



# Age-related changes in multisensory integration of emotions in an audiovisual face-prosody-semantic Stroop task

Yi Lin<sup>1,2\*</sup>, Shumeng Ni<sup>1,2</sup>, Yangfan Lu<sup>1,2</sup>

<sup>1</sup>Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup>National Research Centre for Language and Well-being, Shanghai, China

carol.y.lin@sjtu.edu.cn, colormebblue@sjtu.edu.cn, roseeee\_lyf@sjtu.edu.cn

## Abstract

The present study examined age-related changes in multisensory integration of emotions across facial, prosodic, and semantic channels. Older and younger adults performed an audiovisual Stroop task, in which happy or sad emotions were expressed simultaneously through the three channels. They selectively attended to one of the emotional channels while ignoring congruent or incongruent emotions from other channels. Results indicated that older adults showed an overall decline in emotion integration with greater deficits in the nonverbal channels (particularly prosody) compared to verbal semantics. These channel-specific patterns of age differences were more prominent when information was incongruent across channels compared to the cross-channel congruent condition. To sum up, age differences in multisensory emotion perception are shaped by the complex interactions between channel asymmetry and information congruity, which may be linked to age-related declines in cognitive control.

**Index Terms:** multisensory integration of emotion, channel asymmetry, information congruity, age-related changes, face-prosody-semantic Stroop task

## 1. Introduction

To achieve high-quality interpersonal communication, one needs to integrate emotions from various verbal (i.e., semantic) and nonverbal (e.g., prosodic and facial) sensory channels. With advancing age, individuals often experience increasing difficulty in decoding emotions from lexical semantics [1], vocal prosody [2-4] and facial expressions [5, 6]. These age-related declines in emotion perception can be differentiated across sensory channels. Age differences are typically more pronounced in the lexical task compared to the facial task [1], although older adults may perform better on these two sensory channels than on emotional prosody [7, 8].

One potential explanation for the variations across channels lies in the sensory dominance effect, which highlights the more efficient and effective processing of one modality/channel over another [11]. Visual facial cues have often been reported to facilitate or bias the processing of auditory information, thus holding greater perceptual advantages during audiovisual integration [9, 10]. However, this sensory dominance effect is not always consistent across development [11]. For the relationships between prosody and semantics, older individuals tend to distribute relatively balanced attention to the two channels with slight-to-moderate semantic advantages [2, 7, 12], though prosodic dominance has also been observed especially under noise settings [13]. Similarly, younger adults can give more weights to either the

semantic or prosodic cues according to contextual modulations or individual preferences [14-16]. Given these mixed findings on sensory dominance effects, the magnitude of age-related emotional declines requires careful investigation across verbal and nonverbal channels.

In addition to sensory asymmetry, complex interactions across channels conveying congruent or incongruent information also influence multisensory emotion perception. According to the congruity effect, if information is congruent across channels, it enhances the integration performance; otherwise, emotion perception will be disrupted. There is a general tendency for older adults to show a larger congruity effect in emotional or non-emotional contexts [12, 17]. Nevertheless, existing studies often focus on pairwise comparisons between sensory channels, with limited exploration of interactions involving more than two channels [15, 18]. This approach poses constraints on its real-world applicability, as everyday communication often requires the integration of information concurrently presented through multiple verbal and nonverbal channels.

Two influential accounts have been proposed to explain age differences in multisensory emotion perception. One is the information degradation hypothesis, which associates the perceptual differences in stimulus modalities/channels with age-related declines in the corresponding sensory systems [19]. The other account is the inhibition deficit hypothesis [20], which attributes the increased congruity effects observed in older adults to their cognitive declines, particularly in their capacity to inhibit interference from irrelevant information. Although age-related differences in emotion perception may be associated with both sensory and cognitive factors, these two sources are likely to play distinct roles under different contextual conditions. It remains to ascertain under what circumstances age-related emotional deficits stem from sensory, cognitive or both sources in a complex multisensory setting.

The purpose of this study is to investigate the perceptual mechanisms of multisensory integration of emotions among older and younger adults. We administered a multi-channel Stroop task to examine the extent of age-related declines in integrating emotions conveyed through facial, prosodic, and semantic channels. The Stroop paradigm allows for simultaneous presentation of multisensory cues, thus being widely employed in previous research on sensory asymmetry and information congruity [21]. In our three-dimensional audiovisual Stroop task, semantically happy and sad words were spoken with a happy or sad prosody, and presented simultaneously with a happy and sad visual facial expression. Participants were asked to attend to emotions from one of the facial, prosodic and semantic channels while ignoring the other channels with either congruent or incongruent emotions. Based

\* Corresponding author

on previous literature, we expected to observe a general decline in the ability to integrate emotions across multisensory channels among older adults compared to their younger counterparts (Hypothesis 1). However, the extent of age-related declines would vary across different verbal and nonverbal channels (Hypothesis 2). Older adults would also show a greater congruity effect during multisensory integration of emotions (Hypothesis 3). Furthermore, our Stroop design can shed light on the sources of reduced ability in multisensory emotion perception among older populations. If the age-related changes across the three sensory channels differed between congruent and incongruent conditions, this would indicate a cognitive basis for the deficits that would be sensitive to inhibitory control; otherwise, age differences in emotion perception would show channel-specific modulations that would predominantly arise from sensory sources.

## 2. Methods

### 2.1. Participants

The present study was approved by the Institutional Review Boards (IRB) and adhered to the principles outlined in the Declaration of Helsinki at Shanghai Jiao Tong University (SJTU). The study included a total of 65 participants. They were divided into two groups: 33 older adults (11 men and 22 women, mean age  $\pm$  SD:  $68.61 \pm 4.74$ ) and 32 younger adults (10 men and 22 women, mean age:  $19.81 \pm 2.07$ ). No significant difference in years of education was observed between the two groups,  $t(36.16) = -1.22, p = 0.231$ . In terms of hearing ability, younger adults exhibited significantly better hearing than older adults; however, both groups fell within the clinically normal hearing range for pure tones at frequencies of 500, 1000, and 2000 Hz. Specifically, the average hearing thresholds for these designated frequencies were no more than 15 dB HL for younger adults and 25 dB HL for older adults with a difference between ears of less than 20 dB HL. All participants were assessed for cognitive functioning and showed clinically normal cognitive performance in the Chinese version of the Montreal Cognitive Assessment (MoCA) [22].

### 2.2. Stimuli

The auditory and visual stimuli used in this study were derived from an audiovisual emotion database established in our prior research on cross-channel and cross-modal emotion perception in young, healthy adults [15, 23, 24].

Emotions of happiness and sadness were simultaneously expressed through three communication channels: semantic, prosodic and facial. The semantic stimuli consisted of 32 Mandarin adjectives, all synonymous with “happy” or “sad,” sourced from the Chinese Affective Words System (CAWS) [25]. These adjectives, composed of two characters each, were uttered by two male and two female amateur actors in either a happy or sad tone as the prosodic stimuli. The facial stimuli consisted of 32 black-and-white images of emotional faces sourced from the Chinese Affective Picture System (CAPS) [26].

For each voice stimulus, four facial expressions (two happy and two sad) by four actors of the same gender were paired together. These emotional expressions from all three channels were recognized by over 90% of 24 Chinese university students in a two-choice task (either happy or sad). A norming study also revealed that emotional intensity ratings for these stimuli were consistent, with an average score exceeding 3 on a 7-point Likert scale (0 = *not intense*, 6 = *very intense*).

### 2.3. Procedures

The experiment was conducted in a soundproof booth, with the participant seated comfortably in a chair approximately 70 cm from an LCD monitor. The stimulus presentation was programmed using PsychoPy (Version 2022.2.4). Auditory stimuli were delivered binaurally through Sennheiser HD280 PRO headphones at a sound pressure level of 70 dB SPL.

The participants familiarized themselves with the testing procedure through eight practice trials, and they were required to achieve at least 75% accuracy to proceed to the formal test. The test consisted of three counterbalanced blocks: prosody, semantics, and face. Within each block, the target channel remained consistent across all trials. Participants needed to selectively attend to the target channel while disregarding information from the other two channels. For instance, in the prosody block, participants were asked to focus only on the emotional tone conveyed by the speaker’s voice while ignoring the semantic and facial expression channels. Each block included 32 trials, resulting in a total of 96 trials for the entire test. The trial order was pseudo-randomized.

In each trial, two channels had the same emotion, while the emotion of the third channel could either match or differ from the other two, creating four congruence conditions (as shown in Figure 1). The purpose was to investigate how the emotions in the two congruent channels interact with the third congruent or incongruent channel. After completing 32 trials, participants could take a short break and continue by pressing the spacebar when ready. Each trial began with a 1000 ms fixation point. Then, auditory stimuli with semantic and prosodic emotional content were presented simultaneously in both ears, while a facial expression image appeared in the center of the screen. The initial presentation time for both visual and auditory stimuli was aligned. Participants were required to quickly press a key to indicate the emotion category of the target channel, while maintaining accuracy. The response keys were “f” for happy and “j” for sad, and the key positions were counterbalanced across participants. Accuracy and response times were measured from the start of the stimulus presentation, with a 1000 ms inter-trial interval after each response. The detailed protocol is shown in Figure 1.

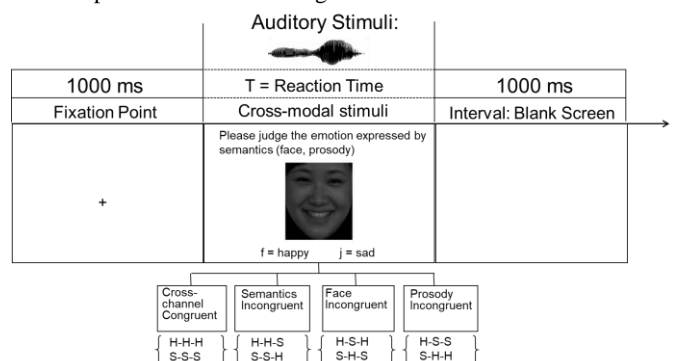


Figure 1: Schematic illustration of the cross-modal emotion perception test protocol. The three-letter sequences of “H” (happy) and “S” (sad) represent emotions in prosody, face and semantics respectively. For example, “H-S-S” refers to a word spoken in a happy tone, paired with a sad facial expression, and with a sad semantic meaning.

### 2.4. Statistical Analysis

Data analysis was performed using a series of linear mixed-effects models in R (version 4.4.1) with the lme4 package. Accuracy data were first averaged across channel and

congruence conditions for each participant and the mean scores were transformed into rationalized arcsine units (RAU) to reduce ceiling effects [27]. For reaction time analysis, only data points within two standard deviations from the mean and correct responses were included for both groups [28]. The final exclusion rates of response time data were 10.10% for the older group and 4.17% for the younger group. To address positive skewness, reaction time data were log-transformed [28].

Mean accuracy in RAUs and the logarithm of reaction time were used as the dependent variables. Task type (semantic, prosodic, and facial) and congruence (cross-channel congruent, semantics-prosody congruent [face incongruent], semantics-face congruent [prosody incongruent], and prosody-face congruent [semantics incongruent]), as well as group (old and young) were treated as categorical fixed factors. Listener subjects was included as random factors for intercepts for accuracy analysis and test items was further added for reaction time analysis. Tukey's post hoc tests were applied when significant main or interaction effects were found, using the emmeans package. Since we primarily focused on age differences in multisensory emotion perception in the present study, we only detailed the main effect of group, the two-way interactions between group and the other fixed factor (e.g., task or congruence), and the highest-level multivariate interaction. The younger group was set as baseline in the pairwise comparison of group effects.

### 3. Results

#### 3.1. Accuracy

Linear mixed-effects analyses on accuracy data (Figure 2A) suggested a significant main effect of group ( $\chi^2(1) = 20.25, p < .001$ ). Overall, older adults performed the tasks with lower accuracy than younger adults ( $\beta_3 = -11.2$ , Standard error (SE) = 2.33,  $t = -4.80, p < .001$ , Cohen's  $d = -0.45$ ). There was a significant interaction between task and group ( $\chi^2(2) = 45.56, p < .001$ ). Group differences were significant in the facial ( $\beta_5 = -8.56, SE = 2.9, t = -2.95, p = .004, d = -0.35$ ) and prosodic ( $\beta_5 = -22.38, SE = 2.9, t = -7.71, p < .001, d = -0.93$ ) tasks but not in the semantic one ( $p = .373$ ).

There was also a significant interaction between congruence and group ( $\chi^2(3) = 24.66, p < .001$ ). No significant group difference was found for the cross-channel congruent

condition. However, older adults produced significantly lower accuracy than younger adults in semantics incongruent ( $\beta_6 = -16.12, SE = 3.14, t = -5.14, p < .001, d = -0.67$ ), prosody incongruent ( $\beta_6 = -16.60, SE = 3.14, t = -5.29, p < .001, d = -0.69$ ) and face incongruent ( $\beta_6 = -10.34, SE = 3.14, t = -3.30, p < .001, d = -0.43$ ) conditions.

Most importantly, we observed a significant three-way interaction among task, congruence and group ( $\chi^2(6) = 31.96, p < .001$ ). While no significant group differences were found in all three types of task when emotional information was congruent across channels (all  $p$  values  $> .05$ ), age differences emerged in the facial and prosodic tasks under certain incongruent conditions. In the facial task, older adults were less accurate in emotion recognition when prosody ( $\beta_7 = -10.43, SE = 4.62, t = -2.26, p = .025, d = -0.44$ ) and semantics ( $\beta_7 = -20.45, SE = 4.62, t = -4.42, p < .001, d = -0.86$ ) presented incongruent information against the target facial channel, whereas the group difference was not significant in the semantics incongruent condition ( $p = .583$ ). In the prosodic tasks, older adults showed lower accuracy across the face incongruent ( $\beta_7 = -23.03, SE = 4.62, t = -4.98, p < .001, d = -0.97$ ), prosody incongruent ( $\beta_7 = -41.36, SE = 4.62, t = -8.94, p < .001, d = -1.74$ ), and semantics incongruent ( $\beta_7 = -20.94, SE = 4.62, t = -4.53, p < .001, d = -0.88$ ) conditions. In the semantic task, no significant difference was found for all congruent or incongruent conditions (all  $p$  values  $> .05$ ).

#### 3.2. Reaction time

Linear mixed-effects analyses on the logarithm of reaction time (Figure 2B) revealed a main effect of age group ( $\chi^2(1) = 50.93, p < .001$ ), suggesting that older adults were generally slower for the processing of emotional information than younger adults ( $\beta_3 = 0.41, SE = 0.05, t = 8.67, p < .001, d = 1.25$ ). There was also a significant task and group interaction ( $\chi^2(2) = 482.70, p < .001$ ). Older adults produced slower responses in all three types of tasks, but the effect size was largest in the prosodic task ( $\beta_5 = 0.59, SE = 0.05, t = 11.98, p < .001, d = -1.45$ ), followed by the facial ( $\beta_5 = 0.50, SE = 0.05, t = 10.27, p < .001, d = -0.83$ ) and semantic ( $\beta_5 = 0.16, SE = 0.05, t = 3.17, p = .002, d = 0.62$ ) ones. There were no significant interactions between congruence and group ( $\chi^2(3) = 5.27, p = .153$ ), and among task, congruence and group ( $\chi^2(6) = 11.18, p = .083$ ).

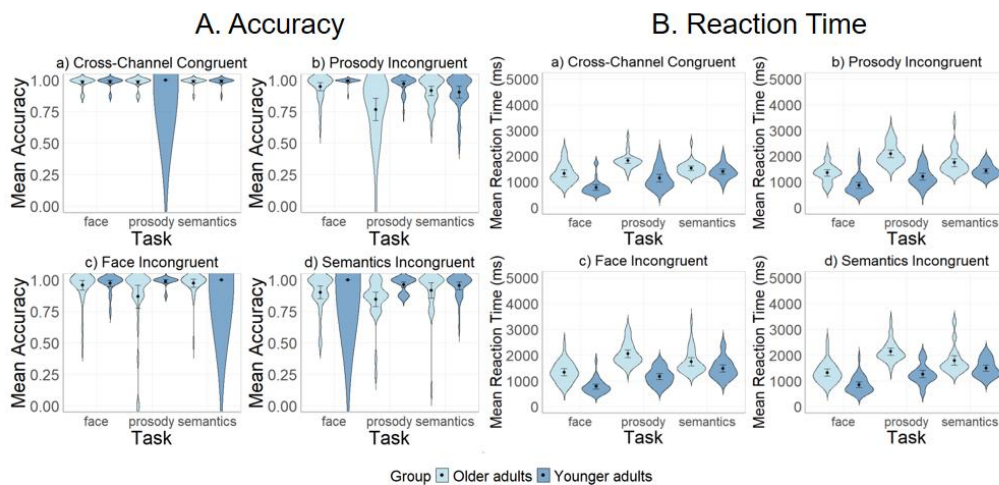


Figure 2: Violin plots for (A) mean identification accuracy and (B) mean reaction time (ms) displaying group differences under different congruence and task conditions. Data distribution shape was indicated by the density plots. Mean values were represented by the black dots and 95% confidence intervals were shown by the error bars.

## 4. Discussion

The present study compared how older and younger adults integrated emotions simultaneously presented through multisensory channels in a three-dimensional Stroop task. Overall, older adults exhibit a reduced ability to integrate facial, prosodic, and semantic emotional cues compared to younger adults. However, their performance declines are not uniform across these three channels. They demonstrate more significant declines in nonverbal emotion processing than in verbal semantics, with prosody showing the most pronounced impairment across all three channels. These channel-specific declines in older adults become more pronounced during nonverbal emotion perception when information across channels is incongruent rather than congruent. These findings suggest that age-related impairments in multisensory emotion perception are sensitive to specific communication channels and their interactions related to information congruence.

In line with our first hypothesis, older individuals demonstrated generally lower accuracy and slower response times in multisensory integration of emotions than their younger counterparts. This aligns with previous studies demonstrating age-related declines in the ability to understand emotions conveyed through various verbal and nonverbal channels [1, 5, 6, 8, 29, 30].

In accordance with our second hypothesis, the extent of these declines varies across different communication channels. While older adults made slower responses to verbal semantic emotions than younger adults, the two groups showed comparable accuracy in response to the verbal cues. In contrast, older adults' ability to process nonverbal emotional cues displayed more significant declines, with prosody revealing greater age-related differences in both accuracy and reaction time compared to facial expressions. This indicates that age-related deterioration of emotion perception is not symmetric across modalities but rather specific to individual sensory channels.

Similar patterns of channel asymmetry have been observed in our recent unisensory study, where emotions were expressed through face, prosody and semantics as separate channels [7]. One possible account for the channel-specific emotion deficits in older adults lies in stimulus characteristics. Our semantic stimuli consisted of emotion-labeled adjectives that express emotions in an explicit manner, thus yielding no accuracy differences between the two groups. In contrast, emotional expressions from facial expressions and prosody involve subtle variations in muscle movements and acoustic signals [31], which can be more ambiguous and subjective than verbal semantics so that their processing accuracy and response time can be affected. Another conceivable account involves age-related sensory declines in vision and hearing, as indicated by the information degradation hypothesis [19]. While semantic processing tends to be strengthened with age as it relies more on higher-level integration and assessment skills [32], interpreting emotions from facial expressions and vocal prosody depends on perceptual sensitivity to finer details. Consequently, age differences in emotion perception tend to be exacerbated in the nonverbal context compared to the verbal one.

Interestingly, while sensory imbalance may contribute to channel-specific performance declines in older adults, our study suggests that cognition plays a more significant role. This is because these observed age-related differences are evident only under incongruent conditions (but not the congruent ones) and are specific to attention modulation. As indicated by the three-way interaction in the accuracy measure, there was no significant group difference in all three tasks when information was congruent across channels. Older adults experienced interference in the face-oriented task when the two unattended

auditory channels were incongruent with the target facial channel. They showed even greater interference in the prosody-oriented task not only when prosody (the attended channel) was incongruent with face and semantics but also when incongruent information was presented through either the semantic or facial channels as the unattended channels. Conversely, no significant group difference was observed under all incongruent conditions in the semantics-oriented task. Consistent with our third hypothesis, these results suggest that age-related declines in multisensory emotion perception are more vulnerable to incongruent information during nonverbal than verbal processing, with prosody perception more severely affected than facial expression perception. In view of the inhibition deficit hypothesis, aging is linked to cognitive declines in inhibitory control [20]. Thus, older adults may have greater difficulty suppressing dominant information that is task-irrelevant [12, 33, 34], particularly when processing nonverbal signals, which express emotions through intricate cues, compared to verbal semantics that explicitly denotes emotions through literal meaning.

There are some limitations in the present study. On one hand, visual stimuli were static photos of facial expressions whereas the auditory stimuli were dynamic speech materials of emotional words. This cross-modal asymmetry may potentially affect the way emotions are perceived in each channel. Thus, future studies can employ more ecologically valid audiovisual inputs such as dynamic video materials with more complex linguistic contexts (e.g., sentences) and extralinguistic messages (e.g., gestures, postures) from individuals of different age groups [35]. On the other hand, our study suggests that age-related deterioration in nonverbal emotion perception differs between congruent and incongruent conditions, highlighting the potential influence of inhibitory control. However, the precise correlations between emotion perception and other cognitive abilities are yet to be determined. Future studies can further explore how emotion perception is associated with specific measures of perceptual and cognitive ability (e.g., pure-tone hearing thresholds, MoCA scores), which can quantify the relative contribution of basic sensory and cognitive functions [29, 31]. We also encourage future works to systematically investigate how individual variations in emotional speech perception within unisensory settings are related to their performance in multisensory contexts. This can provide a comprehensive picture of the interplay between individual and contextual factors in multisensory emotion perception, which has the potential to facilitate the development of rehabilitation and intervention techniques for aging populations with socio-communicative challenges.

## 5. Acknowledgments

This study was supported by the youth program of National Social Science Foundation of China (23CYY051).

## 6. References

- [1] D. M. Isaacowitz *et al.*, "Age differences in recognition of emotion in lexical stimuli and facial expressions," *Psychol. Aging*, vol. 22, no. 1, pp. 147-59, Mar 2007.
- [2] K. Dupuis and M. K. Pichora-Fuller, "Use of affective prosody by young and older adults," *Psychol. Aging*, vol. 25, no. 1, pp. 16-29, Mar 2010.
- [3] A. Sen, D. Isaacowitz, and A. Schirmer, "Age differences in vocal emotion perception: on the role of speaker age and listener sex," (in eng), *Cogn. Emot.*, vol. 32, no. 6, pp. 1189-1204, Sep 2018.
- [4] X. Fan, E. Tang, M. Zhang, Y. Lin, H. Ding, and Y. Zhang, "Decline of affective prosody recognition with a positivity bias among older adults: A systematic review and meta-analysis," *J. Speech. Lang. Hear. Res.*, vol. 67, no. 10, pp. 3862-3879, Oct 2024.

- [5] A. R. Goncalves, C. Fernandes, R. Pasion, F. Ferreira-Santos, F. Barbosa, and J. Marques-Teixeira, "Effects of age on the identification of emotions in facial expressions: a meta-analysis," *PeerJ*, vol. 6, p. e5278, 2018.
- [6] G. S. Hayes *et al.*, "Task characteristics influence facial emotion recognition age-effects: A meta-analytic review," *Psychol. Aging*, vol. 35, no. 2, pp. 295-315, Mar 2020.
- [7] Y. Lin *et al.*, "Category-sensitive age-related shifts between prosodic and semantic dominance in emotion perception linked to cognitive capacities," *J. Speech. Lang. Hear. Res.*, vol. 67, no. 12, pp. 4829-4849, Dec 2024.
- [8] L. Lambrecht, B. Kreifelts, and D. Wildgruber, "Age-related decrease in recognition of emotional facial and prosodic expressions," *Emotion*, vol. 12, no. 3, pp. 529-39, Jun 2012.
- [9] C. Spence, "Explaining the Colavita visual dominance effect," *Prog. Brain Res.*, vol. 176, pp. 245-258, 2009.
- [10] P. Garrido-Vasquez, M. D. Pell, S. Paulmann, and S. A. Kotz, "Dynamic facial expressions prime the processing of emotional prosody," *Front. Hum. Neurosci.*, vol. 12, p. 244, Jun 12 2018, Art no. 244.
- [11] W. R. Barnhart, S. Rivera, and C. W. Robinson, "Different patterns of modality dominance across development," *Acta Psychol. (Amst.)*, vol. 182, pp. 154-165, Jan 2018.
- [12] B. M. Ben-David, S. Gal-Rosenblum, P. H. H. M. van Lieshout, and V. Shakuf, "Age-related differences in the perception of emotion in spoken language: The relative roles of prosody and semantics," *J. Speech. Lang. Hear. Res.*, vol. 62, no. 4S, pp. 1188-1202, Apr 2019.
- [13] Y. I. Dor, D. Algom, V. Shakuf, and B. M. Ben-David, "Age-related differences in processing of emotions in speech disappear with babble noise in the background," *Cogn. Emot.*, pp. 1-10, May 2024.
- [14] K. Ishii, J. A. Reyes, and S. Kitayama, "Spontaneous attention to word content versus emotional tone: Differences among three cultures," *Psychol. Sci.*, vol. 14, no. 1, pp. 39-46, 2003, doi: 10.1111/1467-9280.01416.
- [15] Y. Lin, H. Ding, and Y. Zhang, "Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal Stroop effects," *J. Speech. Lang. Hear. Res.*, vol. 63, no. 3, pp. 896-912, 2020.
- [16] M. D. Pell, A. Jaywant, L. Monetta, and S. A. Kotz, "Emotional speech processing: Disentangling the effects of prosody and semantic cues," *Cogn. Emot.*, vol. 25, no. 5, pp. 834-53, Aug 2011.
- [17] M. Zurrón, M. Lindin, S. Galdo-Alvarez, and F. Diaz, "Age-related effects on event-related brain potentials in a congruence/incongruence judgment color-word Stroop task," *Front. Aging Neurosci.*, vol. 6, p. 128, 2014.
- [18] P. Filippi *et al.*, "More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing," *Cogn. Emot.*, vol. 31, no. 5, pp. 879-891, Aug 2017.
- [19] B. A. Schneider and M. K. Pichora-Fuller, "Implications of perceptual deterioration for cognitive aging research," in *The Handbook of Aging and Cognition*, F. I. M. C. T. A. Salthouse Ed., 2nd ed. Mahway, NJ: Erlbaum. : Erlbaum., 2000, pp. 155-219.
- [20] L. Hasher and R. T. Zacks, "Working Memory, Comprehension, and Aging: A Review and a New View," in *Psychology of Learning and Motivation*, vol. 22, G. H. Bower Ed.: Academic Press, 1988, pp. 193-225.
- [21] R. D. Melara and D. Algom, "Driven by information: A tectonic theory of Stroop effects," *Psychol. Rev.*, vol. 110, no. 3, pp. 422-471, 2003.
- [22] K. L. Chen *et al.*, "Validation of the Chinese version of Montreal Cognitive Assessment Basic for screening mild cognitive impairment," *J. Am. Geriatr. Soc.*, vol. 64, no. 12, pp. e285-e290, Dec 2016.
- [23] Y. Lin and H. Ding, "Multisensory integration of emotions in a face-prosody-semantic Stroop task." In *Proceedings of ICMI '19: NeuroManagement and Intelligent Computing Method on Multimodal Interaction*, 2019, pp. 1-5.
- [24] Y. Lin, H. Ding, and Y. Zhang, "Unisensory and multisensory Stroop effects modulate gender differences in verbal and nonverbal emotion perception," *J. Speech. Lang. Hear. Res.*, vol. 64, no. 11, pp. 4439-4457, Nov 2021.
- [25] Y. Wang, L. Zhou, and Y. Luo, "The pilot establishment and evaluation of Chinese affective words system," *Chinese Mental Health Journal*, vol. 22, no. 8, pp. 608-612, 2008.
- [26] L. Bai, H. Ma, Y. X. Huang, and Y. J. Luo, "The development of native Chinese affective picture system: A pretest in 46 college students," *Chinese Mental Health Journal*, vol. 19, no. 11, pp. 719-722, 2005.
- [27] G. A. Studebaker, "A rationalized arcsine transform," *J. Speech Hear. Res.*, vol. 28, pp. 455-462, 1985.
- [28] R. H. Baayen and P. Milin, "Analyzing reaction times," *International Journal of Psychological Research*, vol. 3, no. 2, pp. 12-28, 2010.
- [29] R. L. C. Mitchell and R. A. Kingston, "Age-related decline in emotional prosody discrimination: acoustic correlates," *Exp. Psychol.*, vol. 61, no. 3, pp. 215-23, 2014.
- [30] T. Ruffman, J. D. Henry, V. Livingstone, and L. H. Phillips, "A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging," *Neurosci. Biobehav. Rev.*, vol. 32, no. 4, pp. 863-81, 2008.
- [31] E. S. Dmitrieva and V. Y. Gelman, "The relationship between the perception of emotional intonation of speech in conditions of interference and the acoustic parameters of speech signals in adults of different gender and age," *Neurosci. Behav. Physiol.*, vol. 42, no. 8, pp. 920-928, Oct 2012.
- [32] T. H. Gollan and M. Goldrick, "Aging deficits in naturalistic speech production and monitoring revealed through reading aloud," *Psychol. Aging*, vol. 34, no. 1, pp. 25-42, 2019.
- [33] Y. I. Dor, D. Algom, V. Shakuf, and B. M. Ben-David, "Age-related changes in the perception of emotions in speech: Assessing thresholds of prosody and semantics recognition in noise for young and older adults," *Front. Neurosci.*, vol. 16, p. 846117, 2022.
- [34] A. Vallesi, D. T. Stuss, A. R. McIntosh, and T. W. Picton, "Age-related differences in processing irrelevant information: evidence from event-related potentials," *Neuropsychologia*, vol. 47, no. 2, pp. 577-86, Jan 2009.
- [35] N. Hedgecoth, N. Strand, and R. B. Adams, "The Intersection of Race, Gender/Sex, and Age in Emotion Perception from Faces and Bodies," in *Emotion Communication by the Aging Face and Body: A Multidisciplinary View*, U. Hess, J. R. B. Adams, and R. E. Kleck Eds., (Studies in Emotion and Social Interaction. Cambridge: Cambridge University Press, 2023, pp. 106-139.