

# Valence and arousal perception among first language users, foreign language users, and naïve listeners of Mandarin across various communication modalities

## Research Article

**Cite this article:** Lorette, P., & Dewaele, J-M (2024). Valence and arousal perception among first language users, foreign language users, and naïve listeners of Mandarin across various communication modalities. *Bilingualism: Language and Cognition*, 27, 874–885. <https://doi.org/10.1017/S1366728923000925>



Received: 30 October 2022  
Revised: 28 November 2023  
Accepted: 28 November 2023  
First published online: 10 January 2024

### Keywords:

emotion perception; valence and arousal; modalities; first and foreign language speakers; self-report

### Author for correspondence:

Pernelle Lorette;  
Email: [p.lorette@uni-mannheim.de](mailto:p.lorette@uni-mannheim.de)

Pernelle Lorette<sup>1</sup>  and Jean-Marc Dewaele<sup>2,3</sup> 

<sup>1</sup>Department of English linguistics, University of Mannheim, Mannheim, Germany; <sup>2</sup>Institute of Education, University College London, London, United Kingdom and <sup>3</sup>Birkbeck University of London, London, United Kingdom

## Abstract

This online study investigates how first (L1) and foreign language (LX) users, and naïve (L0) listeners of Mandarin perceive the valence and arousal level of a Chinese interlocutor in various communication modalities. The 1485 participants (651 L1, 292 LX, and 542 L0 Mandarin users) were presented with 12 recordings of a Chinese actor conveying emotional events in the visual-vocal-verbal, vocal-verbal, visual-only, or vocal-only modality. Valence and arousal perceptions were collected via the 2DAFS (Lorette, 2021). Disregarding the vocal-only modality which led to neutral perceptions, bootstrapped regression models suggest that modality does not affect L1 users' valence perceptions. LX and L0 users perceive markedly more neutral valence levels in the absence of visual cues, and in the case of positive stimuli, slightly lower arousal levels. This calls for a more nuanced conceptualisation of valence and arousal as universal features of emotions and stress the significance of modality for intercultural communication.

## 1. Introduction

In 1963, President John F. Kennedy had a direct hotline installed with Nikita Khrushchev in the Kremlin to facilitate rapid communication in times of crisis (Clavin, 2013). Historians agree that this so-called “red phone” – which was in fact a duplex telegraph circuit for transmitting text and later a fax machine – contributed to averting a nuclear war. Diplomatic phone calls on secure lines between world leaders are common (Baker, 2016). One may wonder whether videocalls may not be preferable. Indeed, it is harder to perceive the emotional state of one's interlocutor over the phone than in face-to-face communication, when one can “read the space” and infer cues from the other's face. Interpreting how one's interlocutor is feeling might not only be particularly challenging when the communication modality restricts access to certain cues, but also when the communication happens with someone who does not share the same first language (or culture). Kennedy and Khrushchev had been raised and socialised in different contexts and were thus likely to assign meaning to verbal and nonverbal information in different ways. When it came to the communication of emotions specifically, early emotion perception studies focused on emotions conveyed in the face – typically presenting participants from different cultures with static photographs of actors' face and requiring them to choose one emotion label corresponding to the configuration of facial movements displayed in the picture. Findings of these early studies (e.g., Ekman, 1972; Ekman & Friesen, 1971; Ekman et al., 1969) tend to support the idea that (basic) emotions can be recognised universally based on their facial displays, although there seems to be a cultural in-group advantage for the recognition of the intended emotions. However, in everyday life, faces are rarely seen in isolation, as communication most typically involves the integration of information collected from various modalities simultaneously, including verbal information. Multisensory context – e.g., voices, different types of visual information, cultural orientation, words – influence how a face is perceived (Barrett et al., 2011). Moreover, communication can also happen without seeing the interlocutor's face, as during a phone call. In such a context, it becomes apparent that visual information can be particularly helpful when interlocutors do not share the same first language. Thus, it is crucial to consider the cross-linguistic communication of emotions across different modalities to refine our understanding of interpersonal emotion perception.

## 2. Approaches to emotion research

In the field of psychology, the various emotion theories can be organised along a continuum (Gross & Barrett, 2011). At one extremity, the basic emotion approach conceptualises

emotions as discrete, automatic reactions to a stimulus that can be recognised based on the one-to-one link between each entity and its specific physiological and behavioural manifestations (e.g., Ekman, 1992; Izard, 1971; Keltner & Shiota, 2003; Tomkins, 1962). On the other extremity, the constructionist approach regards emotions as momentary constructions of the mind based on the dimensional interoceptive sensations of valence and arousal – i.e., how pleasant and how activated one is feeling, respectively – and the exteroceptive information that contributes to making sense of these interoceptive sensations. The interpretation of interoceptive and exteroceptive information depends on previous experiences which are embedded in a specific sociocultural context (e.g., Barrett, 2014; Mesquita & Boiger, 2014; Russell, 2003). Finally, the appraisal approach functions as a bridge between these two extremities: it assumes that emotions are discrete categories that are triggered by an external event and bring about a chain of specific reactions – similarly to the basic approach. Contrary to the basic approach, however, the appraisal approach does not conceptualise emotions as reflexes that are independent of the meaning imparted by the experiencer on the trigger. Rather, appraisal theorists postulate that a specific emotion arises depending on APPRAISALS – i.e., the meaning that the experiencer imparts on the stimulus based on their needs, goals and values – which are dimensional in nature (e.g., Arnold, 1960; Scherer, 2001). The crucial role of the meaning imparted on information and the dimensional approach to emotions are common aspects of the appraisal and the constructionist approaches.

These different conceptualisations of emotions have led to different methodological choices to study emotion perception, and hence to different findings and conclusions. While studies conducted in the basic approach support the universality thesis (Nelson & Russell, 2013), i.e., the idea of universal emotions that are assumed to be experienced and expressed universally in the same way – and thus universally ‘recognisable’ from their manifestations (e.g., Ekman et al., 1969; Pell et al., 2009; Tombs et al., 2014), the constructionists assume that emotions cannot be universal since they are individual, momentary constructions embedded in a specific context. They suggest that an emotion category is an heterogeneous, fuzzy entity, which may be linked to a multitude of manifestations, and can thus not be ‘recognized’ as a perceiver-independent object (Barrett, 2017). The “glue” that holds this heterogeneous entity together – despite the lack of perceptual regularity or consistency among different instances of that category – is language: namely, the label used to refer to the multiple instances of that category (Barrett & Lindquist, 2008; Lindquist et al., 2015). Constructionists support the minimal universality hypothesis (Russell, 2003), which assumes that the only universal aspects of emotions are the dimensions of valence and arousal, which they call “core affect”. According to constructionists, the universality thesis emerged from methodological artefacts. One of the most influential artefacts is the use of forced-choice response format, which boosts interpersonal and intercultural agreement in perceptions (Gendron et al., 2014b, 2014a, 2018).

### 3. Emotion perception across modalities

Several studies comparing emotion perception in the visual modality and in the vocal modality found that the intended emotions are recognised less ‘accurately’ when presented in the vocal-only modality compared to the (dynamic) visual-only

modality (Burns & Beier, 1973; Collignon et al., 2008; Paulmann & Pell, 2011). However, when visual and vocal modalities are available simultaneously, vocal cues seem to contribute more heavily than visual cues to the overall emotion perception (Bänziger et al., 2009). Generally, the availability of dynamic visual cues also yields higher perceptions of emotional intensity compared to when only vocal and/or vocal-verbal modalities are available (Collignon et al., 2008; Dewaele & Moxsom-Turnbull, 2020; Lorette & Dewaele, 2022). Studies in the empathic accuracy research paradigm (Gesn & Ickes, 1999; Hall & Schmid Mast, 2007; Zaki et al., 2009) also found that visual cues, when available simultaneously with verbal ones, do not necessarily boost empathic accuracy – i.e., agreement between the experiencer and the perceiver about the feeling and/or thought in question. Verbal cues seem to affect empathic accuracy most, although someone else’s feelings can still be inferred (to some extent) based on visual cues only (Gesn & Ickes, 1999; Hall & Schmid Mast, 2007).

Overall, multimodality seems to facilitate emotion perception compared to unimodal contexts (Bänziger et al., 2009; Collignon et al., 2008; Kreifelts et al., 2007). However, the vast majority of previous studies investigating the facilitating effect of additional modalities for emotion perception focused exclusively on nonverbal communication. One exception is a study conducted in the discrete approach by Paulmann and Pell (2011) comparing emotion ‘recognition accuracy’ in English under six different conditions, i.e., visual-only, vocal-only, verbal-only, visual-vocal, vocal-verbal, and visual-vocal-verbal modality. The authors interpreted their results as evidence for multimodal encoding of emotions yielding better ‘recognition accuracy rates’ than bimodal encoding, and bimodal encoding in turn yielding better ‘recognition accuracy rates’ than unimodal encoding. However, the significance values of these multiple comparisons and the considerable overlap of the standard error bars in their graph (p. 198) suggests that these results might only hold for their particular sample and need to be confirmed in future research before allowing any generalisation.

The studies reviewed above did not take into consideration potential cultural or linguistic differences in the way perceivers integrate information from various modalities. However, Kitayama and Ishii (2002) found that, when confronted with vocal-verbal stimuli, English-speaking American are more attuned to verbal emotion cues while Japanese participant are more attuned to vocal emotion cues. Similarly, Tanaka et al. (2010) demonstrated cultural mediation of multimodal integration of emotional information. Japanese participants’ emotion perception was more influenced by the auditory modality, while Dutch participants were more influenced by visual information. These effects can also lead to shifts in how people integrate multisensory emotion information when they are exposed to a new culture or a new language later in life (Chen et al., 2022).

### 4. Emotion perception across languages (and cultures)

The field of emotion perception has been dominated by studies comparing emotion perception in different cultures, with the vast majority conducted in the basic approach and focusing on only one modality: be it the visual (e.g., Crivelli et al., 2016; Ekman & Friesen, 1971; Gendron et al., 2014b; Tombs et al., 2014), the vocal (e.g., Gendron et al., 2014a; Pell et al., 2009; Sauter et al., 2010; Scherer et al., 2001; Thompson & Balkwill, 2006) or the verbal modality (e.g., Pérez-García & Sánchez,

2020). To our best knowledge, only two studies (Koeda et al., 2013; Sneddon et al., 2011) investigated the cross-cultural perception of the valence and arousal level of someone else's internal state, moving away from a forced-choice response format and focusing on interpersonal emotion communication, but they both focused exclusively on cultural differences, without taking language into account. Both studies found slight differences between participants from different nationalities in terms of valence (Sneddon et al., 2011), but no difference in terms of arousal perception (Koeda et al., 2013). These two studies provide a first indication that the perception of the valence level of someone else's internal state might demonstrate slight cross-cultural variation, contrary to the theoretical assumption that core affect is a universal aspect of emotion (Russell, 2003).

However, although communication most typically involves verbal information, the studies reviewed so far took neither language nor multilingualism into account, as they only included L1 users or compared them with participants who do not have any knowledge of the language of communication – which we call L0 users. However, first and foreign language users (L1 and LX users) arguably behave differently from each other. Moreover, various aspects of language – relating to sounds, words, grammar and discourse – have been shown to interact differently with emotion depending on the language in question (see Majid, 2012 for a review). When communication happens between interlocutors who do not share the same L1, the extraction of relevant cues from the input might be trickier or cues might be interpreted differently (e.g., Irvine, 1982). On the verbal level, translation equivalents in language A and in language B will not necessarily completely match in terms of conceptual representation (De Groot, 1993, 2002). Especially in the case of abstract words, and arguably even more so in the case of emotion words that refer to deeply personal experiences shaped in the course of socialisation, specific connotations or nuances might be lost in translation. Indeed, the emotion lexicons of different languages have been shown to vary both in terms of structure and in terms of conceptual organisation (Panayiotou, 2006; Pavlenko, 2008; Semin et al., 2002; Wierzbicka, 1999). On the vocal level, languages differ in the (prototypical) social and affective meanings usually associated with pitch levels (Ohara, 2001; Valentine & Saint Damian, 1988; Van Bezooijen, 1995, 1996), shifts in pitch (Fant, 1973; Swan & Smith, 2001; Yuasa, 2002), intonation contours (Gibson, 1997; Gumperz, 1982; Holden & Hogan, 1993; Swan & Smith, 2001), stress, length and volume (Holden & Hogan, 1993; Swan & Smith, 2001), tempo (Besnier, 1990; Ron & Scollon, 2001), and rhythm (Swan & Smith, 2001). Therefore, verbal and vocal cues to someone else's internal state might be interpreted differently by L1 and LX users.

To our best knowledge, all studies comparing L1 and LX users' emotion perception have been conducted in the basic approach, i.e., with constraining tasks and focussing on the 'recognition' of a specific emotion in a stimulus rather than on the broader (and arguably universal) dimensions of valence and arousal. In a pioneering study, Rintell (1984) found that L1 English participants were better able to 'recognise' the intended emotions conveyed in 11 vocal-verbal stimuli than LX English users. Rintell interpreted her results as evidence for a cultural distance effect, as Chinese learners of English had significantly more difficulties to 'recognise' the intended emotions than Arabic and Spanish learners of English. This trend of Asian participants having more difficulty to perceive emotions than participants from other cultures has been echoed in numerous subsequent studies:

be it in the vocal modality or in other modalities (Lorette & Dewaele, 2015, 2020; Scherer et al., 2001; Thompson & Balkwill, 2006; Tombs et al., 2014). Moving beyond a single modality, Riviello et al. (2011) investigated how American L1 users of English and Italian and French LX users of English categorised emotions expressed by American actors and compared perceptions in the visual-only modality and in the vocal-verbal modality. Modality did not affect emotion recognition for American L1 participants nor for French LX participants, but Italian LX participants recognised the intended emotion more accurately in the visual-only than in the vocal-verbal modality. Moreover, while the visual-only modality did not yield any difference between American, Italian and French participants, Italian LX participants had significantly more difficulty to 'recognise' the intended emotions than the French LX and American L1 participants in the vocal-verbal modality. This suggests that language, culture, and communication modality interact in various ways in their influence on emotion perception.

More recently, several studies – influenced by the basic approach and implementing a forced-choice response format – were conducted as part of a project investigating emotion perception in L1 and LX English. In a first study (Lorette & Dewaele, 2015), L1 and LX English users categorised the main emotion conveyed in each of the six stimuli in which a British-English-speaking actress intended to convey an emotion in a short monologue. Contrary to Rintell's (1984) findings, 'recognition accuracy' was similar for L1 and LX users of English. This difference in findings might come from the different nature of the LX learners and from the different modalities investigated in both studies – i.e., young English Foreign Language learners who had mostly acquired English in a formal context versus older LX users using English both in naturalistic and formal contexts. Moreover, Rintell's participants heard vocal-verbal stimuli, while Lorette and Dewaele's participants were presented with visual-vocal-verbal stimuli. Following-up on this, the database was enriched with data from L1 and LX participants who were presented with the same stimuli used in the previous study, except that the visual modality had been made unavailable – they were presented with the audio recordings only (Lorette & Dewaele, 2020). Results showed that participants who could only rely on vocal-verbal cues had more difficulty to 'recognise' the intended emotions than participants who had been presented with the visual-vocal-verbal stimuli, which expands previous findings (e.g., Collignon et al., 2008; Kreifelts et al., 2007; Paulmann & Pell, 2011). Moreover, consistently with Rintell's (1984) findings, L1 users outperformed LX users in the vocal-verbal modality while no difference was found in the visual-vocal-verbal modality.

L1 and LX users do not only interpret the emotions experienced by their interlocutor differently, they also seem to perceive the intensity of the emotions experienced by their interlocutor differently. In another study of this project (Lorette & Dewaele, 2022), L1 users rated the intensity of the emotion experienced by the actress significantly lower than LX users, regardless of the modality in which they were presented with the stimuli. This finding was surprising in the light of the existing literature on the so-called 'detachment effect' (coined by Marcos, 1976). This refers to reduced emotionality in an LX compared to a L1 and has been supported in several experimental studies (e.g., Caldwell-Harris & Ayçiçeği-Dinn, 2009; Thoma & Baum, 2019; Wu & Thierry, 2012) while various other studies failed to replicate this effect (e.g., Conrad et al., 2011; Eilola et al., 2007; Opitz & Degner, 2012). However, following Dewaele and Moxsom-Turnbull

(2020), one might speculate that LX users are actually aware of the detachment effect and inflate their ratings as they realise that they might underrate emotions in their LX. By doing so, they would overcompensate for their conscious lack of emotionality, leading to even higher ratings than L1 users' ratings. This speculation regarding a potential 'LX overcompensation effect' has also been formulated in the only study comparing L1 and LX users' perception of valence and arousal rather than the perception of a specific emotion (Mavrou & Dewaele, 2020). In this study, Spanish L1 and LX users indicated their perception of "emotionality" (which could be understood as arousal) and "pleasantness" (which could be understood as valence) of a story. This story was presented either in the audiovisual modality (original version of an award-winning animated short film) or in the visual modality (sequence of 30 images representing the most important events of the story). As the authors explain, "the scenes presented in the story can elicit joy and sadness, as well as admiration, melancholy, empathy, confusion, etc." (p. 320). The audiovisual modality yielded higher pleasantness ratings than the visual modality for both L1 and LX users. Moreover, although no significant difference appeared between both groups in the audiovisual modality, LX users did perceive higher emotionality than L1 users in the visual modality. Similarly to Dewaele and Moxsom-Turnbull (2020) and to Lorette and Dewaele (2022), the authors argue that this might be due to LX users overcompensating their ratings due to their awareness of reduced emotionality in a LX, although the design of this study does not allow to draw conclusions about the mechanisms underlying the participants' perceptions. Furthermore, American LX users found the audiovisual stimuli more pleasant than the Asian LX users, echoing Sims et al.'s (2015) findings that Americans tend to focus more on the positive than Asian participants. However, in this study, modality was confounded with static versus dynamic stimuli and the ratings were based on a single measurement, calling for caution of interpretation.

In summary, it is striking that very few studies on interpersonal emotion perception across modalities have considered multilingualism, and most studies comparing L1 and LX users' emotion perception across modalities have been conducted in the basic approach, using a forced-choice response format with only a very limited number of response options and regarding the emotion intended to be conveyed in the stimulus as the accurate response. This limits the generalisability or the validity of conclusions on emotion perception across modalities, cultures, and languages. There thus is an urgent need for more studies in the constructionist approach. Moreover, the majority of previous studies have focused on the perception of emotions expressed by Westerners.

## 5. Research question and hypothesis

The research question of the present contribution is:

- Does communication modality affect the perception of the emotional state of a Chinese interlocutor similarly for L1, LX and L0 users of Mandarin?

Given initial evidence for cultural variation in valence and arousal perception (Koeda et al., 2013; Sneddon et al., 2011) and evidence for higher emotionality (Dewaele, 2011, 2013; Pavlenko, 2005) and for an 'emotion recognition advantage' in an L1 compared to an LX (Graham et al., 2001; Lorette & Dewaele, 2020; Rintell, 1984), we expect that, compared to the

L1 users, the less familiar users of Mandarin (i.e., LX and L0 users) will perceive the internal state of their interlocutor as less activated in terms of arousal and more neutral in terms of valence perception, especially in the absence of visual cues (Burns & Beier, 1973; Collignon et al., 2008; Paulmann & Pell, 2011).

## 6. Materials and methods

### 6.1. Stimuli

Stimuli were twelve 10- to 17-second-long recordings presented in one of the four modalities included in this study – namely, visual-vocal-verbal, vocal-verbal, visual-only, and vocal-only. For each stimulus, a different scenario was imagined, depicting a situation which could typically trigger the emotion in question for a Chinese person – see Chinese transcriptions and English translations of scenarios in the Appendix S1. These scenarios were imagined together with two researchers who were born and raised in China in order to guarantee the plausibility of these situations in a Chinese context and avoid a Western bias. These were then conveyed by a 27-year-old Mandarin L1 user from Beijing and recorded in the visual-vocal-verbal modality, i.e., as audiovisual recordings. Three additional versions of each stimulus were then created by making one or two modalities unavailable at a time – i.e. vocal-verbal stimuli (audio without visuals), visual-only stimuli (visuals without audio), and vocal-only stimuli (low-pass-filtered audio recordings making the words indecipherable but retaining prosodic information such as intonation and rhythm). The stimuli can be found in all four modalities online on OSF (<https://osf.io/dgys4/>). The intended emotions in the stimuli were happiness, sadness, disgust, (positive) surprise, fear, anger, embarrassment, contempt, pride, hope, and *jiu jié* 纠结 – which might be translated to feeling tangled together or in a knot, feeling confusion and chaos due to a difficult situation in which one cannot take a decision – and *wěi qu* 委屈 – which might be translated to feeling wronged or feeling unfairly treated. The focus of this investigation is not on agreement between emotion experiencer and emotion perceiver – i.e., not on "accurate recognition" of intended emotions – but on how different emotional states would be interpreted by different groups of participants. Therefore, the inclusion of these emotional states simply served to guarantee the inclusion of various levels of valence and arousal across the stimuli, regardless of whether the actor really succeeded in conveying these emotions<sup>1</sup>. Each stimulus is still labelled by the intended emotion to give an idea of the intended valence and arousal of the content of the stimulus. Each participant saw each of the 12 stimuli once, in randomised order. For each stimulus individually, the modality of presentation was randomly assigned for each participant, meaning that each participant was presented with different modalities throughout the emotion perception task, and that not all participants saw the same stimulus in the same modality.

### 6.2. Response format

After seeing and/or hearing each stimulus, participants indicated the perceived level of valence and arousal of the Chinese interlocutor using the Two-Dimensional Affect and Feeling Space (2DAFS, Lorette, 2021), a dynamic response format allowing to collect dimensional valence and arousal ratings as continuous data in a first phase and categorical perceptions of specific feelings in a second phase – see Figure 1. In phase 1, participants see a



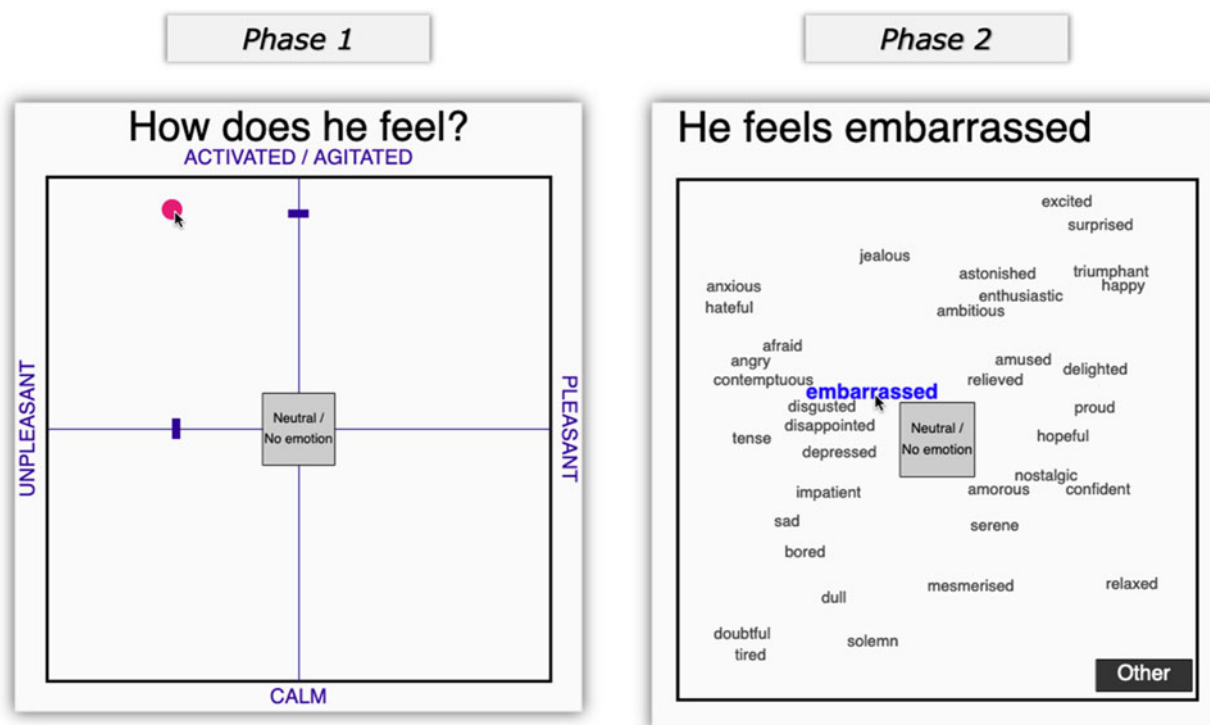


Figure 1. Phase 1 and phase 2 of the 2DAFS – data collected in phase 2 are not discussed in this contribution.

two-dimensional space characterized by a horizontal axis  $x$  going from “unpleasant” to “pleasant” – i.e., valence – and a vertical axis  $y$  going from “calm” to “activated/agitated” – i.e., arousal. The cursor – highlighted in pink – can freely move in the space, and its position is simultaneously reflected with a pointer on both axes. To indicate their perception, the participants click on the spot corresponding to the perceived level of valence and arousal and the instrument collects the data in terms of the  $(x, y)$  coordinate of the chosen spot – ranging between 0 (extremely unpleasant / calm) and 800 (extremely pleasant / activated) for each axis. Participants can also click in a square in a middle of the space to indicate that they perceived neutral / no emotion.<sup>2</sup> Rating differences smaller than 60 are considered meaningless and thus negligible, because pre-tests of the instrument have shown that when users are asked to make the smallest shift they can with their pointer in the two-dimensional space, the smallest adjustment in location of their cursor that they can perceive as located at a significantly different spot on the axis is 60 – i.e., everything below a 60-point difference is thus regarded as meaningless.

Data collected in phase 2 about the perception of specific feelings were also collected during the same study but are not discussed in this article. A more detailed description of this instrument is provided in Lorette (2021).

### 6.3. Participants

Data were collected from 1485 participants (630 males, 828 females, 4 others, 23 no information) aged between 14 and 88 years old (*mean* age = 28, *SD* = 13). Six hundred fifty-one of them reported being L1 users of Mandarin (*mean* age = 22, *SD* age = 8), 292 reported being LX users (*mean* age = 31, *SD* age = 14), and 542 reported being L0 users (*mean* age = 36, *SD* = 13).<sup>3</sup>

However, the number of observations varies between 1298 to 1312 per stimulus and between 210 and 387 per modality for each stimulus because stimulus order and modality were randomised, and 274 participants dropped out before the end of the task (but their responses before dropping out were still included in the analysis). Gender and age distribution differed between groups, but Mann-Whitney  $U$  analyses revealed non-significant differences<sup>4</sup> in valence and arousal ratings between male and female participants, neither for positive nor for negative stimuli. On the other hand, a significant Spearman's correlation between age and arousal ratings was shown for positive stimuli ( $p < .001$ ) and a significant relationship between age and valence ratings for negative stimuli ( $p < .001$ ), but these correlations were both very weak (Spearman's  $\rho = -.109$  and  $\rho = .036$ , respectively).<sup>5</sup> Participants come from all inhabited continents, with the most represented nationality being Chinese ( $n = 623$ ), followed by French ( $n = 107$ ), Belgian ( $n = 95$ ), British ( $n = 88$ ), American ( $n = 69$ ), and Italian ( $n = 43$ ).

### 6.4. Procedure

The stimuli were embedded in an online questionnaire shared via mailing lists and social media in English, simplified Chinese, and traditional Chinese. The questionnaire – including the response format – was also available in all three languages<sup>6</sup>, with participants having the possibility to choose the language they preferred and to complete it on both desktop and mobile devices – although they were encouraged to use a desktop device. However, in all three versions, the stimuli were the same, i.e., a Chinese speaker interacting in Mandarin. The questionnaire opened with sociodemographic-background questions, followed by a tutorial video presenting the interlocutor in the upcoming stimuli and introducing the response format. After the emotion perception

task, the last part of the questionnaire was a Mandarin proficiency test – the data of which are disregarded here.

### 6.5. Data analysis

The two outcome variables, i.e., valence ratings and arousal ratings, have been analysed separately. The relationships between each outcome variable and the nominal independent variables language group and communication modality were investigated by means of separate linear regression models for the stimuli intending to convey positively-valenced emotions and the stimuli intending to convey negatively-valenced emotions, because potential valence effects would be expected in different directions. Non-parametric bootstrapping – implementing the boot package (v1.3-27, Canty & Ripley, 2021) in R – was used in order to overcome the unreliability of standard errors, the significance values of the parameters, and the overall fit of the entire models due to heteroskedastic and non-normal data (Davison & Hinkley, 1997). The method adopted in the analyses of this study allows a comparison of the amount of variance explained by a first model including  $n$  parameters with the amount of variance explained by a second model including  $n + 1$  parameters. This difference of explained variance is then bootstrapped, resulting in a bootstrapped (adjusted)  $r^2$  difference. Once this method establishes a significant effect of the independent variable on the outcome variable, one needs to find out which levels of the independent variable lead to significant differences in the outcome variable. Therefore, bootstrapped regression models were run several times with releveling, i.e., with a different factor level forced as baseline level in order to compare each level against all other levels.<sup>7</sup> In order to correct for the multiple comparisons computed between each of the three levels of the variable language user group and each of the four levels of the variable modality, 99% CIs will be examined<sup>8</sup>.

## 7. Results

As the research question pertains to whether communication modality affects valence and arousal perception similarly across

language groups, the crucial effect of interest is an interaction of modality and language user group. However, we first discuss the main effect of modality and the main effect of language group on valence and arousal perception before turning to their interaction.

### 7.1. Main effects of language user group and modality

Overall, valence and arousal were perceived rather similarly by L1, LX, and L0 users. As summarised in Table 1, language user group did not significantly explain any variance in valence perception. Moreover, it explained less than 0.2% of variance in arousal perception for negative stimuli, which we regard as negligible. In the case of positive emotions, language user group explained 3% of variance in arousal perception, with L1 users perceiving significantly higher arousal levels than LX and LI users – see Figure S1, panel B in the supplementary materials. Modality, on the other hand, significantly explained between 2% and 30% variance in valence and in arousal ratings for both positive and negative stimuli. For positive stimuli, modality explained 6% of variance in arousal ratings and 30% of variance in valence ratings, with the vocal-verbal modality yielding lower perceptions of both valence and arousal compared to the visual-vocal-verbal and the visual-only modality – see Figure S2, panel A and B respectively, in the supplementary materials. For negative stimuli, modality explained 6% of variance in valence ratings, with the visual-vocal-verbal and the visual-only modality yielding lower (i.e., more negative) valence perceptions than the vocal-verbal modality, and the visual-vocal-verbal modality yielding even lower valence ratings than the visual-only modality – see Figure S2, panel C. Finally, for negative stimuli, the 2% variance explained in arousal by modality was solely driven by the vocal-only modality. We decided to disregard differences between the vocal-only modality and any other modality, since the average valence and arousal ratings in the vocal-only modality was neutral in all cases. In other words, the sole presence of prosodic cues did not yield consistency between the perceivers regarding the level of valence and arousal of the internal state of the Chinese speaker.

**Table 1.** Amount of variance in valence ratings (panel A) and arousal ratings (panel B) explained by language user group, modality, and the interaction term language user group \* modality in positive and negative stimuli (significant adjusted  $r^2$  in bold, significant but negligible adjusted  $r^2$  in italics, based on 95% BCa CIs for 20000 resampling).

A. VALENCE									
	Language user group			Modality			Language user group * modality		
	95% Bca CI (R = 20000)			95% Bca CI (R = 20000)			95% Bca CI (R = 20000)		
	Adj. $r^2$	Lower	Upper	Adj. $r^2$	Lower	Upper	Adj. $r^{22}$	Lower	Upper
Positive	<.001	-.000	.001	<b>.308</b>	<b>.285</b>	<b>.331</b>	<b>.013</b>	<b>.007</b>	<b>.018</b>
Negative	.007	.004	.01	<b>.062</b>	<b>.054</b>	<b>.071</b>	<b>.011</b>	<b>.007</b>	<b>.015</b>
B. AROUSAL									
	Language user group			Modality			Language user group * modality		
	95% Bca CI (R = 20000)			95% Bca CI (R = 20000)			95% Bca CI (R = 20000)		
	Adj. $r^2$	Lower	Upper	Adj. $r^2$	Lower	Upper	Adj. $r^2$	Lower	Upper
Positive	<b>.025</b>	<b>.017</b>	<b>.033</b>	<b>.06</b>	<b>.048</b>	<b>.072</b>	<b>.015</b>	<b>.009</b>	<b>.021</b>
Negative	.002	.000	.003	<b>.019</b>	<b>.014</b>	<b>.024</b>	.002	.000	.003

In summary, modality explains more variance in arousal and in valence perception than language user group. The presence of visuals tends to lead to higher arousal ratings for both positive and negative stimuli, and to yield more extreme valence perceptions (i.e., more pleasant for positive stimuli and more unpleasant for negative stimuli) compared to when visual cues are absent.

## 7.2. Interaction effects of language user group and modality

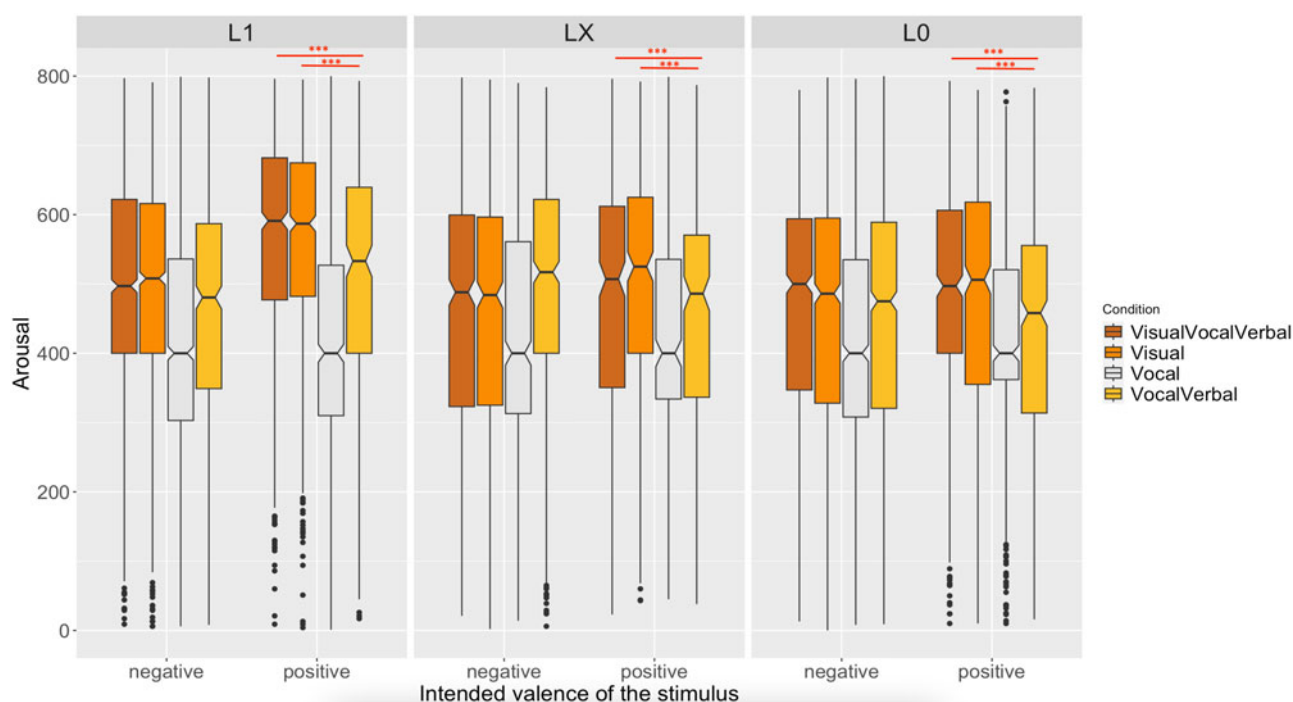
The bootstrapped regression models revealed a significant interaction effect of language user group and modality on arousal and perception ratings for both positive and negative stimuli, with between 0.2% and 2.5% of variance explained by the interaction term. In the case of negative stimuli, the amount of variance in arousal ratings explained by the interaction of language user group and modality is negligible (0.2%). Given that the main effect of language user group was non-significant and the main effect of modality was solely driven by the vocal-verbal modality – which yields neutral arousal ratings – these results overall suggest that arousal ratings for negative stimuli are not affected by modality nor language user group. The picture is somewhat similar for positive stimuli. Although the interaction of language user group and modality accounted for 3% of variance in arousal perception, this interaction was solely driven by dissimilar patterns across the language user groups regarding differences between the vocal-only modality (which yielded neutral ratings) and other modalities. Therefore, we disregard this interaction, since for other modalities, the same pattern holds for all three language groups – namely, that the vocal-verbal modality yielded lower arousal perceptions compared to the visual-only and the visual-vocal-verbal modality – see Figure 2.

Turning to valence perception, negative stimuli yielded a (small) interaction of language user group and modality, explaining 1% of variance in valence perception. As visualised in

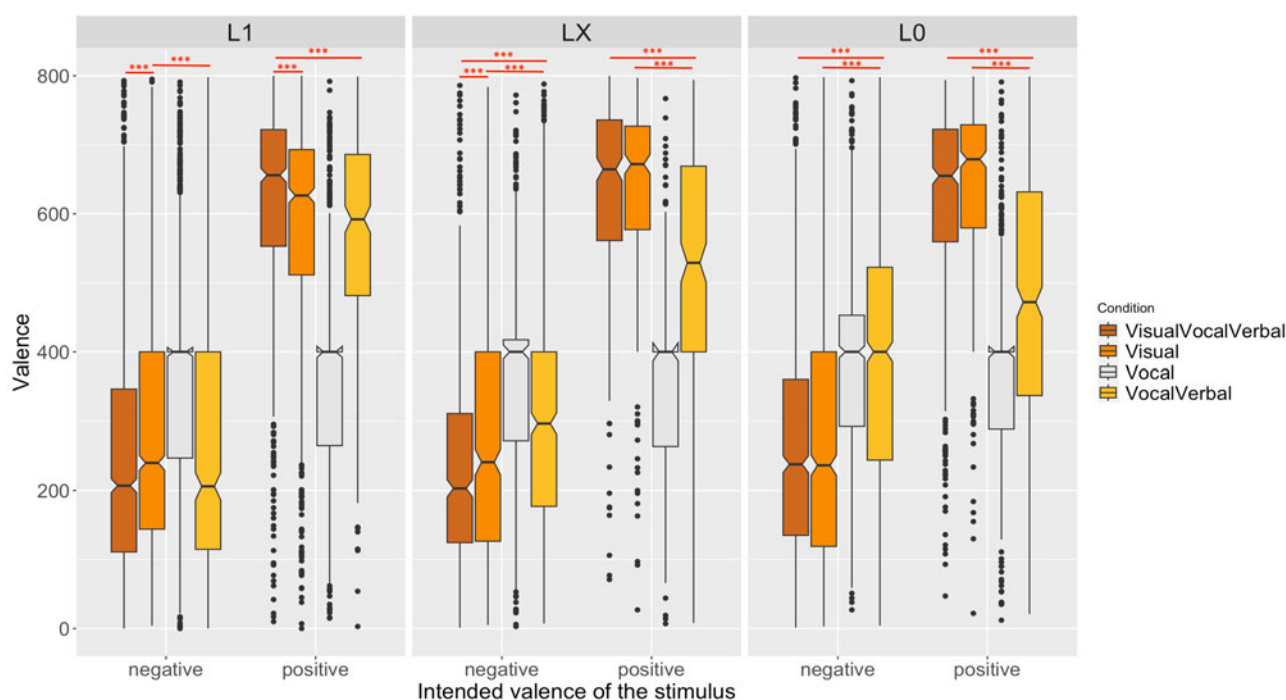
Figure 3, the presence of visual cues led LX and L0 users to perceive their Chinese interlocutor as feeling even more unpleasant compared to the vocal-verbal modality, while the absence of vocal-verbal cues (i.e., in the visual-only modality) yielded slightly less negative perceptions among L1 users. For positive stimuli, the interaction term also explained 1% of variance in valence perception. While the absence of visuals yielded all three language user groups to perceive their interlocutor as feeling less positive, this was especially marked for LX users and (even more so) for L0 users. Moreover, the visual-only and the visual-vocal-verbal modalities did not lead to different valence perceptions among LX and L0 users, while L1 users perceived even more positive valence levels in the visual-vocal-verbal modality compared to when visual cues were not accompanied by vocal-verbal ones.

## 8. Discussion

In this paper, we aimed to investigate whether communication modality affects emotion perception to the same extent for L1, LX and L0 users of a language. Specifically, we compared L1, LX, and L0 Mandarin users' perceptions of how (un)pleasant – i.e., valence – and how (un)activated – i.e., arousal – a Chinese speaker feels when they can rely on visual and/or verbal and/or vocal cues. Although the psychological constructionist approach assumes that valence and arousal are universal features of emotions (Russell, 2003) and would thus not predict differences in valence and arousal perception, there is some evidence for cultural variation in valence and arousal perception (Koeda et al., 2013; Sneddon et al., 2011), but the linguistic background of compared groups has rarely been taken into account. Moreover, previous self-report studies have highlighted an overall reduced LX emotionality compared to L1 (Dewaele, 2011, 2013; Pavlenko, 2005) and an 'emotion recognition advantage' for L1 users compared to LX or L0 users (Graham et al., 2001; Lorette & Dewaele,



**Figure 2.** Boxplot showing the arousal ratings per language group in each modality for positive and negative stimuli, with significant differences at .01 level indicated with red starred lines.



**Figure 3.** Boxplot showing the valence ratings per language group in each modality for positive and negative stimuli, with significant differences at .01 level indicated with red starred lines.

2020; Rintell, 1984). It was hypothesised that the less familiar one is with Mandarin, the closer to neutral their valence perception of the Mandarin speaker would be and the lower their arousal perception would be, especially in the absence of visual cues.

Analyses indicated that, except for positive stimuli, perceptions of the arousal level of one's interlocutor are stable for L1, LX and L0 users irrespective of the communication modality. In other words, regardless of whether one uses a language as a L1, a LX, or a L0, being able to only hear, only see, or simultaneously hear and see one's interlocutor does not lead to different interpretations of how activated this interlocutor feels. Only in the case of positive stimuli did the presence of visual cues lead to higher arousal perceptions, for all three language user groups. On the other hand, valence perception shows more variation. While modality yields to small differences in L1 users' valence perception, LX and L0 users perceive their interlocutor's valence level very differently depending on whether they can see him or not. When they don't have access to visual cues, LX and L0 users' perception of the valence level of their interlocutor is much closer to neutral – i.e., less pleasant for positive stimuli and less unpleasant for negative stimuli, which is only true to a much smaller extent for L1 users. This highlights the importance of visual (i.e., non-verbal) cues when one is less familiar with a language, and thus also with a culture.<sup>9</sup> As language can be seen as a doorway to culture, being less familiar with Mandarin might also imply being less familiar with the cultural mandates in China, i.e., the “cultural norms, ideals, or goals for how to be a good person, how to interact, how to build good relationships, or even more specifically, how to feel” (Mesquita et al., 2017, p. 97). This boosting effect of visual cues on valence perception echoes the facilitating effect of visual cues found in previous research into emotion categorisation conducted in the basic approach (Burns & Beier, 1973; Collignon et al., 2008; Lorette & Dewaele, 2020; Paulmann & Pell, 2011). On the other hand, our findings do not echo the

facilitating effect of multimodality over unimodality found in previous research (Bänziger et al., 2009; Collignon et al., 2008; Kreifelts et al., 2007; Lorette & Dewaele, 2020, 2022; Paulmann & Pell, 2011), even for the LX and L0 users. LX and L0 users' valence perceptions in the visual-only modality did mostly not differ from perceptions in the visual-vocal-verbal modality but were more extreme than in the vocal-verbal modality – i.e., higher in the case of positive stimuli and lower in the case of negative stimuli. In other words, for LX and L0 users, the availability of vocal and verbal cues on top of visuals does not alter the perception of the valence of the speaker's internal state compared to perceptions solely based on visual cues<sup>10</sup>. These findings, however, are limited by the rather low proficiency level of our LX users. It could be the case that the valence perceptions of higher proficient LX users may approximate those of L1 users even in the absence of visuals, which will have to be addressed in further research.

The fact that visuals are particularly influential for LX and L0 users' valence perception but less important for arousal perception might indicate that arousal is overall more consistently perceived than valence regardless of the environment and the information available, while vocal and/or verbal cues might be important to fine-tune valence perceptions. This suggests a more universal character of arousal compared to valence (see also Bhatara et al., 2016; Koeda et al., 2013; Sneddon et al., 2011), although further research is needed to confirm this claim. These findings would also chime with previous work demonstrating that arousal is processed earlier and more automatically than valence (e.g., Dresler et al., 2009; Kensinger & Corkin, 2004; Schimmack, 2005). The minimal universality hypothesis does not offer a theoretical framework that can account for differences between valence and arousal, neither in terms of perception nor in terms of processing. Panksepp's (1998) hierarchical model of emotions, however, offers one possible framework to account



for our findings. It posits that arousal is processed at the so-called primary level, which, according to him, involves subcortical circuits and is shared with all mammals. Valence processing – i.e., categorising an event as positive or negative – would be a high-order operation that is processed later, at the tertiary, neocortical level (Panksepp, 2006), and would be more dependent on the availability of cognitive resources. Accordingly, when language processing is cognitively more challenging, such as in the case of LX, valence processing would be more limited, especially when language processing cannot be supported by nonverbal information such as visuals.

Finally, although the perception differences revealed in this study are limited, they do suggest that the minimal universality hypothesis (Russell, 2003), claiming that valence and arousal are universal, needs more nuance. Although it seems that, in most cases, people can universally interpret *WHETHER* their interlocutor is feeling pleasant or unpleasant and *WHETHER* their interlocutor is feeling activated or calm, perception differences do exist regarding *HOW STRONGLY* someone is feeling (un)pleasant and (de)activated. Therefore, it is crucial to use continuous ratings of valence and arousal – rather than dichotomous judgements (e.g., Crivelli et al., 2017) – to refine our claims about the universality of emotion perception.

Therefore, interlocutors need to be aware of the fact that they interpret the emotional state of their interlocutor through the prism of their own linguistic (and cultural) background, which might not reflect the actual emotions and illocutions of the other interlocutor(s) in the communication. It is especially pivotal to stress this in domains in which (intercultural) communication has crucial implications, such as business or conflict resolution. Moreover, these findings have specific implications for LX learning and teaching. Since few – if any – aspects of emotion perception can be regarded as completely universal, LX learners should be made aware of the linguistic (and cultural) differences in emotion communication, echoing the development witnessed in LX learning “from the focus of teaching culture towards the development of intercultural competence” (Coffey, 2013, p. 279). Despite the absence of ‘universal truths’ one can rely on, LX learners can be made aware of certain trends within a linguistic or cultural group and of certain cultural mandates – although variation within linguistic and cultural groups should be emphasised – and of factors that tend to affect emotion perception. Such practices could contribute to the establishment of the learning process as “an authentic social practice, [...] which appears to be valued as more meaningful and leading to more sustainable engagement with the language learning project” (Coffey, 2009, p. 10).

Our findings are limited by some methodological choices. First, our conclusions only pertain to emotion perception – rather than processing – as they are solely based on self-report data. Further behavioural and processing studies could share more light on the actual mechanisms accounting for these perception differences. Second, the fact that the survey was available in three languages introduces potential issues related to translation equivalents. However, this is a limitation which is inherent to multilingual research. Moreover, as the survey was only available in three languages, some participants have filled in the survey in their L1, while others have filled in the survey in a LX. The added cognitive load of responding to a survey in a LX might also have affected emotion perceptions, since research suggests that the LX detachment effect, leading to lower reactions to emotional stimuli in a LX compared to a L1, might result from reduced processing automaticity potentially due to LX use (Thoma & Baum, 2019).

Finally, our L1, LX and L0 subsamples were rather heterogeneous in terms of linguistic and cultural background. Even within our L1 sample, valence and arousal perception differences emerged between different regions of China (Lorette, in prep.). Although this study provides a first indication for variation in valence and arousal perception, further research is needed with more homogeneous groups to gain more confidence about our conclusions. In the future, the cultural background of the participants, and their familiarity with the culture of their interlocutor, will need to be better controlled, although that necessarily has implications for the easiness to access large samples.

## 9. Conclusion

We started the introduction referring to phone conversations between world leaders – with the use of interpreters – where interlocutors risk misjudging how the person they were speaking to was feeling in the absence of visuals. Based on the current study we can suggest that not seeing each other may affect the interpretation of the positivity of the interaction, particularly so if the interlocutors have a low proficiency level in their LX. Our findings showed that communication modality had close to no effect on arousal perception, but that modality had a (limited) effect on valence perception, particularly for LX and L0 users of Mandarin who perceive more neutral valence levels in the absence of visuals. Although further research is needed to confirm this, the findings can be interpreted as an indication that verbal cues are not necessarily crucial to roughly interpret the valence and arousal level of the internal state of one’s interlocutor, and that visual cues are particularly helpful for LX and L0 users to approximate L1 users’ perceptions, although they might then lack the verbal cues necessary to fine-tune their perceptions (see Lorette, 2021). This contribution to our understanding of emotion perception in different modalities is innovative as it relies on empirical evidence that is exclusively based on dynamic modalities, including both nonverbal and verbal modalities – contrary to previous studies in which verbal modalities (Bänziger et al., 2009; Collignon et al., 2008) or nonverbal modalities (Lorette & Dewaele, 2020) were excluded, or in which visual, vocal, and verbal modalities were confounded with dynamic versus static modalities (Paulmann & Pell, 2011).

To conclude, heads of state may want to install a “red video-phone” to talk to each other in times of crisis.

**Supplementary Material.** For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728923000925>

**Data availability statement.** Data are available from the authors upon request.

**Funding statement.** The project was partly funded by the Bloomsbury Studentship and Birkbeck College SSHP Postgraduate Support Fund.

**Competing interests.** The author(s) declare none

## Notes

<sup>1</sup> Each recording was assessed in a small-scale pilot study for the credibility of both the scenario and the acting of the speaker by four L1 users of Chinese on a nine-point Likert-type scale. Overall, the recordings obtained a mean rating of 8.3 ( $SD = .33$ ) out of 10 for the scenario’s credibility and 8.1 ( $SD = .48$ ) for the acting credibility.

<sup>2</sup> The delimitation of the middle “neutral” square ranges from coordinates 365 to 435. Therefore, ratings with a value between 365 and 435 are recoded to 400

to eliminate meaningless variation in the data since variation in the coordinates of any click within this range is arbitrary.

<sup>3</sup> On a scale from 1 (no mastery) to 5 (very good mastery) in four different language skills, the L1 users reported a higher proficiency ( $mean = 3.1$ ,  $SD = .7$ ) than the LX users ( $mean = 1.4$ ,  $SD = .9$ ). Note that data from 114 LX users being Chinese but having been raised in another Chinese variety different from Mandarin were also collected but are left out of the analyses for this contribution.

<sup>4</sup> Or significant but not meaningful – i.e., with a difference in location smaller than 60

<sup>5</sup> The inferential analyses reported below were also conducted with age as a factor besides language user group and modality, but no significant effect of age was revealed.

<sup>6</sup> Two L1 Mandarin translators translated the questionnaire from English to Chinese and then compared their translations during collaborative discussions until agreement was reached. A third translator finally reviewed this translation.

<sup>7</sup> Non-parametric bootstrapping of regression models was chosen rather than non-parametric Kruskal-Wallis tests because the distributions of valence or arousal ratings observed in each group or modality do not only differ in centrality, but also in shape. In such cases, rank-based tests such as Kruskal-Wallis tests only indicate if participants from different groups were drawn from different populations, but cannot be used to make any comparisons of means nor of medians between different groups (Dinno, 2015).

<sup>8</sup> 99% CIs correspond to an  $\alpha$  level of .01. This corresponds to the rounded adjusted  $\alpha$  level recommended in the case of 6 comparisons – i.e., one comparison between each of the four levels of the *Modality* variable. This adjusted  $\alpha$  level is obtained by dividing the original  $\alpha$  level – i.e., .05 – by the number of comparisons being performed (Loewen & Plonsky, 2015); in this case  $.05/6 = .008$ , rounded up to .01. The same alpha level is kept for comparisons between each of the three levels of the *Language user group* variable.

<sup>9</sup> Given the overall low level of (and low variation in) proficiency of the LX users in our sample, we did not include proficiency as a variable in our main analyses. However, proficiency is likely to play a role in the understanding of verbal cues, and thus have an effect on verbal emotion perception. Correlations performed on our data indeed suggest that higher proficiency is (weakly) related to lower valence perception for negative stimuli (Spearman's  $\rho = -0.6$ ,  $p = .003$ ). This significant relationship was not found in the case positive stimuli, potentially due to a lower statistical power since eight stimuli were negative and only four were positive.

<sup>10</sup> Our results suggest that this is more the case for L1 users, but future research will need to confirm this.

## References

- Arnold, M. B. (1960). *Emotion and personality. Vol. I. Psychological aspects*. Columbia University Press.
- Baker, V. (2016, December 5). Presidential phone calls: How do world leaders talk to each other? *BBC News*. <https://www.bbc.com/news/world-us-canada-38202271>
- Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion*, 9(5), 691–704. <https://doi.org/10.1037/a0017088>
- Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, 6(4), 292–297.
- Barrett, L. F. (2017). *How Emotions Are Made. The Secret Life of the Brain*. Houghton Mifflin Harcourt.
- Barrett, L. F., & Lindquist, K. A. (2008). The embodiment of emotion. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 237–262). Cambridge University Press.
- Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, 20(5), 286–290.
- Besnier, N. (1990). Language and affect. *Annual Review of Anthropology*, 19, 419–451.
- Bhatara, A., Laukka, P., Boll-Avetisyan, N., Granjon, L., Elfenbein, H. A., & Bänziger, T. (2016). Second language ability and emotional prosody perception. *PloS One*, 11(6), e0156855.
- Burns, K. L., & Beier, E. G. (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *Journal of Communication*, 23(1), 118–130.
- Caldwell-Harris, C. L., & Ayçiçeği-Dinn, A. (2009). Emotion and lying in a non-native language. *International Journal of Psychophysiology*, 71(3), 193–204.
- Canty, A., & Ripley, B. D. (2021). *boot: Bootstrap R (S-Plus) Functions*. [Computer software].
- Chen, P., Chung-Fat-Yim, A., & Marian, V. (2022). Cultural Experience Influences Multisensory Emotion Perception in Bilinguals. *Languages*, 7(12), 1–17.
- Clavin, T. (2013, June 18). *There Never Was Such a Thing as a Red Phone in the White House*. Smithsonian Magazine. <https://www.smithsonianmag.com/history/there-never-was-such-a-thing-as-a-red-phone-in-the-white-house-1129598/>
- Coffey, S. (2009). Representations of language learning: From talking about a school subject to talking about “authentic” contexts. *Encuentro*, 18, 1–11.
- Coffey, S. (2013). Strangerhood and intercultural subjectivity. *Language and Intercultural Communication*, 13(3), 266–282.
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, 1242, 126–135.
- Conrad, M., Recio, G., & 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. *Frontiers in Psychology*, 2(351), 1–16. <https://doi.org/doi:10.3389/fpsyg.2011.00351>
- Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, 145(7), 830–843. <https://doi.org/10.1037/xge0000172>
- Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2017). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. *Emotion*, 17(2), 337.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application* (Vol. 1). Cambridge University Press.
- De Groot, A. M. (1993). Word-type effects in bilingual processing tasks. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon*. (pp. 27–51). John Benjamins.
- De Groot, A. M. (2002). Lexical representation and lexical processing in the L2 user. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 32–63). Multilingual Matters.
- Dewaele, J.-M. (2011). Self-reported use and perception of the L1 and L2 among maximally proficient bi-and multilinguals: A quantitative and qualitative investigation. *International Journal of the Sociology of Language*, 208, 25–52.
- Dewaele, J.-M. (2013). *Emotions in multiple languages* (2nd ed.). Palgrave Macmillan.
- Dewaele, J.-M., & Moxsom-Turnbull, P. (2020). Visual cues and perception of emotional intensity among L1 and LX users of English. *International Journal of Multilingualism*, 17(4), 499–515.
- Dinno, A. (2015). Nonparametric Pairwise Multiple Comparisons in Independent Groups Using Dunn's Test. *Stata Journal*, 15, 292–300.
- Dresler, T., Mériaux, K., Heekeren, H. R., & van der Meer, E. (2009). Emotional Stroop task: Effect of word arousal and subject anxiety on emotional interference. *Psychological Research PRPF*, 73(3), 364–371. <https://doi.org/10.1007/s00426-008-0154-6>
- Eilola, T. M., Havelka, J., & Sharma, D. (2007). Emotional activation in the first and second language. *Cognition and Emotion*, 21(5), 1064–1076.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska symposium on motivation*, 1971 (Vol. 19, pp. 208–283). University of Nebraska Press.
- Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3–4), 169–200.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17(2), 124–129.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164, 86–88.
- Fant, G. (1973). *Speech sounds and features*. MIT Press.

- Gendron, M., Crivelli, C., & Barrett, L. F. (2018). Universality Reconsidered: Diversity in Making Meaning of Facial Expressions. *Current Directions in Psychological Science*, 27(4), 211–219.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014a). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, 25(4), 911–920.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014b). Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion*, 14(2), 251–262. <https://doi.org/10.1037/a0036052>
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology*, 77(4), 746–761.
- Gibson, M. (1997). Non-native perception and production of English attitudinal intonation. In J. Leather & A. James (Eds.), *New Sounds 97: Proceedings of the Third International Symposium on the Acquisition of Second-Language Speech* (pp. 96–102). University of Klagenfurt.
- Graham, C. R., Hamblin, A. W., & Feldstein, S. (2001). Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *IRAL - International Review of Applied Linguistics in Language Teaching*, 39(1), 19–37. <https://doi.org/10.1515/iral.39.1.19>
- Gross, J. J., & Barrett, L. F. (2011). Emotion Generation and Emotion Regulation: One or Two Depends on Your Point of View. *Emotion Review*, 3(1), 8–16. <https://doi.org/10.1177/1754073910380974>
- Gumperz, J. (1982). *Discourse strategies*. Cambridge University Press.
- Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7(2), 438–446.
- Holden, K. T., & Hogan, J. T. (1993). The emotive impact of foreign intonation: An experiment in switching English and Russian intonation. *Language and Speech*, 36(1), 67–88.
- Irvine, J. T. (1982). Language and affect: Some cross-cultural issues. In H. Byrnes (Ed.), *Contemporary Perceptions of Language: Interdisciplinary Dimensions* (pp. 31–47). Georgetown University Press.
- Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts.
- Keltner, D., & Shiota, M. N. (2003). New displays and new emotions: A commentary on Rozin and Cohen (2003). *Emotion*, 3, 86–91. <https://doi.org/10.1037/1528-3542.3.1.86>
- Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences*, 101(9), 3310–3315. <https://doi.org/10.1073/pnas.0306408101>
- Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion*, 16(1), 29–59.
- Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-Cultural Differences in the Processing of Non-Verbal Affective Vocalizations by Japanese and Canadian Listeners. *Frontiers in Psychology*, 4(105), 1–8. <https://doi.org/10.3389/fpsyg.2013.00105>
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage*, 37(4), 1445–1456. <https://doi.org/10.1016/j.neuroimage.2007.06.020>
- Lindquist, K. A., MacCormack, J. K., & Shaback, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, 6(444), 1–17.
- Loewen, S., & Plonsky, L. (2015). *An A-Z of applied linguistics research methods*. Macmillan International Higher Education.
- Lorette, P. (2021). Investigating emotion perception via the Two-Dimensional Affect and Feeling Space: An example of a cross-cultural study among Chinese and non-Chinese participants. *Frontiers in Psychology*, 12 (662610), 2597. <https://doi.org/10.3389/fpsyg.2021.662610>
- Lorette, P., & Dewaele, J.-M. (2015). Emotion recognition ability in English among L1 and LX users of English. *International Journal of Language and Culture*, 2(1), 62–86. <https://doi.org/10.1075/ijolc.2.1.03lor>
- Lorette, P., & Dewaele, J.-M. (2020). Emotion recognition ability across different modalities: The role of language status (L1/LX), proficiency and cultural background. *Applied Linguistics Review*, 11(1), 1–26. <https://doi.org/10.1515/applirev-2017-0015>
- Lorette, P., & Dewaele, J.-M. (2022). Interpersonal perception of emotional intensity by English first (L1) and foreign (LX) language users in audio(visual) communication. *International Journal of Multilingualism*, 1–18. <https://doi.org/10.1080/14790718.2022.2144326>
- Majid, A. (2012). Current Emotion Research in the Language Sciences. *Emotion Review*, 4(4), 432–443. <https://doi.org/10.1177/1754073912445827>
- Marcos, L. R. (1976). Bilinguals in psychotherapy: Language as an emotional barrier. *American Journal of Psychotherapy*, 30(4), 552–560.
- Mavrou, I., & Dewaele, J. (2020). Emotionality and pleasantness of mixed-emotion stimuli: The role of language, modality, and emotional intelligence. *International Journal of Applied Linguistics*, 30(2), 313–328. <https://doi.org/10.1111/ijal.12285>
- Mesquita, B., & Boiger, M. (2014). Emotions in Context: A Sociodynamic Model of Emotions. *Emotion Review*, 6(4), 298–302. <https://doi.org/10.1177/1754073914534480>
- Mesquita, B., Boiger, M., & De Leersnyder, J. (2017). Doing emotions: The role of culture in everyday emotions. *European Review of Social Psychology*, 28(1), 95–133.
- Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, 5(1), 8–15.
- Ohara, Y. (2001). Finding one's voice in Japanese: A study of the pitch levels of L2 users. In A. Pavlenko, A. Blackledge, A. Piller, & M. Teutsch-Dwyer (Eds.), *Multilingualism, second language learning, and gender*. (pp. 231–254). Mouton de Gruyter.
- Opitz, B., & Degner, J. (2012). Emotionality in a second language: It's a matter of time. *Neuropsychologia*, 50(8), 1961–1967. <http://dx.doi.org/10.1016/j.neuropsychologia.2012.04.021>
- Panayiotou, A. (2006). Translating guilt: An endeavor of shame in the Mediterranean? In A. Pavlenko (Ed.), *Bilingual Minds: Emotional Experience, Expression, and Representation* (pp. 183–208). Multilingual Matters.
- Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.
- Panksepp, J. (2006). *The Core Emotional Systems of the Mammalian Brain: The Fundamental Substrates of Human Conditions. About a Body: Working With the Embodied Mind in Psychotherapy*. Routledge.
- Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, 35(2), 192–201. <https://doi.org/10.1007/s11031-011-9206-0>
- Pavlenko, A. (2005). *Emotions and multilingualism*. Cambridge University Press.
- Pavlenko, A. (2008). Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. *The Mental Lexicon*, 3(1), 92–121.
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120.
- Pérez-García, E., & Sánchez, M. J. (2020). Emotions as a linguistic category: Perception and expression of emotions by Spanish EFL students. *Language, Culture and Curriculum*, 33(3), 274–289.
- Rintell, E. M. (1984). But How Did You FEEL about That?: The Learner's Perception of Emotion in Speech1. *Applied Linguistics*, 5(3), 255–264. <https://doi.org/10.1093/applin/5.3.255>
- Riviello, M. T., Esposito, A., Chetouani, M., & Cohen, D. (2011). Inferring Emotional Information from Vocal and Visual Cues: A Cross-Cultural Comparison. *2nd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 1–4.
- Ron, S., & Scollon, S. W. (2001). *Intercultural communication: A discourse approach* (2nd ed.). Blackwell Publishing.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172. <https://doi.org/10.1037/0033-295X.110.1.145>
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, 107(6), 2408–2412. <https://doi.org/www.pnas.org/cgi/doi/10.1073/pnas.0908239106>
- Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal*

- processes in emotion: Theory, methods, research (pp. 92–120). Oxford University Press.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92.
- Schimmack, U. (2005). Attentional Interference Effects of Emotional Pictures: Threat, Negativity, or Arousal? *Emotion*, 5(1), 55–66. <https://doi.org/10.1037/1528-3542.5.1.55>
- Semin, G. R., Görts, C. A., Nandram, S., & Semin-Goossens, A. (2002). Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition & Emotion*, 16(1), 11–28.
- Sims, T., Tsai, J. L., Jiang, D., Wang, Y., Fung, H. H., & Zhang, X. (2015). Wanting to maximize the positive and minimize the negative: Implications for mixed affective experience in American and Chinese contexts. *Journal of Personality and Social Psychology*, 109(2), 292–315. <https://doi.org/10.1037/a0039276>
- Sneddon, I., McKeown, G., McRorie, M., & Vukicevic, T. (2011). Cross-Cultural Patterns in Dynamic Ratings of Positive and Negative Natural Emotional Behaviour. *PLoS ONE*, 6(2), e14679. <https://doi.org/10.1371/journal.pone.0014679>
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge University Press.
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de Gelder, B. (2010). I Feel Your Voice: Cultural Differences in the Multisensory Perception of Emotion. *Psychological Science*, 21(9), 1259–1262. <https://doi.org/10.1177/0956797610380698>
- Thoma, D., & Baum, A. (2019). Reduced Language Processing Automaticity Induces Weaker Emotions in Bilinguals Regardless of Learning Context. *Emotion*, 19(6), 1023–1034. <https://doi.org/10.1037/emo0000502>
- Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica*, 2006(158). <https://doi.org/10.1515/SEM.2006.017>
- Tombs, A. G., Russell-Bennett, R., & Ashkanasy, N. (2014). Recognising emotional expressions of complaining customers: A cross-cultural study. *European Journal of Marketing*, 48(7/8), 1354–1374.
- Tomkins, S. (1962). *Affect Imagery Consciousness: Volume I: The Positive Affects*. Springer Publishing Company.
- Valentine, C. A., & Saint Damian, B. (1988). Gender and culture as determinants of the 'ideal voice'. *Semiotica*, 71(3–4), 285–304.
- Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253–265.
- Van Bezooijen, R. (1996). The effect of pitch on the attribution of gender related personality traits. In N. Warner, J. Ahlers, J. Bilmes, M. Oliver, S. Wertheim, & M. Chen (Eds.), *Gender and Belief Systems. Proceedings of the Fourth Berkeley Women and Language Conference, April 19–21, 1996* (pp. 755–766).
- Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.
- Wu, Y. J., & Thierry, G. (2012). How reading in a second language protects your heart. *Journal of Neuroscience*, 32(19), 6485–6489. <https://doi.org/10.1523/JNEUROSCI.6119-11.2012>
- Yuasa, I. P. (2002). Empiricism and emotion: Representing and interpreting pitch ranges. In S. Benor, M. Rose, D. Sharma, J. Sweetland, & Q. Zhang (Eds.), *Gendered practices in language* (pp. 193–209). CSLI Publications.
- Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, 9(4), 478–487.