L2-categorisation of linguistic and emotional prosody in German: an EEG study

Article *in* Language, Cognition and Neuroscience · July 2025

DOI: 10.1080/23273798.2025.2526466

5 authors, including:

Huan Wei
Philipps University of Marburg
7 PUBLICATIONS 7 CITATIONS

Yifei He
Philipps University of Marburg
55 PUBLICATIONS 446 CITATIONS

Christina Kauschke
Philipps University of Marburg
134 PUBLICATIONS 1,853 CITATIONS

Mathias Scharinger
Philipps University of Marburg
82 PUBLICATIONS 1,054 CITATIONS

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: www.tandfonline.com/journals/plcp21

L2-categorisation of linguistic and emotional prosody in German: an EEG study

Huan Wei, Yifei He, Christina Kauschke, Mathias Scharinger & Ulrike Domahs

To cite this article: Huan Wei, Yifei He, Christina Kauschke, Mathias Scharinger & Ulrike Domahs (03 Jul 2025): L2-categorisation of linguistic and emotional prosody in German: an EEG study, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2025.2526466

To link to this article: https://doi.org/10.1080/23273798.2025.2526466

Published online: 03 Jul 2025.

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2025.2526466

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

L2-categorisation of linguistic and emotional prosody in German: an EEG study

Huan Wei[a,b], Yifei He[b,c], Christina Kauschke[a], Mathias Scharinger[a,b] and Ulrike Domahs[a,b]

[a]Institute of German Linguistics, Marburg University, Marburg, Germany; [b]Center for Mind, Brain, and Behavior (CMBB), Marburg, Germany; [c]Department of Psychiatry and Psychotherapy, Marburg University, Marburg, Germany

ABSTRACT
Prosodic information such as intonation patterns can have linguistic or emotional functions and may be expressed differently cross-linguistically. This study investigates whether speakers of typologically distinct languages – Chinese L2 learners of German and German L1 speakers – differ in their ability to distinguish between linguistic and emotional prosody. In a judgment task combined with EEG recordings, participants' sensitivity to linguistic and emotional prosody was assessed. Behavioural responses to emotional conditions differed between groups, suggesting that both groups may use different cues during the prosodic evaluation processes. The ERP results indicate a higher sensitivity to linguistic prosody in the L1 group compared to the L2 group, but no group differences in the processing of emotional prosody. The present study, therefore, provides evidence for distinctive L2 effects on the processing of linguistic and emotional prosody.

ARTICLE HISTORY
Received 21 June 2024
Accepted 19 June 2025

KEYWORDS
Prosody perception; linguistic prosody; emotional prosody; cross-cultural processing; second language learning; Chinese German L2

Introduction

Prosody, expressed by the modulation of acoustic-phonetic parameters such as duration, intensity, and fundamental frequency (F0) in spoken language, fulfils a multifaceted role. It specifies, for instance, linguistic distinctions between sentence types (statements vs. questions) and conveys semantic or emotional meanings (Larrouy-Maestri et al., 2024; Pell & Kotz, 2021) in non-tone languages. In tone languages like Mandarin Chinese, however, prosody can also be used to express lexical distinctions, because lexical words are differentiated by means of pitch contours. Given the subtle modifications of prosodic parameters in expressions and utterances, the question arises whether linguistic and emotional prosody pose comparable difficulties in second language (L2) learning. This might be particularly the case for learners with a tone native language (L1). The studies reported in the present paper aim at investigating neurocognitive processes underlying German linguistic and emotional prosody perception in L1 speakers (L1: German) and L2 (L1: Mandarin Chinese) learners. Specifically, it examines processing differences between L1 German (an intonational language) speakers and L2 learners of German with Mandarin Chinese as L1 (a tone language).

Linguistic and emotional prosody in German and Mandarin

Linguistic prosody, encompassing the encoding of lexical stress or sentence mode, constitutes an important interface to grammar that is needed to interpret the meaning of sentences in an intonational language like German. In German, statements are realised with a falling and questions with a rising intonation contour (Gibbon, 1997). In Mandarin, however, due to its status as tone language, the dual functions of pitch result in an interaction of lexical tones and sentence modes. For instance, "ke4 fang2 (guest room)" ends with tone 2 and maintains a rising pitch in both question and statement modes, while "kai1 fang4 (open)" ends with tone 4 and exhibits a falling pitch, as visually illustrated in Figure 1 of M. Liu et al. (2022). Polar questions, which are yes/no questions, are marked in Mandarin by sentence-final particles such as "ma" or "me", along with a falling pitch (Yuan et al., 2002; Zahner-Ritter et al., 2022). This divergence in sentence mode marking necessitates the inclusion of linguistic prosodic features in teaching materials for Mandarin learners of German. For example, the syllabus for the "Basic German I" at the School of Foreign Languages UJN, V.R. China (G. Liu, 2016), explicitly incorporates prosodic markers of sentence mode.

CONTACT Huan Wei vera-huan.wei@mailbox.org Institute of German Linguistics, Marburg University, Pilgrimstein 16, 35032, Marburg, Germany; Center for Mind, Brain, and Behavior (CMBB), Hans-Meerwein-Straße 6, 35032 Marburg, Germany
This article has been corrected with minor changes. These changes do not impact the academic content of the article.

For emotional prosody, Wei et al. (2023) found that positive emotions in German and Mandarin Chinese tend to be produced with a higher pitch, and negative emotions with longer duration, while the pitch contours and durations differ between the two languages. More specifically, in German, positive emotions are typically expressed with a higher pitch and intensity and increased duration compared to the neutral mode, while negative emotions are characterised by decreased intensity and have longer duration than positive emotions (Kotz & Paulmann, 2007; Scherer, 2003; Wickens & Perry, 2015). Emotional prosody in Mandarin systematically modulates the acoustic properties and perception of lexical tones, with emotional cues exerting a stronger influence on tone identification than lexical tones do on emotional recognition (H.-S. Chang et al., 2023). Under multi-talker babble noise, lexical tone perception is prioritised over emotional prosody for Mandarin listeners, reflecting greater perceptual robustness of linguistic tones in noisy environments (Zhang et al., 2023). Acoustic and articulatory distinctions further differentiate emotional categories: positive emotions (e.g. happiness, pleasant surprise) are characterised by higher F0 means, wider F0 variation, forward tongue positioning, and faster speech rates, while negative emotions (e.g. sadness, fear, disgust) exhibit slower speech rates, smaller F0 variation, longer durations, and retracted tongue position (Erickson et al., 2016; Gong et al., 2023).

In contrast to the acquisition of linguistic prosody in L2, emotional prosody is predominantly shaped by contextual and social factors, learned rather implicitly, and can be subject to modification within the classroom environment, often influenced by the teacher's demeanour (Smith & King, 2018). Moreover, the emotional experience and expression of both teachers and students in L2 learning are profoundly influenced by the social climate within the classroom and their willingness to communicate (Joe et al., 2017; Wang et al., 2021). Given that it is rather unlikely that L2 learners learn emotional prosody explicitly in class, the question arises whether they can perceive, process, and classify the emotions expressed in the second language in the same way as L1 listeners.

Neurolinguistic insights into linguistic and emotional prosody

Studies that measure electroencephalography (EEG) responses are able to identify the influence of phonetic parameters on different processing steps in time. In the following, we will review neurolinguistic studies of linguistic and emotional prosody obtained from L1 processing before turning to L2 processing. In an EEG study (Tomasello et al., 2022), investigating linguistic prosody in Italian, sentences ended in words with either a rising pitch for questions or a falling pitch for statements. Italian native listeners displayed distinct neurophysiological responses for questions and statements 100 milliseconds (ms) after the noun onset. The authors identified this response as a P50/N100 event-related potential (ERP) component, reflecting greater processing demands for questions compared to statements and the brain's recognition of prosodic cues signalling speaker intentions. Control tests with non-vocal musical sounds showed no significant differences between rising and falling pitch contours, confirming that the observed responses were specific to linguistic prosody rather than general pitch differences. Lu et al. (2015) explored the processing of intonation (*statement* or *question*) in emotional Chinese words by asking participants to match the intonation of two words. The ERPs of Mandarin Chinese speakers showed an N100 component for congruent intonation and an increased N200 component for incongruent intonation, indicating that processing incongruent intonation is more demanding than processing congruent intonation.

In EEG studies comparing linguistic and emotional prosodic processing (Paulmann et al., 2012; Zora et al., 2020), researchers investigated whether linguistic and emotional prosodies are based on similar time courses of processing. Paulmann et al. (2012) found prosodic expectancy positivity (PEP) elicited by linguistic and emotional prosodic expectancy violations. PEP differed in latency and distribution, occurring for linguistic prosody at 1000–1400 ms (~620 ms post-stimulus onset) in the frontal region and for emotional prosody at 850–1400 ms (~470 ms post-stimulus onset) in the posterior region. This suggests that emotional and linguistic prosody processing rely not only on different neural mechanisms but also on different timing. Zora et al. (2020) found that changes in Swedish linguistic lexical contour patterns (low tone vs. high tone) elicited a stronger mismatch negativity (MMN) at 300–350 ms than in pseudowords, indicating the presence of long-term memory traces linked to words and their prosodic features. Emotional prosody (angry vs. neutral) evoked a larger positive response (P3a) at 350–400 ms compared to linguistic prosody, reflecting that the motivational salience of affective prosody requires more processing resources. Stimuli with combined affective and emotional prosody generated a late positive complex (LPC) response compared to linguistic or emotional prosody, providing neural evidence for the integration of emotional and linguistic prosody.

Building on neurocognitive models (Schirmer & Kotz, 2006) and previous neurophysiological data on the time course of prosodic processing, Pell and Kotz (2021) proposed a "three-stage model" for processing of socioemotive aspects of prosody. The first stage involves the basic level of sensory encoding, which structures the incoming acoustic stream, shaping it into an auditory gestalt that facilitates the discrimination of human voices from other auditory stimuli, such as music or other sound categories (Rigoulot et al., 2015). Following this, the second stage entails salience detection, a process that directs attention to motivationally significant vocal features within the unfolding stimulus, including aspects such as emotionality, high arousal, and other relevant characteristics. In this regard, ERPs reveal amplitude differences in the N100 component, which is associated with basic level sensory encoding, and P200 component, which is linked to salience detection (Paulmann et al., 2013). The third stage marks the higher cognitive analysis of vocal expressions, commencing approximately 300 ms post-onset of the vocal stimulus and persists in a sustained manner as the vocal expression continues, reflected in an increased LPC in the 300–900 ms time window post stimulus onset for emotional expressions (Kotz & Paulmann, 2011; Pell et al., 2015). The integration of word meaning and prosody is reflected in the N400 component (Schirmer & Kotz, 2003), while the P600 effect indicates how this meaning is incorporated into the overall interpretation of the utterance (Brouwer et al., 2017), especially when the prosody conflicts with word meaning.

In an EEG study investigating emotional prosody (sadness, anger, happiness, and surprise) with varying speech durations (short: 0.5–1 s, medium 1.5–2 s, and long 2.5–3 s) in Mandarin Chinese, J. Chang et al. (2018) identified neurophysiological components corresponding to the stages of the "three-stage model." The N100 component was elicited by happiness prosody in short-duration stimuli and by sadness prosody in long-duration stimuli, aligning with the model's first stage of sensory encoding. The P200 component, observed for emotional prosody in short-duration stimuli, corresponded to the second stage, highlighting the attentional focus on emotionally salient features. Lastly, the N300 component at 300–400 ms, observed for emotional speech with short-duration, reflected cognitive analysis processes associated with the model's third stage, emphasising the sustained evaluation of vocal expressions. In another study on differentiating emotions (anger, happiness, and neutral) in pseudowords, Steber et al. (2020) found a larger P200 and an enhanced late positivity (LPP, 600–900 ms) for happiness, along with greater negativities (500–550 ms) for angry and neutral prosody compared to happiness. These findings suggest that emotional prosody is processed differently based on emotional valence and arousal. Vergis et al. (2020) investigated vocal stances in interpersonal requests (e.g. "Lend me a nickel") with polite versus impolite attitudes. Impolite prosody increased early anterior positivity (P200) compared to polite attitudes. ERPs time-locked to the target word ("nickel") showed that impolite prosody affected word retrieval, as indicated by the N400, and the contextual integration of emotional intent with meaning, as reflected in P600-like effects occurring 500–800 ms post-onset of the critical word.

Given the findings obtained from native language processing, the question arises whether the reported effects may be also evoked in second language perception. Yu et al. (2021) investigated word stress processing in L1 Chinese and L2 English among native Chinese listeners, using disyllabic Chinese and English pseudowords as speech conditions and their hummed versions as nonspeech conditions. As a result, Yu and colleagues found that the native Chinese speech stimuli elicited a larger late negative response (LNR, 300–800 ms post-stimulus onset) than the L2 English stimuli, highlighting language-specific effects in prosodic phonology processing. However, no such effect was observed in the phoneme-free prosodic acoustic control conditions, indicating that the LNR is linked to linguistic features rather than acoustic properties. Similarly, Jiang et al. (2019) examined how social group membership (in-group speakers, out-group regional speakers, and out-group foreign speakers) influence believability judgments using vocal stimuli with confident and doubtful prosody. Their ERPs showed that in-group speaker's vocal confidence was differentiated early (N100 and P200), emphasising the motivational importance of doubtful voices for believability. For out-group speakers, believability judgments involved greater cognitive effort, with increased N400 and late negativity responses (LNR, 550–900 ms post-stimulus onset) reflecting challenges in integrating accented speech. These findings demonstrated that both language-specific features and social group membership influence the cognitive processes underlying speech perception and believability judgments.

Aims of the present study

So far, it is not clear whether the processing of linguistic and emotional prosody differs between L1 speakers and L2 learners or whether linguistic and/or emotional prosody lead to difficulties for L2 learners, particularly if L1 and L2 express linguistic and emotional prosody differently. To fill this gap, the present study aimed to investigate how L2 listeners of German with the tone language Mandarin Chinese as L1 process linguistic and emotional prosodic information of L2 German. We ran two EEG experiments in which stimuli differing in either linguistic or emotional prosody had to be classified. In Experiment I, we presented semantically neutral German single words that were modified regarding linguistic sentence mode (*STATEMENT* vs. *QUESTION*). In Experiment II the similar stimuli were presented with the emotional prosodies (*LIKE*, *DISGUST*, and *NEUTRAL*). The same groups of German native speakers (L1 group) and Chinese-German learners (L2 group) participated in both experiments.

The studies aimed to determine whether Mandarin listeners classify linguistic and emotional prosody similarly to the L1 group or whether the L2 group exhibits more effortful processing, as reflected in stronger ERP components. Specifically, increased N100 amplitudes would indicate increased sensory encoding demands, while enhanced P200 or late components (such as N400, P600, or LPC) would suggest greater attentional and cognitive resource allocation. Regarding emotional prosody, one account suggests that its acoustic-phonetic features may be shared across languages (Larrouy-Maestri et al., 2024). Based on this, we hypothesise that L2 learners may find it easier to perceive emotional prosody than linguistic prosody in their L2. Consequently, their ERP components might resemble those of German native speakers during the integration and reanalysis stage, such as LPP or LPC (Kotz & Paulmann, 2011; Pell et al., 2015; Steber et al., 2020). Alternatively, if the processing of emotional prosody is more demanding in L2 learners compared to L1 listeners, this could result in stronger ERP magnitudes of the N400 or LNR component (Jiang et al., 2019; Yu et al., 2021). Additionally, if language status has an impact on prosodic processing, a further question was whether processing of linguistic and emotional prosody was affected in similar or different ways.

Experiment I: linguistic prosody

Materials and methods

Participants

The L1 group consisted of 18 native German speakers (11 females, mean age = 25.11 years, SD = 2.99), and the L2 group of 19 native Chinese speakers (10 females, mean age = 25.89 years, SD = 4.27). The Chinese participants learned German in adulthood (mean = 19.95 years old, SD = 2.55) in China for about 2.35 years (SD = 1.7). At the time of the experiment, they studied and lived in Germany for more than four years (SD = 2.97). All Chinese participants performed German proficiency tests at level C1 (or equivalent) according to the CEFR (Common European Reference Framework), corresponding to an advanced proficiency level. Participants of both groups were healthy, right-handed, and did not report any hearing impairments.

The Prosody-Analysis-Test (Walther & Otten, 2016) was used to roughly assess the L2 group's receptive prosodic abilities at a behavioural level. Originally designed to evaluate German-speaking children's prosodic abilities in the context of speech and language therapy, this computer-assisted instrument was adapted for use with adults learning German as their second language. Specifically, two subtests were selected for the present study. In these subtests, the L2 group was asked to categorise linguistic prosody (sentence modes: *STATEMENT* and *QUESTION*) and emotional prosody (emotions: *FEAR* and *JOY*) of single German nouns. Results indicated a high level of accuracy in both tasks (sentence modes: mean = 95.92%, SD = 12%; emotions: mean = 98.38%, SD = 11%). Furthermore, the paired samples *t*-test showed no significant difference in accuracy between the tasks, suggesting comparable receptive prosodic abilities for both linguistic and emotional prosody categories in the L2 group.

Stimulus material

40 trisyllabic German nouns from the word categories food, animal, place, and hobby, without reduced syllables and with identical stress patterns (e.g. *Feldsalat* "*corn salad*"), were comprised as stimuli in Experiment I. According to the EmoInt-2017-Database[1] (Köper et al., 2017), the mean emotional valence of the stimuli is 5.62 (SD = .82), which means the stimuli are semantically neutral with regard to emotionality. Two female native German speakers recorded the selected stimuli with two different prosodies correlated with *STATEMENT* and *QUESTION*. To ensure the validity of the intended prosodic distinctions, a group of four healthy volunteers participated in a forced-choice task, classifying each prosodic stimulus. They rated each audio stimuli on a three-level typicality scale (1 = very atypical, 2 = typical, 3 = very typical). Only stimuli that were correctly classified by at least three out of four volunteers and received a typicality rating of at least 2.5 were selected. In total, 80 stimuli were selected, with 20 stimuli per condition from each speaker, ensuring a well-balanced distribution.

The acoustic properties of stimuli (Figure 1) (F0, intensity, and word duration) were analysed using Praat

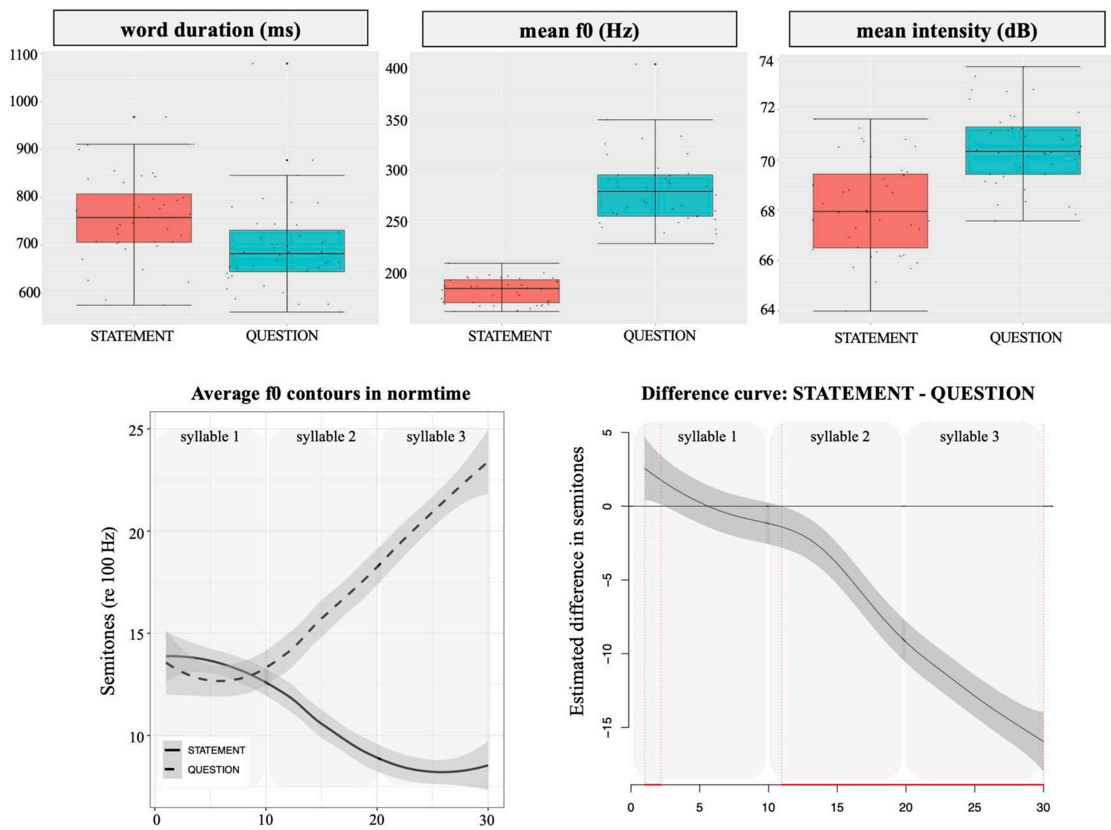

Figure 1. Boxplots of word duration, mean F0, mean intensity, F0 contours, and difference curve of linguistic prosody (*STATEMENT* and *QUESTION*) in German. These significant differences are indicated by red vertical lines in the difference curve plot and the dark grey shading displays the 95% CI of the predicted mean difference.

(version 6.0.37) (Boersma & Weenink, 2018). The F0 contour of each utterance was time-normalized (10 points per syllable) using the Praat script of ProsodyPro (Xu, 2013). The mean F0, mean intensity, and word duration of stimuli were analysed with linear mixed models and the *lmer* function provided in the *lme4* package (version 1.1–26) (Bates et al., 2014) for the statistical programme R (version 4.3.1) (R Core Team, 2023). The base model included random intercept SPEAKERS and random slope (1 + CONDITION | SPEAKERS), with the variable CONDITION (*STATEMENT* and *QUESTION*) as a fixed effect. Effect size estimates for mixed model predictors were obtained using the *MuMIn* package (Bartoń, 2022). The results were presented with the respective test statistics, corresponding *p*-values, and the effect sizes (Cohen's *d*). We used generalised additive mixed models (GAMMs) with the *mgcv* package (Wood, 2006, 2017) to compare the F0 contours (in semitones) between two conditions over time, in order to align these differences to potential neural differences as indexed by different ERP components. The models incorporated a pre-specified number of base functions of different shapes, including splines and smooth functions (Sóskuthy, 2021; van Rij et al., 2019; Wieling, 2018). The results were plotted using the *itsadug* package (van Rij et al., 2017). This procedure reveals when two F0 contours significantly differ from each other over time. When the 95% confidence intervals (CIs) for the smooth terms do not include zero, it indicates a significant difference between the F0 contours at those time points. These significant differences are indicated by red vertical lines in the difference curve plots. The lists of the stimuli with their emotional valences, the audio recordings, and the R codes can be accessed via the Open Science Framework (OSF, https://osf.io/4xqn7/?view_only=2d03a41872b44f9aaacfa13fcae1fc78).

The analysis revealed significant differences between *QUESTION* and *STATEMENT* conditions in mean F0 ($ß = 81.78$, $SE = 36.45$, $d = 1.51$, $t = 2.24$, $p < .05$), mean intensity ($ß = 1.73$, $SE = .43$, $d = .19$, $t = 4.036$ $p < .001$), and word duration ($ß = -62.99$, $SE = 23.6$, $d = -.69$, $t = -2.67$, $p < .01$). Specifically, the *QUESTION* condition had a higher F0 and more vigorous intensity than the *STATEMENT* condition (as shown in Figure 1). However, the *STATEMENT* condition had a longer word duration (mean = 759.4 ms, SD = 88.55) than the *QUESTION* condition (mean = 696.4 ms, SD = 95.07). The GAMMs analyses showed that the model with the smooth term of CONDITION over time was significantly better than the model without the CONDITION ($\chi^2 = 73.58$, $p < .001$), suggesting that the two conditions differ in F0 contours. The final model (with the scat-linking function), corrected for autocorrelation, accounted for 89.2% of the variance. As shown in Figure 1, the differences between *STATEMENT* and *QUESTION* in the 2nd and 3rd syllables are significant, with the F0 contour of the *QUESTION* condition being higher than that of the *STATEMENT* condition.

Procedure

The participants were seated comfortably at a fixed distance of 80 cm in front of a 21.5-inch Samsung monitor and loudspeakers, situated within a soundproof cabin to minimise external noise and electromagnetic interference during the experiments. The participants were instructed to listen to the auditory stimuli while disregarding their semantic meaning, focusing instead on identifying the linguistic prosody of the words as either *STATEMENT* or *QUESTION*. Participants indicated their decision by pressing the corresponding symbols on the keyboard (green dot for *STATEMENT* and red dot for *QUESTION*). Each trial began with a white fixation point, followed by a 300 ms blank screen. The auditory stimulus coincided with the white fixation point, maintaining an inter-stimulus interval (ISI) of 300 ms. A visual question mark then appeared on the screen, prompting the participants to respond within a maximum of 4000 ms. A 500 ms interval after each button press allowed participants to blink, indicated by the display of an eye symbol for 1000 ms. The experiment started with a practice block consisting of 10 trials, followed by eight test blocks with a total of 80 trials. The sequence of the trials was randomised for each participant. Moreover, the assignment of responses to either the right or left index finger was counterbalanced across participants. No feedback was provided during or after the experiment.

EEG recording

The electroencephalogram (EEG) was recorded via 26 active Ag/AgCl electrodes (actiCAP) with positions following the 10–20 system (F7, F3, Fz, F4, F8, FC1, FC2, T7, C3, Cz, C4, T8, CP5, CP6, P7, P3, Pz, P4, P8, PO9, PO10, FC5, FC6, CP1, CP2, Oz) on a BrainAmp standard amplifier (Brain Products GmbH). FPz was used as ground and FCz as the online reference, where TP9 and TP10 served for offline re-referencing (approximating the linked mastoids). For the electrooculogram, two electrodes were used to register left right eye movements, and two electrodes were used to register vertical eye movements.

Behavioural data analysis

Due to the experiment's design, the participants had to wait until the visual question mark appeared to press one of the responses to avoid movement artifacts in the EEG data. For this reason, we did not analyse the reaction time of the behavioural data. Instead, the accuracy scores were calculated using generalised logistic mixed-effects regression models provided in the *lme4* package in R. The base model included random intercepts PARTICIPANTS and SPEAKERS, as fixed effects the variables CONDITION (*STATEMENT* and *QUESTION*) and GROUP (*L1 group* and *L2 group*).

ERP data analysis

The electrophysiological data were pre-processed using the EEGLAB toolbox (Delorme & Makeig, 2004) within MATLAB (R2021b). First, the EEG data were downsampled to 250 Hz and then referenced to the average of the electrodes TP9 and TP10. To eliminate line noise effectively, the Zapline Plus function (Klug & Kloosterman, 2022) was applied. The data were then subjected to an extended infomax Independent Component Analysis (ICA). The ICA outcomes were then applied to the unfiltered data, and to further refine the dataset, a bandpass filter was applied within the range of 0.1 and 30 Hz. Subsequent analyses were carried out using the FieldTrip toolbox (Oostenveld et al., 2011) for EEG/MEG analysis within MATLAB (R2021b). ERP averages were computed over epochs with a 200 ms pre-stimulus baseline and 1000 ms time window post-stimulus onset. On average, 7.9% (SD = 3.8%) of trials from the L1 group and 9.7% (SD = 4.5%) from the L2 group were rejected as artifacts.

Based on previous literature and visual inspection of the grand mean ERP data, four time windows (TW1: N100, 80–120 ms; TW2: P200, 165–215 ms; TW3: centro-parietal N400, 250–500 ms, TW4: centro-parietal late positivity, 500–800 ms post stimulus onset) were selected for the analysis. Within these time intervals, the mean amplitudes from selected electrodes underwent statistical analyses using linear mixed regression models, implemented through the *lme4* package (Bates et al., 2014). Post-hoc tests were performed using the *emmeans* package (Lenth et al., 2023), and effect size estimates for mixed model predictors were obtained with the *MuMIn* package (Bartoń, 2022). The dependent variable in the model was the mean voltage amplitude for each condition within a certain time window. The fixed factors included the variables CONDITION and GROUP, and the base model included both PARTICIPANTS and SPEAKERS (the two female native German speakers: *speaker1*, *speaker2*) as random intercepts. The reference level was set as *STATEMENT* for the factor CONDITION and *L1 group* for GROUP.

The factor REGION consisted of the levels anterior (*F3, Fz, F4*), central (C3, Cz, C4), and posterior (*P3, Pz, P4*). Over each time window, voltage amplitudes were averaged for each participant and electrode. Statistical analyses were performed using the Type II Wald-Chi-Square test to determine the significance of the fixed effects. Post-hoc comparisons were conducted using planned *t*-tests to compare relevant contrasts within each time window, with corresponding Bonferroni-corrected *p*-values and the effect sizes (Cohen's *d*) reported. The R codes and the anonymized data can be accessed via OSF.

Results

Behavioural results

Participants of the L1 group categorised the condition *STATEMENT* correctly in 96.81% of cases (SD = 18%) and *QUESTION* in 97.22% (SD = 16%). The L2 group recognised *STATEMENT* with an accuracy of about 96.05% (SD = 19%) and *QUESTION* with about 96.32% (SD = 19%). The between-group analysis showed no significant effects or interactions.

ERP results

The ERP results are illustrated in Figure 2. During TW1 (N100: 80–120 ms post stimulus onset), the linear mixed regression analysis revealed neither significant main effects nor interactions.

In TW2 (P200: 165–215 ms post stimulus onset), significant effects of CONDITION ($\chi^2 = 10.95$, $p < .001$) and GROUP ($\chi^2 = 5.73$, $p < .05$) were found, with the brain responses evoked by the *STATEMENT* condition were more positive than the brain responses for the *QUESTION* condition ($ß = .9$, SE = .41, $d = .26$, $t = 2.21$, $p < .05$) and

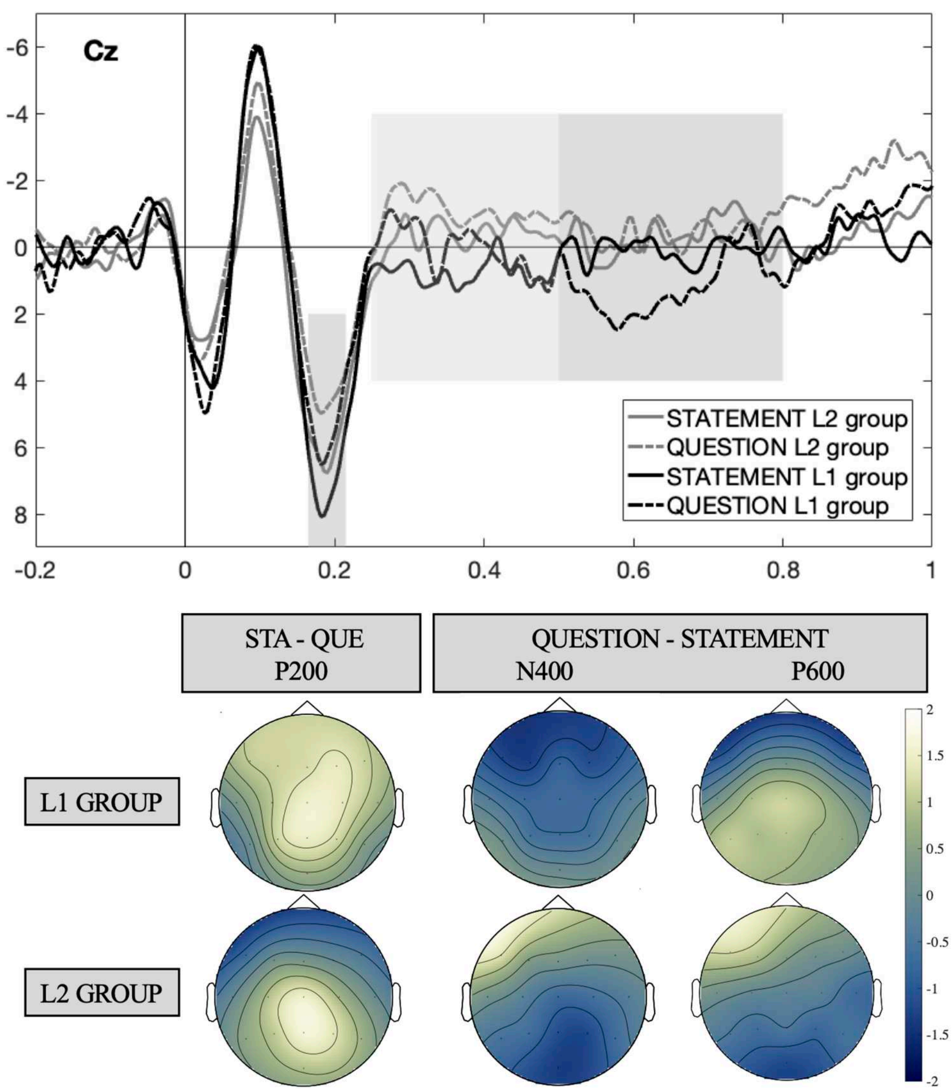

Figure 2. Grand-average ERPs for both groups during the processing of linguistic prosody at selected Cz electrode 200 ms before stimulus onset up to 1000 ms post-stimulus onset. The time windows with significant effects are indicated. Scalp maps display amplitude differences for both groups at P200, N400, and late positivity in 500–800 ms time windows.

the difference between *STATEMENT* and *QUESTION* conditions was larger in L1 group than in the L2 group ($ß = 1.21$, SE = 0.41, $d = 0.35$, $t = 2.94$, $p < .01$). There was no interaction between GROUP and CONDITION, or between CONDITION and REGION.

During TW3 (centro-parietal N400: 250–500 ms post stimulus onset), there were significant effects of CONDITION ($\chi^2 = 5.79$, $p < .05$) and GROUP ($\chi^2 = 9.26$, $p < .01$). The brain responses of the *QUESTION* condition were more negative than the brain responses of the *STATEMENT* condition ($ß = .69$, SE = .29, $d = .28$, $t = 2.37$, $p < .05$) in both groups and the difference was larger in L2 group than in the L1 group ($ß = .89$, SE = .12, $d = .36$, $t = 3.02$, $p < .01$). There was no interaction between GROUP and CONDITION.

In TW4 (centro-parietal positivity P600: 500–800 ms post stimulus onset), the analysis revealed a significant interaction between CONDITION and GROUP ($\chi^2 = 5.97$, $p < .05$). Further post-hoc analyses were conducted to examine the effects of the CONDITION within each group separately. The results revealed a significant effect of CONDITION in the L1 group ($\chi^2 = 12.62$, $p < .001$) with the brain responses of the *QUESTION* condition being more positive than the brain responses of the *STATEMENT* condition ($ß = 1.13$, SE = .32, $d = .59$, $t = 3.55$, $p < .001$), but no significant effect of CONDITION in the L2 group ($\chi^2 = .86$, $p > .1$).

Discussion of experiment I

In the first experiment, L1 German speakers and Chinese-German L2 learners categorised the linguistic prosody of *STATEMENT* and *QUESTION* in one-word stimuli in German. The behavioural responses of the two groups showed a high accuracy rate and no significant difference. The main EEG findings are as follows: (1) A significantly larger positivity occurred in the time window from 165 to 215 ms for the *STATEMENT* condition compared to the *QUESTION* condition for both groups; (2) a significantly larger centro-parietal negativity effect was observed in the time window from 250 to 500 ms for the *QUESTION* condition compared to the *STATEMENT* condition for both groups, suggesting that both groups relied on the same prosodic cues during the acoustic analysis and interpretation of linguistic prosody. (3) In the later time window (500–800 ms post stimulus onset), only the L1 group showed a significantly larger centro-parietal positivity for the *QUESTION* condition compared to the *STATEMENT* condition, indicating that L1 listeners engage in processes to update their mental representation of the communicated message based on the linguistic prosodic context. Following Vergis et al. (2020), we interpret this component as an instance of the P600 effect. In sum, these findings altogether imply that L2 learners were able to map the prosodic form onto the intended meaning, although their judgement about the linguistic category were not as certain as those of the native speakers.

Unlike linguistic prosody, which is learned more explicitly, emotional prosody is predominantly shaped by contextual and social factors and is learned rather implicitly (Filippa et al., 2022). This distinction leads to expectations that Chinese L2 learners of German may process emotional prosody differently compared to native German speakers. It is essential to examine whether L2 learners process emotional prosodic cues in the same way as native speakers and how this might differ from their processing of linguistic prosody.

Experiment II: emotional prosody

Methods

Participants

In the second experiment, the same group of participants as in Experiment I took part after a short break of approximately 5 min. Due to the loss of data of one female Chinese-German learner, the L2 group in Experiment II consisted of 18 native Chinese speakers (9 females, mean age = 26 years, SD = 4.38). The participants of the L1 group were identical to Experiment I.

Materials

We selected 40 bi-syllabic German words from the semantic categories food, animal, and school subject as stimuli, without reduced syllables and with identical stress patterns (e.g. Knoblauch "garlic"). According to the EmoInt-2017-Databank, the mean emotional valence of the stimuli is 5.64 (SD = .68), which means they are semantically neutral with regard to emotionality.

The stimuli were recorded by three female native German speakers with three different prosodies correlated with *NEUTRAL*, *LIKE* as positive, and *DISGUST* as negative emotions. Similar to Experiment I, the same volunteers classified the emotional prosodic stimuli in a forced-choice task. Audio stimuli classified correctly by three out of four ratings were included in the final stimulus set of Experiment II. In total, 120 stimuli were selected, with 40 stimuli per condition. The complete set of stimuli included an equal number of recordings taken from each of the three speakers. The lists of the stimuli providing also information about their emotional valences as well as the audio recordings are available on OSF.

As in Experiment I, the acoustic properties (mean F0, mean intensity, and word duration) of the stimuli (Figure 3) were analysed with linear mixed models in R. The base models included SPEAKERS as a random intercept, a random slope (1 + CONDITION | SPEAKERS), and the variable CONDITION (*NEUTRAL*, *LIKE*, and *DISGUST*) as a fixed effect. The normtime F0 contours (in semitones) for the three emotional conditions were analysed using GAMMs, as was done in the Experiment I. The results were reported in the same way in Experiment I.

The results of the linear mixed models' analysis indicated significant differences between the *LIKE* and *NEUTRAL* conditions in F0 ($ß = 55.63$, $SE = 27.95$, $d = 1.07$, $t = 1.99$, $p < .05$) and intensity ($ß = 1.42$, $SE = 0.59$, $d = .14$, $t = 2.41$, $p < .05$), but not in word duration. Specifically, the *LIKE* condition was associated with a higher F0 and higher intensity compared to the *NEUTRAL* condition. The *DISGUST* condition significantly differed from the *NEUTRAL* condition in word duration ($ß = 251.2$, $SE = 43.03$, $d = 1.89$, $t = 5.84$, $p < .001$), but not in F0 or intensity. In addition, the *DISGUST* condition (mean = 972.3 ms, SD = 152.46) had a significantly longer word duration than the *NEUTRAL* condition (mean = 718.05 ms, SD = 135.01). Between the *LIKE* and *DISGUST* conditions, there was a significant difference in F0 ($ß = 60.07$, $SE = 32.57$, $d = 1.15$, $t = 1.84$, $p < .05$), with the *LIKE* condition having a higher F0 than the *DISGUST* condition, but not in intensity or word duration.

The GAMM analysis showed that the model with the smooth term of CONDITION over time was significantly better than the model without the CONDITION ($\chi^2 = 60.46$, $p < .001$), suggesting that the CONDITION variable explains sufficient variance. The final model, corrected for autocorrelation and using the scat-linking function, accounted for 84.3% of the variance. As shown in Figure 3, the F0 contour differences between the *NEUTRAL* and *LIKE* conditions are significant over time, with the F0 contour of the *LIKE* condition being significantly higher than that of the *NEUTRAL* condition. Additionally, the differences between *NEUTRAL* and *DISGUST* conditions are significant in the 2nd syllable, with the F0 contour of the *NEUTRAL* condition being significantly higher than that of the *DISGUST* condition. Furthermore, the differences between the *DISGUST* and *LIKE* conditions are significant over time, with the F0 contour of the *LIKE* condition being significantly higher than that of the *DISGUST* condition. This raises the question whether there is evidence that the distinct time course of ERP components is based on the time-varying differences in the intonation contours.

Procedure

The procedure was identical to the one in Experiment I. Participants' task was to categorise the emotional prosody of each presented word. Their responses were indicated by pressing the corresponding smiley symbol on the keyboard. The response for the emotion *NEUTRAL* was assigned to the keyboard letter "B", and

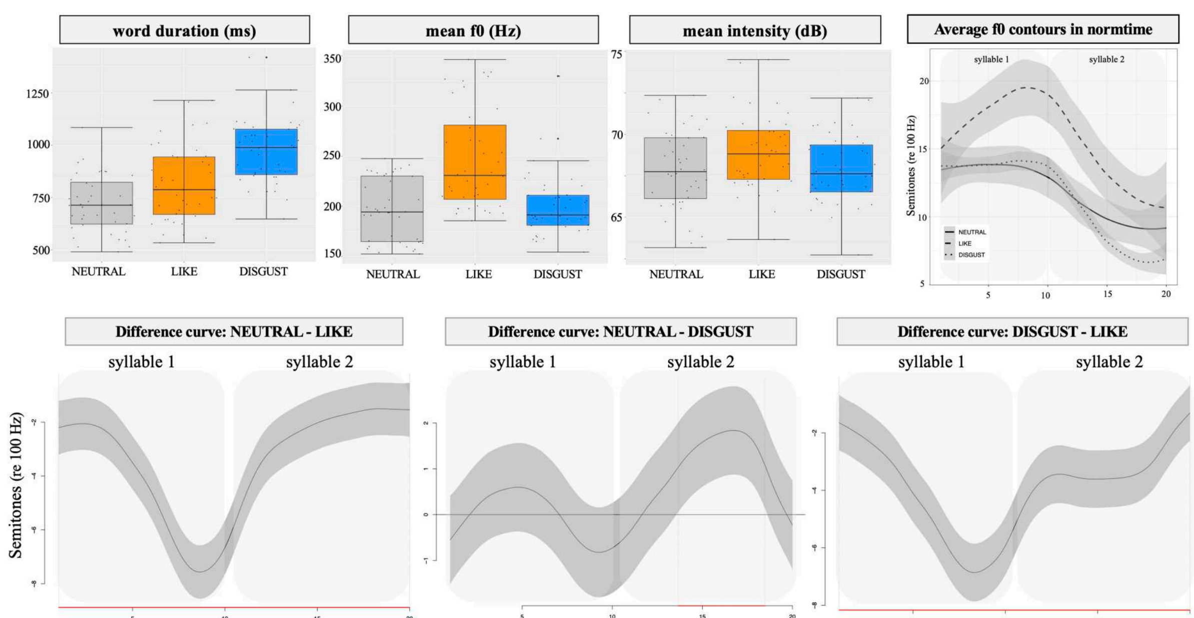

Figure 3. Boxplots of word duration, mean F0, mean intensity, F0 contours, and difference curves for emotional prosodies (*NEUTRAL*, *LIKE*, and *DISGUST*) in German. These significant differences are indicated by red vertical lines in the difference curve plot and the dark grey shading displays the 95% CI of the predicted mean difference.

the responses for *LIKE* and *DISGUST* to “C” and “M” with either the right or left index finger in a counterbalanced way across participants. The experiment started with a practice block consisting of 10 trials, followed by twelve blocks adding up to 120 trials (40 per condition). The sequence of the trials was randomised for each participant. No feedback was provided during or after the experiment.

Behavioural analysis

As in Experiment I, we did not analyse the reaction time due to the experiment design. The accuracy scores were calculated using the same generalised logistic mixed-effects regression models as in Experiment I. The variable CONDITION (*NEUTRAL*, *LIKE*, *DISGUST*) and GROUP (*L1 group* and *L2 group*) were entered as fixed effects.

ERP analysis

The electrophysiological data underwent pre-processing and analysis identical to that of Experiment I. Statistical comparisons between conditions are based on time windows around the functionally relevant ERP components (N100, P200, N400, and LPC) and visual inspection of the grand mean ERP data. We calculated the averaged ERPs obtained from four selected time windows (TW1: N100, *80–120* ms, TW2: P200, *165–215* ms, TW3: N400, *300–500* ms, TW4: LPC, *500–900* ms *post stimulus onset*) using linear mixed regression models as in Experiment I.

To construct the base model, PARTICIPANTS and SPEAKERS (the three female native German speakers: *speaker1*, *speaker2*, *speaker3*) were considered as a random intercept. Meanwhile, the fixed factors included CONDITION (*NEUTRAL*, *LIKE*, *DISGUST*) and GROUP (*L1 group* and *L2 group*). As in Experiment I, distributional differences were calculated by the factor REGION (*anterior*, *central*, and *posterior*). Trials containing eye blinks or movement artifacts were omitted from the ERP analysis. An average of 4.3% (SD = 3.2%) of trials from the L1 group and 12.6% (SD = 5.7%) from the L2 group were rejected as artifacts. To maintain sufficient trial numbers for reliable ERP averaging, trials were not excluded based on response accuracy. The R codes and the anonymized data can be accessed via OSF.

Results

Behavioural results

Participants of the L1 group categorised the emotional prosody *LIKE* correctly in 92% of cases (SD = 27%), *DISGUST* in 93% (SD = 26%), and *NEUTRAL* in 92% (SD = 27%). The L2 group performed differently, and they recognised *NEUTRAL* with an accuracy of about 88% (SD = 32%), *LIKE* about 70% (SD = 46%), and *DISGUST* with about 61% (SD = 49%). Interestingly, the L2 group misinterpreted *DISGUST* as a neutral emotion in 24% of cases and as a positive emotion in 15% of cases (see Figure 4).

The logistic mixed-effects regression analysis showed that there was a significant effect of GROUP ($\chi^2 = 5.795$, $p < .05$) and an interaction between CONDITION and GROUP ($\chi^2 = 53.74$, $p < .001$). The L1 group performed better than the L2 group in categorising the emotional prosodies ($ß = -2.21$, SE = .17, $z = -12.99$, $p < .001$). There were significant differences between the two groups in categorising *LIKE* ($ß = -1.70$, SE = .16, $z = -10.38$, $p < .001$) and *DISGUST* ($ß = -2.21$, SE = .17, $z = -12.99$, $p < .001$). Furthermore, within-group analysis revealed no significant differences between conditions in the L1 group, while the L2 group exhibited significant

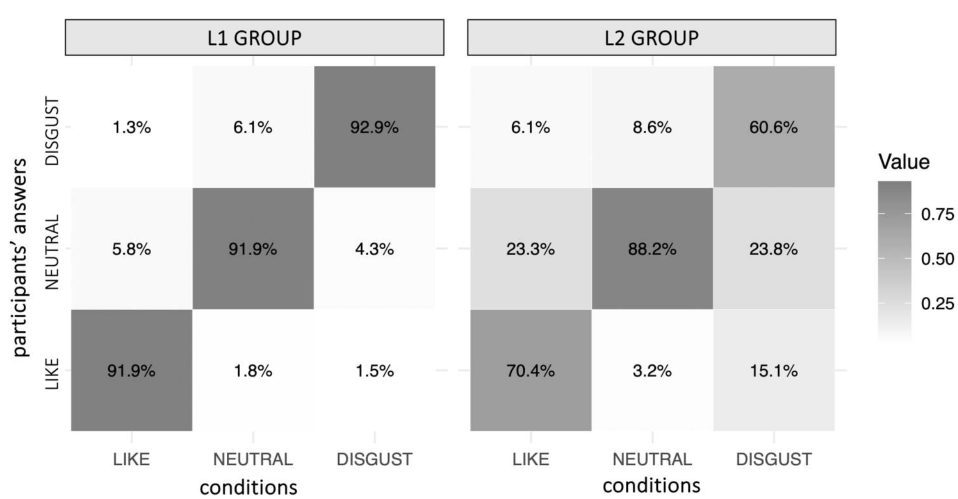

Figure 4. Matrix of behavioural results of categorising emotional prosodies for both groups.

differences between conditions in the categorisation of *LIKE* and *NEUTRAL* (ß = −1.27, SE = .15, $z = -8.57$, $p < .001$), *DISGUST* and *NEUTRAL* (ß = −1.77, SE = .15, $z = -12.01$, $p < .001$), and *LIKE* and *DISGUST* (ß = .50, SE = .12, $z = 4.15$, $p < .001$).

ERP results

The ERP results are illustrated in Figure 5. For the TW1 (N100: 80–120 ms post stimulus onset), the linear mixed regression models revealed significant effects of CONDITION ($\chi^2 = 8.38$, $p < .05$) and of GROUP ($\chi^2 = 5.11$, $p < .05$), with the brain responses evoked by the *DISGUST* condition being more negative than the brain responses for the *LIKE* condition (ß = .79, SE = .29, $d = .22$, $t = -2.74$, $p < .05$). This difference was larger in L1 group than in the L2 group (ß = .52, SE = .23, $d = .15$, $t = -2.24$, $p < .05$). There was no interaction between GROUP and CONDITION, or between CONDITION and REGION.

During the TW2 (P200: 165–215 ms post stimulus onset), linear mixed regression models revealed a significant effect of CONDITION ($\chi^2 = 7.94$, $p < .05$), with the brain responses for the *NEUTRAL* condition showing a more positive waveform in both groups than the brain responses for the *LIKE* condition (ß = −.86, SE = .34, $d = .86$, $t = -2.5$, $p < .05$). Additionally, there was a significant effect of GROUP ($\chi^2 = 7.65$, $p < .01$), indicating the effect was larger in the L1 group than in the L2 group (ß = .73, SE = .28, $d = .17$, $t = 2.6$, $p < .01$). However, there were no significant interactions between GROUP and CONDITION, or between CONDITION and REGION.

For the TW3 (N400: 300–500 ms post stimulus onset), the analyses revealed no significant effects or interactions.

Finally, regarding the TW4 (500–900 ms post stimulus onset), linear mixed regression models revealed a significant effect of CONDITION ($\chi^2 = 10.89$, $p < .01$), but no interactions. A significant difference in brain responses between *NEUTRAL* and *DISGUST* conditions was observed (ß = .66, SE = .27, $d = .2$, $t = 2.5$, $p < .05$), with the brain responses for the *NEUTRAL* condition showing a more positive waveform in both groups. The difference in brain responses between *NEUTRAL* and *LIKE* conditions was significant (ß = .83, SE = .27, $d = .83$, $t = 3.1$, $p < .01$), with the *LIKE* condition eliciting a larger negative deflection compared to the *NEUTRAL* condition in both groups.

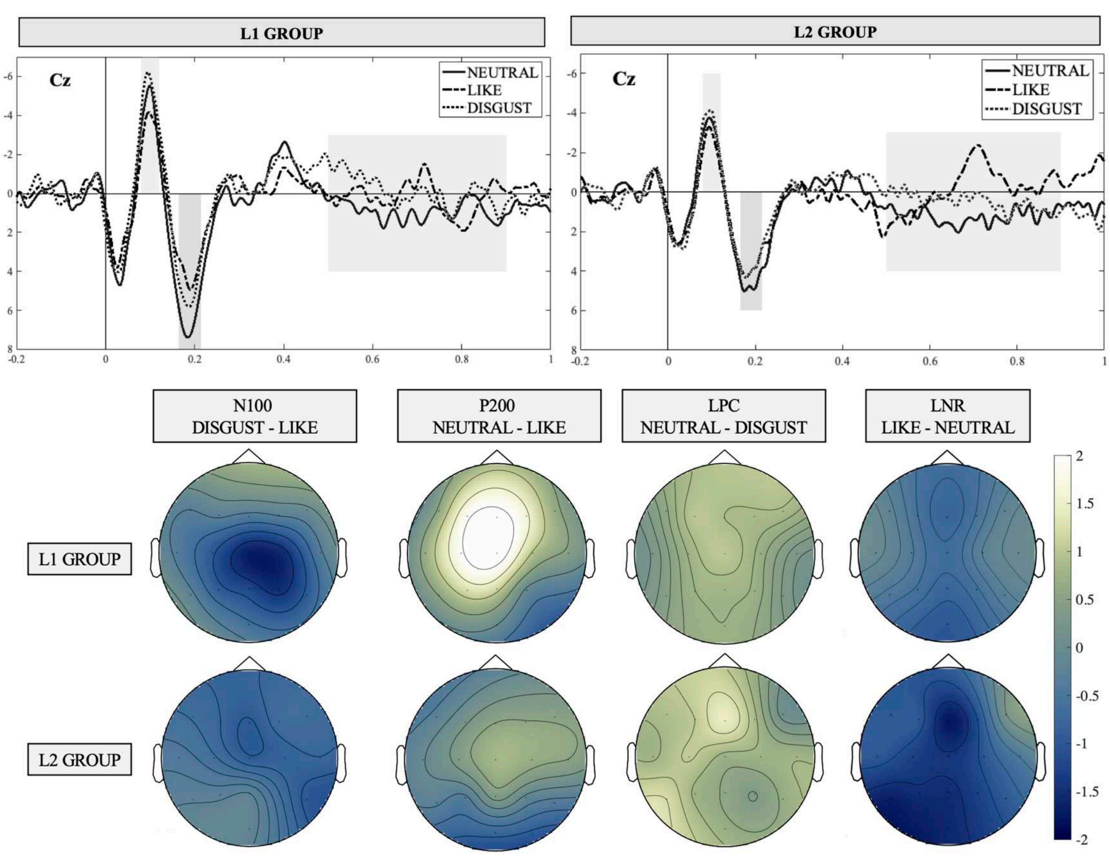

Figure 5. Grand-average ERPs for both groups for the classification of emotional prosody in an epoch ranging from 200 ms before stimulus onset up to 1000 ms post-stimulus onset at the Cz electrode. The time windows with significant effects are indicated. Scalp maps display amplitude differences for both groups at P200 and LPC/LNR time windows.

Discussion of experiment II

In the second experiment, we tested the emotional prosodic processing between L1 German speakers and Chinese-German L2 learners. The main findings were as follows: (1) The *DISGUST* condition exhibited significantly greater negativity in the 80–120 ms time window compared to the *LIKE* condition; (2) in the 165–215 ms time window, the *NEUTRAL* condition showed a significantly larger positivity relative to the *LIKE* condition; and (3) a late positive complex (LPC, 500–900 ms) was significantly larger for the *NEUTRAL* condition compared to the *DISGUST* condition, suggesting that *NEUTRAL* condition conveys minimal emotional charge and reduces cognitive load. Conversely, the *LIKE* condition evoked a significantly larger late negative response (LNR, 500–900 ms) than the *NEUTRAL* condition, reflecting more extensive evaluative processing. The late positive complex and increased negative response highlight processes involved in identification of emotional meaning and evaluation of emotional prosody processing (Kotz & Paulmann, 2011; Pell & Kotz, 2021; Schirmer et al., 2002). Notably, the findings revealed no interactions between CONDITION and GROUP, suggesting that emotional prosody processing does not significantly differ between L1 speakers and L2 learners. Nevertheless, distinct behavioural patterns between the two groups imply that they may rely on different cues during the evaluation processes.

General discussion

Comparison of linguistic and emotional prosody processing

The present study investigated whether linguistic prosody (*QUESTION, STATEMENT*) and emotional prosody (*NEUTRAL, LIKE, DISGUST*) processing exhibit comparable temporal patterns in native German speakers (L1 group) and Chinese L2 learners of German (L2 group). In the linguistic prosody perception study, the *QUESTION* condition exhibited shorter word duration but higher intensity, as well as higher F0 contours in the 2nd and 3rd syllables compared to the *STATEMENT* condition, as shown in the Figure 1. In the emotional prosodic perception study, the *LIKE* condition had higher F0 contour compared to the *NEUTRAL* and the *DISGUST* conditions during the utterances, while the *NEUTRAL* condition had higher pitch contour than the *DISGUST* condition in the 2nd syllable. Additionally, the *DISGUST* condition had longer word duration than the *NEUTRAL* condition (see Figure 3).

Our EEG results provide support for the "three-stage model" for vocal expression processing (Pell & Kotz, 2021). The first stage of prosody processing involves the basic level of sensory encoding and is represented by the N100 component, which is sensitive to changes in the sensory environment and imparts structure to the incoming acoustic stream (Näätänen & Picton, 1987; Rigoulot et al., 2015). The N100 component observed during the processing of emotional prosody showed significantly larger amplitudes for the *DISGUST* condition compared to the *LIKE* condition. This amplitude difference may have been elicited by the significant F0 contour differences in the first syllable between these two conditions. This finding aligns with earlier studies (Frank et al., 2020; Rinke et al., 2023) that reported a reverse or negative correlation between F0 and N100 amplitude: higher F0 values correspond to lower N100 amplitudes. In contrast, the absence of F0 contour differences in the first syllable between the *STATEMENT* and *QUESTION* conditions may be the main reason for the lack of the N100 component during the processing of linguistic prosody.

The second stage involves the processing of vocal features of the stimuli, as indicated by the enhanced P200 component for the processing of the linguistic prosody *STATEMENT* compared to the *QUESTION* condition, and for the *NEUTRAL* condition compared to the *LIKE* condition. The P200 component is associated with attentional processes and perceptual aspects of auditory stimuli and is involved in the salience detection, differentiating between various sound characteristics (Paulmann et al., 2011, 2013; Steber et al., 2020). The P200 component reflects higher-order auditory processing and demonstrates sensitivity to the acoustic salience of prosodic patterns in both groups. Specifically, the enhanced P200 response observed for linguistic prosody in the *STATEMENT* condition suggests the salience detection and differentiation of prosodic features related to sentence mode. Similarly, the increased P200 component observed for the *NEUTRAL* condition in the emotional prosody categorisation task indicates early differentiation of prosodic cues linked to emotional valence. Results from both studies highlight the ability to detect and process modulations in acoustic properties, such as F0 contours and intensity in linguistic and emotional prosody.

The third stage marks the cognitive analysis and interpretation of vocal expressions. During linguistic prosody processing, both groups in our study exhibited an enhanced N400 component for the *QUESTION* condition compared to the STATEMENT condition, reflecting increased cognitive effort required to integrate prosodic information for identifying interrogative intent. These integration challenges may be attributed to F0 contour differences between the *QUESTION* and *STATEMENT* conditions in the 2nd and 3rd syllables, as illustrated in Figure 1. Following the N400 component, we observed a larger P600-like positivity (500–800 ms) for the *QUESTION* condition compared to the *STATEMENT* condition in the L1 group. This finding suggests that L1 listeners engage in detailed processing and reanalysis of prosodic cues to confirm the interrogative prosody in the *QUESTION* condition. Similar P600-like positivity driven by prosodic context have been reported by Vergis et al. (2020) for polite stance prosody compared to rude prosody, and by Rigoulot et al. (2014), who identified a P600 effect for sincere prosody compared to insincere prosody during a judgement task.

During emotional prosody processing, both groups evoked an enhanced LPC for the *NEUTRAL* condition compared to the *DISGUST* condition, and a larger LNR for the *LIKE* condition compared to the *NEUTRAL* condition. These effects reflect higher-order processing of emotional ambiguity or reanalysis of prosodic cues. The LPC is typically associated with emotional or motivational processing, and reflects sustained attention, memory, and evaluative processes (Bayer & Schacht, 2014; Lin et al., 2022; Pell et al., 2015). The *NEUTRAL* condition, conveying minimal emotional charge compared to *DISGUST* or *LIKE* conditions, resulted in reduced cognitive load. The increased LNR observed for the *LIKE* condition suggests more extensive evaluative processing due to its higher emotional salience (Jiang et al., 2019; Yu et al., 2021).

Our results suggest that the late effects are not solely due to the acoustic differences between the conditions but reflect higher cognitive reanalysis of the incoming information. These findings align with previous research (Paulmann et al., 2012; Zora et al., 2020), indicating that linguistic and emotional prosody processing unfolds differently over time and involves distinct underlying processes.

Language- and culture-specific effects on prosody processing

In this part of the discussion section, we aim to develop an explanation for the findings that (i) the processing of linguistic prosody differs between L1 speakers and L2 learners, (ii) the categorisation of emotional prosody is demanding for L2 learners, and (iii) the language status affects the processing of linguistic and emotional prosody differently.

In recognising linguistic prosody, behavioural data revealed high accuracy scores for both groups and showed no significant interactions or differences between the two groups. However, during the cognitive identification and interpretation process of linguistic prosody, a significant GROUP difference in the N400 time window and an interaction between CONDITION and GROUP in the P600 time window were observed. Specifically, the L2 group exhibited an enlarged N400 for the *QUESTION* condition compared to the L1 group. The N400 amplitude decreases when word or stimuli are contextually predictable, as it reflects prediction errors arising from deviations between expected and actual linguistic input (Eddine et al., 2024; Lau et al., 2013). In our study, listeners encountered one-word stimuli with *QUESTION* intonation, which deviated from the expected statement intonation. This expectation may be stronger in German (an intonation language) than Mandarin (a tone language), where question intonation interacts with lexical tone. The L2 group's increased N400 for the *QUESTION* condition suggests difficulties in integrating L2 prosodic features with their existing linguistic knowledge. Furthermore, only the L1 group elicited a larger P600-like positivity for the *QUESTION* condition, suggesting the engagement of "higher-level" processes potentially influenced by language usage. The absence of the late positivity in the L2 group may be because their proficiency in German is still not comparable to that of native German speakers, and they could not use the prosodic cues to reanalyse or repair the morpho-syntactic structures in German as the native German speakers do (Kaan et al., 2024; Pélissier, 2020). Similar results between L1 speakers and L2 learners were also shown in EEG studies comparing the processing of native and second language at the lexico-semantic level (Bermúdez-Margaretto et al., 2022) and ambiguity in sentence processing with auditory stimuli (Pélissier, 2020).

In the categorisation of emotional prosody in German, the behavioural results showed significant differences between L1 and L2 groups and interactions between CONDITION and GROUP. The group effect in the categorisation of emotions demonstrated a clear advantage for the German L1 listeners, who recognised all emotion conditions with high accuracy scores, while the group of Chinese-German L2 learners showed particular difficulties in categorising the emotional prosodies *DISGUST* and *LIKE* (as shown in Figure 4). This result contrasts findings from the Prosody-Analysis-Test, where L2 learners achieved high accuracy (98.38%) in categorising the emotions *FEAR* and *JOY*. However, in the emotional categorisation task, the L2 learners showed a strong bias towards the neutral option. Specifically, they misinterpreted *DISGUST* in almost 24% of cases as *NEUTRAL* and 15% as *LIKE* and miscategorised *LIKE* in 23% of cases as *NEUTRAL* and 6% as *DISGUST*. This discrepancy between the high accuracies observed in the Prosody-Analysis-Test and the emotional categorisation task may be attributable to differences in task and condition demands. The Prosody-Analysis-Test involved only two conditions (positive vs. negative). In contrast, the EEG study introduced an additional *NEUTRAL* condition increasing the complexity of the categorisation task. In fact, many misclassifications of *LIKE* and *DISGUST* involve the *NEUTRAL* response.

A further possible explanation for the high miscategorisation of *DISGUST* may be the culture-specific, emotionally relevant rule to hide negative feelings (Paulmann & Uskul, 2013), which may have reduced the number of hit responses for the negative emotions. This cultural influence may have exacerbated L2 learners' challenges in distinguishing between *DISGUST*, *LIKE*, and *NEUTRAL* conditions compared to *FEAR* and *JOY*. Additionally, phonetic differences in expressing emotional prosodies between German and Chinese may have led to difficulties in categorising and interpreting emotional prosody *LIKE* in L2 German. In Chinese, expressions of positive emotion are produced with a higher pitch but shorter duration than neutral or negative emotions (Erickson et al., 2016; Wei et al., 2023). The *LIKE* condition in the present study was characterised by a higher pitch compared to the *DISGUST* and *NEUTRAL* conditions (see Figure 3), but its duration did not show a significant difference from the *NEUTRAL* condition, which is unusual for the respective conditions in Chinese. Furthermore, L2 learners may have been less emotionally engaged with the L2 material at an experiential level. While the arousal of the stimuli may have been sufficient for L1 listeners, it may not have elicited the same emotional resonance in L2 listeners, leading to categorisation difficulties in the behavioural data. On the other hand, no interaction between CONDITION and GROUP was found in the ERP results. Given the inconsistent behavioural and electrophysiological data, the present article cannot contribute to the disembodied account of L2 processing (Keysar et al., 2012), in which emotional prosody in a second language may not evoke the same internal resonance as in the native language.

It remains to be seen whether the culture- and language-specific paralinguistic patterns observed at the behavioural level in decoding vocal emotions also manifest as electrophysiological reflections in processing of emotional prosody. In our study, no significant interactions between CONDITION and GROUP were observed across any time window in the electrophysiological data. Both Chinese L2 learners of German and German native speakers elicited the same three components (N100, P200, and LPC/LNR) in response to the same conditions, though the amplitudes varied between the two groups. In the early stage, the emotional meaning of the auditory input is derived based on acoustic cues, including a valence tagging process. The later stage, reflected in the LPC/LNR components, involves the construction of emotional meaning.

ERP measures can be highly sensitive to the time course of cognitive processes and can reveal similarities in neural processing that are not evident in behavioural performance (Luck, 2014; Luck & Kappenman, 2013). The conflicting results of the electrophysiological data and the behavioural results of the emotional prosody in the L2 group suggest that both L1 and L2 groups might be engaging similar neural mechanisms when processing emotional prosody in German. The L2 group's worse behavioural performance might be due to less experience with emotional prosody in German and reduced ability to use the prosodic cues as effectively as native German speakers. The performance in the behavioural task is not reflected in ERPs with similar brain responses for emotional prosody cues for both groups.

Limitations

While our study sheds light on the differences in linguistic and emotional prosody processing between German L1 speakers and Chinese-German L2 learners with advanced German proficiency, some limitations of our study should be pointed out and addressed in further research. Firstly, we tested only one group of German L2 speakers with a tone L1, Mandarin Chinese. Therefore, it is important to investigate L2 learners of German whose native language is non-tone to assess which role the tone contrast between Chinese and German has played in our experiment. Is there a confound of tone processing at the lexical and intonational level that impedes the processing of linguistic and emotional prosody in the L2 group? Secondly, the Chinese-German L2 learners in our study lived in Germany and had an advanced German proficiency level. The question that follows is how prosodic abilities develop in the L2. Testing learners with different L2 proficiency levels will help us understand the prosodic learning processes and whether there are correlations between L2 proficiency and ERPs, which might reflect a "learning effect" on the processing level of prosody. Third, the order of the two experiments was not counterbalanced. This may have led to increased familiarity with the task during the emotional prosody experiment. Future studies could address this by counterbalancing the task order to control for potential order effects. Additionally, our study only examined three emotional prosodies.

It would be beneficial to expand the range of emotional prosodies tested in future studies. This would allow for a more comprehensive understanding of how emotional prosody processing may differ across languages and cultures. Furthermore, we acknowledge that the lack of explicit arousal measurements represents a limitation of the present study. Although the basic words were selected to be neutral in emotional valence and were validated for their intended emotional category, future research should include systematic arousal assessments of the recordings to complement valence and prosodic categorisation. Finally, while our study focused on L1 and L2 speakers of German, it would be interesting to explore how foreign speakers without knowledge of German may process the emotional prosody of the German language. This exploration would enable us to examine how consistent early and late processes of identification of emotional prosody are across different linguistic backgrounds.

Conclusion

The present study compared the perception of linguistic and emotional prosody between L1 German speakers and Chinese-German L2 learners. Our results indicate that the L1 speakers exhibited distinct neural responses when processing linguistic and emotional prosody, suggesting that these two types of prosodic functions are processed differently and rely on distinct neural mechanisms and temporal encoding in native speakers. On the other hand, the L2 learners showed somewhat different patterns of neural responses compared to the L1 group when processing the linguistic prosody. Specifically, the L2 group exhibited an increased N400 component for the *QUESTION* condition compared to the *STATEMENT* condition, indicating difficulties in integrating prosodic features with their existing linguistic knowledge of *QUESTION* intonation. In contrast, only the L1 group showed a larger P600-like positivity for the *QUESTION* condition compared to the *STATEMENT* condition, suggesting that native speakers engage in a reanalysis process, allowing them to re-evaluate the prosodic information. However, during the processing of emotional prosody, they showed similar responses during the early and late stages when inferring emotional meanings from the acoustic input, indicating shared aspects in the identification and evaluation processes of emotional prosody. Still, the behavioural differences between the two groups indicate that the L2 group's ability to use the emotional prosodic cues in German is not as effective as that of L1 speakers, possibly due to less experience with emotional prosody in German. In summary, these findings highlight the complex interplay between language and culture in prosodic processing. Differences between behavioural and electrophysiological results further underscore the importance of investigating distinct measures to better understand the processes of L2 prosody.

Note

1. The EmoInt-2017-Database provides information about the emotional valence of German words ranging from very negative to very positive on a 9-point scale, with "1" referring to very negative and "9" to very positive.

Acknowledgements

The study was approved by the ethics committee of the German Society of Linguistics (DGfS). All participants provided informed consent prior to their participation, in accordance with the ethical standards of the committee and the principles of the Declaration of Helsinki. We thank Stefanie Türk, Jonas Gerards, and Seung-Goo Kim for their helpful discussion on the ERP data analysis, Katharina Zahner-Ritter for her help in the GAMM analysis, as well as two anonymous reviewers for their valuable comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research is supported by the University of Marburg and a doctoral fellowship (Konrad-Adenauer-Stiftung awards scholarships) awarded to Huan Wei. The funding source was not involved in the design or conduct of this study.

References

Bartoń, K. (2022). *MuMIn: Multi-model inference*. R package version 1.47.1.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv*.

Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces – a cross-domain comparison. *Frontiers in Psychology*, *5*, 1106. https://doi.org/10.3389/fpsyg.2014.01106

Bermúdez-Margaretto, B., Gallo, F., Novitskiy, N., Myachykov, A., Petrova, A., & Shtyrov, Y. (2022). Ultra-rapid and automatic interplay between L1 and L2 semantics in late bilinguals: EEG evidence. *Cortex*, *151*, 147–161. https://doi.org/10.1016/j.cortex.2022.03.004

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.37. https://www.fon.hum.uva.nl/praat/

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*(Suppl 6), 1318–1352. https://doi.org/10.1111/cogs.12461

Chang, H.-S., Lee, C.-Y., Wang, X., Young, S.-T., Li, C.-H., & Chu, W.-C. (2023). Emotional tones of voice affect the acoustics and perception of Mandarin tones. *PLoS One*, *18*(4), e0283635. https://doi.org/10.1371/journal.pone.0283635

Chang, J., Zhang, X., Zhang, Q., & Sun, Y. (2018). Investigating duration effects of emotional speech stimuli in a tonal language by using event-related potentials. *IEEE Access*, *6*, 13541–13554. https://doi.org/10.1109/access.2018.2813358

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Eddine, S. N., Brothers, T., Wang, L., Spratling, M., & Kuperberg, G. R. (2024). A predictive coding model of the N400. *Cognition*, *246*, 105755. https://doi.org/10.1016/j.cognition.2024.105755

Erickson, D., Zhu, C., Kawahara, S., & Suemitsu, A. (2016). Articulation, acoustics and perception of mandarin Chinese emotional speech. *Open Linguistics*, *2*, 620–635. https://doi.org/10.1515/opli-2016-0034

Filippa, M., Lima, D., Grandjean, A., Labbé, C., Coll, S. Y., Gentaz, E., & Grandjean, D. M. (2022). Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. *Scientific Reports*, *12*, 17144. https://doi.org/10.1038/s41598-022-21554-0

Frank, M., Muhlack, B., Zebe, F., & Scharinger, M. (2020). Contributions of pitch and spectral information to cortical vowel categorization. *Journal of Phonetics*, *79*, 1–13. https://doi.org/10.1016/j.wocn.2020.100963

Gibbon, D. (1997). *Intonation in German* (pp. 78–95). University Bielefeld.

Gong, B., Li, N., Li, Q., Yan, X., Chen, J., Li, L., Wu, X., & Wu, C. (2023). The Mandarin Chinese auditory emotions stimulus database: A validated set of Chinese pseudo-sentences. *Behavior Research Methods*, *55*(3), 1441–1459. https://doi.org/10.3758/s13428-022-01868-7

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2019). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79. https://doi.org/10.1177/1747021819865833

Joe, H.-K., Hiver, P., & Al-Hoorie, A. H. (2017). Classroom social climate, self-determined motivation, willingness to communicate, and achievement: A study of structural relationships in instructed second language settings. *Learning and Individual Differences*, *53*, 133–144. https://doi.org/10.1016/j.lindif.2016.11.005

Kaan, E., Dai, H., & Xu, X. (2024). Adaptation in L2 sentence processing: An EEG study. *Second Language Research*, *40*(4), 887–910. https://doi.org/10.1177/02676583231192169

Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological Science*, *23*(6), 661–668. https://doi.org/10.1177/0956797611432178

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, *43*(9), 2743–2758. https://doi.org/10.1002/hbm.25832

Köper, M., Kim, E., & Klinger, R. (2017, September). IMS at EmoInt-2017: Emotion intensity prediction with affective norms, automatically extended resources and deep learning. In *Proceedings of the 8th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pp. 50–57.

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

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*(3), 108–125. https://doi.org/10.1111/j.1749-818x.2010.00267.x

Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2024). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*, 1–16. https://doi.org/10.1177/17456916231217722

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*(3), 484–502. https://doi.org/10.1162/jocn_a_00328

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). emmeans: Estimated marginal means, aka least-squares means. R package version 1.8.5. https://github.com/rvlenth/emmeans

Lin, Y., Fan, X., Chen, Y., Zhang, H., Chen, F., Zhang, H., Ding, H., & Zhang, Y. (2022). Neurocognitive dynamics of prosodic salience over semantics during explicit and implicit processing of basic emotions in spoken words. *Brain Sciences*, *12*, 1706. https://doi.org/10.3390/brainsci12121706

Liu, G. (2016). The Syllabus of school of foreign languages UJN. 《基础德语1》教学大纲-济南大学外国语学院. http://wyxy.ujn.edu.cn/info/1036/3694.htm

Liu, M., Chen, Y., & Schiller, N. O. (2022). Context matters for tone and intonation processing in mandarin. *Language and Speech*, *65*(1), 52–72. https://doi.org/10.1177/00238309920986174

Lu, X., Ho, H. T., Liu, F., Wu, D., & Thompson, W. F. (2015). Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: An ERP study. *Frontiers in Psychology*, *6*, 385. https://doi.org/10.3389/fpsyg.2015.00385

Luck, S. (2014). *An introduction to the event-related potential technique*. MIT Press.

Luck, S., & Kappenman, E. (2013). *The Oxford handbook of event-related potential components*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.001.0001

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://doi.org/10.3389/fpsyg.2013.00345

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, *50*(7), 1609–1620. https://doi.org/10.1016/j.neuropsychologia.2012.03.014

Paulmann, S., Ott, D. V. M., & Kotz, S. A. (2011). Emotional speech perception unfolding in time: The role of the Basal Ganglia. *PLoS One*, *6*(3), e17694. https://doi.org/10.1371/journal.pone.0017694

Paulmann, S., & Uskul, A. K. (2013). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition and Emotion*, *28*(2), 230–244. https://doi.org/10.1080/02699931.2013.812033

Pélissier, M. (2020). Comparing ERPs between native speakers and second language learners: Dealing with individual variability. In A. Edmonds, P. Leclercq, & A. Gudmestad (Eds.), *Interpreting language-learning data* (pp. 39–69). Language Science Press. https://doi.org/10.5281/zenodo.4032282

Pell, M. D., & Kotz, S. A. (2021). Comment: The next frontier: Prosody research gets interpersonal. *Emotion Review*, *13*(1), 51–56. https://doi.org/10.1177/1754073920954288

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rigoulot, S., Fish, K., & Pell, M. D. (2014). Neural correlates of inferring speaker sincerity from white lies: An event-related potential source localization study. *Brain Research*, *1565*, 48–62. https://doi.org/10.1016/j.brainres.2014.04.022

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184. https://doi.org/10.1016/j.neuroscience.2015.01.033

Rinke, P., Lavan, N., & Scharinger, M. (2023). Interactions of speech and speaker processing in early evoked components. In R. Skarnitzl, & J. Volin (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 3992–3996). Guarant International.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/s0167-6393(02)00084-5

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*(8), 1135–1148. https://doi.org/10.1162/089892903322598102

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

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233. https://doi.org/10.1016/s0926-6410(02)00108-8

Smith, L., & King, J. (2018). *Silence in the foreign language classroom: The emotional challenges for L2 teachers* (pp. 323–339). Emotions in second language teaching: theory, research and teacher education. https://doi.org/10.1007/978-3-319-75438-3_18

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017. https://doi.org/10.1016/j.wocn.2020.101017

Steber, S., König, N., Stephan, F., & Rossi, S. (2020). Uncovering electrophysiological and vascular signatures of implicit emotional prosody. *Scientific Reports*, *10*, 5807. https://doi.org/10.1038/s41598-020-62761-x

Tomasello, R., Grisoni, L., Boux, I., Sammler, D., & Pulvermüller, F. (2022). Instantaneous neural processing of communicative functions conveyed by speech prosody. *Cerebral Cortex*, *32*(21), 4885–4901. https://doi.org/10.1093/cercor/bhab522

van Rij, J., van Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 1–22. https://doi.org/10.1177/2331216519832483

van Rij, J., van Wieling, M., Baayen, R. H., & van Rijn, D. (2017). *itsadug: Interpreting time series and autocorrelated data using GAMMs*. https://research.rug.nl/en/publications/itsadug-interpreting-time-series-and-autocorrelated-data-using-ga

Vergis, N., Jiang, X., & Pell, M. D. (2020). Neural responses to interpersonal requests: Effects of imposition and vocally-expressed stance. *Brain Research*, *1740*, 146855. https://doi.org/10.1016/j.brainres.2020.146855

Walther, W., & Otten, M. (2016). *ProsA: Prosodie-Analyse: ein computergestütztes Verfahren zur Erfassung rezeptiver prosodischer Fähigkeiten: Manual*. Hogrefe.

Wang, H., Peng, A., & Patterson, M. M. (2021). The roles of class social climate, language mindset, and emotions in predicting willingness to communicate in a foreign language. *System*, *99*, 102529. https://doi.org/10.1016/j.system.2021.102529

Wei, H., Scharinger, M., & Domahs, U. (2023). Production and perception of vocal emotions: A comparison of mandarin Chinese and German emotional prosody. In R. Skarnitzl, & J. Volin (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 4086–4090). Guarant International.

Wickens, S., & Perry, C. (2015). What do you mean by that?! An electrophysiological study of emotional and attitudinal prosody. *PLoS One*, *10*(7), e0132947. https://doi.org/10.1371/journal.pone.0132947

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. N. (2006). *Generalized additive models*. Chapman and Hall/CRC. https://doi.org/10.1201/9781420010404

Wood, S. N. (2017). *Generalized additive models, An introduction with R*. Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279

Xu, Y. (2013). *ProsodyPro - A tool for large-scale systematic prosody analysis*. TRASP.

Yu, L., Zeng, J., Wang, S., & Zhang, Y. (2021). Phonetic encoding contributes to the processing of linguistic prosody at the word level: Cross-linguistic evidence from event-related potentials. *Journal of Speech, Language, and Hearing Research*, *64*(12), 4791–4801. https://doi.org/10.1044/2021_jslhr-21-00037

Yuan, J., Shen, L., & Chen, F. (2002). *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, pp. 2025-2028. Denver, Colorado, USA. https://doi.org/10.21437/icslp.2002-556

Zahner-Ritter, K., Chen, Y., Dehé, N., & Braun, B. (2022). The prosodic marking of rhetorical questions in standard Chinese. *Journal of Phonetics*, *95*, 101190. https://doi.org/10.1016/j.wocn.2022.101190

Zhang, M., Zhang, H., Tang, E., Ding, H., & Zhang, Y. (2023). Evaluating the relative perceptual salience of linguistic and emotional prosody in quiet and noisy contexts. *Behavioral Sciences*, *13*(10), 800. https://doi.org/10.3390/bs13100800

Zora, H., Rudner, M., & Magnusson, A. K. M. (2020). Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *European Journal of Neuroscience*, *51*(11), 2236–2249. https://doi.org/10.1111/ejn.14658