Motivation and Emotion (2022) 46:719–734
https://doi.org/10.1007/s11031-022-09953-2

ORIGINAL PAPER

Audio-Visual Interactions during Emotion Processing in Bicultural Bilinguals

Ashley Chung-Fat-Yim[1] · Peiyao Chen[2] · Alice H. D. Chan[3] · Viorica Marian[1]

Accepted: 16 May 2022 / Published online: 23 August 2022

Abstract

Despite the growing number of bicultural bilinguals in the world, the way in which multisensory emotions are evaluated by bilinguals who identify with two or more cultures remains unknown. In the present study, Chinese-English bicultural bilinguals from Singapore viewed Asian or Caucasian faces and heard Mandarin or English speech, and evaluated the emotion from one of the two simultaneously-presented modalities. Reliance on the visual modality was greater when bicultural bilinguals processed Western audio-visual emotion information. Although no differences between modalities emerged when processing East-Asian audio-visual emotion information, correlations revealed that bicultural bilinguals increased their reliance on the auditory modality with more daily exposure to East-Asian cultures. Greater interference from the irrelevant modality was observed for Asian faces paired with English speech than for Caucasian faces paired with Mandarin speech. We conclude that processing of emotion in bicultural bilinguals is guided by culture-specific norms, and that familiarity influences how the emotions of those who speak a foreign language are perceived and evaluated.

Keywords Biculturalism · Bicultural bilinguals · Emotion · Emotion perception · Modality dominance · Cultural frame switching

One of the ways in which we can tell how another person is feeling is through their facial and vocal expressions. The ability to read the emotions of others allows us to build stronger connections, navigate new relationships and friendships, or capture moments of deceit. We use subtle cues expressed in the faces and voices of others to regulate our own emotions and behaviors. Despite the universality in recognizing basic emotions across cultures (Ekman, 1972; Ekman et al., 1969; Izard, 1971), cross-cultural differences in emotion perception exist. For example, American participants tend to rate emotional expressions more intensely compared to Japanese participants (Matsumoto & Ekman, 1989; Matsumoto, 1990). This occurs in part because culture serves as a roadmap and shapes the way we perceive, interact with, and respond to emotional information in our social world (Chiu et al., 2013; Kashima, 2001).

In addition to culture, language plays a role in emotion perception. According to the constructionist Conceptual Act Theory, pre-existing conceptual knowledge (e.g., mental representation of the emotion "fear") is accessed through language to provide meaning to exteroceptive sensations in our environment (e.g., the sound of an approaching rattlesnake) and our physiological states (e.g., increased heart rate and palpitation; Barrett et al., 2007; Lindquist & Barrett, 2008; Lindquist & Gendron, 2013; Lindquist et al., 2015). Differences in emotionality have been observed in bilinguals (see Pavlenko et al., 2012 for a review), with emotional phrases (e.g., "I love you"; Dewaele 2008), taboo words (Dewaele, 2004), and advertising slogans (Puntoni et al., 2009) perceived to carry more emotional weight in a bilingual's first language compared to their second language.

In recent years, two trends around the world have led to greater cultural and linguistic diversity, with implications for how culture, language, and emotional processing interact. These trends include the increase rate of worldwide migration (International Organization for Migration, 2020) and the rapid development of several language learning

[1] Department of Communication Sciences and Disorders, Northwestern University, Evanston, Illinois, United States

[2] Department of Psychology, Swarthmore College, Swarthmore, Pennsylvania, United States

[3] Linguistics and Multilingual Studies, School of Humanities, Nanyang Technological University S639818, Nanyang avenue, Singapore

apps and online language tutoring platforms (Andress et al., 2020; Fox, 2020).

A natural by-product of the rapid growth in migration is that our social circle has expanded to include faces and languages from a wide variety of cultural and linguistic backgrounds. In addition, many individuals report feeling a sense of belonging to two (or more) cultural groups. These individuals, known as bicultural bilinguals (or multicultural multilinguals), adopt new cultural norms after migrating to a new country or while living in a culturally diverse community where multiple cultural traditions and norms exist, such as Singapore or São Paolo (Grosjean, 2015; LaFromboise et al., 1993; Nguyen & Benet-Martínez,2007; 2013). Although multiple definitions of biculturalism exist, in the present paper, we define biculturalism as the internalization and synthesis of two or more cultures (Nguyen & Benet-Martínez, 2007; 2013). According to Grosjean (2015), bicultural bilinguals adapt to the attitudes, behaviors, and values of each culture by assimilating into both cultures and taking aspects of each culture to find a balance between the two. In some cases, depending on their understanding of the cultural environment, bicultural bilinguals engage in culture-specific behaviors and cognitive thought processes (Berry, 1980, 1997). Furthermore, individuals high in bicultural competence are better able to express, identify, and understand their emotions during stressful events compared to those low in bicultural competence (Eng et al., 2005). Despite the growing research demonstrating that cultural background influences multisensory emotion perception and integration (e.g., Liu et al., 2015a, 2015b; Tanaka et al., 2010), our understanding of the cognitive and emotional consequences of being bicultural remains limited. The current study aims to investigate whether bicultural bilinguals demonstrate culture-specific behaviors when perceiving multisensory emotions.

With the increase in language learning apps and tools, technology has revolutionized the ways people learn new languages by making language learning more accessible. Many individuals can now learn a foreign language that does not necessarily align with their own culture. Because language and culture are intertwined (Allwright & Bailey, 1991; Byram, 1989), learning a foreign language has the potential to provide individuals with a new social and cultural frame of reference, leading to more complex and nuanced cultural representations. To our knowledge, no study to date has examined multisensory emotion perception when a misalignment in culture exists between the auditory modality (i.e., emotional tone of voice) and visual modality (i.e., facial expressions). Due to strong economic ties between the East and West, interacting with individuals who speak a foreign language that is different from the one spoken by their own community is quite common (e.g., Caucasian person speaking Mandarin). The ability to correctly identify the emotions in speakers from other backgrounds is especially important in contexts that may have lasting repercussions, such as international trade, economic transactions, and political agreements. Hence, another aim of the current study will be to examine multisensory emotion perception when the auditory and visual modalities are from different cultures.

A growing body of research has found cross-cultural differences in how individuals from Eastern and Western cultures process emotional faces (Barrett et al., 2011; Caldara, 2017). On emotion recognition tasks, Easterners tend to focus more on the information from the eyes, whereas Westerners tend to focus more on the information from the mouth to interpret the emotions of others (Jack et al., 2012; Yuki et al., 2007). According to Yuki et al. (2007), Easterners rely on the eye regions to interpret the mental state of others because the muscles around the eyes require more effort to control than the muscles around the mouth, and therefore the eyes reveal the person's internal thoughts and feelings more accurately. In contrast, Westerners rely on the mouth region because smiles and frowns are distinct facial features, and overt expressions of emotions are highly encouraged in individualistic societies. In another study, Masuda et al. (2008) observed that Japanese participants inferred the mental state of a person by looking at the facial expressions of the people surrounding the central person, whereas Americans did not. Hence, Easterners perceive the emotions of a person as inseparable from the group. These findings demonstrate that culture guides the fixation patterns, scanning strategies, and processing styles during emotion recognition.

However, the emotions of others are evaluated not only by extracting sensory information from the visual modality, but also by combining sensory information from the visual and auditory modalities. In studies that combine facial and vocal cues, participants are typically asked to judge the emotion in one modality, while ignoring the emotion in the other modality. The emotion in the non-target modality could either be congruent (e.g., cheerful voice) or incongruent (e.g., sad voice) to the emotion in the target modality (e.g., happy face). The difference in accuracy between incongruent and congruent trials has been referred to by Takagi and colleagues (2015) as modality dominance. Participants are typically more accurate at identifying the target emotion when the emotional content from both modalities elicits the same emotion than when it elicits different emotions (Collignon et al., 2008; de Gelder & Vroomen, 2000; Föcker et al., 2011; Vroomen et al., 2001). However, when the face-voice pair evokes different emotions, one sensory modality often interferes with the processing of the other, leading to less accurate judgments of the target emotion. Furthermore,

people are generally more accurate at judging facial expressions than vocal expressions (visual dominance; e.g., Collignon et al., 2008; Hawk et al., 2009; Paulmann & Pell, 2011), illustrating the larger impact of facial cues over vocal cues. These studies reveal that emotional input from the auditory and visual modalities interact automatically, and that we tend to rely more on the information from the visual modality.

Previous research has shown that the perceiver's cultural background impacts how individuals use information from each modality to evaluate multisensory emotions (Liu et al., 2015a, b; Tanaka et al., 2010). Whereas individuals from East-Asian cultures are more impacted by the tone of voice than by the facial expression of a speaker (auditory dominance; Ishii et al., 2003; Liu et al., 2015a; Tanaka et al., 2010), individuals from Western cultures show the opposite pattern and are more impacted by the facial expression than the tone of voice of a speaker (visual dominance; Liu et al., 2015b; Tanaka et al., 2010). Note that an exception to these patterns has been observed, with Chinese participants equally impacted by the auditory and visual modalities (Liu et al., 2015b).

Engelmann & Pogosyan (2013) explained that display rules, which are cultural norms that regulate how emotions should be expressed (Matsumoto, 1990), can alter our attentional biases, mental representations, and cognitive styles through learning and repeated exposure. In a similar vein, the East-West differences in modality dominance have been attributed by Liu and colleagues (2015a,b) to the display rules prescribed by each culture. Easterners in collectivistic societies prioritize the needs of others before themselves and as a result are often told to control or conceal their emotions to maintain group harmony (e.g., Markus & Kitayama 1991; Matsumoto et al., 2008). Therefore, individuals from Eastern cultures tend to direct their attention to the less explicit auditory modality (Sanchez-Burks et al., 2003; Yum, 1988). In contrast, Westerners in individualistic societies prioritize their own needs before others and are encouraged to convey emotions through direct and explicit means (e.g., McCarthy et al., 2006, 2008). The most direct way to achieve this is through facial expressions.

In the studies discussed thus far, modality dominance effects have only been examined in participants who were monocultural and identified with a single culture. It remains unknown whether biculturalism can influence emotion perception in individuals who identify with two (or more) cultures that consist of different social norms and schemas. For individuals who identify with two cultures, do bicultural bilinguals switch mental schemas in response to the cultural context or do they combine both cultures to form a new schema?

According to the Cultural Frame Switching hypothesis (CFS; Hong et al., 1997, 2000), bicultural individuals can access specific mental models for each culture and flexibly switch in response to the social context. Hong and colleagues (2000) proposed a dynamic constructivist approach, in which the internalized culture is a network of discrete domain-specific knowledge structures. An individual can acquire more than one cultural system, even if they contain conflicting or opposite schemas (e.g., collectivist versus individualist cultures). Specific knowledge from each culture is accessed when bicultural individuals are primed with a cultural image, such as a national flag, famous face, or place of interest. For instance, Chinese-American biculturals are more likely to make external attributions (i.e., inferring that a person's behavior is due to situational or external factors) after being exposed to Chinese images (e.g., dragon), and more internal attributions (i.e., inferring that a person's behavior is due to dispositional or personal reasons) after being exposed to American images (e.g., American flag) (Hong et al., 2000), demonstrating that bicultural bilinguals can shift attributions depending on the cultural context. Though visual primes are more commonly used, language has also been shown to elicit culture-specific behaviors. Bilinguals can switch personalities (e.g., Chen & Bond 2010; Ramírez-Esparza et al., 2006, 2008), access different information in response to the same question (Marian & Kaushanskaya, 2007), display different emotional reactions after reading the same story (Panayiotou, 2004), or show different affective patterns (Perunovic et al., 2007), depending on the language they are using.

According to the CFS hypothesis, we predict that bicultural bilinguals will rely more on the visual modality (i.e., Caucasian face) than auditory modality (i.e., English speech) when presented with Western audio-visual emotional information, and more on the auditory modality (i.e., Mandarin speech) than visual modality (i.e., Asian face) when presented with Eastern audio-visual emotional information during multisensory emotion perception. To our knowledge, only two studies to date have examined the CFS hypothesis in emotion processing. Kreitler & Dyson (2016) asked Mexican-American bicultural bilinguals to rate emotions on a scale from 0 (Never Experience) to 6 (Always Experience) based on how much they were currently or generally experiencing each emotion. Those primed with Mexican cultural images reported less negative affect and more positive affect than those primed with American cultural images. The authors attributed the overall positive affect when presented with Mexican cultural images to Hispanic and Latin cultural scripts (Díaz-Loving & Draguns, 1999; Triandis et al., 1984). In another study, Chinese-English bilinguals who migrated to the United States from China performed an emotion recognition task, in which they viewed facial

expressions and heard emotional speech from their old Eastern culture (Asian face and Mandarin speech) and new Western culture (Caucasian face and English speech) (Chen et al., 2022). Chinese-English bilinguals experienced a larger modality dominance in the voice task than in the face task when evaluating emotional information from the West (i.e., visual dominance), but no difference in modality dominance between the face and voice tasks when evaluating emotional information from the East. Though these studies provide some support for the CFS hypothesis in emotional processing, it remains unknown whether similar effects can be observed when perceiving the emotions of others in bilinguals who are bicultural since birth.

Bicultural bilinguals from Singapore are an ideal sample to study how biculturalism impacts emotion perception because the Singaporean identity is a unique combination and hybrid of multiple ethnic backgrounds. Singapore is a multi-ethnic country with four official languages, including English, Malay, Mandarin, and Tamil. The largest ethnic group in Singapore is Chinese (73.9%), followed by Malay (13.8%), and then Indian (9.1%). In Singapore, students are required to learn English and one of the three other official languages because of the Bilingual Policy. Furthermore, Singapore's history of colonization from Western countries and the integration of Western media in society has increased the prevalence of the English language. According to Singapore's Census of Population data in 2020 (Department of Statistics, Ministry of Trade and Industry, Republic of Singapore, 2021), 47.6% of the population reported they spoke English most frequently at home, while 40.2% reported Mandarin. Therefore, Singaporeans are often interacting with individuals who speak both Mandarin and English. This unique attribute affords us the ability to examine multisensory emotion perception in instances when the auditory and visual inputs are from the same culture (e.g., Asian face with Mandarin speech), and in instances when the auditory and visual inputs are from different cultures (e.g., Asian face paired with English speech). The latter will provide a more nuanced examination of audio-visual integration processes in bicultural bilinguals.

When the two modalities of input convey different cultures, familiarity with the audio-visual pairing may play a role in how the emotion from each modality is perceived. In Singapore, interacting with an Asian person speaking English is more common than interacting with a Caucasian person speaking Mandarin. In fact, only 3.2% of the population self-identified as "Other," which consists of Eurasians, Caucasians, Japanese, Filipino, and Vietnamese (Department of Statistics, Ministry of Trade and Industry, Republic of Singapore, 2021). Furthermore, given that the majority of the population self-identified as Asian, Singaporeans may recognize the emotions of those from the same ethnic group (i.e., Asian faces) more easily than the emotions of those from a different ethnic group (i.e., Caucasian faces), which would be consistent with the findings from a meta-analysis on cross-cultural differences in emotion recognition (Elfenbein & Ambady, 2002). As a result, emotions elicited by an Asian person may be evaluated differently than emotions elicited by a Caucasian person, due to variability in the amount and type of previous experience.

In the present study, we examined how bicultural bilinguals living in Singapore integrate audio-visual information during emotion perception. Based on the assumption that multisensory perception of emotion may be subject to Cultural Frame Switching, we predicted that bicultural bilinguals will show different patterns of modality dominance across Eastern and Western cultures. Specifically, participants will have a larger auditory dominance when presented with emotional input from Eastern cultures, and a larger visual dominance when presented with emotional input from Western cultures. Alternatively, there is the possibility that participants will find it easier to evaluate the emotions from members of the same race (i.e., Asian faces) than members of a different race (i.e., Caucasian faces). We also predict there will be greater interference from the irrelevant modality for the audio-visual pairing most familiar to participants (i.e., Asian face with English speech) than the audio-visual pairing that is least familiar to participants (i.e., Caucasian face with Mandarin speech) when the auditory and visual inputs are from different cultures. In other words, the Asian face paired with English speech will produce a larger modality dominance effect than the Caucasian face paired with Mandarin speech.

Methods

Participants

Thirty-seven bicultural bilinguals between the ages of 21 and 31 were recruited through posters around a university campus in Singapore or through word of mouth. Participants had to be born and raised in Singapore, be living in Singapore at the time of testing, and have English and Mandarin as their two most dominant languages to participate in the study. Using G*Power 3.1 (Faul et al., 2009), power analyses were performed to establish the sample size. Based on the effect size of $r = .35$ (Liu et al., 2015b), an $\alpha = 0.05$, and power = 0.85, the minimum number of participants needed was N = 14. To increase power even further and to account for potentially larger variance among remote participants, we more than doubled our sample size to 37 participants. Informed consent was obtained from all participants at the beginning of the study.

Table 1 Linguistic and Cultural Background Information. Standard deviations are in parentheses

	Language	
	English	Mandarin
Age of Acquisition (in years)	1.55 (1.80)	1.97 (2.32)
Proficiency (1 to 10)	9.00 (1.2)	7.13 (1.91)
Daily Usage (%)	65.77 (17.11)	30.84 (16.45)
	Culture	
	Western	East-Asian
Identity Score (0 to 10)	4.00 (1.97)	7.65 (3.46)
Daily Exposure (%)	38.23 (22.68)	60.90 (22.86)

Note. English and Mandarin proficiencies were rated from 1 = very low to 10 = perfect; Western and East-Asian cultural identity scores were rated from 0 = no identification to 10 = complete identification

Three participants did not complete the task and were removed from the analyses. Another three participants were removed after conducting an outlier analysis, as their accuracy rates were 3 standard deviations below the group's mean. The remaining 31 participants (10 males, 21 females; M_{age} = 24.00 years, SD_{age} = 3.33) lived in Singapore for approximately 22.5 years (SD = 2.45). Of these participants, 19 learned English and Mandarin simultaneously, 9 learned English first, and 3 learned Mandarin first. All participants acquired both languages before the age of 7. Language and cultural background information of the participants are summarized in Table 1.

Seventeen of the 31 participants reported knowing three or more languages, but only seven participants rated their proficiency in these languages greater than 3. The list of non-English and non-Mandarin languages included French, German, Hakka, Hokkien, Italian, Japanese, Khmer, Korean, Malay, Spanish, and Swedish. All participants had normal or corrected-to-normal vision, no hearing impairments, and no previous history of neuropsychological disorders.

Materials

Language Experience and Proficiency Questionnaire (LEAP-Q). The LEAP-Q (Marian et al., 2007) was used to obtain background information about each participant's language and cultural history. Participants answered questions about their language use, age of acquisition, and proficiency for all known languages, including any non-native languages. For each language listed, level of proficiency in speaking was rated on a scale from 1 to 10 (1 = very low and 10 = perfect) and the percentage of time spent speaking each language daily was reported on a scale from 0 to 100%, with the percentages across all three languages adding up to 100% (e.g., 50% Mandarin, 30% English, and 20% Spanish = 100%). For cultural background information, participants listed and rated the extent to which they identified with each culture on a scale from 0 to 10 (0 = no identification and 10 = complete identification). They then listed the percentage of time on average exposed to US-American and East-Asian cultures, including interactions with friends, watching TV shows, listening to music, reading, etc.. Additionally, participants answered demographic questions about their age, gender, handedness, years of formal education, and history of hearing or vision problems.

Emotion Recognition Task

The auditory stimuli consisted of 20 Mandarin and 20 English pseudo-sentences from two validated vocal emotion databases (Mandarin: Liu & Pell 2012; English: Pell et al., 2009) spoken by four different native speakers of each language (2 females and 2 males) in five different emotions (happiness, sadness, disgust, fear, and anger). The English and Mandarin pseudo-sentences were matched on recognition rate (Mandarin: M = 86%, SD = 7.3%; English: M = 88%, SD = 7.4%), emotional intensity (Mandarin: M = 3.3 out of 5, SD = 0.6; English: M = 3.4 out of 5, SD = 0.4), and duration (Mandarin: M = 1.78 s, SD = 0.26; English: M = 1.79 s, SD = 0.19), ts < 1.

The visual stimuli were 20 Asian faces and 20 Caucasian faces from the Taiwanese Facial Expression Image Database (Chen & Yen, 2007) and the Karolinska Directed Emotional Faces Database (Lundqvist et al., 1998), respectively. Four actors (2 females and 2 males) portraying five different emotions (happiness, sadness, disgust, fear, and anger) were selected from each database. The Asian and Caucasian faces were matched on recognition rate (Asian: M = 83%, SD = 12.1%; Caucasian: M = 84%, SD = 12.6%) and emotional intensity (Asian: M = 5.6 out of 9, SD = 0.6; Caucasian: M = 5.7 out of 9, SD = 1.0), ts < 1. To control for brightness and contrast between datasets, all images were formatted to the same dimension (345 pixels wide x 430 pixels high) and resolution (300 dpi), and converted to grayscale using GIMP 2 (GIMP Development Team, 2018).

For each culture, bimodal stimuli were created by presenting the auditory and visual stimuli simultaneously. The pairing of the auditory and visual stimuli varied along two

Table 2 Match and Mismatch Conditions by Culture, Task, Speaker, and Gender

	Match Condition	
Culture	Face Stimuli	Voice Stimuli
Eastern Face/Eastern Voice	Asian Face 1 (Male)	Mandarin Speech 1 (Male)
	Asian Face 2 (Male)	Mandarin Speech 2 (Male)
	Asian Face 3 (Female)	Mandarin Speech 3 (Female)
	Asian Face 4 (Female)	Mandarin Speech 4 (Female)
Western Face/Western Voice	Caucasian Face 1 (Male)	English Speech 1 (Male)
	Caucasian Face 2 (Male)	English Speech 2 (Male)
	Caucasian Face 3 (Female)	English Speech 3 (Female)
	Caucasian Face 4 (Female)	English Speech 4 (Female)
	Mismatch Condition	
	Face Stimuli	Voice Stimuli
Eastern Face/Western Voice	Asian Face 1 (Male)	English Speech 1 (Male)
	Asian Face 2 (Male)	English Speech 2 (Male)
	Asian Face 3 (Female)	English Speech 3 (Female)
	Asian Face 4 (Female)	English Speech 4 (Female)
Western Face/Eastern Voice	Caucasian Face 1 (Male)	Mandarin Speech 1 (Male)
	Caucasian Face 2 (Male)	Mandarin Speech 2 (Male)
	Caucasian Face 3 (Female)	Mandarin Speech 3 (Female)
	Caucasian Face 4 (Female)	Mandarin Speech 4 (Female)

factors: emotional congruency (congruent, incongruent) and match (match, mismatch). For emotional congruency, the auditory and visual stimuli either portrayed the same emotion (e.g., happy face and happy voice; bimodal congruent trial) or different emotions (e.g., happy face and sad voice; bimodal incongruent trial). Each facial expression was paired once with the voice of the same emotion and once with the four other emotions for a total of 20 bimodal congruent trials and 80 bimodal incongruent trials. For match, the auditory and visual stimuli could either be from the same culture (e.g., Caucasian face and English speech or Asian face and Mandarin speech; bimodal match condition) or different cultures (e.g., Caucasian face and Mandarin speech or Asian face and English speech; bimodal mismatch condition). For the analyses, we determined which culture the mismatch condition belonged to by the culture of the target modality. For example, if the participant was asked to judge the emotion from the Asian face (face task) or Mandarin speech (voice task), the trial would be labeled as East. If the participant was asked to judge the emotion from the Caucasian face (face task) or English speech (voice task), the trial would be labeled as West. The stimuli used in the match condition from the East and West were swapped between auditory and visual modalities for the mismatch condition. In other words, the bimodal stimuli for the mismatch condition were created by pairing the Asian face stimuli with the English speech stimuli from the match condition and by pairing the Caucasian face stimuli with the Mandarin speech stimuli from the match condition. Refer to Table 2 for a breakdown of the match and mismatch conditions by culture, task, speaker, and gender. For both the match and mismatch conditions, each unique voice was always paired with the same unique face of the same gender to maintain consistency between face and voice identity. An illustration of the bimodal stimuli by emotional congruency and cultural match can be found in Fig. 1.

Experimental design

The emotion recognition task was programmed in HTML, JavaScript, and CSS, and conducted online[1]. For each culture, two unimodal lists were created containing 20 faces (unimodal visual) and 20 voices (unimodal auditory). A total of 80 unimodal trials were presented, half of which were from the Eastern culture (4 Asian faces x 5 emotions = 20 Eastern face trials; 4 Mandarin pseudo sentences x 5 emotions = 20 Eastern voice trials) and the other half were from the Western culture (4 Caucasian faces x 5 emotions = 20 Western face trials; 4 English pseudo sentences x 5 emotions = 20 Western voice trials).

In addition, due to the ratio of incongruent to congruent trials (4:1), four different bimodal lists were created in

[1] A feature of using JavaScript for online data collection is that RT data can be collected and recorded locally. Therefore, only the response device (i.e., computer mouse) affects absolute RTs. Connection speed only impacts the speed at which the data collected are sent to the server, MySQL. Several validation studies have noted that web-based tasks programmed using a combination of HTML, CSS, and JavaScript have very good reliability, with standard deviations less than 10 ms for RTs and stimulus presentation durations (Reimers & Stevens, 2014) and RTs comparable to those obtained in lab-based experiments (de Leeuw & Motz, 2015).

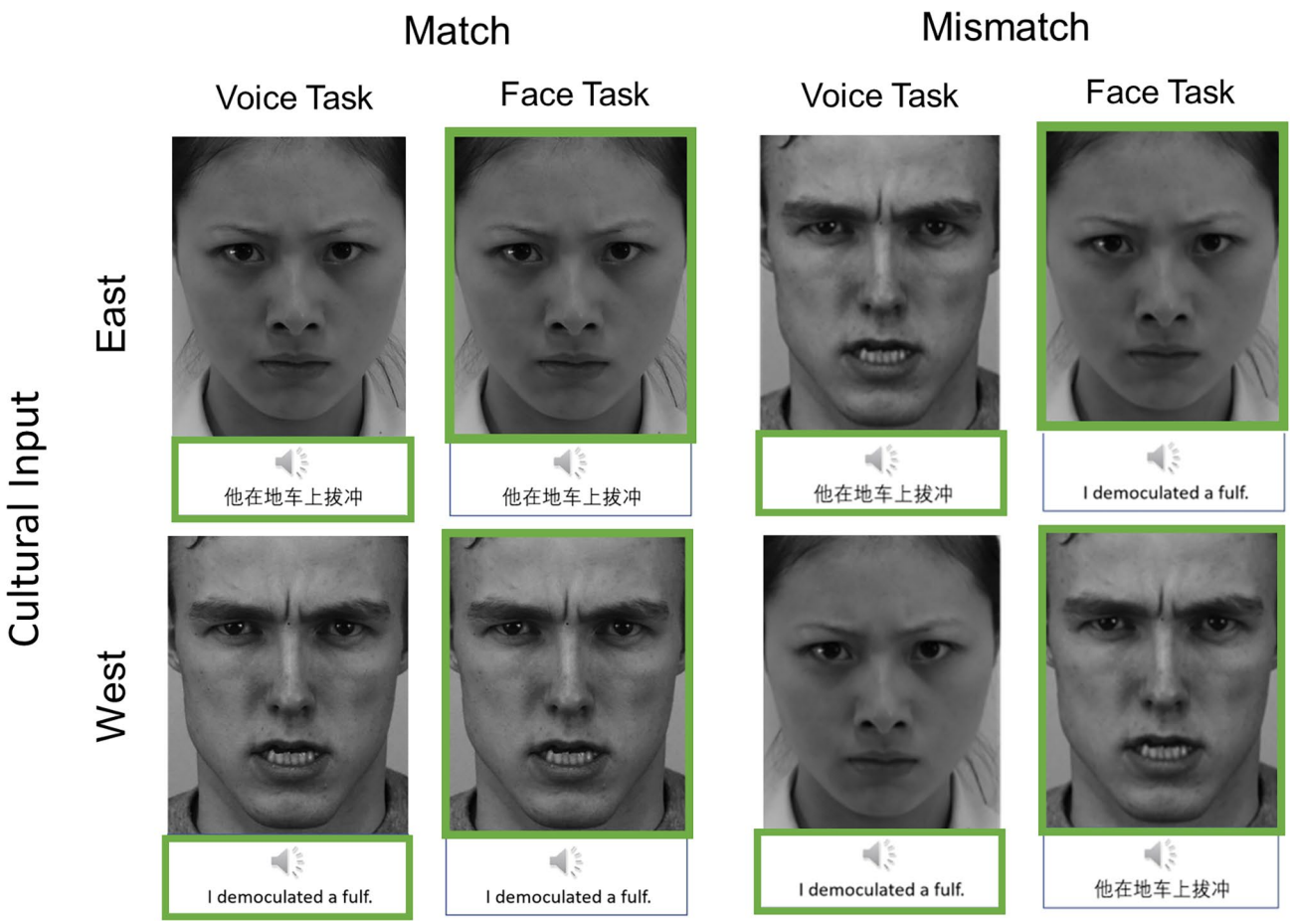

Fig. 1 Examples of bimodal stimuli by task (voice vs. face), culture (East vs. West), and match (match vs. mismatch). The green box denotes the target modality that participants were instructed to respond to in each condition. For both match and mismatch conditions, trials that required participants to judge the emotion from the Asian face (face task) or Mandarin speech (voice task) were labeled as East, whereas trials that required participants to judge the emotion from the Caucasian face (face task) or English speech (voice task) were labeled as West. (Asian face—Image ID ang118: Adapted with permission from Chen & Yen (2007); Mandarin pseudo-sentence: Adapted with permission from Liu & Pell (2012); Caucasian face—Image ID AM08ANS: Adapted with permission from Lundqvist et al., (1998); English pseudo-sentence: Adapted with permission from Pell et al., (2009))

each culture containing 20 congruent trials and 20 of the 80 incongruent trials. Thus, the same face (or voice) appeared once in the congruent condition and once in the incongruent condition. The emotions were equally distributed across congruent and incongruent conditions. Each participant was assigned to one of the four bimodal lists and both unimodal lists from each culture (so that they received emotional stimuli from both cultures). The same bimodal list was presented once in the face task and once in the voice task. A total of 320 bimodal trials were presented to participants, half of which were from the match condition and half from the mismatch condition. Within each condition, there were 20 congruent Eastern face trials, 20 incongruent Eastern face trials, 20 congruent Eastern voice trials, 20 incongruent Eastern voice trials, 20 congruent Western face trials, 20 incongruent Western face trials, 20 congruent Western voice trials, and 20 incongruent Western voice trials, for a total of 160 trials. In addition to the unimodal and bimodal trials, twelve bimodal filler trials with new faces and voices were inserted into each task to minimize the likelihood that participants would develop a response strategy (e.g., closing their eyes or muting the sound).

Within each task, the bimodal, unimodal, and filler trials were randomly presented. The order of receiving the face or the voice task first was counterbalanced across participants. In addition, the mismatch condition was intermixed with the match condition. There were three breaks embedded within each task.

Procedure

At the start of the experiment, participants were asked to complete the task in a quiet environment using web browser

Chrome[2] and to adjust the sound to their level of comfort. On each trial, a prompt appeared instructing participants in English to either "Judge the Voice Emotion" for the voice task, in which participants had to identify the emotion from the actor's tone of voice, or "Judge the Face Emotion" for the face task, in which participants had to identify the actor's facial expression. After clicking on the prompt, the stimuli appeared in the center of the screen. In the bimodal trials, the face and voice appeared simultaneously, and the face remained on the screen for the duration of the speech. In the unimodal face trials, the face appeared anywhere between 1500 and 2000 ms in 100 ms interval (i.e., 1500, 1600, 1700, 1800, 1900, and 2000 ms), consistent with the duration of 95% of the pseudo-sentences. In the unimodal voice trials, a fixation cross appeared and remained on the screen for the duration of the speech. Upon stimulus presentation, participants were instructed to select one emotion from five emotions listed in English (happiness, sadness, disgust, fear, and anger) as quickly and accurately as possible, and then rate the perceived intensity of the emotion on a scale from 0 (not intense at all) to 6 (extremely intense). The emotion word choices were shown in the same order across participants and trials. On half of the filler trials, a red dot (20 mm in size) on the cheek of the face or a beep in the speech stream was presented for 500 ms within the last 600 ms or 700 ms of a trial, respectively. Participants clicked "Yes" or "No" to report whether they saw a flashing red dot on the face or heard a beep in the speech stream.

At the beginning of each task, participants were given eight practice trials, including two filler trials. The experiment took approximately 60 to 90 min to complete. Participants were then debriefed about the purpose of the study and compensated with a gift card of their choice in the amount of 10 Singapore dollars per hour for their time.

Statistical analyses

Mean accuracy rates, response times (RTs), and intensity ratings were obtained for each condition and participant (Table 3). RTs were measured from the onset of a trial until a response was made. Only correct trials were included in the analyses for RTs and intensity ratings. In addition, RTs less than 200 ms and above 5000 ms were removed as outliers (< 200 ms: 0% of the data; > 5000 ms: 3.42% of the data). A cut-off of 5000 ms was used to remove long response times based on the study by Rigoulot & Pell (2014). This timeframe accounts for the additional time required to discriminate between five emotions rather than only two emotions. For accuracy rates and intensity ratings, modality dominance was calculated by subtracting the incongruent from the congruent trials. For RTs, modality dominance was calculated by subtracting the congruent from the incongruent trials. A larger modality dominance in the face task reflects greater interference from the voice, whereas a larger modality dominance in the voice task reflects greater interference from the face. Each dependent variable was subjected to a three-way repeated-measures ANOVA of task (voice and face), culture (East and West), and match (match and mismatch)[3]. The ANOVAs were reported with uncorrected degrees of freedom, and Greenhouse-Geiser correction was applied in instances where the assumption of sphericity was violated. Pairwise comparisons were corrected for multiple comparisons using the Bonferroni method.

Results

Accuracy rates

A three-way ANOVA on accuracy rates revealed a main effect of task, $F(1,30) = 5.44$, $p = .027$, $\eta_p^2 = 0.15$. The modality dominance in accuracy rates was larger in the voice task ($M = 0.10$, $SE = 0.017$) than the face task ($M = 0.061$, $SE = 0.010$), 95% CI [0.005, 0.080], suggesting that participants were overall more impacted by facial cues than vocal cues. There was also a main effect of match, $F(1,30) = 4.26$, $p = .048$, $\eta_p^2 = 0.12$, such that the match condition ($M = 0.092$, $SE = 0.013$) produced a larger modality dominance than the mismatch condition ($M = 0.072$, $SE = 0.011$), 95% CI [0.00, 0.040]. The interaction between culture and task was marginally significant, $F(1,30) = 3.97$, $p = .056$, $\eta_p^2 = 0.12$. Follow-up analyses by culture (East vs. West) revealed that the modality dominance was larger in the voice task ($M = 0.12$, $SE = 0.019$) than face task ($M = 0.044$, $SE = 0.013$) when the auditory and visual inputs were from the West, $p = .022$,

[2] In comparison to Safari, Firefox and Edge, the Chrome browser yields the smallest measurement error for stimulus presentation times and response times across operating systems (Anwyl-Irvine et al., 2021; Henninger et al., 2021).

[3] As suggested by a Reviewer, we examined the effect of emotion type on modality dominance by performing repeated measures ANOVAs on accuracy rates, response times, and intensity ratings with culture (East, West), task (face, voice), and emotion type (anger, sadness, disgust, fear, and happiness) as within-subject factors. For all three dependent variables, the task by emotion interaction was significant, $ps < 0.019$, such that a visual dominance was observed for happiness, $ps < 0.003$. The other emotions did not differ in modality dominance between the face task and voice task. For accuracy rates, a culture by emotion interaction emerged, $F(4,120) = 3.46$, $p = .015$, $\eta_p^2 = 0.10$. A larger modality dominance for disgust was found in the East-Asian culture than the Western culture, $p = .047$. Based on these findings, the effect of emotion type on modality dominance appears to be minimal. These findings should be interpreted with caution considering the low number of trials per emotion when broken down by congruency, culture, and task (5 trials each).

Table 3 Mean Accuracy Rates (ACC), Response Times (RT), and Intensity Ratings (IR) by Match, Culture, and Task for each Condition. (Standard Deviations are in Parentheses.)

Match	Measure	Culture	Task	Unimodal	Congruent	Incongruent	Modality Dominance
Match	ACC	East	Face	0.87 (0.094)	0.90 (0.086)	0.81 (0.12)	0.090 (0.10)
			Voice	0.79 (0.13)	0.87 (0.076)	0.76 (0.11)	0.11 (0.14)
		West	Face	0.87 (0.071)	0.89 (0.077)	0.83 (0.11)	0.052 (0.10)
			Voice	0.75 (0.15)	0.83 (0.10)	0.72 (0.15)	0.12 (0.13)
	RT	East	Face	1392 (244)	1349 (244)	1452 (292)	104 (265)
			Voice	1539 (312)	1467 (356)	1614 (398)	147 (252)
		West	Face	1443 (275)	1394 (265)	1455 (287)	61 (238)
			Voice	1608 (383)	1470 (314)	1717 (395)	247 (267)
	IR	East	Face	3.45 (0.66)	3.63 (0.74)	3.46 (0.71)	0.17 (0.32)
			Voice	3.66 (0.75)	3.85 (0.64)	3.50 (0.70)	0.35 (0.48)
		West	Face	3.47 (0.63)	3.58 (0.65)	3.39 (0.63)	0.19 (0.30)
			Voice	3.23 (0.77)	3.51 (0.70)	3.13 (0.64)	0.38 (0.43)
Mismatch	ACC	East	Face	-	0.90 (0.084)	0.83 (0.088)	0.066 (0.10)
			Voice	-	0.83 (0.088)	0.76 (0.10)	0.07 (0.12)
		West	Face	-	0.90 (0.072)	0.86 (0.10)	0.036 (0.093)
			Voice	-	0.83 (0.11)	0.72 (0.15)	0.12 (0.12)
	RT	East	Face	-	1358 (240)	1426 (305)	68 (257)
			Voice	-	1467 (346)	1628 (429)	162 (330)
		West	Face	-	1373 (253)	1450 (297)	77 (228)
			Voice	-	1546 (377)	1687 (429)	141 (316)
	IR	East	Face	-	3.66 (0.72)	3.35 (0.72)	0.31 (0.35)
			Voice	-	3.75 (0.65)	3.60 (0.68)	0.16 (0.38)
		West	Face	-	3.61 (0.61)	3.45 (0.58)	0.16 (0.31)
			Voice	-	3.43 (0.71)	3.07 (0.67)	0.36 (0.42)

95% CI [0.029, 0.12], but no differences between tasks emerged when the auditory and visual inputs were from the East, $p = .64$, 95% CI [-0.041, 0.065] (Fig. 2). All other effects and interactions were not significant, $ps > 0.31$.

Intensity ratings

The analyses on intensity ratings yielded a significant effect of task, $F(1,30) = 4.72$, $p = .038$, $\eta_p^2 = 0.14$, such that there was a larger modality dominance in the voice task ($M = 0.37$, $SE = 0.072$) than the face task ($M = 0.18$, $SE = 0.039$), 95% CI [0.006, 0.20]. Moreover, the culture by task, $F(1,30) = 9.20$, $p = .005$, $\eta_p^2 = 0.24$, match by task, $F(1,30) = 6.79$, $p = .014$, $\eta_p^2 = 0.19$, and culture, match, and task, $F(1,30) = 6.09$, $p = .020$, $\eta_p^2 = 0.17$, interactions were all significant. To understand the three-way interaction, separate two-way ANOVAs were performed by match (match, mismatch). When the auditory and visual inputs were both from the same culture (match), a main effect of task emerged, $F(1,30) = 11.91$, $p = .002$, $\eta_p^2 = 0.28$, such that the modality dominance was larger in the voice task ($M = 0.37$, $SE = 0.072$) than the face task ($M = 0.18$, $SE = 0.039$), 95% CI [0.075, 0.29]. In other words, when both emotional inputs were from the same culture, bicultural bilinguals were more distracted by facial expressions. The main effect of culture and the culture by task interaction were not significant, $Fs < 1$.

When the auditory and visual inputs were from different cultures (mismatch), there was a significant culture by task interaction, $F(1,30) = 14.60$, $p = .001$, $\eta_p^2 = 0.33$. Specifically, when rating the emotional intensity of audio-visual inputs from the West, the modality dominance in intensity ratings was larger for Asian faces ($M = 0.36$, $SE = 0.075$) than Mandarin speech ($M = 0.16$, $SE = 0.056$), $p = .017$, 95% CI [0.040, 0.37], whereas the opposite pattern emerged when rating the emotional intensity of audio-visual inputs from the East. In this case, the modality dominance in intensity ratings was larger for English speech ($M = 0.31$, $SE = 0.062$) than Caucasian faces ($M = 0.16$, $SE = 0.069$), $p = .040$, 95% CI [0.008, 0.30]. In both cases, the irrelevant modality was more difficult to ignore when the Asian face was paired with English speech than the Caucasian face paired with Mandarin speech. The main effects of task and culture were not significant, $Fs < 1$ (Fig. 3).

Response Times (RTs)

For RTs, only a main effect of task emerged, $F(1,30) = 4.17$, $p = .050$, $\eta_p^2 = 0.12$. Specifically, the modality dominance was larger in the voice task ($M = 174.20$, $SE = 35.37$) than

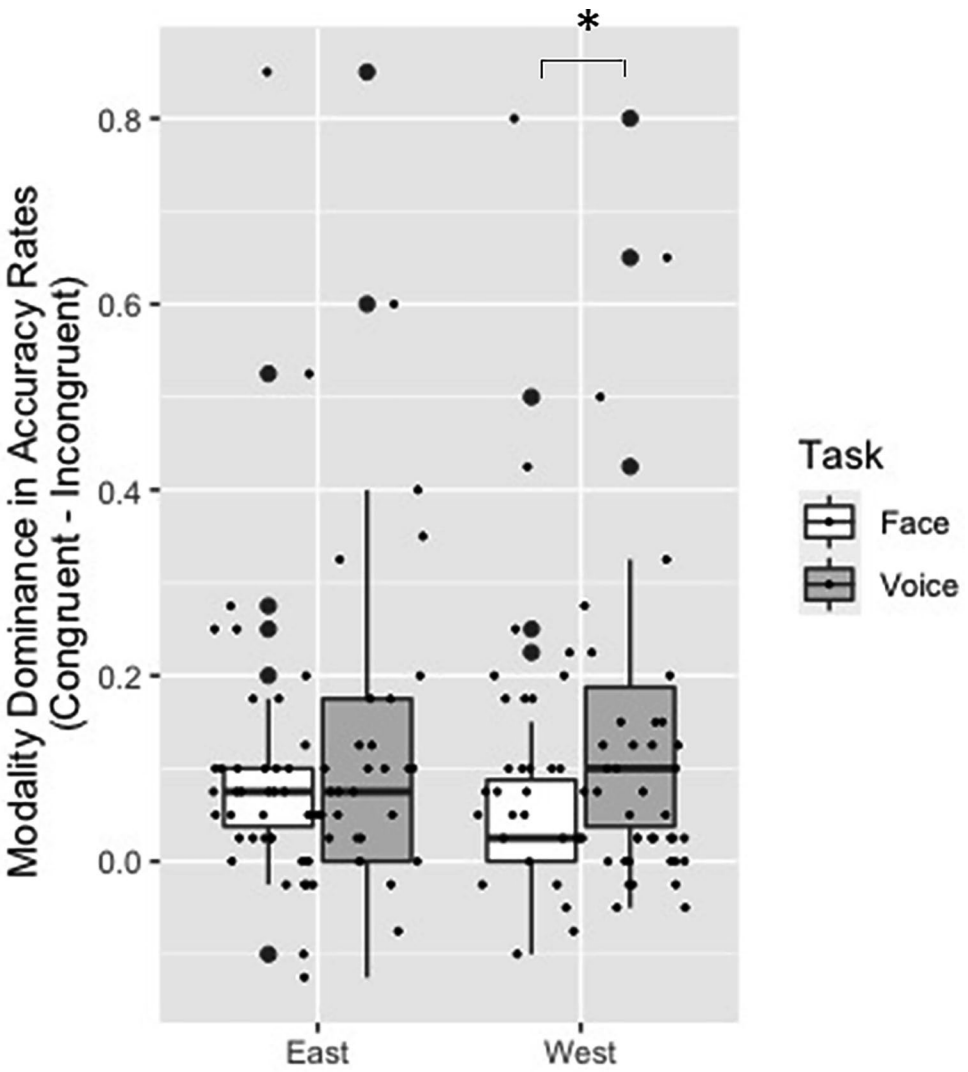

Fig. 2 The effect of culture (East vs. West) and task (face vs. voice) on modality dominance in accuracy rates. Modality dominance was calculated by subtracting the raw accuracy score of the incongruent condition from the congruent condition. A larger modality dominance in the face task reflects greater interference from the voice, whereas a larger modality dominance in the voice task reflects greater interference from the face. The voice task (grey) produced a larger modality dominance than the face task (white) for audio-visual emotional information from the West (i.e., visual dominance). No difference in modality dominance between tasks emerged for emotional audio-visual information from the East. *$p < .05$

the face task ($M = 77.34$, $SE = 28.47$), 95% CI [-0.055, 193.77], again demonstrating that facial cues are more distracting than vocal cues. All other effects and interactions were not significant, $ps > 0.12$.

Correlations between daily exposure to each culture, exposure to each Language, and modality dominance

To examine the Cultural Frame Switching hypothesis more closely, correlations between daily exposure to each culture (East or West) as well as daily exposure to each language (Mandarin or English) were performed separately on the size of the modality dominance (auditory or visual) in the match condition. There was a small positive correlation between daily exposure to East-Asian cultures and auditory dominance to Mandarin speech on accuracy rates, $r(31) = 0.37$, $p = .043$ (Fig. 4a), suggesting that when daily exposure to East-Asian cultures increased, the reliance on the auditory modality increased. The correlations between daily exposure to Mandarin and auditory dominance to Mandarin speech were not significant for RTs, accuracy rates, and intensity ratings, $ps > 0.35$. There was also a marginally significant positive correlation between daily exposure to Western cultures and visual dominance to Caucasian faces on accuracy rates, $r(31) = 0.35$, $p = .053$ (Fig. 4b). As daily exposure to Western cultures increased, the reliance on the

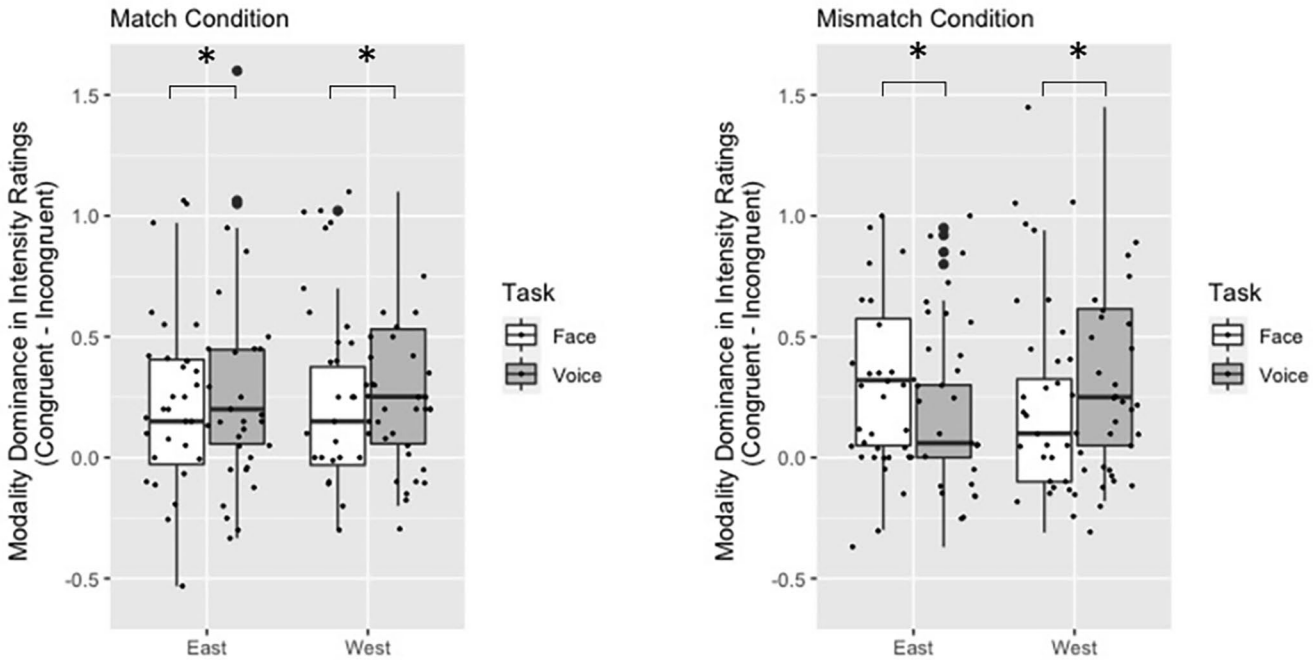

Fig. 3 The interaction of culture (East vs. West), task (face vs. voice), and match (match vs. mismatch) on modality dominance in intensity ratings. Modality dominance was calculated by subtracting the intensity ratings of the incongruent condition from the congruent condition. A larger modality dominance in the face task reflects greater interference from the voice, while a larger modality dominance in the voice task reflects greater interference from the face. When both the auditory and visual inputs were from the same culture, the voice task produced a larger modality dominance than the face task. In the cultural mismatch conditions, where the auditory and visual inputs were from different cultures, the modality dominance changed depending on the target's culture. When the target's culture was Eastern, the face task (i.e., Asian face) produced a larger modality dominance than the voice task (i.e., Mandarin speech). In contrast, when the target's culture was Western, the voice task (i.e., English speech) produced a larger modality dominance than the face task (i.e., Caucasian face). *$p < .05$

visual modality increased. The same correlations performed on intensity ratings and RTs were not significant, $ps > 0.23$. The correlations between daily exposure to English and visual dominance to Caucasian faces were not significant for all dependent measures, $ps > 0.17$.

Discussion

The current study examined how bicultural bilinguals integrate and perceive audio-visual emotions. Across three dependent measures (intensity ratings, response times, and accuracy rates), bicultural bilinguals were more sensitive to the visual modality than to the auditory modality when both emotional inputs were from the West, consistent with the findings previously reported among native English speakers from North America (Liu et al., 2015b). In contrast, there was no clear preference for either the auditory modality or visual modality when both emotional inputs were from the East, consistent with the findings by Chen et al. (2022) and Liu et al. (2015b). However, a significant correlation between daily exposure to East-Asian cultures and auditory dominance on accuracy rates emerged, providing some evidence in favor of the Cultural Frame Switching hypothesis.

We also examined modality dominance when the auditory and visual inputs were from different cultures, a likely scenario in Singapore and many other countries in Asia and elsewhere (e.g., Asian person speaking English and Caucasian person speaking Mandarin). When the cultures were different across modalities, bicultural bilinguals had greater difficulty ignoring the irrelevant modality when the audio-visual pairing consisted of an Asian face with English speech than a Caucasian face with Mandarin speech on intensity ratings, suggesting that familiarity and prior experiences shape affective perceptions.

When the audio and visual inputs were from the East, bicultural bilinguals did not show greater sensitivity towards the auditory modality. The lack of a significant difference between the face and voice task when presented with emotional stimuli from the East is consistent with the behavioral findings reported by Liu and colleagues (2015b), who also found no differences between the face and voice task on an emotional Stroop task in Mandarin speakers from China. There is the possibility that the English task instructions in the current study served as a language prime eliciting Western culture-specific behaviors when evaluating emotional information from the East, thereby increasing the influence of the visual modality. However, the level of dominance to

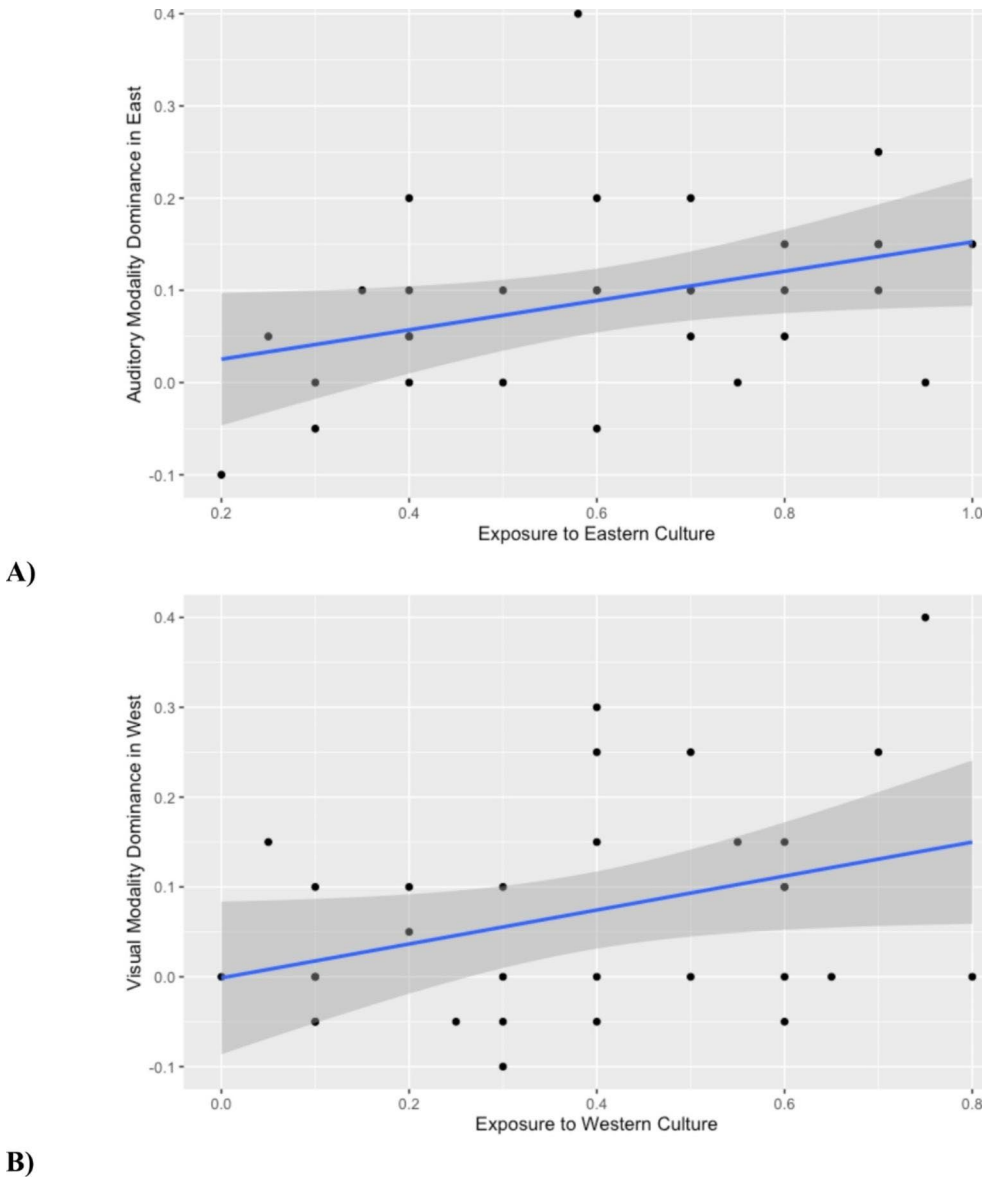

Fig. 4 Scatter plots of the correlations between auditory dominance and exposure to East-Asian cultures when both the auditory and visual inputs were from the East (A), and correlations between visual dominance and exposure to Western cultures when both the auditory and visual inputs were from the West (B). The shaded gray band around the regression line (in blue) is the 95% confidence interval

the visual modality was comparable across cultures, suggesting that bicultural bilinguals in the current study were equally distracted by the Asian and Caucasian faces.

Other explanations for the lack of an auditory dominance effect when perceiving emotional stimuli from the East could be due to the participant sample and the sociolinguistic context of Singapore. For example, after Singapore's independence in 1965, English was implemented as the language of instruction, public administration, commerce, and law because it was perceived to be the language of modernity, connecting Singapore to the rest of the world (Ng & Cavallaro, 2021). Compared to Mandarin, Singaporeans rated English higher in prestige and power (Xu et al., 1998) as well as importance (Leong, 2014). The perceived higher status of English compared to Mandarin may have led participants to implicitly adopt the behaviors from the Western culture more readily than the behaviors from the Eastern culture.

Moreover, Singapore is considered a regional center where East meets West, and Singaporeans routinely interact with individuals from various nationalities. This mix may have converged the social norms and cultural values of bicultural bilinguals. Therefore, it is possible that language and race did not act as strong cultural primes, leading to less consistent mapping between the voice and face. In other words, the boundary between Eastern and Western cultures

is not as well defined among bicultural bilinguals from Singapore, considering the country is a melting pot of cultures, races, ethnicities, and languages. Because the two cultures can be thought of as intersecting rather than separate cultures, bicultural bilinguals from Singapore may be open to the possibility that a person of a certain race can belong to, represent, or practice two different cultures. For instance, it is entirely possible that a bicultural bilingual in Singapore practices East-Asian social norms while speaking English. In line with this possibility, an interesting observation about our sample is that their language exposure did not align with their cultural exposure. While their language exposure was higher for English (66%) than Mandarin (31%), their cultural exposure was higher for Eastern (61%) than Western (38%) cultures. This misalignment between cultural and language exposure indicates that cultural knowledge is not as tightly associated to a specific language among the bicultural bilinguals in our sample and presents a rare opportunity to dissociate and compare the effects of culture vis-à-vis those of language. Based on the results from the correlations, cultural exposure appears to have a stronger effect on modality dominance than language exposure.

A potential reason why the Cultural Frame Switching hypothesis (Hong et al., 1997, 2000) was only partially observed could be that it is harder to switch between cultural mindsets when assessing the emotions of *others*. In previous studies where frame switching has been observed, the tasks usually involved higher-order cognitive processes, such as describing one's personality, affect, or identity (e.g., Cheng et al., 2006; Kreitler & Dyson, 2016; Luna et al., 2008; Ramírez-Esparza et al., 2006; West et al., 2018). These tasks were related to one's thinking style and involved reflecting upon oneself rather than upon others. A study on multisensory integration by Serino and colleagues (2008) reported that individuals who perceived their own face being touched had heightened tactile experiences compared to those viewing other people's faces being touched. Thus, the ability to recognize our own emotions may be embedded more deeply into our cultural knowledge because the information is highly relevant to us. Alternatively, it has been found that evaluating other people's emotions involves additional processing steps relative to evaluating one's own emotions, including the judgments about the identity of the speaker, one's relationship to the speaker, and the intention behind the emotion (Hess & Fischer, 2013).

As previously noted, the participants in the current study reported being more familiar with the English language and East-Asian culture. It is therefore not surprising that in the mismatch condition, participants experienced greater interference from the irrelevant modality when presented with an Asian face and English speech than a Caucasian face and Mandarin speech. Our findings suggest that perceptual or attentional processes related to audio-visual binding are not the only contributing factors to modality dominance, but that top-down processes, like familiarity, also contribute to modality dominance effects. Singaporeans encounter Asian people speaking English across a wide variety of contexts (e.g., different facial expressions, emotional tones, and settings). Mental representations of familiar faces are formed by taking an average across encounters (e.g., Burton et al., 2005), resulting in a rich and stable mental representation that is well-integrated across modalities. Repeated exposure has also been shown to selectively enhance the features of a face. Carr et al. (2017) found that people tend to perceive familiar faces as happier than unfamiliar faces. Hence, the less frequent exposure to Caucasian faces in Singapore may have led to less integration of the Caucasian face and Mandarin speech, enabling participants to focus on each modality separately. Although Western media has infiltrated many parts of the world and portrayed Caucasian people speaking foreign languages, for example through audio dubbing, the level of engagement may be low because emotional processing does not occur in person. Future studies should examine audio-visual integration in a group of bicultural bilinguals who have had equal levels of exposure and experience with both mismatch conditions.

Future directions and conclusions

Grosjean (2015) noted that there is very little research that highlights the experience of "bicultural bilinguals," even though this is the reality for many individuals. Although our study attempts to shed light on how bicultural bilinguals perceive multisensory emotions, it is restricted to only bicultural bilinguals living in Singapore. Other types of bicultural bilinguals, such as those who migrated from one country to another (e.g., Eastern immigrants living in the West or Western immigrants living in the East) or those who grew up with parents from two or more cultures, may show different patterns of results. Bicultural children with parents from a culture that is different from that of their community may be more likely to keep their cultural identities separate by trying to fit in with the mainstream culture. In such cases, they would need to actively shift cultural mindsets when they are at home and when they are interacting with members of the community. To increase the generalizability of the findings to other cultures and bicultural groups, future research will need to test multisensory emotion perception in participants from diverse bicultural backgrounds.

In conclusion, the current study examined multisensory emotion perception in bicultural bilinguals in Singapore. We found that bicultural bilinguals were more distracted by the face than by the voice of a speaker (i.e., visual dominance)

when presented with multisensory emotions from the West, which is consistent with the patterns of modality dominance observed in Western cultures. As daily exposure to East-Asian cultures increased, bicultural bilinguals relied more on the voice of a speaker (i.e., auditory dominance). These findings demonstrate that bicultural bilinguals exhibit culture-specific behaviors when perceiving multisensory emotions. We also found that prior experience and familiarity influence how bicultural bilinguals evaluate the emotions of individuals who speak a foreign language. Bilinguals face the complicated task of negotiating two cultural worlds, and their experience provides a unique lens for examining the complex relationship between culture and emotional processing (Marian, 2023).

Acknowledgements The authors thank Dr. Marc Pell for providing the vocal emotion stimuli. We also thank Editor-in-Chief Michael Richter and three anonymous reviewers for their insightful suggestions and comments. Research reported in this publication was supported in part by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R01HD059858 to Viorica Marian. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Allwright, D., & Bailey, K. M. (1991). *Focus on the language learner*. Cambridge: Cambridge University Press

Andress, M., Givant Star, M., & Balshem, D. (2020, April 15). Language learning apps are seeing a surge in interest during the COVID-19 pandemic. *Forbes*.https://www.forbes.com/sites/mergermarket/2020/04/15/language-learning-apps-are-seeing-a-surge-in-interest-during-the-covid-19-pandemic/?sh=b39bd6c48f4c

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavioral Research*, *53*, 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Science*, *11*, 327–332. https://doi.org/10.1016/j.tics.2007.06.003

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science*, *20*(5), 286–290. https://doi.org/10.1177/0963721411422522

Berry, J. W. (1980). Acculturation as varieties of adaptation. In A. M. Padilla (Ed.), *Acculturation: Theory, models, and some new findings* (pp. 9–25). Boulder, CO: Westview

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied Psychology: An International Review*, *46*(1), 5–68. https://doi.org/10.1111/j.1464-0597.1997.tb01087.x

Byram, M. (1989). *Cultural studies in foreign language education*. Clevedon: Multilingual Matters Ltd.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: The power of averages. *Cognitive Psychology*, *51*(3), 256–284. https://doi.org/10.1016/j.cogpsych.2005.06.003

Caldara, R. (2017). Culture reveals a flexible system for face processing. *Current Directions in Psychological Science*, *26*(3), 249–255. https://doi.org/10.1177/0963721417710036

Carr, E. W., Brady, T. F., & Winkielman, P. (2017). Are you smiling, or have I seen you before? Familiarity makes faces look happier. *Psychological Science*, *28*(8), 1087–1102. https://doi.org/10.1177/0956797617702003

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528. https://doi.org/10.1177/0146167210385360

Chen, P., Chung-Fat-Yim, A., & Marian, V. (2022). Cultural experience influences multisensory emotion perception in bilinguals. *Languages*, *7*(1), 12. https://doi.org/10.3390/languages7010012

Chen, L. F., & Yen, Y. S. (2007). *Taiwanese Facial Expression Image Database.* Taipei: National Yang-Ming University. Retrieved from http://bml.ym.edu.tw/~download/html/

Cheng, C., Lee, F., & Benet, V. (2006). Assimilation and contrast effects in cultural frame switching: Bicultural identity integration and valence of cultural cues. *Journal of Cross-Cultural Psychology*, *37*(6), 742–760. https://doi.org/10.1177/0022022106292081

Chiu, C. Y., Ng, S. S. L., & Au, E. W. M. (2013). Culture and social cognition. In D. Carlston (Ed.), *The Oxford Handbook of social cognition* (pp. 767–785). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199730018.013.0037

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. https://doi.org/10.1016/j.brainres.2008.04.023

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311. https://doi.org/10.1080/026999300378824

de Leeuw, J. R., & Motz, B. A. (2015). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behavioral Research Methods*, *48*, 1–12. https://doi.org/10.3758/s13428-015-0567-2

Department of Statistics, Ministry of Trade and Industry, Republic of Singapore (2021). *Singapore Census of Population 2020, Statistical Release 1: Demographic Characteristics, Education, Language and Religion.* Retrieved from https://www.singstat.gov.sg/-/media/files/publications/cop2020/sr1/cop2020sr1.pdf

Dewaele, J. M. (2004). The emotional force of swearwords and taboo words in the speech of multilinguals. *Journal of Multilingual and Multicultural Development*, *25*, 204–222. https://doi.org/10.1080/01434630408666529

Dewaele, J. M. (2008). The emotional weight of I love you in multilinguals' languages. *Journal of Pragmatics*, *40*, 1753–1780. https://doi.org/10.1016/j.pragma.2008.03.002

Díaz-Loving, R., & Draguns, J. G. (1999). Culture, meaning, and personality in Mexico and in the United States. In Y. T. Lee, C. R. McCauley, & J. G. Draguns (Eds.), *Personality and person perception across cultures* (pp. 103–126). Lawrence Erlbaum Associates Publishers

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotions. In J. Cole (Ed.), *Nebraska symposium on motivation* (pp. 207–282). Lincoln, NB: University of Nebraska Press

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86–88. https://doi.org/10.1126/science.164.3875.86

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. https://doi.org/10.1037//0033-2909.128.2.203

Eng, T. C., Kuiken, D., Temme, K., & Sharma, R. (2005). Navigating the emotional complexities of two cultures: Bicultural competence, feeling expression, and feeling change in dreams. *Journal of Cultural and Evolutionary Psychology*, *3*(3–4), 267–285. https://doi.org/10.1556/jcep.3.2005.3-4.4

Engelmann, J. B., & Pogosyan, M. (2013). Emotion perception across cultures: The role of cognitive mechanisms. *Frontiers in Psychology*, *4*, 118. https://doi.org/10.3389/fpsyg.2013.00118

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica*, *137*(1), 36–47. https://doi.org/10.1016/j.actpsy.2011.02.004

Fox, A. (2020, December 18). 30 million people attempted to learn a new language in 2020, according to Duolinguo – and this was the most popular. *Travel and Leisure*.https://www.travelandleisure.com/travel-tips/mobile-apps/duolingo-most-popular-languages

GIMP Development Team (2018). *GIMP*. Retrieved from https://www.gimp.org

Grosjean, F. (2015). Bicultural bilinguals. *International Journal of Bilingualism*, *19*(5), 572–586. https://doi.org/10.1177/1367006914526297

Hawk, S. T., van Kleef, G. A., Fischer, A. H., & van der Schalk, J. (2009). "Worth a thousand words": Absolute and relative decoding of nonlinguistic affect vocalizations. *Emotion*, *9*(3), 293–305. https://doi.org/10.1037/a0015178

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2021). lab.js: A free, open, online study builder. *Behavioral Research Methods*. https://doi.org/10.3758/s13428-019-01283-5

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review*, *17*(2), 142–157. https://doi.org/10.1177/1088868312472607

Hong, Y. Y., Chiu, C. Y., & Kung, T. M. (1997). Bringing culture out in front: Effects of cultural meaning system activation on social cognition. In K. Leung, Y. Kashima, U. Kim, & S. Yamaguchi (Eds.), *Progress in Asian social psychology* (1 vol., pp. 135–146). Singapore: Wiley

Hong, Y. Y., Morris, M. W., Chiu, C. Y., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, *55*, 709–720. https://doi.org/10.1037/0003-066X.55.7.709

International Organization for Migration (2020). World Migration Report 2020. Retrieved from https://www.un.org/sites/un2.un.org/files/wmr_2020.pdf

Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46. https://doi.org/10.1111/1467-9280.01416

Izard, C. E. (1971). *The face of emotion*. Appleton-Century-Crofts

Jack, R. E., Garrod, O. G., Yu, B., Caldara, H., R., & Schyns, R. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences, 109*(19), 7241–7244. https://doi.org/10.1073/pnas.1200155109

Kashima, Y. (2001). Culture and Social Cognition: Toward a Social Psychology of Cultural Dynamics. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp. 325–360). Oxford University Press

Kreitler, C. M., & Dyson, K. S. (2016). Cultural frame switching and emotion among Mexican Americans. *Journal of Latinos and Education*, *15*(2), 91–96. https://doi.org/10.1080/15348431.2015.1066251

LaFromboise, T., Coleman, H. L. K., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin*, *114*(3), 395–412. https://doi.org/10.1037/0033-2909.114.3.395

Leong, N. C. (2014). A study of attitudes towards the Speak Mandarin Campaign in Singapore. *Intercultural Communication Studies*, *23*(3), 54–65

Lindquist, K. A., & Barrett, L. F. (2008). Constructing emotion: The experience of fear as a conceptual act. *Psychological Science*, *19*(9), 898–903. https://doi.org/10.1111/j.1467-9280.2008.02174.x

Lindquist, K. A., & Gendron, M. (2013). What's in a word: Language constructs emotion perception. *Emotion Review*, *5*, 66–71. https://doi.org/10.1177/1754073912451351

Lindquist, K. A., MacCormack, J. K., & Shablack, H. (2015). The role of language in emotion: Predictions from psychological constructionism. *Frontiers in Psychology*, *6*, 444. https://doi.org/10.3389/fpsyg.2015.00444

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods*, 1042–1051. https://doi.org/10.3758/s13428-012-0203-3

Liu, P., Rigoulot, S., & Pell, M. D. (2015a). Cultural differences in on-line sensitivity to emotional voices: Comparing East and West. *Frontiers in Human Neuroscience*, *9*, 311. https://doi.org/10.3389/fnhum.2015.00311

Liu, P., Rigoulot, S., & Pell, M. D. (2015b). Culture modulates the brain response to human expressions of emotion: Electrophysiological evidence. *Neuropsychologia*, *67*, 1–13. https://doi.org/10.1016/j.neuropsychologia.2014.11.034

Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One individual, two identities: Frame switching among biculturals. *Journal of Consumer Research*, *35*(2), 279–293. https://doi.org/10.1086/586914

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska Directed Emotional Faces—KDEF (CD ROM)*. Stockholm: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech Language and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marian, V., & Kaushanskaya, M. (2007). Language context guides memory content. *Psychonomic Bulletin and Review*, *14*, 925–933. https://doi.org/10.3758/BF03194123

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology*, *94*(3), 365–381. https://doi.org/10.1037/0022-3514.94.3.365

Matsumoto, D. (1990). Cultural similarities and differences in display rules. *Motivation and Emotion*, *14*, 195–214. https://doi.org/10.1007/BF00995569

Matsumoto, D., & Ekman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, *13*(2), 143–157. https://doi.org/10.1007/BF00992959

Matsumoto, D., Yoo, S. H., Fontaine, J., Anguas-Wong, A. M., Arriola, M., Ataca, B. … Grossi, E. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology*, *39*(1), 55–74. https://doi.org/10.1177/0022022107311854

McCarthy, Lee, K., Itakura, S., & Muir, D. W. (2006). Cultural display rules drive eye gaze during thinking. *Journal of Cross-Cultural Psychology*, *37*(6), 717–722. https://doi.org/10.1177/0022022106292079

McCarthy, Lee, K., Itakura, S., & Muir, D. W. (2008). Gaze display when thinking depends on culture and context. *Journal of Cross-Cultural Psychology*, *39*(6), 716–729. https://doi.org/10.1177/0022022108323807

Ng, B. C., & Cavallaro, F. (2021). The Case of Mandarin Chinese in Singapore. In J. Ritu (Ed.), *Multilingual Singapore: Language Policies and Linguistic Realities* (pp. 159–178). New York: Routledge

Nguyen, A. M. D., & Benet-Martínez, V. (2007). Biculturalism unpacked: Components, measurement, individual differences, and outcomes. *Social and Personality Psychology Compass*, *1*(1), 101–114. https://doi.org/10.1111/j.1751-9004.2007.00029.x

Nguyen, A., & Benet-Martínez, V. (2013). Biculturalism and adjustment: A meta-analysis. *Journal of Cross-Cultural Psychology*, *44*(1), 122–159. https://doi.org/10.1177/0022022111435097

Panayiotou, A. (2004). Switching codes, switching code: Bilinguals' emotional responses in English and Greek. *Journal of Multilingual and Multicultural Development*, *25*(2–3), 124–139. https://doi.org/10.1080/01434630408666525

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. (2012). Affective processing in bilingual speakers: Disembodied cognition? *International Journal of Psychology*, *47*, 405–428. https://doi.org/10.1080/00207594.2012.743665

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417–435. https://doi.org/10.1016/j.wocn.2009.07.005

Perunovic, W. Q. E., Heller, D., & Rafaeli, E. (2007). Within-person changes in the structure of emotion: The role of cultural identification and language. *Psychological Science*, *18*(7), 607–613. http://www.jstor.org/stable/40064742

Puntoni, S., De Langhe, B., & Van Osselaer, S. (2009). Bilingualism and the emotional intensity of advertising language. *Journal of Consumer Research*, *35*, 1012–1025. https://doi.org/10.1086/595022

Ramírez-Esparza, N., Gosling, S. D., Benet-Martínez, V., Potter, J. P., & Pennebaker, J. W. (2006). Do bilinguals have two personalities? A special case of cultural frame switching. *Journal of Research in Personality*, *40*, 99–120. https://doi.org/10.1016/j.jrp.2004.09.001

Ramírez-Esparza, N., Gosling, S. D., & Pennebaker, J. W. (2008). Paradox Lost: Unraveling the puzzle of Simpatía. *Journal of Cross-Cultural Psychology*, *39*(6), 703–715. https://doi.org/10.1177/0022022108323786

Reimers, S., & Stevens, N. (2014). Presentation and response timing accuracy in Adobe Flash and HTML5/JavaScript web experiments. *Behavior Research Methods*, *47*, 309–327. https://doi.org/10.3758/s13428-014-0471-1

Rigoulot, S., & Pell, M. D. (2014). Emotion in the voice influences the way we scan emotional faces. *Speech Communication*, *65*, 36–49. https://doi.org/10.1016/j.specom.2014.05.006

Sanchez-Burks, J., Lee, F., Choi, I., Nisbett, R., Zhao, S., & Koo, J. (2003). Conversing across cultures: East-West communication styles in work and nonwork contexts. *Journal of Personality and Social Psychology*, *85*(2), 263–372. https://doi.org/10.1037/0022-3514.85.2.363

Serino, A., Pizzoferrato, F., & Làdavas, E. (2008). Viewing a face (especially one's own face) being touched enhances tactile perception on the face. *Psychological Science*, *19*(5), 434–438. https://doi.org/10.1111/j.1467-9280.2008.02105.x

Takagi, S., Hiramatsu, S., Tabei, K. I., & Tanaka, A. (2015). Multisensory perception of six basic emotions is modulated by attentional instruction and unattended modality. *Frontiers in Integrative Neuroscience*, *9*, 1. https://doi.org/10.3389/fnint.2015.00001

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

Triandis, H. C., Marín, G., Lisansky, J., & Betancourt, H. (1984). Simpatía as a cultural script of Hispanics. *Journal of Personality and Social Psychology*, *47*(6), 1363–1375. https://doi.org/10.1037/0022-3514.47.6.1363

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive Affective and Behavioral Neuroscience*, *1*(4), 382–387. https://doi.org/10.3758/CABN.1.4.382

West, A. L., Zhang, R., Yampolsky, M., & Sasaki, J. (2018). The potential cost of cultural fit: Frame switching undermines perceptions of authenticity in Western contexts. *Frontiers in Psychology*, *9*, 2622. https://doi.org/10.3389/fpsyg.2018.02622

Yum, J. O. (1988). The impact of Confucianism on interpersonal relationships and communication patterns in east Asia. *Communication Monographs*, *55*(4), 374–388. https://doi.org/10.1080/03637758809376178

Yuki, M., Maddux, W. W., & Masuda, T. (2007). Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. *Journal of Experimental Social Psychology*, *43*(2), 303–311. https://doi.org/10.1016/j.jesp.2006.02.004

Marian, V. (2023). The Power of Language: How the Codes We Use to Think, Speak, and Live Transform our Minds. New York: Dutton. ISBN: 9780593187074.

Xu, D., Chew, C. H., & Chen, S. (1998). Language use and language attitudes in the Singapore Chinese community. In S. Gopinathan, A. Pakir, H. W. Kam, and V. Saravanan (Eds.), Language, Society and Education in Singapore (pp. 133–155). Singapore: Times Academic Press.