Title	Take your word or tone for it? European American and Chinese children's attention to emotional cues in speech
Author(s)	Yang Yang, Li Wang, and Qi Wang

Running head: CULTURE AND EMOTION RECOGNITION IN SPEECH

Take Your Word or Tone for It? Attention to Emotional Cues in Speech

Yang Yang[1], Li Wang[2], and Qi Wang[3]

[1] National Institute of Education, Nanyang Technological University

[2] Peking University

[3] Cornell University

This research was supported by C.V. Starr Fellowship from Cornell East Asia Program to Yang Yang. We thank members of the Social Cognition Development Lab at Cornell University for their assistance. Special thanks go to the participants who made the study possible.

Correspondence may be addressed to Yang Yang, National Institute of Education, Nanyang Technological University, 1 Nanyang walk, Singapore; telephone +65 62196252; email yang.yang@nie.edu.sg.

Take Your Word or Tone for It? European American and Chinese Children's Attention to Emotional Cues in Speech

Abstract

Cultural experiences can influence how people attend to different emotional cues. Whereas semantic content explicitly describes feelings, vocal tone conveys implicit information regarding emotions. This cross-cultural study examined children's attention to emotional cues in spoken words. The sample consisted of 121 European American and 120 Chinese children (4-9 years old). Each child played two computer games in which they listened to spoken words and judged the pleasantness of either the word meaning (Word game) or the vocal tone (Tone game) while ignoring the other aspect. Chinese children paid more spontaneous attention to vocal tones and less to word meanings than did European American children. These findings shed critical light on the role of culture in shaping affective cognitive processes during development.

Key words: Emotion recognition, culture, vocal tone, emotional cues in speech, emotional development

Take Your Word or Tone for It? European American and Chinese Children's Attention to Emotional Cues in Speech

Emotion understanding is critical for children's psychosocial development, affecting their social competence, psychological adjustment, and school achievement (Izard et al., 2001; Trentacosta & Fine, 2010). Research on children's emotion understanding has focused on their abilities to recognize facial expressions and understand emotional situations (Yang & Wang, 2019). Importantly, much emotional information is conveyed through words and tones in everyday contexts. Yet little is known about how children attend to different emotional cues in speech and what factors influence the affective cognitive processes in the course of development. Extensive research has further suggested the important role of culture in shaping emotional development (Wang, 2006; Wang & Yang, 2019; Yang & Wang, 2019). Against this backdrop, the present study examined European American and Chinese Children's attention to emotional cues in spoken words.

When processing emotional utterances, there are at least two common emotional cues beyond facial and body expressions. Whereas semantic content explicitly describes feelings, vocal tone conveys implicit information regarding emotions. Culturally predominant communication styles may influence how people attend to these emotional cues. In Western cultures, verbal content plays a primary role in conveying information during communication, whereas in East Asian cultures much information is conveyed through contextual cues (Ambady, Koo, Lee, & Rosenthal, 1996; Markus & Kitayama, 1991). Hall (1976) proposed that in low-context cultures, such as European American culture, individuals' thoughts need to be explicitly expressed in words to facilitate understanding. Accordingly, verbal content contains the major proportion of information in communication and a smaller proportion of information is carried in

contextual cues. In contrast, in high-context cultures, such as Chinese culture, individuals' ideas can be understood without explicit expression in words as long as enough contextual information is given. Contextual cues thus play a larger role in communication, whereas explicit verbal content conveys a relatively smaller proportion of information.

Research with adults have revealed the influence of cultural communication styles on the processing of emotional cues in speech. Kitayama, Ishii, and colleagues conducted a series of cross-cultural studies on adults' attention to word meanings and vocal tones in utterances using Stroop tasks (Ishii, Reyes, & Kitayama, 2003; Kitayama & Ishii, 2002). In these studies, participants listened to emotional words with either a congruent (e.g., a pleasant word with a pleasant tone) or an incongruent vocal tone (e.g., a pleasant word with an unpleasant tone). They then judged the pleasantness of the utterance either by the word meaning while ignoring the vocal tone, or by the vocal tone while ignoring the word meaning. The researchers found that Japanese participants were more distracted by the vocal tone than the verbal content, whereas Americans were more distracted by the verbal content than the vocal tone. Similar differences are found between Chinese and North Americans (Liu, Rigoulot, & Pell, 2015b, 2015a). Taken together, it appears that East Asians pay more attention to the context (vocal tone) than the verbal content of an emotional utterance, whereas Westerners are more sensitive to emotional cues from the word meaning than vocal tones.

Whether children from different cultures show differential sensitivities to emotional cues in speech remains an empirical question. Research with Western children has shown that children gradually develop the ability to recognize emotions from auditory cues. Infants are able to respond to variations in speech prosody (Fernald, 1993; Mastropieri & Turkewitz, 1999). By age 4 or 5 years, children are able to accurately label vocal tones while judging speech with

obscured semantic language (foreign language) or without any semantic information (low pass filtered speech) (Morton & Trehub, 2001). However, when vocal tones are in conflict with the verbal content or when they are presented with contextual information (e.g. drawings of scenes), toddlers and young preschoolers are unable to consistently use pitch cues to infer emotions (Aguert, Laval, Lacroix, Gil, & Bigot, 2013; Morton & Trehub, 2001; Quam & Swingley, 2012). Children's ability to judge emotions based on prosodic tones continues to develop well into adolescence, when they achieve the adult-level of performance (Grosbras, Ross, & Belin, 2018; Morton & Trehub, 2001; Sauter, Panattoni, & Happé, 2013).

Importantly, this developmental pattern may not be the same in other cultural contexts that embed values and practices different from Western cultures. Indeed, research on family socialization suggests that culture-specific ways of attention to emotional cues may occur early in development. In line with the characteristics of their low-context culture (Hall, 1976), European American parents frequently help their children to verbally articulate emotions to fulfill their needs and wishes (Chao, 1995). In contrast, in line with the characteristics of their high-context culture (Hall, 1976), Chinese parents rarely discuss or explain emotions with their children (Doan & Wang, 2010). Accordingly, Chinese and Chinese immigrant preschoolers exhibit a lower level of explicit emotion knowledge than their European American peers, such as the ability to explain the situations that elicit specific emotions (Doan & Wang, 2010; Wang, 2008). On the other hand, Chinese parents often use subtle ways to help children attend to and infer others' feelings without being told (Chao, 1995; Wang, 2006, 2013). Given the different socialization practices, Chinese children may come to be more sensitive to implicit cues of emotions and better able to understand others' emotions from contextual cues (i.e., vocal tones) than European American children.

Cross-cultural studies on children's attention and perception further shed light on this issue. It has been found that East Asian children pay more attention to contextual cues (Imada, Carlson, & Itakura, 2013; Kuwabara & Smith, 2016; Senzaki, Masuda, Takada, & Okada, 2016) than do their Western peers as a result of socialization. For example, in one study, Japanese and Canadian children aged 4-9 years watched videos of underwater scenes containing a focal fish and background objects. When asked to independently recall what they saw in the videos, children in the two cultures made similar references to the focal fish and the background objects. However, when children discussed and recalled the videos together with their mothers, the older Japanese children referred more to the background and less to the focal fish than did their Canadian peers. Thus, different cultural patterns of attention emerged by age 7 to 9 years, when children were scaffolded by their parents (Senzaki et al., 2016).

To further understand cultural influences on affective cognitive processing in children, we examined in the current study 4- to 9-year-old European American and Chinese children's attention to emotional cues using the Stroop task adapted from Ishii et al. (2003). We expected that Chinese children would pay more attention to the vocal tone of the emotional utterance than European American children, who would focus more on the word meaning. Specifically, in the Stroop task, Chinese children were expected to be more distracted by the vocal tone, whereas European American children would show a greater interference from the word meaning. Furthermore, we expected older children to be more aligned with the culture-specific pattern of attention, given that they were socialized and immersed longer in their respective cultural contexts than younger children.

Method

Participants

The study protocol (project title: The Development of Implicit Emotion Knowledge Among Chinese and European American Children) was approved by the Cornell University's Institutional Review Board for Human Participants in the U.S. (protocol number: IRB # 1511005942) and the Peking University's Committee for Protecting Human and Animal Subjects in China (IRB #2015-12-07). Power analysis suggested a target sample size of 198 children to achieve a power of .8 for detecting medium effect sizes (d = .4). To ensure sufficient power to test interaction effects, we recruited approximately 40 children (half boys and half girls) per age (4-5, 6-7, and 8-9 years) and cultural group, with a total of 121 European American (EA) children from the U.S. and 120 Chinese children from China (See Table S1 in supplementary materials). EA Children were recruited in local schools and a science center in upstate New York. Chinese children were recruited in schools in Beijing.

Procedure

We modified the Stroop task developed by Ishii et al. (2003) to test children's sensitivity to emotional content and tones in spoken words. In a quiet room, children were introduced to the tasks as two "computer games" in which they listened to utterances and responded to them. The utterances were spoken in the participants' native language. Children were told that they would hear many words that differed in both emotional meanings and the emotional tones of voice. Some of the spoken words were congruent utterances (i.e., a pleasant word with a pleasant tone, an unpleasant word with an unpleasant tone), while others were incongruent (i.e., a pleasant word with an unpleasant tone, an unpleasant word with a pleasant tone). In the Word game, children were instructed to judge the meaning of each word as good or bad while ignoring the sound of the utterance (to-be-ignored aspect of the utterance in the Word task). In the Tone game, children were instructed to judge the utterance as either "sounds happy" or "sounds sad or

mad" merely according to the vocal tone while ignoring the word meaning (to-be-ignored aspect of the utterance in the Tone task). To make sure that children understood the concept of tone, the experimenter asked children before the Tone game, "Do you know what tone means?" and gave one example of the child's name with a happy tone and a sad/mad tone to explain the concept of tone. The order of the two games was counterbalanced.

To facilitate children's motor response, we used large JellyBean buttons (2.5-in. diameter) in this task. On each button, there was either a pleasant cartoon face or an unpleasant cartoon face. The locations of the buttons were fixed with the pleasant button on the left and the unpleasant button on the right. Children were instructed to use one hand for each button, where they should press the pleasant-cartoon-face button when they judged an utterance as pleasant and press the unpleasant-cartoon-face button when they judged an utterance as unpleasant. They were asked to play as quickly as they could while also maintaining response accuracy. Response time was measured from the offset of the stimuli in milliseconds.

All children were presented with 44 trials (12 practice trials, followed by 32 testing trials) in each game, and 88 trials in total. In the first four practice trials in each game, the to-be-ignored aspect of the utterances was neutral. For example, at the beginning of the Word game, we presented four utterances with pleasant or unpleasant meaning and neutral tone to facilitate children's understanding of the game. In contrast, the first four practice trials in the Tone game were neutral words with pleasant or unpleasant tone. The second set of 4 practice trials contained utterances with congruent word meanings and vocal tones. The last 4 practice trials presented utterances with incongruent word meanings and vocal tones. The 32 testing trials of each game included 16 congruent and 16 incongruent utterances. The order of the utterances was fixed in the practice trials, but randomized in the testing trials.

Before each utterance, a "+" appeared in the center of the screen on each trial to warn children that the utterance was coming. The utterances were only presented from the headphone rather than the screen. After each utterance, a yellow smiley face (left) and a red frowny face (right) were presented on the screen to remind children to press one of the two buttons. After a button was pressed, children were given both visual and auditory feedback for the practice trials. If the answer was correct, a green check mark, "√" appeared on the screen, along with a correct sound effect presented in the headphone. Otherwise, a red question mark, "?", and a wrong sound effect were presented. During the testing trials, children did not receive any feedback.

Materials

We developed stimulus utterances in four steps. First, we selected 28 pairs of translation-equivalent Chinese and English words from McArthur Communication Development Inventory (MCDI) which contains a list of words commonly produced by toddlers. The selected words include nouns, adjectives, and verbs with different meanings, consisting 14 pairs of words with pleasant meaning and 14 pairs with unpleasant meaning. The 28 Chinese and 28 English words were rated by 21 Chinese native speakers and 10 English native speakers respectively on the pleasantness of word meaning (*1= very unpleasant, 7 = very pleasant*) (see Table S4 in supplementary materials for the ratings of the words in the final set).

Secondly, one female balanced Mandarin-English bilingual research assistant, who grew up acquiring Mandarin and English simultaneously, read all the English and Chinese words in two different tones of voice: a pleasant tone in which a smooth and round tone was used and an unpleasant tone in which a harsh and constricted tone was used. It yielded 112 utterances in total (28 words × 2 languages × 2 tones). T-tests showed that, compared with pleasant tones, unpleasant tones had lower levels of mean pitch, as would be expected, and lasted shorter (The

length and mean pitch of the utterance stimuli for each language and each tone valence are presented in Table S5 in supplementary materials).

Thirdly, the 112 utterances from the second step were listened to and judged by 11 bilingual adults on the pleasantness of the vocal tones (*1= very unpleasant, 7 = very pleasant*). Finally, we selected utterances for the experiment based on ratings on the pleasantness of both vocal tones (from the third step) and word meanings (from the first step). The final set of stimuli included 64 utterances in testing trials (8 utterances × 2 language × 2 word valence × 2 tone valence) and 16 utterances (2 utterances × 2 language × 2 word valence × 2 tone valence) in the practice trials. In addition, the bilingual research assistant also recorded neutral utterances for each task, including 4 utterances with pleasant or unpleasant meaning but a neutral tone for the Word game, and 4 utterances with pleasant or unpleasant tones but neutral meanings for the Tone game. In the final set of utterances used in testing trials, pleasant and unpleasant vocal tones were manipulated independent of language and word meaning, and vocal tone and word meaning were equally extreme in the two languages (see Table S4 in supplementary materials). Research materials can be accessed at https://osf.io/a5erq/?view_only=1c2a1dbad56f4fba878a6dc5c10e4556.

Results

Preliminary analyses

We first calculated the mean accuracy for each utterance. Two words had lower than 0.5 mean accuracy in the Word task. One word was “Hungry” in English, and the other was “坏了” (broken) in Chinese. Results with and without these two words showed similar patterns. We reported the results without “Hungry” in English and “坏了” (broken) in Chinese in the Word task. The mean accuracies for the rest of utterances in both Word and Tone tasks were .84

(range: .60 - .97, *SD* = 0.10), and .89 (range: .68 - .98, *SD* = 0.06) for English and Chinese, respectively. The length of the utterance was negatively correlated with response time within each cultural group (EA: $r = -.16$, $p < .001$; Chinese: $r = -.19$, $p < .001$); it was therefore included in relevant analyses as a covariate.

Accuracy

The percentage of accurate responses in each task for each participant was calculated. In general, children performed well in congruent testing trials (EA: *M* = .90, range: .42 – 1; *SD* = .11; Chinese: *M* = .92, range: .42-1; *SD* = .11; see Table S6, S7, and S9 in supplementary materials). The accuracies in congruent trials were positively correlated with age, $r = .48$, $p < .001$. Analyses with and without participants who performed below .90 in congruent testing trials yielded identical patterns of results. Results with a different cutoff, .80 accuracy in congruent testing trials, remained the same. We reported the results with the entire sample below (See Table S8 and S10 in supplementary materials for results with only participants with accuracies above .90 or .80 in congruent trials respectively).

We first used R and *lme4* to conduct a Generalized Linear Mixed Model (GLMM) analysis on the trial-level accuracy scores, with task (Word vs. Tone), culture (Chinese vs. European American), gender, age, word valence (pleasant vs. unpleasant), tone valence (pleasant vs. unpleasant), length of the utterance, and mean pitch of the utterance as fixed effects, and subject as a random effect. As expected, there was a significant interference effect on accuracy, as indicated by a significant Word valence × Tone valence interaction, $Z = 14.87$, $p < .001$. Judgment accuracies were higher for congruent than incongruent utterances. Further, there were other significant interaction effects, such as Culture × Task × Word × Tone, $Z = 3.04$, $p = .002$, Task × Age × Word × Tone, $Z = 4.21$, $p < .001$, and Culture × Age × Word × Tone, $Z = 2.87$, $p =$

.004. Descriptive data are presented in Table S2 and full results of the trial-level GLMM are presented in Table S12 and S13 in supplementary materials.

To further understand the interference effects in Stroop tasks, we computed an interference index for each child in each task by subtracting the percentage of correct responses for incongruent utterances from the percentage of correct responses for congruent utterances. Positive scores indicated the interference by the to-be-ignored information in the utterances. In other words, the interference score in the Word task indicated the interference by the vocal tone, whereas the interference score in the Tone task indicated the interference by the word meaning. Descriptive data are presented in Table S2. We used R and *lme4* to conduct a Linear Mixed Model (LMM) analysis on the interference score, with task, culture, gender, and age as fixed factors, and subject as a random factor. Age was included as a continuous variable in the models.

The interference effects on accuracy varied by both culture and task, as shown in a significant Culture × Task interaction, $t = -3.21$, $p = .001$ (See Figure 1). Simple effects tests (separate ANOVAs for EA and Chinese children) showed that, for EA children, the interference was greater for Tone task (interference by the word meanings) than for Word task (interference by the vocal tones), $F(1, 120) = 24.54$, $p < .001$, $\eta_p^2 = .17$. However, for Chinese children, there was no significant difference in the interference on accuracies between the two tasks, $F(1, 119) = 1.86$, $p = .175$, $\eta_p^2 = .02$. Further simple effects tests (separate ANOVAs for Word task and Tone task) showed that in the Word task, the effect of culture was not significant, $F(1, 239) = .72$, $p = .399$, $\eta_p^2 = .003$. In the Tone task, EA children were influenced more by the word meaning than were Chinese children, $F(1, 239) = 10.10$, $p = .002$, $\eta_p^2 = .04$. Moreover, the 3-way interaction Task × Culture × Age on the interference score was significant, $t = 2.16$, $p = .031$, whereby cultural differences were only significant for older children (8-9 years) in the Word task

but significant for the two younger groups (4-5 and 5-7 years) in the Tone task (see Figure 2 and supplementary materials). The complete results of the LMM analysis are presented in Table S3 in supplementary materials.

Response time

Similar analyses were performed on the response time for correct responses. We first submitted the response time to a trial-level LMM analysis in R. The interference effect was confirmed by a significant Word valence × Tone valence interaction, $Z = -9.23$, $p < .001$ (see Table S12 and S13 in supplementary materials for additional results from the trial-level LLM analyses). Response time was shorter for congruent utterances than for incongruent utterances. The interference scores of response time were calculated by subtracting the mean response time of congruent utterances from the mean response time of incongruent utterances. Therefore, positive scores still indicated an interference effect of the to-be-ignored information in the utterances. Descriptive data are presented in Table S2. The distribution of the interference scores of response time was slightly right skewed, but the patterns of results with and without log transformation of the interference scores of response time did not differ. We reported the results without log transformation below for easier interpretation (see results with log-transformed response time in Table S11). LMM analyses on the interference score showed that the interactions of Task × Culture, $t = -1.14$, $p = .261$, and Task × Culture × Age, $t = 1.17$, $p = .242$, were not significant. However, there was a significant main effect of culture on the interference, $t = 3.51$, $p < .001$, whereby EA children had lower interference scores of response time than did Chinese children across tasks (see Table S3 for additional results from the LLM analyses).

We further examined cultural differences in the two tasks separately to test our hypotheses. In the Word task, Chinese children experienced greater vocal-tone interference than

did EA children on the response time measure, $F(1, 238) = 16.22$, $p < .001$, $\eta_p^2 = .06$. However, there was no cultural difference in the Tone task, $F(1, 235) = 1.94$, $p = .165$, $\eta_p^2 = .01$ (see Figure 3). Separate analyses for the three age groups showed significant cultural differences for the youngest and oldest groups in the Word task, but no difference for any age group in the Tone task (see Figure 4 and supplementary materials).

Discussion

This study was the first to examine children's attention to different emotional cues in speech across cultures. The results provide insights into cultural influences on affective cognitive processing in the course of development. Consistent with our hypotheses, there were differences in the patterns of attention to word meanings and vocal tones between European American and Chinese children. European American children were more sensitive to word meanings but less sensitive to vocal tones than were Chinese children. The cultural differences in attention to vocal tones, as reflected in processing accuracy, were more pronounced among 8-9 years old children than younger age groups.

Cultural effects on processing of emotional cues

As predicted, we found cultural differences in the interference effects in the two tasks. European American children experienced more interference by word meanings and less by vocal tones than did Chinese children in their processing accuracy or response time. Alternatively, word meanings were more interfering than vocal tones for European American children only, but not for Chinese children. These results are in line with the theoretical notion that people in high-context cultures like China often express and detect emotional cues through contextual information (Ambady et al., 1996; Hall, 1976). In this cultural context, children are socialized to attend to contextual information to understand social situations (e.g., Senzaki et al., 2016). In

contrast, in low-context cultures like the U.S., people tend to explicitly express and communicate emotions to be understood by others (Hall, 1976; Wang, 2003). Accordingly, European American children are encouraged to discuss their thoughts and feelings during family interactions (Doan & Wang, 2010; Wang, 2008). As a result, Chinese children come to be relatively more sensitive to vocal tones, whereas European American children attend more to explicit word meanings. Would linguistic differences play a role in the current findings? Perhaps Chinese children paid more attention to vocal tones than European American children because word meanings in Chinese might more strongly depend on tones than those in English. However, studies with Japanese adults showed that Japanese attended to vocal tones more than word meanings even though Japanese is not a tonal language (Ishii & Kitayama, 2003). Also, Tagalog-English bilingual Philippians paid more attention to vocal tones than word meanings even when they completed the tasks in English (Ishii et al., 2003). These findings suggest that the attention biases reflect largely cultural rather than linguistic influences. Future research among English-speaking Asian children will further clarify this issue.

Age effect, limitations, and future directions

Some interesting age effects emerged. For accuracy in the Word task, Chinese 8-9-year-old children, but not the two younger groups, experienced greater vocal tone-interference than did their European American counterparts. Similarly, for response time, the oldest Chinese children had higher interference scores than did European American children of the same age. The results suggest that with enculturation and socialization, older Chinese children come to focus their attention more on vocal tones and thus become more sensitive to vocal tones during communication than their European American counterparts. This age pattern further reflects the process of cultural learning during which children become increasingly socialized into their

cultural modes of thinking and understanding (Wang & Kushnir, 2019). In the Tone task, however, the two younger European American groups experienced greater word meaning-interference than did their Chinese counterpart. This result may suggest that European American children become more sensitive to word meanings than Chinese children from early ages. Unexpectedly, there was no significant cultural difference in the Tone task in the oldest age group. Perhaps given their greater experience with schooling, word meanings are too salient to ignore for older children regardless of culture. Taken together, the current patterns of age-related findings are inconclusive and require corroboration of additional research, especially longitudinal studies.

Intriguingly, interference for accuracy was generally greater among European American children, whereas interference for response time was greater among Chinese children. Whereas European American children experienced greater word-meaning interference than Chinese on the accuracy measure only, Chinese children experienced greater vocal-tone interference mainly on the response time measure. Conceivably, given Chinese parents' high expectations for children's test performance (Campbell & Mandel, 1990; Chao, 1996), Chinese children might focus more on accuracy relative to speed in the current tasks than did European American children. Moreover, the cultural patterns of results pertaining to accuracy were clearer than those of response time, especially for younger children, which may reflect a variety of age-related factors such as motor skills, general attention, inhibition, and processing speed (Ling, Wong, & Diamond, 2016; Simpson et al., 2012). Therefore, for young children in particular, accuracy might be a more appropriate index to understand the processing of emotional cues across cultures. Additional studies using different methods among different populations are required to corroborate the current findings. Studies using the same stimuli with adults will also be helpful.

In conclusion, this study provides the first demonstration of cultural influences on children's sensitivity to emotional cues in speech. It sheds light on our knowledge about the role of culture in children's emotion recognition and delineates a more complete picture of affective cognitive processes during development.

References

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Bigot, L. Le. (2013). Inferring emotions from speech prosody: Not so easy at age five. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0083657

Ambady, N., Koo, J., Lee, F., & Rosenthal, R. (1996). More than words: Linguistic and nonlinguistic politeness in two cultures. *Journal of Personality and Social Psychology*, *70*(5), 996–1011. https://doi.org/10.1037/0022-3514.70.5.996

Campbell, J. R., & Mandel, F. (1990). Connecting math achievement to parental influences. *Contemporary Educational Psychology*, *15*(1), 64–74. https://doi.org/10.1016/0361-476X(90)90006-M

Chao, R. K. (1995). Chinese and European American cultural models of the self reflected in mothers' childrearing beliefs. *Ethos*, *23*(3), 328–354. https://doi.org/10.1525/eth.1995.23.3.02a00030

Chao, R. K. (1996). Chinese and European American mothers' beliefs about the role of parenting in children's school success. *Journal of Cross-Cultural Psychology*, *27*(4), 403–423. https://doi.org/10.1177/0022022196274002

Doan, S. N., & Wang, Q. (2010). Maternal discussions of mental states and behaviors: relations to emotion situation knowledge in European American and immigrant Chinese children. *Child Development*, *81*(5), 1490–1503. https://doi.org/10.1111/j.1467-8624.2010.01487.x

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*(3), 657–674. https://doi.org/10.1111/j.1467-8624.1993.tb02934.x

Grosbras, M. H., Ross, P. D., & Belin, P. (2018). Categorical emotion recognition from voice

improves during childhood and adolescence. *Nature*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-32868-3

Hall, E. T. (1976). *Beyond culture*. New York: Anchor Books.

Imada, T., Carlson, S. M., & Itakura, S. (2013). East-West cultural differences in context-sensitivity are evident in early childhood. *Developmental Science*, *16*(2), 198–208. https://doi.org/10.1111/desc.12016

Ishii, K., & Kitayama, S. (2003). Selective attention to contextual information in Japan. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 1358–1358).

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., Fine, S. E., Schultz, D., Mostow, A. J., Ackerman, B. P., & Youngstrom, E. (2001). Emotion knowledge as a predictor of social behavior and academic competence in children at risk. *Psychological Science*, *12*(1), 18–23. https://doi.org/10.1111/1467-9280.00304

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion*, *16*(1), 29–59. https://doi.org/10.1080/0269993943000121

Kuwabara, M., & Smith, L. B. (2016). Cultural differences in visual object recognition in 3-year-old children. *Journal of Experimental Child Psychology*, *147*, 22–38. https://doi.org/10.1016/j.jecp.2016.02.006

Ling, D. S., Wong, C. D., & Diamond, A. (2016). Do children need reminders on the day-night task, or simply some way to prevent them from responding too quickly? *Cognitive Development*, *37*, 67–72. https://doi.org/10.1016/j.cogdev.2015.10.003

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*. 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*. https://doi.org/10.1016/j.neuropsychologia.2014.11.034

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

Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology*, *35*(3), 204–214. https://doi.org/10.1002/(SICI)1098-2302(199911)35:3<204::AID-DEV5>3.0.CO;2-V

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834–843. https://doi.org/10.1111/1467-8624.00318

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250. https://doi.org/10.1111/j.1467-8624.2011.01700.x

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97–113. https://doi.org/10.1111/j.2044-835X.2012.02081.x

Senzaki, S., Masuda, T., Takada, A., & Okada, H. (2016). The communication of culturally dominant modes of attention from parents to children: A comparison of Canadian and Japanese parent-child conversations during a joint scene description task. *PloS One*, *11*(1),

e0147199–e0147199. https://doi.org/10.1371/journal.pone.0147199

Simpson, A., Riggs, K. J., Beck, S. R., Gorniak, S. L., Wu, Y., Abbott, D., & Diamond, A. (2012). Refining the understanding of inhibitory processes: how response prepotency is created and overcome. *Developmental Science*, *15*(1), 62–73. https://doi.org/10.1111/j.1467-7687.2011.01105.x

Trentacosta, C. J., & Fine, S. E. (2010). Emotion knowledge, social competence, and behavior problems in childhood and adolescence: A meta-analytic review. *Social Development*, *19*(1), 1–29. https://doi.org/10.1111/j.1467-9507.2009.00543.x

Wang, Q. (2003). Emotion situation knowledge in American and Chinese preschool children and adults. *Cognition and Emotion*, *17*(5), 725–746. https://doi.org/10.1080/02699930302285

Wang, Q. (2006). Developing emotion knowledge in cultural contexts. *International Society for the Study of Behavioral Development Newsletter*, *1*(49), 8–12.

Wang, Q. (2008). Emotion knowledge and autobiographical memory across the preschool years: A cross-cultural longitudinal investigation. *Cognition*, *108*(1), 117–135. https://doi.org/10.1016/j.cognition.2008.02.002

Wang, Q. (2013). Chinese socialization and emotion talk between mothers and children in native and immigrant Chinese families. *Asian American Journal of Psychology*, *4*(3), 185.https://doi.org/10.1037/a0030868

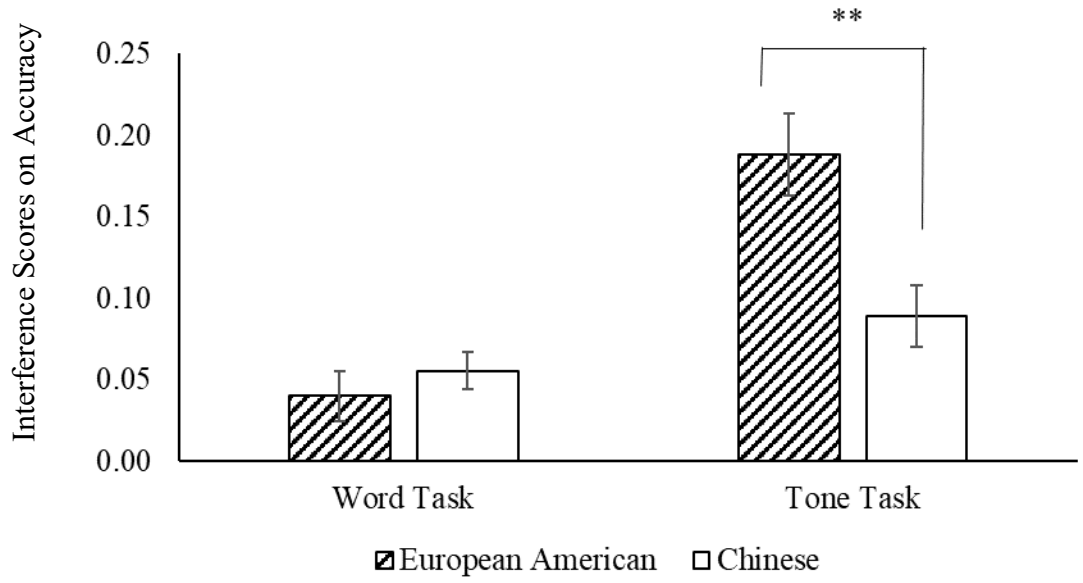
Wang, Q., & Kushnir, T. (2019). Cultural Pathways in Cognitive Development: Introduction to the Special Issue. *Cognitive Development*, 52, 100816. https://doi.org/10.1016/j.cogdev.2019.100816

Wang, Q., & Yang, Y. (2019). Culture and emotional development: introduction to the special issue. *Culture and Brain*, *7*(2), 95–98. https://doi.org/10.1007/s40167-019-00088-9

Yang, Y., & Wang, Q. (2019). Culture in emotional development. In V. LoBue, K. Perez-Edgar, & K. Buss (Eds.), *Handbook of Emotional Development* (pp. 569–593). Springer, Cham. https://doi.org/10.1007/978-3-030-17332-6_22

Figure 1

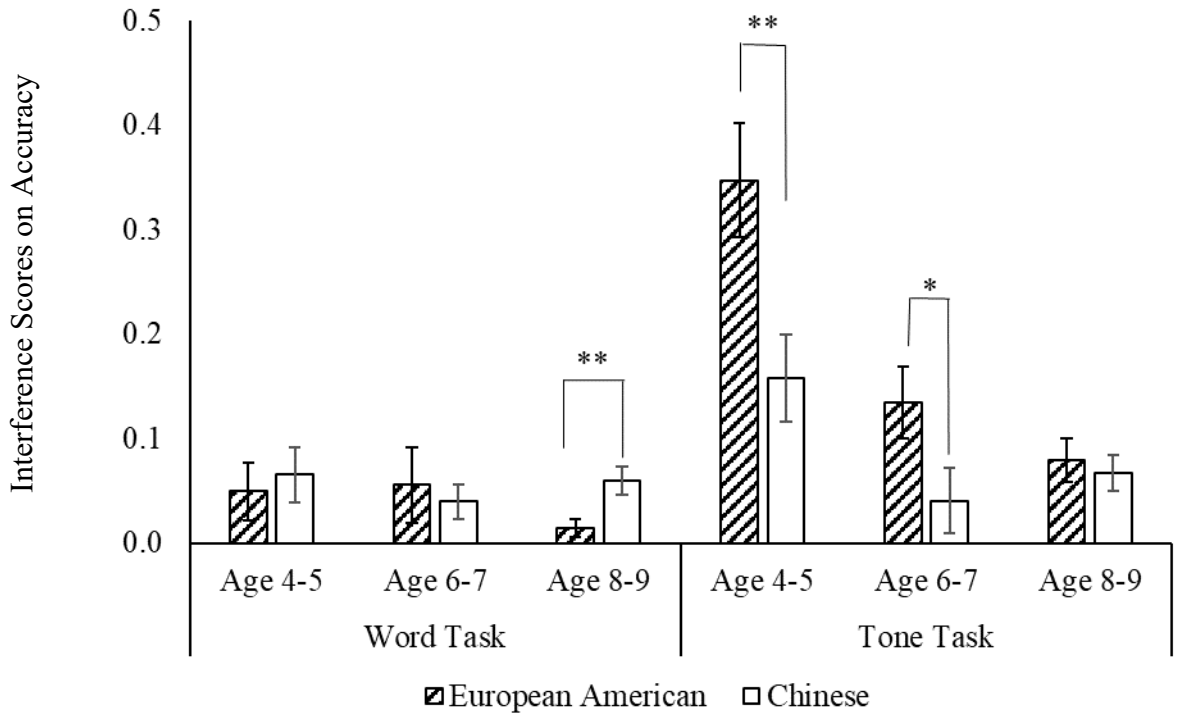
Mean Interference Scores on Accuracy as a Function of Culture and Task



Note. Error bars denote ± one standard error of the mean. ** $p < .01$

Figure 2

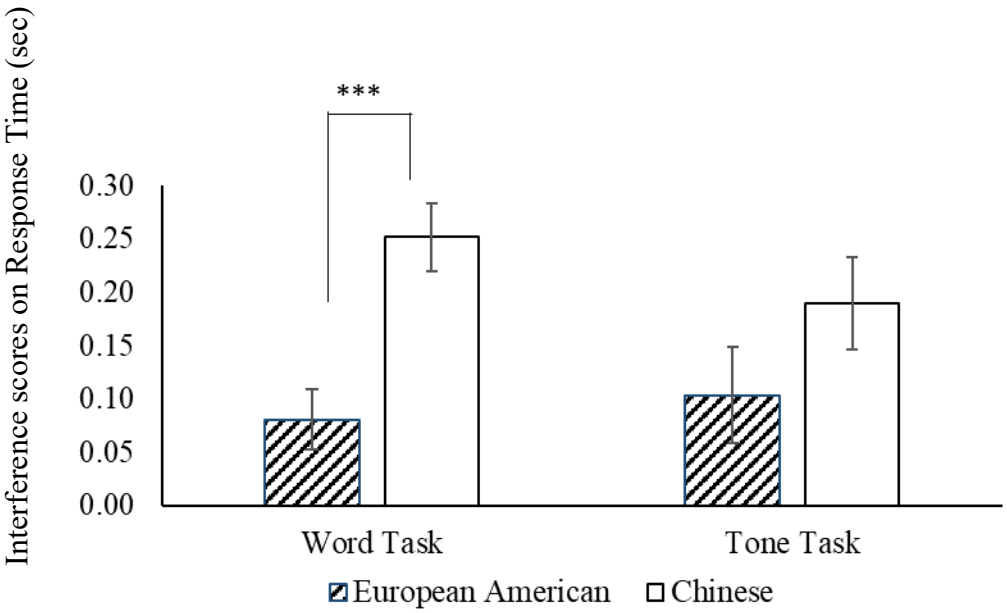
Mean Interference Scores on Accuracy as a Function of Age Group, Task, and Culture



Note. Error bars denote ± one standard error of the mean. ** $p < .01$ * $p < .05$

Figure 3

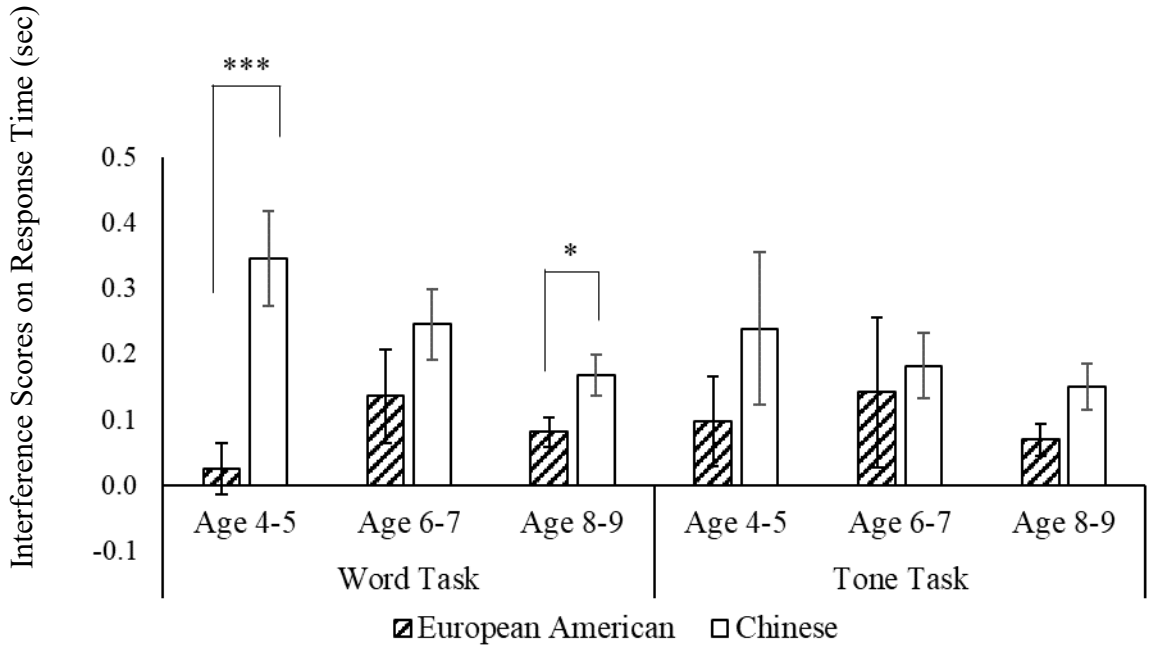
Mean Interference Scores on Response Time as a Function of Task and Culture



Note. Error bars denote ± one standard error of the mean. *** $p < .001$

Figure 4

Mean Interference Scores on Response Time as a Function of Age Group, Task, and Culture



Note. Error bars denote ± one standard error of the mean. *** $p < .001$ * $p < .05$