *Cogn. Neurodyn.* **2015**, *9*, 351-358.
32. Chen, X.; Yang, J.; Gan, S.; Yang, Y. The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PLoS One* **2012**, *7*, e30278.
33. Kotz, S.A.; Paulmann, S. Emotion, language, and the brain. *Lang. Linguist. Compass* **2011**, 108-125.
34. Paulmann, S.; Kotz, S.A. An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain Lang.* **2008**, *105*, 59-69.
35. Schirmer, A.; Kotz, S.A. Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.* **2006**, *10*, 24-30.
36. Pinheiro, A.P.; Del Re, E.; Mezin, J.; Nestor, P.G.; Rauber, A.; McCarley, R.W.; Goncalves, O.F.; Niznikiewicz, M.A. Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation. *Psychol. Med.* **2012**, *43*, 603-618.
37. Pinheiro, A.P.; Rezaii, N.; Rauber, A.; Liu, T.; Nestor, P.G.; McCarley, R.W.; Goncalves, O.F.; Niznikiewicz, M.A. Abnormalities in the processing of emotional prosody from single words in schizophrenia. *Schizophr. Res.* **2014**, *152*, 235-241.
38. Paulmann, S.; Seifert, S.; Kotz, S.A. Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Soc. Neurosci.* **2010**, *5*, 59-75.
39. Pinheiro, A.P.; Galdo-Alvarez, S.; Rauber, A.; Sampaio, A.; Niznikiewicz, M.; Goncalves, O.F. Abnormal processing of emotional prosody in Williams syndrome: an event-related potentials study. *Res. Dev. Disabil.* **2011**, *32*, 133-147.
40. Diamond, E.; Zhang, Y. Cortical processing of phonetic and emotional information in speech: A cross-modal priming study. *Neuropsychologia* **2016**, *82*, 110-122.
41. Liu, T.; Pinheiro, A.P.; Deng, G.; Nestor, P.G.; McCarley, R.W.; Niznikiewicz, M.A. Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport* **2012**, *23*, 108-112.
42. Pinheiro, A.P.; Niznikiewicz, M. Altered attentional processing of happy prosody in schizophrenia. *Schizophr. Res.* **2019**, *206*, 217-224.
43. Paulmann, S.; Bleichner, M.; Kotz, S.A. Valence, arousal, and task effects in emotional prosody processing. *Front. Psychol.* **2013**, *4*, 345.
44. Kotz, S.A.; Paulmann, S. When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Res.* **2007**, *1151*, 107-118.
45. Gaillard, R.; Del Cul, A.; Naccache, L.; Vinckier, F.; Cohen, L.; Dehaene, S. Nonconscious semantic processing of emotional words modulates conscious access. *Proceedings of the National Academy of Sciences* **2006**, *103*, 7524-7529.
46. Herbert, C.; Junghofer, M.; Kissler, J. Event related potentials to emotional adjectives during reading. *Psychophysiology* **2008**, *45*, 487-498.
47. Kissler, J.; Herbert, C.; Peyk, P.; Junghofer, M. Buzzwords: Early cortical responses to emotional words during reading. *Psychol. Sci.* **2007**, *18*, 475-480.
48. Schacht, A.; Sommer, W. Time course and task dependence of emotion effects in word processing. *Cogn. Affect. Behav. Neurosci.* **2009**, *9*, 28-43.
49. Zhang, D.; He, W.; Wang, T.; Luo, W.; Zhu, X.; Gu, R.; Li, H.; Luo, Y.J. Three stages of emotional word processing: an ERP study with rapid serial visual presentation. *Soc. Cogn. Affect. Neurosci.* **2014**, *9*, 1897-1903.
50. Lin, Y.; Ding, H.; Zhang, Y. Gender differences in identifying facial, prosodic, and semantic emotions show category- and channel-specific effects mediated by encoder's gender. *Journal of Speech, Language & Hearing Research* **2021**, *64*, 2941-2955.
51. Lin, Y.; Ding, H.; Zhang, Y. Unisensory and multisensory Stroop effects modulate gender differences in verbal and nonverbal emotion perception. *J. Speech. Lang. Hear. Res.* **2021**, *64*, 4439-4457.
52. Paulmann, S.; Ott, D.V.; Kotz, S.A. Emotional speech perception unfolding in time: the role of the basal ganglia. *PLoS One* **2011**, *6*, e17694.
53. Symons, A.E.; El-Deredy, W.; Schwartze, M.; Kotz, S.A. The functional role of neural oscillations in non-verbal emotional communication. *Front. Hum. Neurosci.* **2016**, *10*, 239.
54. Kanske, P.; Plitschka, J.; Kotz, S.A. Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia* **2011**, *49*, 3121-3129.
55. Paulmann, S.; Pell, M.D. Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective, & Behavioral Neuroscience* **2010**, *10*, 230-242.
56. Liebenthal, E.; Silbersweig, D.A.; Stern, E. The language, tone and prosody of emotions: Neural substrates and dynamics of spoken-word emotion perception. *Front. Neurosci.* **2016**, *10*, 506.
57. Ackerman, B.P.; Abe, J.A.A.; Izard, C.E. Differential emotions theory and emotional development. In *What Develops in Emotional Development?*, Mascolo, M.F., Griffin, S., Eds.; Springer US: Boston, MA, 1998; pp. 85-106.

58. Hofmann, M.J.; Kuchinke, L.; Tamm, S.; Vo, M.L.; Jacobs, A.M. Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not positive words. *Cogn. Affect. Behav. Neurosci.* **2009**, *9*, 389-397.
59. Warriner, A.B.; Kuperman, V. Affective biases in English are bi-dimensional. *Cogn. Emot.* **2015**, *29*, 1147-1167.
60. Delaney-Busch, N.; Wilkie, G.; Kuperberg, G. Vivid: How valence and arousal influence word processing under different task demands. *Cogn. Affect. Behav. Neurosci.* **2016**, *16*, 415-432.
61. Wambacq, I.J.A.; Shea-Miller, K.J.; Abubakr, A. Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. *Neuroreport* **2004**, *15*, 555-559.
62. Crowley, K.E.; Colrain, I.M. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin. Neurophysiol.* **2004**, *115*, 732-744.
63. Pinheiro, A.P.; Barros, C.; Dias, M.; Niznikiewicz, M. Does emotion change auditory prediction and deviance detection? *Biol. Psychol.* **2017**, *127*, 123-133.
64. Näätanen, R.; Teder, W.; Alho, K.; Lavikainen, J. Auditory attention and selective input modulation: A topographical ERP study. *Neuroreport* **1992**, *3*.
65. Lenz, D.; Schadow, J.; Thaerig, S.; Busch, N.A.; Herrmann, C.S. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *Int. J. Psychophysiol.* **2007**, *64*, 31-38.
66. Ullsperger, P.; Freude, G.; Erdmann, U. Auditory probe sensitivity to mental workload changes – an event-related potential study. *Int. J. Psychophysiol.* **2001**, *40*, 201-209.
67. Weiss, S.; Mueller, H.M. The contribution of EEG coherence to the investigation of language. *Brain Lang.* **2003**, *85*, 325-343.
68. Iredale, J.M.; Rushby, J.A.; McDonald, S.; Dimoska-Di Marco, A.; Swift, J. Emotion in voice matters: neural correlates of emotional prosody perception. *Int. J. Psychophysiol.* **2013**, *89*, 483-490.
69. Kim, K.H.; Kim, J.H.; Yoon, J.; Jung, K.Y. Influence of task difficulty on the features of event-related potential during visual oddball task. *Neurosci. Lett.* **2008**, *445*, 179-183.
70. Schirmer, A.; Kotz, S.A.; Friederici, A.D. On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research* **2005**, *24*, 442-452.
71. Schirmer, A.; Lui, M.; Maess, B.; Escoffier, N.; Chan, M.; Penney, T.B. Task and sex modulate the brain response to emotional incongruity in Asian listeners. *Emotion* **2006**, *6*, 406-417.
72. Fruhholz, S.; Ceravolo, L.; Grandjean, D. Specific brain networks during explicit and implicit decoding of emotional prosody. *Cereb. Cortex* **2012**, *22*, 1107-1117.
73. Ekman, P. Are there basic emotions? *Psychol. Rev.* **1992**, *99*, 550-553.
74. Koerner, T.K.; Zhang, Y. Differential effects of hearing impairment and age on electrophysiological and behavioral measures of speech in noise. *Hear. Res.* **2018**, *370*, 130-142.
75. Wu, H.; Ma, X.; Zhang, L.; Liu, Y.; Zhang, Y.; Shu, H. Musical experience modulates categorical perception of lexical tones in native Chinese speakers. *Front. Psychol.* **2015**, *6*, 436.
76. Liu, Y.; Liang, N.; Wang, D.; Zhang, S.; Yang, T.; Jie, C.; Sun, W. *A Dictionary of the Frequency of Commonly Used Modern Chinese Words (Alphabetical Sequence Section)*; Astronautic Publishing House: Beijing, 1990.
77. Boersma, P.; Weenink, D. *Praat: Doing phonetics by computer*, 6.1.41; 2021.
78. Psychology Software Tools *E-Prime 2.0*, Pittsburgh, PA, 2012.
79. Chaumon, M.; Bishop, D.V.; Busch, N.A. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J. Neurosci. Methods* **2015**, *250*, 47-63.
80. Pinheiro, A.P.; Rezaii, N.; Nestor, P.G.; Rauber, A.; Spencer, K.M.; Niznikiewicz, M. Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain Lang.* **2016**, *153-154*, 38-49.
81. Delorme, A.; Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **2004**, *134*, 9-21.
82. R Core Team *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria: 2020.
83. Lenth, R.V. emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.5. **2020**.
84. Baayen, R.H.; Milin, P. Analyzing reaction times. *International Journal of Psychological Research* **2010**, *3*, 12-28.
85. Van Petten, C.; Coulson, S.; Rubin, S.; Plante, E.; Parks, M. Time course of word identification and semantic integration. *J. Exp. Psychol. Learn. Mem. Cogn.* **1999**, *25*, 394-417.
86. Paulmann, S. Chapter 88 - The Neurocognition of Prosody. In *Neurobiology of Language*, Hickok, G., Small, S.L., Eds.; Academic Press: San Diego, 2016; pp. 1109-1120.
87. Lin, Y.; Ding, H. Effects of communication channels and actor's gender on emotion identification by native Mandarin speakers. In Proceedings of the Interspeech 2020, 2020; pp. 3151-3155.
88. Liu, P.; Rigoulot, S.; Pell, M.D. Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia* **2015**, *67*, 1-13.
89. Lass, N.J.; Hughes, K.R.; Bowyer, M.D.; Waters, L.T.; Bourne, V.T. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* **1976**, *59*, 675-678.

90. Grieder, M.; Crinelli, R.M.; Koenig, T.; Wahlund, L.O.; Dierks, T.; Wirth, M. Electrophysiological and behavioral correlates of stable automatic semantic retrieval in aging. *Neuropsychologia* **2012**, *50*, 160-171.
91. Spreckelmeyer, K.N.; Kutas, M.; Urbach, T.P.; Altenmuller, E.; Munte, T.F. Combined perception of emotion in pictures and musical sounds. *Brain Res.* **2006**, *1070*, 160-170.
92. Knyazev, G.G.; Slobodskoj-Plusnin, J.Y.; Bocharov, A.V. Event-related delta and theta synchronization during explicit and implicit emotion processing. *Neuroscience* **2009**, *164*, 1588-1600.
93. Mueller, C.J.; Kuchinke, L. Individual differences in emotion word processing: A diffusion model analysis. *Cogn. Affect. Behav. Neurosci.* **2016**, *16*, 489-501.
94. Duncan, S.; Barrett, L.F. Affect is a form of cognition: A neurobiological analysis. *Cogn. Emot.* **2007**, *21*, 1184-1211.
95. Feng, C.; Wang, L.; Liu, C.; Zhu, X.; Dai, R.; Mai, X.; Luo, Y.J. The time course of the influence of valence and arousal on the implicit processing of affective pictures. *PLoS One* **2012**, *7*, e29668.
96. Knyazev, G.G. Motivation, emotion, and their inhibitory control mirrored in brain oscillations. *Neurosci. Biobehav. Rev.* **2007**, *31*, 377-395.
97. Key, A.P.; Dove, G.O.; Maguire, M.J. Linking brainwaves to the brain: An ERP primer. *Dev. Neuropsychol.* **2005**, *27*, 183-215.
98. Luo, Y.; Zhang, Y.; Feng, X.; Zhou, X. Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience* **2010**, *169*, 654-664.
99. van Diepen, R.M.; Mazaheri, A. The Caveats of observing Inter-Trial Phase-Coherence in Cognitive Neuroscience. *Sci. Rep.* **2018**, *8*, 2990.
100. Luck, S.J. *An introduction to the event-related potential technique*; MIT press: 2014.
101. Koerner, T.K.; Zhang, Y. Application of linear mixed-Effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sciences* **2017**, *7*.
102. Thompson, A.E.; Voyer, D. Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cogn. Emot.* **2014**, *28*, 1164-1195.
103. Mittermeier, V.; Leicht, G.; Karch, S.; Hegerl, U.; Moller, H.J.; Pogarell, O.; Mulert, C. Attention to emotion: auditory-evoked potentials in an emotional choice reaction task and personality traits as assessed by the NEO FFI. *Eur. Arch. Psychiatry Clin. Neurosci.* **2011**, *261*, 111-120.
104. Paulmann, S.; Pell, M.D.; Kotz, S.A. How aging affects the recognition of emotional speech. *Brain Lang.* **2008**, *104*, 262-269.
105. Liu, P.; Rigoulot, S.; Pell, M.D. Cultural immersion alters emotion perception: Neurophysiological evidence from Chinese immigrants to Canada. *Soc. Neurosci.* **2017**, *12*, 685-700.
106. Agrawal, D.; Thorne, J.D.; Viola, F.C.; Timm, L.; Debener, S.; Buchner, A.; Dengler, R.; Wittfoth, M. Electrophysiological responses to emotional prosody perception in cochlear implant users. *NeuroImage: Clinical* **2013**, *2*, 229-238.
107. Charpentier, J.; Kovarski, K.; Houy-Durand, E.; Malvy, J.; Saby, A.; Bonnet-Brilhault, F.; Latinus, M.; Gomot, M. Emotional prosodic change detection in autism Spectrum disorder: an electrophysiological investigation in children and adults. *J. Neurodev. Disord.* **2018**, *10*, 28.
108. Garrido-Vasquez, P.; Pell, M.D.; Paulmann, S.; Strecker, K.; Schwarz, J.; Kotz, S.A. An ERP study of vocal emotion processing in asymmetric Parkinson's disease. *Soc. Cogn. Affect. Neurosci.* **2013**, *8*, 918-927.
109. Hawk, S.T.; van Kleef, G.A.; Fischer, A.H.; van der Schalk, J. "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion* **2009**, *9*, 293-305.
110. Jessen, S.; Kotz, S.A. The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage* **2011**, *58*, 665-674.
111. Amodio, D.M.; Bartholow, B.D.; Ito, T.A. Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Soc. Cogn. Affect. Neurosci.* **2014**, *9*, 385-393.