



OPEN ACCESS

EDITED BY

Lars Kuchinke,
International Psychoanalytic University Berlin,
Germany

REVIEWED BY

Isabella Poggi,
Roma Tre University, Italy
Peng Wang,
VU Amsterdam, Netherlands

*CORRESPONDENCE

Mengjiao Wu
✉ mjwu@shmtu.edu.cn

RECEIVED 19 July 2025

ACCEPTED 25 September 2025

PUBLISHED 11 November 2025

CITATION

Wu M and Roche JM (2025) Emotion, proficiency, and arousal: exploring speech and physiological responses in Chinese ESL learners.
Front. Hum. Neurosci. 19:1653894.
doi: 10.3389/fnhum.2025.1653894

COPYRIGHT

© 2025 Wu and Roche. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Emotion, proficiency, and arousal: exploring speech and physiological responses in Chinese ESL learners

Mengjiao Wu^{1*} and Jennifer M. Roche²

¹College of Foreign Languages, Shanghai Maritime University, Shanghai, China, ²Speech Pathology and Audiology, Kent State University, Kent, OH, United States

Introduction: The coordination and expression of cultural specific affective cues during speech production in a second language (L2) reflects pragmatic adaptation, which is a critical step toward learning and achieving broader pragmatic competence. Embodied cognition provides a framework for understanding how cognitive and emotional processes shape L2 expression.

Objective: This study examined how immersive language experience influences pragmatic adaptation through the vocal expression of affect and physiological arousal in Chinese ESL learners.

Methods: Acoustic analysis and electrodermal activity (EDA) measurements were used to assess affectively valenced word production in speakers with varying levels of immersive English experience.

Results: High-immersion speakers exhibited greater pitch, intensity, and duration variation, enhancing emotional expressivity. Low-immersion speakers showed constrained vocal patterns and significantly higher physiological arousal, likely due to increased cognitive demands and anxiety.

Discussion: These findings highlight the impact of L2 proficiency on affective language embodiment and the cognitive challenges faced by L2 learners. This study offers novel insights by considering a pictorial character-based language, broadening our understanding of emotion-language interaction. Findings have implications for second-language education, cross-cultural communication, and bilingual speech therapy.

KEYWORDS

embodied cognition, physiological arousal, L2 speech, electrodermal activity (EDA), acoustic, affective words

Introduction

Learning a language goes beyond the cognitive processes of acquiring words and grammar, and second language learning can be even more challenging, as mastering socio-pragmatic rules is difficult without an immersive context (e.g., Allami and Naeimi, 2011; Bialystok, 1993; Cook and Liddicoat, 2002). Even native speakers may not be aware of their own pragmatics, as formal education rarely teaches pragmatics as explicitly as vocabulary and grammar (Bardovi-Harlig, 2013; Wolfson, 1989). The ability to coordinate affect and prosody in vocal production provides a unique perspective to pragmatic competence in L2 acquisition. In second language acquisition (SLA), the development of pragmatic competence can be particularly challenging in the absence of formal instruction (Bardovi-Harlig and Dörnyei, 1998; Mokoro, 2024; Wyner, 2014), which makes interpreting the pragmatic function of prosody even harder (Levis, 1999; Celce-Murcia et al., 2010).

Nevertheless, L2 learners may develop pragmatic competence through pragmatic transfer and adaptation (Costa et al., 2008; Kasper, 2001; Trofimovich and Isaacs, 2016), which involves adjusting one's communicative behavior to fit the social context and using one's understanding of these cues to guide adaptation in a cross-cultural setting. For instance, adapting one's affective prosody—a key component of pragmatic competence, encompassing intonation modulation, appropriate pausing and intensity, and timing of affective cues—is a skill L2 learners must develop (Kermad, 2021). They may achieve this by leveraging their understanding of culturally shaped affective cues (Kasper, 2001). Therefore, through pragmatic transfer and adaptation, by aligning their vocal cues with those of cultural counterparts, learners gradually acquire the ability to convey and recognize emotions and attitudes in a non-native cultural context—likely being driven by mechanisms of transfer (Kasper, 2001) and interactive alignment (Costa et al., 2008).

Learners may also draw on inherent cognitive processes, such as emotional processing, which integrates the bi-directional nature of cognitive appraisals with physiological reactions (e.g., *appraisal theory*; Ellsworth, 1991, 2013; Moors et al., 2013; Russell, 2003; Scherer, 1999) through embodiment (Barrett and Lindquist, 2008; Titchener, 1914), such that appraisals and bodily responses interact during emotional processing (Jerath and Beveridge, 2020; Manstead and Wagner, 1981; Nummenmaa et al., 2012; Schachter and Singer, 1962). Leveraging the mind–body interaction during language acquisition, especially in an immersive learning context, L2 (second language) learners may more easily express affectively valenced language (Kosmas and Zaphiris, 2020; Lan et al., 2015). For instance, when learners physically and emotionally engage with language—like feeling joy while using joyful words—it can strengthen the connection between emotion and language, helping them understand and use affective words more quickly through an embodied experience (Kissler and Herbert, 2013; Yu et al., 2021; Scott et al., 2012). While extensive research has demonstrated that immersive contexts promote fluency and competence in L2 speech (e.g., Porter and Castillo, 2023; Freed et al., 2004; Nicolay and Poncet, 2013; Segalowitz and Freed, 2004), our study contributes to a growing body of literature that supports the notion that immersion and affect interact during lexical access and vocal production and by examining both the subjective (conscious interpretation and spoken representation) and objective experience (non-conscious physiological response) of emotion, supporting Wang et al.'s (2025a) argument that a dynamic, multimodal approach is essential for understanding emotional processes. Adopting an embodied cognition framework, we investigate how physiological responses and prosodic cues influence language production in an L2 context.

Background

During communicative interactions, pragmatics plays a critical role that shapes language in context, often from a sociocultural framework (Beltrama, 2020; Hasan, 2012; Holmes, 2018). L2 learners may benefit from immersive environments, by learning pragmatic skills from native speakers who facilitate pragmatic rules, thereby strengthening both lexical and contextual use and understanding of the L2 (Bardovi-Harlig and Hartford, 1993; Xiao, 2015). Pragmalinguistic and socio-pragmatic competence is essential for developing culturally and socially appropriate communication skills and enhancing effective cross-cultural exchanges (Byram, 1997; Kinginger and Belz, 2005).

Immersive language environments and real-world interactions not only facilitate the development of pragmatic competence through transfer and adaptation, but also allows language and communication to become embodied—i.e., the mind and body interact to shape our thoughts, actions, and even language (for review, see Barsalou, 2008; Louwerse and Jeuniaux, 2010; Wilson and Golonka, 2013). As language learners learn, they may engage intrinsic cognitive mechanisms (e.g., embodiment, emotional appraisal) that guide the development of strategies to ease language acquisition (Al-Hejin, 2004; Arnold, 2011; Ellis, 2006; MacIntyre and Vincze, 2017). For instance, Louwerse and Jeuniaux (2010) provide evidence for an embodied approach to language processing, demonstrating across four studies that when different facets of a representation—such as semantics and iconicity—are aligned, language processing is facilitated. These findings are consistent with *dual coding theory* (Paivio, 1990), which posits that verbal and non-verbal systems interact to enhance comprehension. Theories of embodied cognition emphasize that cognitive processes are deeply rooted in sensory and motor experiences, suggesting that both concrete and abstract concepts are understood through bodily interactions with the world (Barsalou et al., 2003; Dove, 2014; De Vega et al., 2012). Not only are physical objects understood through our interaction with them, but abstract ideas, like emotions, are grounded in how our body feels and reacts. For instance, abstract words often carry emotional weight, which ties them directly to the body's responses to the environment (Kousta et al., 2011). This integration of perception, action, and emotion highlights the dynamic collaboration between the brain, body, and environment. Language comprehension and production, therefore, are not isolated mental activities but are intertwined with physical and emotional experiences.

Emotional or affective expression (e.g., facial and vocal gestures; Pell et al., 2009; Sauter et al., 2010; Scherer et al., 2001), which develop before language, play a foundational role in how we understand abstract ideas (Bloom, 1998; Hoemann et al., 2020; Ogren and Johnson, 2021). Through embodied experience, emotions provide a direct, physical connection between words and their meanings, making the process of understanding these concepts more intuitive and grounded in real-world interactions (Kosmas and Zaphiris, 2020; Tillman and Louwerse, 2018). This embodied perspective offers a nuanced understanding of how language becomes deeply connected to human experience. These insights are especially relevant for L2 learners, who must integrate new vocabulary and grammar into their existing embodied frameworks so as to acquire the knowledge of form-function-context mappings (Monaco et al., 2019; Pulvermüller et al., 2005), to foster deeper language integration to achieve pragmatic fluency (Atkinson, 2010; Ayedoun et al., 2019; Graesser et al., 2011).

Focusing on the process of perception and action in real-time can help capture the full impact of embodied processing. While concrete words are easier to grasp, affective elements may actually speed up this process, sometimes making affective representations more easily activated in cognition (Kousta et al., 2011; Kousta et al., 2009). A number of studies have found that emotional words in one's native language (L1) are often found to evoke faster and stronger responses compared to neutral words, demonstrating a robust emotional word processing advantage (Chen et al., 2015; Conrad et al., 2011; Kousta et al., 2009). This advantage suggests that these words are more deeply embedded in the cognitive and emotional framework of the speaker (Anooshian and Hertel, 1994; Sheikh and Titone, 2016). Furthermore, emotional words

in L1 elicit stronger physiological responses, such as increased skin conductance (Harris, 2004; Harris et al., 2006) and EMG (electromyographic) activity (Larsen et al., 2003). These physiological indicators provide support that affectively valenced words are embodied, automatically triggering emotional responses that are integrated into the speaker's bodily state. As emotions are biologically grounded and are widely recognized across cultures (Ekman, 1992; Pell et al., 2009; Scherer, 2003), there is the potential that L2 learners' cognition may be strategically coordinating emotion (i.e., one's bodily experience) during processing (i.e., recognition), and action (i.e., speaking).

In bilinguals, there are notable differences in how emotional words are processed in L1 compared to L2. Opitz and Degner (2012) found that affective valence of L2 words are processed in a less immediate way, in the context of a highly integrated L1/L2 lexicon. Bilinguals may also *feel* less affected by some pragmatic forms (e.g., swear words) in L2, making it easier to express taboo words, because they do not hold as much cultural significance (e.g., Dewaele, 2004). Similarly, Harris et al. (2003) found that late Turkish-English bilinguals exhibited stronger skin conductance responses (SCRs) to taboo words in L1 compared to L2, highlighting a physiological difference in emotional processing between the two languages. These differences are sometimes attributed to the “disembodied” nature of L2, where emotional experiences are not as deeply integrated due to the typically formal and less emotionally rich contexts in which L2 is learned (Jończyk, 2016; Pavlenko, 2012). This is likely due to the L1 being developed early on with the emotional regulation systems during affective socialization, thus, embedding vocabulary within specific emotional and contextual frameworks may be more tightly coupled in one's native language (Harris et al., 2006). In contrast, L2 is often acquired in environments that do not foster the same depth of emotional integration, resulting in a larger emotional distance and less embodied emotional responses (Baumeister et al., 2017). This disembodied nature not only affects emotional resonance but may also weaken the pragmatic cues and limits pragmatic transfer and adaptation toward competence in L2 by reducing learners' ability to interpret and produce culturally and contextually appropriate language in social interactions.

On the other hand, L2 proficiency could also affect how embodied cognition influences emotional processing and pragmatic fluency. Studies indicate that higher proficiency in L2 can lead to a more embodied experience of the language, enhancing emotional responses similar to those in L1 (e.g., Baumeister et al., 2017). However, less fluent L2 speakers are more likely to experience higher cognitive load and anxiety when producing emotional speech in L2 (Chen and Chang, 2004; Liu, 2006; Papi and Khajavy, 2023; Wang et al., 2025a), which could result in increased physiological arousal (Shi et al., 2007). This heightened arousal is linked to the greater effort required to coordinate lexical access and emotional regulation in a less proficient language (Altarriba and Basnight-Brown, 2011; McLaughlin et al., 1983). Furthermore, higher proficiency in L2 can diminish the arousal difference between L1 and L2 by making L2 more embodied. As proficiency increases, L2 learners are better able to integrate affective experiences into their L2 repertoire, reducing the emotional distance and enhancing pragmatic fluency (Harris, 2004).

Thus, developing pragmatic competence in a second language is a complex process that requires integrating linguistic forms, cultural norms, and social appropriateness. Embodied cognition offers a compelling framework to explore how cognitive and emotional processes shape L2

pragmatic learning, processing and fluency, especially for multilingual learners navigating diverse cultural contexts. Additionally, the related literature on embodiment mainly focused on alphabetic L1 and L2s, such as Spanish-English (Sutton et al., 2007; Kazanas and Altarriba, 2016), Turkish-English (Harris et al., 2003), and Greek-English (Eilola and Havelka, 2011). Research attention is needed for bilinguals with logographic L1s, such as Chinese-English bilinguals. Tang et al. (2023) argued for the necessity of evaluating emotionality differences in Chinese and Western languages, because emotions are likely understood and conceptualized differently across languages and cultures. When it comes to Chinese and English, Chinese emotion words are embodied more “interoceptively,” associated with internal bodily sensations, whereas English emotion words are embodied more “autonomically,” linked to automatic physiological responses (Zhou et al., 2022). The differences in how these two languages embody emotion concepts might be related to different attitudes about emotional expressions, with Chinese speakers being more introspective and English speakers more emotionally expressive.

Chinese culture values also tend to be different, as they typically reflect more control and restraint in emotional expression (especially negative emotion), while Western culture tends to appreciate more direct expression of feelings (Butler et al., 2007; Murata et al., 2013; Tsai et al., 2006). Chinese and English speakers may differ significantly in their experience of producing affectively valenced words, raising the question of whether Chinese-English bilinguals have similar emotional experiences across both languages. Despite growing interest in bilingual emotionality, relatively few studies have examined these differences among native Chinese speakers (Chen et al., 2015; Tang et al., 2023). The current study seeks to address this gap and contribute to the limited body of research in this area.

In this study, we explore how the interaction between physiology and cognition interacts to support second language (L2) fluency by examining how native Mandarin speakers' immersive language experience express emotionally charged words in their native language and non-native L2 English. This study aims to understand how embodied cognition influences language learning and affective expression in L2 speakers. It was hypothesized that (1) speakers would be more aroused when producing affectively valenced words than neutral words in both L1 and L2 because emotional words are more likely to trigger emotion and result in arousal increase. It was also hypothesized that (2) a higher level of L2 immersion would diminish the difference in arousal between L1 and L2, because L2 becomes more embodied as L1 with the increased experience speaking the language in an immersive context, leading to more similar emotional experience and responses.

Methods

Participants

A total of 22 participants (mean age = 24.8 yrs., *sd* = 4.2 yrs.; women = 12; men = 10) were recruited from international students from a Midwestern University in the United States. Of these participants, approximately 12 participants lived in the USA and enrolled in regular university courses for longer than 12 months (mean stay = 4.65 yrs., *sd* = 2.11 yrs.; high immersive speakers) and 10 participants had been living in the USA for less than 12 months (mean stay = 0.55, *sd* = 0.28 yrs.; low immersive speakers) and were

ESL (English as second language) students at the language center in the same university. According to their self-reports, all participants were born and lived in China until they were 18 yrs. of age. Additionally, 17 participants had complete electrodermal activity (EDA) data, and 16 participants had complete acoustic data—sometimes the devices failed to properly record the sound files and EDA data. All analyses were conducted based on the available complete data for each measure. Participants were compensated with a \$5 gift card for every half hour of participation and all speakers had normal-to-normal corrected vision, with no reports or diagnoses of speech or hearing impairments.

Materials and stimuli

All stimulus presentations and audio recordings were controlled by a Matlab, Psychtoolbox-3 program. All participants were seated in front of a 13-inch Macbook Pro computer and USB CAD U37 Studio Condenser recording microphone. Participants also wore the Empatica E4 sensor, which collected physiological data during the task. Stimuli included bi-syllabic English and Chinese words that shared semantic meaning, which were presented in the middle of the computer screen. The English/Chinese words included 12 negative English/Chinese (e.g., Cancer/癌症—Áizhèng), 12 positive English/Chinese (e.g., Success/成功—Chénggōng), and 24 neutral English/Chinese (e.g., Pencil/铅笔—Qiānbǐ) affectively valenced words chosen from the Affective Norms for English words (ANEW; [Bradley and Lang, 1999](#)). The bi-syllabic English words were chosen from the ANEW database, based on valence ratings; positive (mean = 8.29), negative (mean = 1.78), neutral (mean = 5.17). Once these words were chosen, the English words were translated into the Chinese corollary and characters by a native speaker of Chinese. The Chinese translation of each word was also limited to two syllables. The experimental task also included two short authentic passages of around 150 words that were presented at the start of each experimental block: one passage in Chinese on how to cook rice, and the other in English on how to select teaching materials for reading.

Design and procedure

In this task, participants were presented with a total of 100 affectively valenced Chinese and English words (4 practice trials; 96 experimental trials) in the middle of a computer screen, one word at a time. The word would disappear after being presented for 3 s. Participants would then be instructed to speak the word twice into the microphone after hearing a beep. They were also instructed to press the “spacebar” on the keyboard to end the recording and advance to the next word.

Prior to the start of the task, the experimenter placed the Empatica E4 sensor on the participant's left wrist (all participants were right-handed). For approximately 10 min prior to the experimental task, the participant completed a task on the computer unrelated to the current task. This allowed us to acquire a more accurate reading of the participant's physiological state, as the participant was able to get comfortable and remained in a fixed position prior to the beginning of this experimental task. This was an important methodological consideration, as moving too much and any anxiety from wearing

unfamiliar equipment can impact the measures collected from the E4 sensor. The participant was asked to keep their left hand stable and flat on the computer table, but were allowed to move their right hand to manipulate the computer keyboard to transition through the experimental trials (i.e., ‘spacebar’ keypress when finished recording). To begin an experimental block (language x affective valence), speakers were first presented with a short non-affectively valenced passage in the language condition they were currently in (e.g., a passage in Chinese or a passage in English). This was done as a means to *activate* the L1 or L2 language system, as to control for any physiological or cognitive costs incurred from switching between the language categories. A practice trial consisting of 4 practice words was presented for practice before the main task started.

Participants were randomly assigned to one of four between-subjects conditions: 2 Language Order (English/Chinese or Chinese/English) x 2 Affect Condition Order (Positive first v. Negative first). The Language Order between subjects condition indicates which language condition came first: English/Chinese—English words came first; Chinese/English—Chinese words came first. Additionally, Affect was counterbalanced between subjects, in that participants were randomly assigned to produce the Positive words first, and others were required to produce the Negative words first. This resulted in eight experimental blocks of trials. For example, if a participant was assigned to the Chinese and positive first condition, their trial structure included (1) 12 positive valenced words to be spoken in Chinese, with positive prosody (i.e., tone of voice), (2) 12 neutral valence words to be spoken in Chinese, with neutral prosody, (3) 12 positive valenced words to be spoken in English, with positive prosody (4) 12 neutral words to be spoken in English, with a neutral prosody. This was then repeated with negative prosody in Chinese first and then English. A similar structure was implemented for the other three between subjects Language x Affective Prosody conditions, in which language and affect were counterbalanced between participants.

Measures

Acoustic variation

The most commonly evaluated acoustic correlates of affective expressions include measures of timing, intensity, and pitch ([Juslin and Laukka, 2003](#); [Scherer, 2003](#)). As an estimate of speaker expressiveness, we recorded and composited measures of duration (timing; msec), intensity (amplitude; dB), and pitch (*f₀*; Hz). Each of these measures were collected using the standard aggregating features in Praat ([Boersma and Weenink, 2005](#)). It should be considered that when measuring speech, researchers should consider the type of aggregation method, especially related to pitch. For instance, [Strik and Boves \(1991\)](#) provide a compelling argument to use non-linear aggregation of pitch over time as a means to reduce signal variability caused by differing speaking rates. While this is a common technique, we chose to use linear aggregating methods, because we were explicitly interested in affect and the relation between speaking rate and pitch should be preserved in the signal—because they are both important cues. Using non-linear averaging for pitch across the time series while retaining duration as a predictor allows us to capture interactions between pitch and temporal dynamics that may carry emotional information. Non-linear aggregation might obscure these effects, potentially masking relevant affective signals. Our focus was not on

the specific communicative content of these acoustic variations but rather on the differences in variation across conditions, therefore, a composite was used to evaluate these cues as they are highly correlated and address general variation in the affective expressions.

Electrodermal activity (EDA)

EDA was collected using the Empatica E4 wristband sensor from participants during the course of the experiment. The Empatica E4 sensor is sampled at 4 Hz for EDA, which means it collects EDA data four times per second. EDA, a measure of physiological arousal has been frequently used as a correlate of emotional arousal and cognitive load. EDA is preferred to other measures of arousal because it has been suggested to be very sensitive and under strict control of the sympathetic (involuntary) nervous system (e.g., Cacioppo et al., 2007; Sequeira et al., 2009), therefore being recognized as one of the most sensitive physiological measures of emotional and cognitive activation (Eilola and Havelka, 2011; Hugdahl, 1995). During data collection, a Matlab Psychtoolbox-3 program controlled stimulus presentation, printed a time-stamp of when the stimulus (affectively valenced word) appeared on the computer screen and when the participant pressed the spacebar on the computer's keyboard. This allowed us to time match the participant's trial level data with the EDA data.

Analytic approach

Linear mixed random effects models were used due to their ability to account for both fixed and random effects, because this approach provides more flexibility compared to traditional ANOVA, as it can accommodate variability at multiple levels (e.g., at the subject and item level; Baayen et al., 2008; Barr et al., 2013). To do this, a fully maximal random effects model was always attempted with both fixed effects (predictors of interest) and random effects (random intercepts or slopes for subjects and items). We then employed a backwards removal of random effects until model convergence was met. We then compared the model that converged against the intercept only model, to ensure the model selected produced the best fit to the data.

Results

Acoustic variation

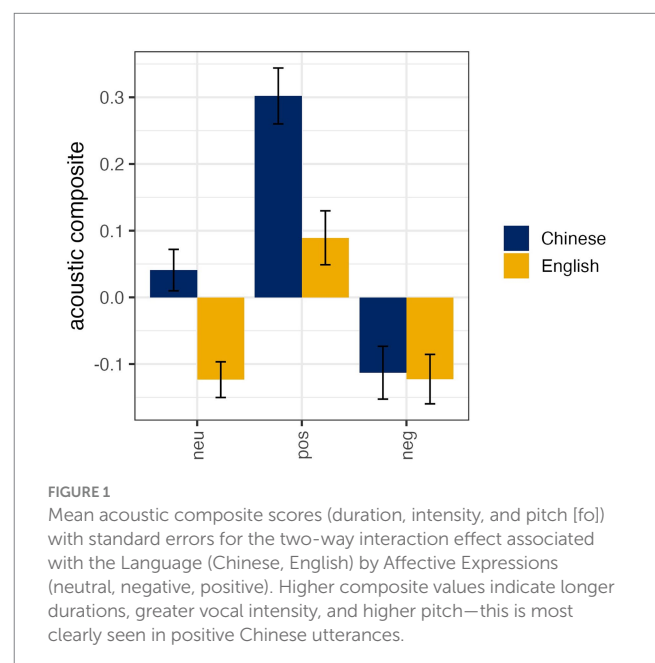
A linear mixed random effects model was used to evaluate the composite of acoustic variation (duration, vocal intensity, and f_0) as a function of language spoken, affective expression, and immersion. Subject and item were set as random intercepts and language spoken and affective expression were modeled as the random slopes on the subject intercept—a fully maximal random effect structure did not permit model convergence. However, this model did produce a significantly better fit than an intercept only model— $\chi^2 = 69.12$, $p < 0.001$, $AIC = 1063.5$. The chosen model accounted for approximately 73.4% (R^2) of the variance in the acoustic composite score. Results indicated a main effect of language spoken, affect type, immersive experience, and an interaction between language spoken and affective expression. In light of the higher order interaction, only the interaction and main effects not involved in the interaction are reported. The main effect of immersive experience ($\beta = -0.51$,

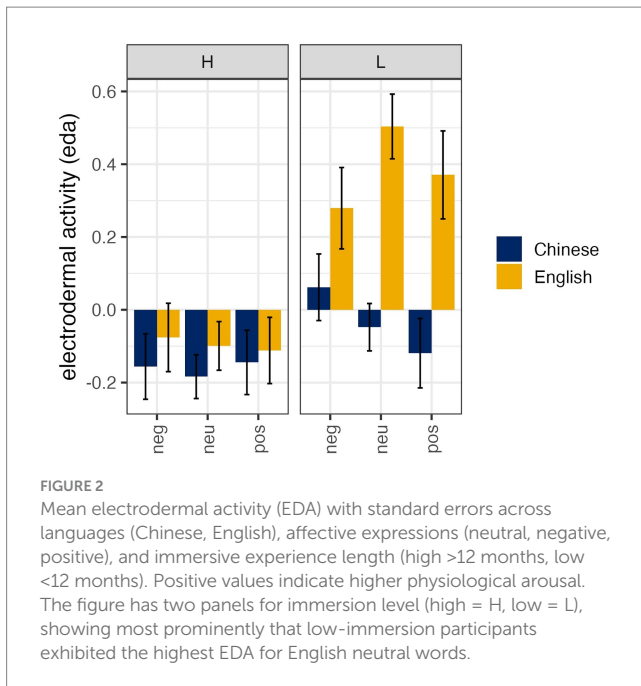
$SE = 0.23$, $t = -2.20$, $p = 0.03$) indicated that participants with longer immersive experiences (i.e., longer than 12 months) had more positive composite scores (i.e., longer durations, higher intensities and pitch) than the speakers with shorter immersive experiences (i.e., less than 12 months). Additionally, speakers varied their acoustics associated with neutral ($\beta = 0.15$, $SE = 0.05$, $t = 3.28$, $p < 0.01$) and positive utterances ($\beta = 0.21$, $SE = 0.07$, $t = 3.19$, $p < 0.01$), but not negative utterances ($\beta = 0.02$, $SE = 0.06$, $t = 0.24$, $p = 0.81$; see Figure 1). This suggests that how people speak—specifically their pitch, intensity, and timing—seems to depend on the language they are using, the emotion they are expressing, and the extent of their immersive experience.

Electrodermal activity (EDA)

A linear mixed random effects model was used to evaluate the electrodermal activity (EDA) as a function of language spoken, affective expression, and length immersion experience. Subject was set as random intercepts and language spoken and affective expression were modeled as the random slopes on the subject intercept. Item was dropped from the model and the fully maximal random effect structure did not permit model convergence. However, the selected model did produce a significantly better fit than an intercept only model— $\chi^2 = 425.68$, $p < 0.001$, $AIC = 1176.6$, and accounted for approximately 91% (R^2) of the variance in the electrodermal activity. Results indicated a 3-way interaction between language spoken, affective expression, and length of experience.

As seen in Figure 2, participants with less immersive experience experienced significantly higher EDA when producing neutral ($\beta = 0.55$, $SE = 0.12$, $t = 4.25$, $p < 0.001$) and positive words ($\beta = 0.50$, $SE = 0.13$, $t = 3.70$, $p < 0.01$) in English, relative to Chinese. Additionally, the participants with less immersive experience experienced higher EDA when producing negative words (relative to neutral) in Chinese ($\beta = 0.13$, $SE = 0.06$, $t = 2.26$, $p = 0.03$), but higher EDA when producing neutral words in English relative to negative





words ($\beta = -0.22$, $SE = 0.06$, $t = -3.84$, $p < 0.001$; see [Figure 2](#)). This suggests that the speakers in this sample with less experience in an immersive language environment showed different physiological responses depending on the language they were speaking and the type of emotional expression they were producing.

Participants with high immersive experience showed no significant difference in EDA between Chinese and English across all affective conditions (negative/neutral/positive). Notably, affective valence (neutral vs. negative) did not modulate EDA in either language among participants with high immersive experience. This suggests that prolonged immersion may reduce physiological arousal differences between the languages they were speaking and among the types of emotional expression they were producing, resulting in uniform EDA responses regardless of linguistic or emotional context.

Discussion

The findings from the current study provided support to the notion that an immersive L2 speaking experience importantly extends beyond the words learned ([Bardovi-Harlig and Hartford, 1993](#); [Kosmas and Zaphiris, 2020](#); [Lan et al., 2015](#); [Porter and Castillo, 2023](#); [Xiao, 2015](#)), as the immersive experience may be critical to shape how emotions are represented, expressed, and experienced (e.g., important aspects of pragmatic competence—[Rafieyan and Rozycki, 2019](#); [Kissler and Herbert, 2013](#); [Yu et al., 2021](#); [Scott et al., 2012](#)). The results of the study showed that length of immersion differentially influenced both outward (vocal affect) and inward (physiological arousal) experiences of emotion during language production (consistent with [Barsalou, 2008](#); [Louwerse and Jeuniaux, 2010](#); [Wilson and Golonka, 2013](#)). The immersive L2 context elicited affective utterances resembling the more vocally expressive style of American speakers ([Ip et al., 2021](#)), potentially reflecting mechanisms of transfer and adaptation that support the development of pragmatic competence.

In fact, speakers having been immersed in an L2 context for a longer period of time tended to be more vocally expressive: more acoustic variation in pitch, intensity, and duration. However, these participants exhibited a more attenuated physiological response, as they exhibited comparable EDA levels between English and Chinese across all affective conditions (negative/neutral/positive). This might indicate that sustained exposure to an L2 language environment may reduce the physiological burden of language switching ([Altarriba and Basnight-Brown, 2011](#); [Chen and Chang, 2004](#); [Liu, 2006](#); [Papi and Khajavy, 2023](#)). L2 speakers with high immersive experience may have developed automated emotional and linguistic integration, minimizing cross-language physiological responses (consistent with [Harris, 2004](#); [Shi et al., 2007](#)).

The opposite occurred for individuals with shorter immersion experiences, such that their vocal expressions were more constrained acoustically (consistent with [Dewaele, 2004](#); [Thoma and Baum, 2019](#)). However, the low immersion participants had a much more pronounced physiological response than the high immersion participants. The low immersion group tended to exhibit heightened arousal when speaking in their non-dominant language. This was particularly evident when they produced neutral and positive utterances in English. When speaking Mandarin Chinese, however, negative utterances elicited the strongest arousal response. This might suggest that cross-cultural differences in emotion suppression norms may be evident, given that Chinese culture may impose stronger social constraints on negative emotional expression than English-speaking Western cultures ([Murata et al., 2013](#); [Tsai et al., 2006](#)). While the literature suggests that negative stimuli elicit heightened arousal across cultures (suggesting arousal to negative stimuli to be universal; [Ho et al., 2015](#); [Järvelä et al., 2021](#); [Naranowicz et al., 2022](#)), the degree of reactivity may be shaped by cultural norms. For instance, US Americans often report higher emotional reactivity to negative visual stimuli than their Chinese counterparts, and in China, cultural norms encourage suppression of overt expressions of negative emotion ([Liddell and Williams, 2019](#); [Huwaë and Schaafsma, 2018](#); [Tyra et al., 2024](#)). Notably, research shows that when overt suppression is required, physiological arousal can increase ([Gross and Levenson, 1993](#); [Peters et al., 2014](#)). This supports the idea that, across cultures, negative affect may universally trigger heightened physiological arousal ([Hermanto et al., 2012](#); [Zhang et al., 2021](#)), even if outward expression is muted. In immersive contexts, however, cultural practices may be carried into the setting ([Kim, 2017](#)), allowing for more overt expression of negative affect, which could facilitate a release of arousal when experiencing negatively valenced stimuli ([Gross and Levenson, 1993](#); [Kennedy-Moore and Watson, 2001](#); [Thakur et al., 2017](#)).

Individuals in the low-immersion group may have experienced heightened arousal when violating these norms. Emotional engagement in speech seems to be shaped not only by one's ability to produce the words of a language, but also by the cultural and linguistic norms that may influence how speakers regulate emotion across languages. Language processing and affective expression are dynamically influenced by the duration of immersion, highlighting the interplay between bodily states and language production ([Kosmas and Zaphiris, 2020](#); [Tillman and Louwerse, 2018](#)).

Assessing variation in both acoustics and EDA offers valuable insights into how the cognitive system represents and supports L2 language learning and pragmatic competence, leading to a better understanding of cross-cultural communication. In early stages of L2 language learning, non-native speakers may experience greater emotional and cognitive effort when expressing affect (Chen and Chang, 2004; Liu, 2006; Papi and Khajavy, 2023), — as both lexical access is generally more difficult and they may experience stress responses associated with public speaking, interpersonal interactions, and even clinical contexts such as speech therapy (Roseberry-McKibbin et al., 2005). While the current study cannot explicitly address the mechanism that elicits greater cognitive load, we do see that the Low immersion group is under greater cognitive strain. Understanding how language and pragmatic competence shape affective expression and physiological response could inform educational approaches for language learning, as well as improving speech recognition models that aim to capture pragmatic nuance in bilingual speakers.

From an embodied cognition view, these findings support the notion that cognitive processes are deeply rooted in the body's sensory and motor systems (Barsalou, 2008; Barsalou et al., 2003; Dove, 2014; Winkielman et al., 2015). The observed acoustic variation highlights how speech production is not just a cognitive act but one that engages sensorimotor mechanisms (Baumeister et al., 2017; Foroni, 2015; Foroni and Semin, 2009; Kousta et al., 2011; Larsen et al., 2003), reinforcing the idea that emotion is not simply encoded abstractly in the brain but is enacted through the body (Baumeister et al., 2017; Dimberg et al., 2000). Similarly, the physiological responses reveal that language processing is intertwined with bodily arousal, suggesting that emotion is not merely understood but physically felt (Chen et al., 2015; Conrad et al., 2011; Harris, 2004; Harris et al., 2006; Kousta et al., 2009).

Ultimately, this research underscores the complex connection between language, pragmatics, cognition, and the body (Baumeister et al., 2017; Louwerse and Jeuniaux, 2010; Wilson and Golonka, 2013). Affective expression in speech is not just a matter of vocal output but reflects a dynamic interaction between linguistic and pragmatic experience, as well as motor control, and physiological states (Barsalou et al., 2003; Dove, 2014; Harris et al., 2006; Kousta et al., 2011; Porter and Castillo, 2023). This deeper understanding of how language and pragmatic competence shapes both vocal expression and bodily responses contributes to broader theories of bilingualism, emotion, and the embodied nature of communication, with meaningful implications for education, technology, and clinical practice.

Limitations and future directions

Like many studies in speech production research, this study had a relatively small sample size, because speech production effects are typically robust within participant; however, it remains consistent with other production studies in the field (e.g., see Ferguson and Kewley-Port, 2007), where detailed acoustic and physiological measures require intensive data collection and analysis. In the current study, because we are also using EDA as a measure, we increased the sample size and based on a power analysis, we were sufficiently powered—with well over 1,500 data points in our sample and the power analysis

calling for a minimum of 12 participants needed for statistical sensitivity. We should note, however, that some data loss occurred due to device failure, a common challenge in studies collecting physiological measures such as electrodermal activity (Boucsein, 2012, p. 245; Braithwaite et al., 2013), but given the sample size and repeated measures component, we believe the findings remain meaningful and interpretable.

We should also draw attention to the fact that our design permitted participants to control their response window (on average 3.9 s (low immersion group); 4.6 (high immersion group)). EDA has a relatively slow rise time, with peak responses typically occurring around 6 s post-stimulus. This means our measured EDA response may under-estimate the true peak amplitude, potentially introducing measurement error and the absolute magnitudes should be interpreted with caution. However, it is still notable that the Low Immersion group exhibited higher EDA despite having shorter trial durations. If measurement truncation biased our data, it would be expected to attenuate rather than inflate group differences. Thus, the direction and robustness of our findings are unlikely to be explained solely by timing limitations.

While the results provide valuable insight into the interaction between language experience, emotion, and physiological response, it is unclear whether these findings would generalize to other languages or if they are specific to the linguistic and cultural background of the sample. Nevertheless, this study provides a framework that evaluates the dynamic interplay between subjective and objective experience of emotion (see Wang et al., 2025b). In addition, we chose a linear aggregation method instead of the sometimes recommended non-linear methods (see Strik and Boves, 1991). This was a strategic move, but aggregating has the potential of leading to oversimplification, especially when trying to understand the dynamics of emotions. In the current study, we deliberately employed aggregation methods to maintain the acoustic integrity of speech and ensure reliable measurement of affect and physiological aspects of embodied emotion. Unlike studies that track emotions across extended real-world contexts—which can introduce additional variability and noise—we focused on short bursts of subjective and objective emotional responses in carefully controlled recording sessions, preserving the natural data stream while balancing ecological validity and experimental rigor. It should be noted, however, that when adopting a multimodal approach, researchers must make careful methodological choices to preserve the richness of the data and avoid oversimplification, as Wang et al. (2025b) emphasize. This is important because oversimplifying can obscure meaningful patterns in the interactions between modalities, potentially leading to inaccurate or incomplete interpretations. To that end, these limitations are typical of experimental research in this area and do not diminish the overall contributions of the study, but rather highlight areas for future investigation.

Conclusion

Understanding how L2 learners represent and express affective language can help clinicians tailor therapy approaches for bilingual individuals. Speech intelligibility and prosody play crucial roles in conveying meaning and emotion, and understanding that shorter immersive experiences may lead to weaker language and pragmatic skills is a useful perspective taking tool for clinicians to use, to help them avoid misinterpretation and implement strategies that reduce

anxiety and enhance confidence in L2 learners. Ultimately, this research bridges cognitive science, bilingualism, and clinical practice, offering valuable perspectives on the importance of immersive experiences on communication.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by The Institutional Review Board of Kent State University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MW: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. JR: Data curation, Formal analysis, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research received partial funding from the 2023 Humanities and Social Sciences Research Youth Foundation of the Ministry of Education of China (Grant No. 23YJC880113).

References

- Al-Hejin, B. (2004). Attention and awareness: evidence from cognitive and second language acquisition research. *Stud. Appl. Ling. TESOL* 4, 1–22.
- Allami, H., and Naeimi, A. (2011). A cross-linguistic study of refusals: an analysis of pragmatic competence development in Iranian EFL learners. *J. Pragmat.* 43, 385–406. doi: 10.1016/j.pragma.2010.07.010
- Allarriba, J., and Basnight-Brown, D. M. (2011). The representation of emotion vs. emotion-laden words in English and Spanish in the affective Simon task. *Int. J. Biling.* 15, 310–328. doi: 10.1177/1367006910379261
- Anoshian, L. J., and Hertel, P. T. (1994). Emotionality in free recall: language specificity in bilingual memory. *Cognit. Emot.* 8, 503–514. doi: 10.1080/02699939408408956
- Arnold, J. (2011). Attention to affect in language learning. *Online Sub.* 22, 11–22.
- Atkinson, D. (2010). Extended, embodied cognition and second language acquisition. *Appl. Linguis.* 31, 599–622. doi: 10.1093/applin/amq009
- Ayedoun, E., Hayashi, Y., and Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *Int. J. Artif. Intell. Educ.* 29, 29–57. doi: 10.1007/s40593-018-0171-6
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Lang. Learn.* 63, 68–86. doi: 10.1111/j.1467-9922.2012.00738.x
- Bardovi-Harlig, K., and Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Q.* 32, 233–259. doi: 10.2307/3587583
- Bardovi-Harlig, K., and Hartford, B. (1993). Learning the rules of academic talk: A longitudinal study of pragmatic development. *Studies in Second Language Acquisition.* 15, 279–304.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Barrett, L. F., and Lindquist, K. A. (2008). “The embodiment of emotion” in Embodied grounding: Social, cognitive, affective, and neuroscientific approaches. eds. G. R. Semin and E. R. Smith (New York, NY: Cambridge University Press), 237–262.
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., and Ruppert, J. A. (2003). Social embodiment. In D. L. Medin (Ed.). *Psychology of learning and motivation: Advances in research and theory* 43, 43–92. doi: 10.1016/S0079-7421(03)01011-9
- Baumeister, J. C., Foroni, F., Conrad, M., Rumati, R. I., and Winkielman, P. (2017). Embodiment and emotional memory in first vs. second language. *Front. Psychol.* 8:394. doi: 10.3389/fpsyg.2017.00394
- Beltrama, A. (2020). Social meaning in semantics and pragmatics. *Lang Ling Compass* 14:e12398. doi: 10.1111/lnc3.12398

Acknowledgments

We would like to express our sincere gratitude to Ke-Jui Yen for her valuable support and assistance with data collection as a lab assistant.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. This document has been refined with the assistance of AI (ChatGPT, OpenAI, 2024) to enhance readability and conciseness while preserving the original meaning and integrity of the content. AI was only used to improve readability of the manuscript – all ideas are unique to the authors.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Bialystok, E. (1993). "Symbolic representation and attentional control in pragmatic competence" in *Interlanguage pragmatics*. eds. G. Kasper and S. Blum-Kulka (Oxford: Oxford University Press), 43–57.
- Bloom, L. (1998). Language development and emotional expression. *Pediatrics* 102, 1272–1277. doi: 10.1542/peds.102.SEI.1272
- Boersma, P., and Weenink, D. (2005). Praat. Doing phonetics by computer (Version 5.1)
- Boucsein, W. (2012). *Electrodermal activity*. New York, NY: Springer Science & Business Media. doi: 10.1007/978-1-4614-1126-0
- Bradley, M. M., and Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings.
- Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1017–1034.
- Butler, E. A., Lee, T. L., and Gross, J. J. (2007). Emotion regulation and culture: are the social consequences of emotion suppression culture-specific? *Emotion* 7, 30–48. doi: 10.1037/1528-3542.7.1.30
- Byram, M. (1997). 'Cultural awareness' as vocabulary learning. *Lang. Learn.* 16, 51–57.
- Cacioppo, J. T., Tassinary, L. G., and Berntson, G. (2007). *Handbook of psychophysiology*. New York: Cambridge University Press. doi: 10.13140/2.1.2871.1369
- Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., and Griner, B. (2010). *Teaching pronunciation: A course book and reference guide*. New York: Cambridge University Press.
- Chen, T. Y., and Chang, G. B. (2004). The relationship between foreign language anxiety and learning difficulties. *Foreign Lang. Ann.* 37, 279–289. doi: 10.1111/j.1944-9720.2004.tb02200.x
- Chen, P., Lin, J., Chen, B., Lu, C., and Guo, T. (2015). Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese–English bilinguals. *Cortex* 71, 34–48. doi: 10.1016/j.cortex.2015.06.002
- Conrad, M., Recio, G., and Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: a cross cultural ERP study with German–Spanish bilinguals. *Front. Psychol.* 2:351. doi: 10.3389/fpsyg.2011.00351
- Cook, M., and Liddicoat, A. J. (2002). The development of comprehension in interlanguage pragmatics: the case of request strategies in English. *Aust. Rev. Appl. Linguist.* 25, 19–39. doi: 10.1075/ara.25.1.02coo
- Costa, A., Pickering, M. J., and Sorace, A. (2008). Alignment in second language dialogue. *Lang. Cogn. Process.* 23, 528–556. doi: 10.1080/01690960801920545
- De Vega, M., Glenberg, A., and Graesser, A. (2012). *Symbols and embodiment: Debates on meaning and cognition*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/978019217274.001.0001
- Dewaele, J.-M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *J. Multiling. Multicult. Dev.* 25, 204–222. doi: 10.1080/01434630408666529
- Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221
- Dove, G. (2014). Thinking in words: language as an embodied medium of thought. *Top. Cogn. Sci.* 6, 371–389. doi: 10.1111/tops.12102
- Eilola, T. M., and Havelka, J. (2011). Behavioural and physiological responses to the emotional and taboo Stroop tasks in native and non-native speakers of English. *International Journal of Bilingualism* 15, 353–369. doi: 10.1177/1367006910379263
- Ekman, P. (1992). Are there basic emotions? *Psychol. Rev.* 99, 550–553. doi: 10.1037/0033-295X.99.3.550
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied linguistics* 27, 1–24. doi: 10.1093/applin/ami038
- Ellsworth, P. C. (1991). "Some implications of cognitive appraisal theories of emotion" in *International review of studies on emotion*. ed. K. Strongman (New York: Wiley), 143–161.
- Ellsworth, P. C. (2013). Appraisal theory: old and new questions. *Emot. Rev.* 5, 125–131. doi: 10.1177/1754073912463617
- Ferguson, S. H., and Kewley-Port, D. (2007). Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J. Speech Lang. Hear. Res.* 50, 1241–1255. doi: 10.1044/1092-4388(2007)087
- Foroni, F. (2015). Do we embody a second language? Evidence for 'partial' simulation during processing of a second language. *Brain Cogn.* 99, 8–16. doi: 10.1016/j.bandc.2015.06.006
- Foroni, F., and Semin, G. R. (2009). Language that puts you in touch with your bodily feelings: the multimodal responsiveness of affective expressions. *Psychol. Sci.* 20, 974–980. doi: 10.1111/j.1467-9280.2009.02400.x
- Freed, B. F., Segalowitz, N., and Dewey, D. P. (2004). Context of learning and second language fluency in French: comparing regular classroom, study abroad, and intensive domestic immersion programs. *Stud. Second. Lang. Acquis.* 26, 275–301. doi: 10.1017/S0272263104262064
- Graesser, A. C., Millis, K., and Graesser, A. (2011). "Discourse and cognition" in *Discourse Studies: A Multidisciplinary Introduction*, ed. T. A. van Dijk London, United Kingdom: SAGE. 126–142.
- Gross, J. J., and Levenson, R. W. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *J. Pers. Soc. Psychol.* 64, 970–986. doi: 10.1037/0022-3514.64.6.970
- Harris, C. L. (2004). Bilingual speakers in the lab: psychophysiological measures of emotional reactivity. *J. Multiling. Multicult. Dev.* 25, 223–247. doi: 10.1080/01434630408666530
- Harris, C. L., Ayçiçeği, A., and Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Appl. Psycholinguist.* 24, 561–579. doi: 10.1017/S0142716403000286
- Harris, C. L., Gleason, J. B., and Ayçiçeği, A. (2006). "When is a first language more emotional? Psychophysiological evidence from bilingual speakers" in *Bilingual minds*. ed. A. Pavlenko (Bristol, UK: Multilingual Matters), 257–283.
- Hasan, R. (2012). A view of pragmatics in a social semiotic perspective. *Linguist. Hum. Sci.* 5, 251–279. doi: 10.1558/lhs.v5i3.251
- Hermanto, N., Moreno, S., and Bialystok, E. (2012). Linguistic and metalinguistic outcomes of intense immersion education: how bilingual? *Int. J. Biling. Educ. Biling.* 15, 131–145. doi: 10.1080/13670050.2011.652591
- Ho, S. M., Mak, C. W., Yeung, D., Duan, W., Tang, S., Yeung, J. C., et al. (2015). Emotional valence, arousal, and threat ratings of 160 Chinese words among adolescents. *PLoS One* 10:e0132294. doi: 10.1371/journal.pone.0132294
- Hoemann, K., Devlin, M., and Barrett, L. F. (2020). Emotions are abstract, conceptual categories that are learned by a predicting brain. *Emot. Rev.* 12, 253–255. doi: 10.1177/1754073919897296
- Holmes, J. (2018). "Sociolinguistics vs pragmatics: where does the boundary lie?" in *Pragmatics and its interfaces*. eds. C. Ilie and N. R. Norrick (Amsterdam: John Benjamins Publishing Company), 11–32.
- Hugdahl, K. (1995). *Psychophysiology: The mind–body perspective*. Cambridge, MA: Harvard University Press.
- Huwaë, S., and Schaafsma, J. (2018). Cross-cultural differences in emotion suppression in everyday interactions. *Int. J. Psychol.* 53, 176–183. doi: 10.1002/ijop.12283
- Ip, K. I., Miller, A. L., Karasawa, M., Hirabayashi, H., Kazama, M., Wang, L., et al. (2021). Emotion expression and regulation in three cultures: Chinese, Japanese, and American preschoolers' reactions to disappointment. *J. Exp. Child Psychol.* 201:104972. doi: 10.1016/j.jecp.2020.104972
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., and Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learn. Instr.* 72:101203. doi: 10.1016/j.learninstruc.2019.04.004
- Jerath, R., and Beveridge, C. (2020). Respiratory rhythm, autonomic modulation, and the spectrum of emotions: the future of emotion recognition and modulation. *Front. Psychol.* 11:1980. doi: 10.3389/fpsyg.2020.01980
- Jończyk, R. (2016). "Affective (Dis)Embodiment in Nonnative Language" in *Affect-language interactions in native and non-native English speakers: A neuropragmatic perspective*. eds. R. R. Heredia and A. B. Cieśliska (Cham: Springer), 149–159. doi: 10.1007/978-3-319-47635-3_7
- Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770
- Kasper, G. (2001). Four perspectives on L2 pragmatic development. *Appl. Linguist.* 22, 502–530. doi: 10.1093/applin/22.4.502
- Kazanas, S. A., and Altarriba, J. (2016). Emotion word processing: effects of word type and valence in Spanish–English bilinguals. *J. Psycholinguist. Res.* 45, 395–406. doi: 10.1007/s10936-015-9357-3
- Kennedy-Moore, E., and Watson, J. C. (2001). How and when does emotional expression help? *Rev. Gen. Psychol.* 5, 187–212. doi: 10.1037/1089-2680.5.3.187
- Kermad, A. (2021). "From the Sound, It Look Like He Said It from the Deep in His Heart": How Do English Learners Make Judgments of Pragma-Prosodic Meaning? *TESL-EJ* 25:n1.
- Kim, Y. Y. (2017). "Cross-cultural adaptation" in *Oxford Research Encyclopedia of Communication*. ed. J. F. Nussbaum (Oxford, UK: Oxford University Press). doi: 10.1093/acrefore/9780190228613.013.21
- Kinginger, C., and Belz, J. A. (2005). Socio-cultural perspectives on pragmatic development in foreign language learning: Microgenetic case studies from telecollaboration and residence abroad. *Intercul. Pragm.* 2, 369–421. doi: 10.1515/ipro.2005.2.4.369
- Kissler, J., and Herbert, C. (2013). Emotion, etmnooi, or emittoon?—faster lexical access to emotional than to neutral words during reading. *Biol. Psychol.* 92, 464–479. doi: 10.1016/j.biopsycho.2012.09.004
- Kosmas, P., and Zaphiris, P. (2020). Words in action: investigating students' language acquisition and emotional performance through embodied learning. *Innov. Lang. Learn. Teach.* 14, 317–332. doi: 10.1080/17501229.2019.1607355
- Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., and Del Campo, E. (2011). The representation of abstract words: why emotion matters. *J. Exp. Psychol. Gen.* 140, 14–34. doi: 10.1037/a0021446
- Kousta, S. T., Vinson, D. P., and Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* 112, 473–481. doi: 10.1016/j.cognition.2009.06.007

- Lan, Y. J., Chen, N. S., Li, P., and Grant, S. (2015). Embodied cognition and language learning in virtual environments. *Educ. Technol. Res. Dev.* 63, 639–644. doi: 10.1007/s11423-015-9401-x
- Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* 40, 776–785. doi: 10.1111/1469-8986.00078
- Levis, J. (1999). Intonation in theory and practice, revisited. *TESOL Q.* 33, 7–63.
- Liddell, B. J., and Williams, E. N. (2019). Cultural differences in interpersonal emotion regulation. *Front. Psychol.* 10:999. doi: 10.3389/fpsyg.2019.00999
- Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System* 34, 301–316. doi: 10.1016/j.system.2006.04.004
- Louwerse, M. M., and Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition* 114, 96–104. doi: 10.1016/j.cognition.2009.09.002
- MacIntyre, P. D., and Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Stud. Second Lang. Learn. Teach.* 7, 61–88. doi: 10.14746/ssl.2017.7.1.4
- Manstead, A. S., and Wagner, H. L. (1981). Arousal, cognition and emotion: an appraisal of two-factor theory. *Curr. Psychol. Rev.* 1, 35–54. doi: 10.1007/BF02979253
- McLaughlin, B., Rossman, T., and McLeod, B. (1983). Second language learning: an information-processing perspective 1. *Lang. Learn.* 33, 135–158. doi: 10.1111/j.1467-1770.1983.tb00532.x
- Mokoro, E. (2024). Pragmatic competence in second language learners. *Eur. J. Linguist.* 3, 15–28. doi: 10.47941/ej.2044
- Monaco, E., Jost, L. B., Gygas, P. M., and Annoni, J. M. (2019). Embodied semantics in a second language: critical review and clinical implications. *Front. Hum. Neurosci.* 13:110. doi: 10.3389/fnhum.2019.00110
- Moors, A., Ellsworth, P. C., Scherer, K. R., and Frijda, N. H. (2013). Appraisal theories of emotion: state of the art and future development. *Emot. Rev.* 5, 119–124. doi: 10.1177/1754073912468165
- Murata, A., Moser, J. S., and Kitayama, S. (2013). Culture shapes electrocortical responses during emotion suppression. *Soc. Cogn. Affect. Neurosci.* 8, 595–601. doi: 10.1093/scan/nss036
- Naranowicz, M., Jankowiak, K., and Behnke, M. (2022). Native and non-native language contexts differently modulate mood-driven electrodermal activity. *Sci. Rep.* 12:22361. doi: 10.1038/s41598-022-27064-3
- Nicolay, A. C., and Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism Lang. Cogn.* 16, 597–607. doi: 10.1017/S1366728912000375
- Nummenmaa, L., Glerean, E., Viinikainen, M., Jääskeläinen, I. P., Hari, R., and Sams, M. (2012). Emotions promote social interaction by synchronizing brain activity across individuals. *Proc. Natl. Acad. Sci.* 109, 9599–9604. doi: 10.1073/pnas.1206095109
- Ogren, M., and Johnson, S. P. (2021). Factors facilitating early emotion understanding development: contributions to individual differences. *Hum. Dev.* 64, 108–118. doi: 10.1159/000511628
- Opitz, B., and Degner, J. (2012). Emotionality in a second language: it's a matter of time. *Neuropsychologia* 50, 1961–1967. doi: 10.1016/j.neuropsychologia.2012.04.021
- Paivio, A. (1990). Mental representations: A dual coding approach. New York: Oxford university press.
- Papi, M., and Khajavy, H. (2023). Second language anxiety: construct, effects, and sources. *Annu. Rev. Appl. Linguist.* 43, 127–139. doi: 10.1017/S0267190523000028
- Pavlenko, A. (2012). Affective processing in bilingual speakers: disembodied cognition? *Int. J. Psychol.* 47, 405–428. doi: 10.1080/00207594.2012.743665
- Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., and Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: a comparison of four languages. *J. Phon.* 37, 417–435. doi: 10.1016/j.wocn.2009.07.005
- Peters, B. J., Overall, N. C., and Jamieson, J. P. (2014). Physiological and cognitive consequences of suppressing and expressing emotion in dyadic interactions. *Int. J. Psychophysiol.* 94, 100–107. doi: 10.1016/j.ijpsycho.2014.07.015
- Porter, S., and Castillo, M. S. (2023). The effectiveness of immersive language learning: an investigation into English language acquisition in immersion environments versus traditional classroom settings. *Res. Stud. English Lang. Teach. Learn.* 1, 155–165. doi: 10.62583/rselt.v1i3.17
- Pulvermüller, F., Hauk, O., Nikulin, V. V., and Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *Eur. J. Neurosci.* 21, 793–797. doi: 10.1111/j.1460-9568.2005.03900.x
- Rafeyan, V., and Rozycki, W. (2019). Development of language proficiency and pragmatic competence in an immersive language program. *World* 9, 10–21.
- Roseberry-McKibbin, C., Brice, A., and O'Hanlon, L. (2005). Serving English language learners in public school settings. *Lang. Speech Hear. Serv. Sch.* 36, 48–61. doi: 10.1044/0161-1461(2005/005)
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145
- Sauter, D. A., Eisner, F., Ekman, P., and Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl. Acad. Sci.* 107, 2408–2412. doi: 10.1073/pnas.0908239106
- Schachter, S., and Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69, 379–399. doi: 10.1037/h0046234
- Scherer, K. R. (1999). "Appraisal theory" in Handbook of cognition and emotion. eds. T. Dalgleish and M. Power (Chichester, UK: Wiley), 637–663. doi: 10.1002/0470013494.ch30
- Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Comm.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5
- Scherer, K. R., Banse, R., and Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross-Cult. Psychol.* 32, 76–92. doi: 10.1177/0022022101032001009
- Scott, G. G., O'Donnell, P. J., and Sereno, S. C. (2012). Emotion words affect eye fixations during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 783–792. doi: 10.1037/a0027209
- Segalowitz, N., and Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Stud. Second. Lang. Acquis.* 26, 173–199.
- Sequeira, H., Hot, P., Silvert, L., and Delplanque, S. (2009). Electrical autonomic correlates of emotion. *Int. J. Psychophysiol.* 71, 50–56. doi: 10.1016/j.ijpsycho.2008.07.009
- Sheikh, N. A., and Titone, D. (2016). The embodiment of emotional words in a second language: an eye-movement study. *Cognit. Emot.* 30, 488–500. doi: 10.1080/02699931.2015.1018144
- Shi, Y., Ruiz, N., Taib, R., Choi, E., and Chen, F. (2007). "Galvanic skin response (GSR) as an index of cognitive load." In CHI'07 extended abstracts on Human factors in computing systems. pp. 2651–2656.
- Strik, H., and Boves, L. (1991). A dynamic programming algorithm for time-aligning and averaging physiological signals related to speech. *J. Phon.* 19, 367–378. doi: 10.1016/S0095-4470(19)30328-6
- Sutton, T. M., Altarriba, J., Gianico, J. L., and Basnight-Brown, D. M. (2007). The automatic access of emotion: emotional Stroop effects in Spanish–English bilingual speakers. *Cognit. Emot.* 21, 1077–1090. doi: 10.1080/02699930601054133
- Tang, D., Fu, Y., Wang, H., Liu, B., Zang, A., and Kärkkäinen, T. (2023). The embodiment of emotion-label words and emotion-laden words: evidence from late Chinese–English bilinguals. *Front. Psychol.* 14:1143064. doi: 10.3389/fpsyg.2023.1143064
- Thakur, E. R., Holmes, H. J., Lockhart, N. A., Carty, J. N., Ziadni, M. S., Doherty, H. K., et al. (2017). Emotional awareness and expression training improves irritable bowel syndrome: a randomized controlled trial. *Neurogastroenterol. Motil.* 29:e13143. doi: 10.1111/nmo.13143
- Thoma, D., and Baum, A. (2019). Reduced language processing automaticity induces weaker emotions in bilinguals regardless of learning context. *Emotion* 19, 1023–1034. doi: 10.1037/em0000502
- Tillman, R., and Louwerse, M. (2018). Estimating emotions through language statistics and embodied cognition. *J. Psycholinguist. Res.* 47, 159–167. doi: 10.1007/s10936-017-9522-y
- Titchener, E. B. (1914). An historical note on the James-Lange theory of emotion. *Am. J. Psychol.* 25, 427–447. doi: 10.2307/1412861
- Trofimovich, P., and Isaacs, T. (2016). "Second language pronunciation assessment: A look at the present and the future" in Second language pronunciation assessment: Interdisciplinary perspectives. eds. T. Isaacs and P. Trofimovich (Bristol, UK: Multilingual Matters), 259–271. doi: 10.21832/ISAACS6848
- Tsai, J. L., Knutson, B., and Fung, H. H. (2006). Cultural variation in affect valuation. *J. Pers. Soc. Psychol.* 90, 288–307. doi: 10.1037/0022-3514.90.2.288
- Tyra, A. T., Fergus, T. A., and Ginty, A. T. (2024). Emotion suppression and acute physiological responses to stress in healthy populations: a quantitative review of experimental and correlational investigations. *Health Psychol. Rev.* 18, 396–420. doi: 10.1080/17437199.2023.2251559
- Wang, P., Ganushchak, L., Welie, C., and van Steensel, R. (2025a). Same anxiety, different faces: shared mechanisms with distinct manifestations in native and non-native speech. *PsyArXiv*.
- Wang, P., Liu, A., and Sun, X. (2025b). Integrating emotion dynamics in mental health: a trimodal framework combining ecological momentary assessment, physiological measurements, and speech emotion recognition. *Interdiscip. Med.* 3:e20240095. doi: 10.1002/INMD.20240095
- Wilson, A. D., and Golonka, S. (2013). Embodied cognition is not what you think it is. *Front. Psychol.* 4:58. doi: 10.3389/fpsyg.2013.00058
- Winkielman, P., Niedenthal, P., Wielgosz, J., Eelen, J., and Kavanagh, L. C. (2015). "Embodiment of cognition and emotion" in APA handbook of personality and social psychology, Volume 1: Attitudes and social cognition. eds. R. F. Krueger, M. J. Lerner and E. T. Higgins (Washington, DC: American Psychological Association), 151–175.
- Wolfson, N. (1989). "The Social Dynamics of Native and Nonnative Variation in Complimenting Behavior" in The Dynamic Interlanguage. Topics in Language and Linguistics. ed. M. R. Eisenstein (Boston, MA: Springer). doi: 10.1007/978-1-4899-0900-8_14
- Wyner, L. (2014). Second language pragmatic competence: individual differences in ESL and EFL environments. *Work. Pap. TESOL Appl. Linguist.* 14, 84–99.
- Xiao, F. (2015). Adult second language learners' pragmatic development in the study-abroad context: a review. *Front. Interdis. J. Study Abroad* 25:349. doi: 10.36366/frontiers.v25i1.349
- Yu, C. S. P., McBeath, M. K., and Glenberg, A. M. (2021). The gleam-glum effect: in/versus/phonemes generically carry emotional valence. *J. Exp. Psychol. Learn. Mem. Cogn.* 47:1173. doi: 10.1037/xlm0001017
- Zhang, H., Diaz, M. T., Guo, T., and Kroll, J. F. (2021). Language immersion and language training: two paths to enhanced language regulation and cognitive control. *Brain Lang.* 223:105043. doi: 10.1016/j.bandl.2021.105043
- Zhou, P., Critchley, H., Nagai, Y., and Wang, C. (2022). Divergent conceptualization of embodied emotions in the English and Chinese languages. *Brain Sci.* 12:911. doi: 10.3390/brainsci12070911