Effects of Aging and Age-Related Hearing Loss on Emotion Identification in Prosodic and Semantic Channels

Yingyang Wang[1], Min Xu[2], Jing Shao[3*], Lan Wang[1], Nan Yan[1]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
[2]Institute of Corpus Studies and Applications, Shanghai International Studies University
[3]Department of English Language and Literature, Hong Kong Baptist University
yingyangwx@gmail.com, xumin@shisu.edu.cn, jingshao@hkbu.edu.hk, lan.wang@siat.ac.cn, nan.yan@siat.ac.cn

ABSTRACT

Accurate perception of emotional status conveyed in speech signal is crucial for communication. Aging is known to have negative effect on speech perception. Age-related hearing loss (ARHL) may exert additional effects on speech processing. In this study, we investigated how these two factors influence emotion identification in prosodic and semantic channels by examining the performance of three groups of participants: younger adults (YA) and older adults (OA) with normal hearing and with hearing loss. The results revealed that OA with hearing loss demonstrated degraded performance than their normal hearing counterparts and YA in both prosodic and semantic channels. Moreover, only OA with hearing loss showed the influence of channel in that their recognition accuracy in the prosodic channel was lower. These findings suggested that ARHL significantly contributes to the inferior emotion perception in OA, and it may especially manifest itself in the conditions where fine-tuning of auditory perception is required.

Keywords: emotional perception, aging, age-related hearing loss, communication channels

1. INTRODUCTION

The processing of emotions plays a crucial role in social interactions [1, 2]. In speech communication, emotions are coded by two main channels: semantic content, which conveys the linguistic meaning behind spoken words; and prosody, which conveys paralinguistic information through acoustic cues such as pitch, intensity, and rhythm. To correctly interpret speakers' communicative intentions, listeners must decode the emotions conveyed in these channels.

Age-related decline of cognitive and social understanding, together with changes of specific neuropsychological in the brain make it more difficult for older adults (OA) to recognize basic emotions across a variety of modalities than younger adults (YA) [3, 4]. These difficulties even exist for the healthy OA with normal hearing. With respect to emotional prosody, a previous study found that subjects' performance began to decline at the age of 45, and the highest error rates occurred over the age of 65 by examining the recognition of emotions across lifespan [5]. Later studies further confirmed an age-related decrease in emotion perception accuracy [6–10]. With respect to emotional semantics, the accuracy of identifying emotions in lexical stimuli was also influenced by aging [11, 12]. Moreover, cross-channel studies revealed aging effects on channel dominance. In the perception of spoken sentences containing different prosody and semantics emotion combination matrixes, YA relied more on prosodic information , whereas OA showed a bias toward semantics [13, 14]. When asked to focus on one channel, OA had difficulty in using the prosodic cues and scored higher in the semantic channel [14]. As these studies required integrating multichannel information simultaneously, it is still unclear whether the semantic biases found in OA are intrinsic or simply result from selective attention during multichannel emotional processing. Examining OA's emotional perception in each channel (prosody vs. semantics) separately is necessary.

Aside from aging, another factor that exacerbates the difficulty of speech perception is age-related hearing loss (ARHL). OA with ARHL have a higher threshold to recognize frequency and temporal information changes [15–17], resulting in deficits in processing fine-tuned auditory cues. However, studies about emotional perception and ARHL are still scarce. Existing studies only focused on emotional prosody, suggesting that OA with hearing loss had a specific mediation effect on the auditory emotion recognition [18]. They also showed reduction in emotional valence, arousal ratings [19] and vocal emotion perception [20]. However, among the studies, none of them has investigated the perception of emotional semantics, which is one of the most important channels for expressing emotions.

As stated above, the effect of aging on emotional perception has been demonstrated in previous studies. However, it is still unclear whether ARHL has an additional impact and, if so, whether and how it manifests itself in different communication channels. In the present study, we aim to fill this research gap. First, we tested three groups of participants, i.e., YA and OA with and without hearing loss, in the emotion identification tasks to investigate the effects of aging and ARHL on emotion processing. Second, we examined how prosodic and semantic channels affect identification accuracy and reaction time (RT). It is expected that aging and ARHL would adversely affect emotional perception, which means that OA, especially those with ARHL, perform worse than YA. Regarding communication channels, it is likely that OA, particularly for those with hearing loss, are worse at recognizing emotional prosody than semantics, due to their declined ability in processing acoustic cues.

2. METHOD

2.1. Participants

Three groups of native Mandarin speakers were recruited as participants. They all scored 26 or higher in the Montreal Cognitive Assessment-Basic test (MoCA-B, Chinese Version March 16, 2019), indicating normal cognitive abilities. All participants' hearing conditions were measured by audiometric thresholds at 6 octave frequencies (from 250 Hz to 8000 Hz) using an audiometer (GSI 18). The group of YA with normal hearing (YNH) contained 20 college and post-graduate students (10 females and 10 males) with an age range of 20–25 years old (Mean = 22.70, SD = 1.51). Their binaural hearing thresholds were below 20 dB HL at all octave frequencies. The group of OA with normal hearing (ONH) contained 18 participants (10 females and 8 males), ranging between 62 and 70 years of age (Mean = 66.39, SD = 2.53). Their binaural audiometric thresholds were lower than 20 dB HL below 4000 Hz and no more than 30 dB HL at 8000 Hz. The last group was OA with moderate or moderate-severe hearing loss (OHL), with 17 participants (9 females and 10 males) ranging in age from 62 to 77 years (Mean = 67.05, SD = 3.52). All of them have binaural hearing thresholds of more than 35 dB HL at 1000 Hz or higher, and most have more than 20 dB HL at 500 Hz or below. Figure 1 shows the audiogram of three groups.

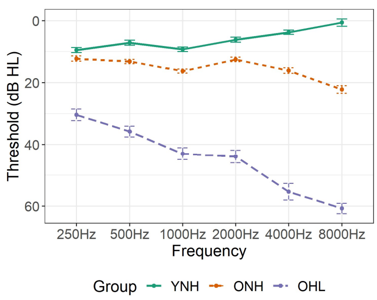

Figure 1: The mean audiometric thresholds (±1 standard error) for binaural hearing.

Each participant signed a written consent form which was approved by the Human Subjects Ethics Subcommittee at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences.

2.2. Stimuli

Two channels were used to convey emotions, namely prosody and semantics. All stimuli were selected from Age-Related Differences in Affective Norms for Chinese Words (AANC) which was the first large-scale Chinese four characters word system on age-related affective norms [21]. Using 9-point scales, the AANC assessed the semantic valence of the pleasantness evoked (positive words: valence > 6; negative words: valence ≤ 4; neutral word: 4 ≤ valence < 6) and the level of familiarity with words. In the present study, we selected stimuli with familiarity levels greater than 7, ensuring that both younger and older participants were familiar with the words.

The prosodic stimuli were 16 semantically neutral concrete words (4 < valence < 6, such as "玩具手枪 /uan^{35} tɕy^{51} ʂəu^{214} tɕʰiɑŋ^{55}/." (toy pistol)) which were enunciated in happy, sad and neutral tone of voice. The semantic stimuli were semantically happy (valence > 6.5, such as "笑口常开. /ɕiɑu^{51} kʰəu^{214} tʂʰɑŋ^{35} kʰai^{55}/" (often laugh)), sad (valence < 3.5, such as "万箭穿心. /uan^{51} tɕiɛn^{51} tʂʰuan^{55} ɕin^{55}/" (in extreme grief)) and neutral (4<valence<6, such as "春去秋来. /tʂʰuən^{55} tɕʰy^{51} tɕʰiəu^{55} lai^{35}/" (time rolls on)) words. The semantic stimuli were 16 words in each emotional state and all were uttered in a neutral tone. Overall, there were a total of 96 stimuli (16 words × 3 emotions × 2 channels). In this experiment, "happy" and "sad" stimuli were targets, while "neutral" stimuli were treated as fillers.

Each stimulus was recorded three times in a sound booth by three female Mandarin speakers sampling at 44100 Hz with a 16 bits resolution. We firstly selected the most emotional and clearly evocative one from 3 repetitions of each speaker. We then evaluated the emotional intensity of the selected tokens by conducting an emotion category identification test (happy or sad), as well as the emotional intensity rating on a 7-point scale (1 = not intense, 7 = very intense), following the procedures used in previous studies [22, 23]. Participants were 8 younger native Chinese speakers (4 males and 4 females) who did not take part in the following perception experiment and are naïve to the experimental design. Only stimuli that were of higher than 88% of identification accuracy and a mean rating of greater than 4.5 were adopted. The mean accuracy of identifying emotional categories and rating emotional intensity for the target stimuli were shown in Table 3.

Channel	Emotional status	Accuracy		Intensity	
		Mean	SD	Mean	SD
Prosody	Happy	99%	0.03	5.46	0.34
	Sad	99%	0.04	5.77	0.15
Semantics	Happy	100%	0.00	5.32	0.38
	Sad	98%	0.04	5.42	0.47

Table 1: The accuracy of identifying emotional categories and rating emotional intensity for the target stimuli selected in the experiment

2.3. Procedure

We implemented the experiment in a standard laboratory soundproof room through E-prime 3.0. Auditory stimuli were presented over Sennheiser HD280 PRO headphones binaurally. In order to preserve the details of the stimuli and to ensure that every listener perceived the same volume, a 1k Hz tone of the same root-mean-square level as all stimuli was generated, and it was used to calibrated the system volume at approximately 60 dB SPL (measured by sound pressure meter (Rion NL-21)).

The tasks of two channels were presented in separate blocks. In prosodic task, 48 stimuli were repeated twice and presented randomly to the participants, resulting in 96 trails. Participants were asked to identify the emotion according to prosody of speaker's tone by pressing one of the three emotion coded keys ("v" for happy, "b" for sad and "n" for neutral) on the keyboard as quickly as possible. In semantics task, trails present in the same way and participants were asked to response according to semantic meaning of the stimuli. The order of these two tasks was counterbalanced across participants. Practices were conducted before each task to help participants familiarize with the experiment.

2.4. Data analysis

Data analyses only included trials involving happy and sad emotions, with fillers excluded. Identification accuracy and RT were analysed. To avoid the influence of ceiling effect, the identification accuracy was converted to rationalized arcsine transform units (RAU) [24]. For RTs, incorrect responses and responses over ± 3 SD from the mean of each subject in each channel were excluded. And then the original data were log-transformed. Transformed accuracy and RT data were entered as dependent variables in a linear mixed effects model, in which group (YNH, ONH and OHL), channel (prosody and semantics), as well as their interactions were treated fixed effects, while subject and item were treated as random effects. Pairwise comparisons with significant interactions were conducted using Tukey's post hoc tests. The data analyses were carried out with R, using the lmeTest package [25] and the emmeans package [26].

3. RESULTS

Figure 2 illustrated the mean identification accuracy in two different communication channels. Results of linear mixed effects model suggested a significant main effect of group (χ^2 (2) = 31.938, $p < 0.001$), an approximate significant main effect of channel (χ^2 (1) = 3.068, $p = 0.080$). The interaction between group and channel was also significant (χ^2 (2) = 15.733, $p < 0.001$). Post-hoc tests indicate that OA with hearing loss identified emotional stimuli less accurately than YA and OA with normal hearing in both prosody (OHL vs YNH: $p < 0.001$; OHL vs ONH: $p < 0.001$) and semantic (OHL vs YNH: $p < 0.001$; OHL vs ONH: $p = 0.003$) channels, whereas there was no significant difference between YA and OA with normal hearing (prosodic channel: $p = 0.873$; semantic channel: $p = 0.628$). Additionally, in the group of OA with hearing loss, the identification accuracy of the prosodic channel was significantly lower than that of the semantic channel ($p < 0.001$).

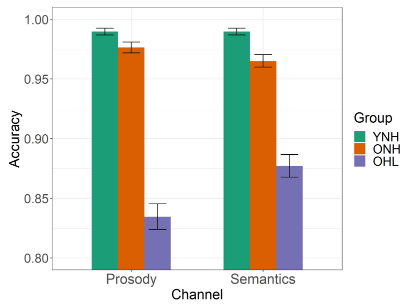

Figure 2: Mean identification accuracy (± 1 standard error) across two channels of three groups.

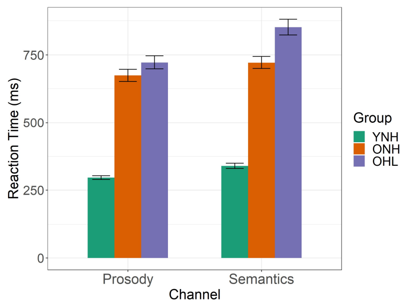

Figure 3: Mean reaction time (± 1 standard error) across two channels of three groups.

The RT of the three groups with respect to prosodic and semantics channels was illustrated in Figure 3. Linear mixed-effects analyses revealed a significant main effect of group (χ^2 (2) = 29.766, $p < 0.001$), an approximate significant main effect of channel (χ^2 (1) = 3.158, p = 0.076), but no significant interaction between two fixed effects (χ^2 (2) = 2.610, p = 0.271). In general, older adults, both those with and without hearing loss, employed longer RT than younger adults (ONH vs YNH: β = 0.250, SE = 0.060, t = 3.930, p < 0.001; OHL vs YNH: β = 0.310, SE = 0.063, t = 4.717, p < 0.001), suggesting greater cognitive efforts in performing the task.

4. DISCUSSION

The present study explored how aging and ARHL influence emotion perception in two channels. The findings showed that OA with hearing loss scored lower than the other two groups. Moreover, their accuracy in prosodic condition was lower than semantic condition, demonstrating the influence of channels. In addition, OA both with and without hearing loss responded slower than YA.

For OA with hearing loss, the most important finding was that their accuracy in recognizing emotional stimuli was lower than both YA and OA with normal hearing, suggesting that hearing ability played a crucial role in emotional speech perception and prosody comprehension. It has been found that there was degeneration of neural coding in the central auditory nervous system in OA with hearing loss [27, 28], which may influence their emotion speech processing. Additionally, previous study also found that when processing emotion information, hearing loss led to more engagement of the prefrontal cortex and less involvement of the amygdala, which may also have resulted in deficient emotion recognition in this population [29].

Another important finding was that OA with hearing loss showed channel effect in that their recognition accuracy in the prosodic channel was lower. A possible explanation was that the comprehension of speech involves a combination of sensory information from external sound and our prior knowledge in the brain. On the one hand, emotional prosody was modulated by detailed changes in acoustic cues. Mandarin speakers encoded happy prosody by employing higher F0, stronger intensity, shorter duration and higher first formant than sad prosody [30, 31]. OA with hearing loss had difficulties in decoding cues in speech signal [19, 20], which would result in deficiency in processing the fine-tuning information contained in the prosodic channel. On the other hand, human brain automatically utilizes prior knowledge to facilitate speech comprehension, especially in adverse hearing environments [32], and OA might rely more heavily on these knowledge than YA [33]. In our study, the stimuli used in the semantic condition were idioms which contain rich semantic information to assist the listeners process the emotion status. OA with hearing loss might be able to use their prior to identify the emotion. In contrast, such information was absent in prosodic condition where the stimuli were semantically neutral. Consequently, they were able to perform better in the semantic channel.

We also found that OA with normal hearing achieved comparable identification accuracy compared with YA, which is inconsistent from previous studies. This might be caused by the different task complexity. Previous studies usually included three or more emotions as targets [5, 7–12], while our study only investigated two (happy vs. sad). The lower task demand may make the group difference (YA vs. OA with normal hearing) less obvious in terms of accuracy. However, it is worth noting that both OA with and without hearing loss took longer RTs than YA. RT is a reliable measurement of listening effort [34]. The longer RT suggests that OA might require more cognitive resources to process speech and connect it to specific emotions compared with YA.

Future studies should increase complexity to reveal the possible group difference between OA with normal hearing and with hearing loss. This might include increasing the numbers of target emotions and investigating the emotion processing across multiple channels. Moreover, the association between cognitive ability and emotion speech perception is also a worthwhile direction for future research.

5. CONCLUSION

In the current study, we found that ARHL significantly affected emotion perception. OA with hearing loss performed worse than their normally hearing peers and YA. Furthermore, they showed channel influence in that they recognized emotion better in semantic channel. These findings broadened our understanding of emotion recognition in OA with hearing loss. It also provided empirical data to better understand the underlying mechanism of the deficient emotional processing in OA with ARHL.

6. ACKNOWLEDGEMENTS

This work was jointly supported by grants from National Key R&D Program of China (2020 YFC2004100), National Natural Science Foundation of China (NSFC 11904381, NSFC 81371900, NSFC 62271477, NSFC U1736202), Shenzhen KQTD Project (KQTD 20200820113106007), Shenzhen Science and Technology Program, Grant Award (JCYJ 20210324115810030) and the Start-up Grant from Hong Kong Baptist University (162646).

7. REFERENCES

[1] Davidson, R. J., Scherer, K. R. & Goldsmith, H. H. 2003. Handbook of Affective Sciences, *Adolescence*, 38, 390–391.

[2] Fischer, A. H.,& Manstead, A. S. R. 2008. Social functions of emotion. *Handb. Emot.* 3, 456–468.

[3] Ruffman, T., Henry, J. D., Livingstone, V., & Phillips, L. H. 2008. A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neurosci. Biobehav. Rev.* 32, 863–881.

[4] Sullivan, S., & Ruffman, T. 2004. Social understanding: How does it fare with advancing years? *Br. J. Psychol.* 95, 1–18.

[5] Brosgole, L., & Weisman, J. 1995. Mood recognition across the ages. *Int. J. Neurosci.* 82, 169–189.

[6] Mitchell, R. L. C. 2007. Age-related decline in the ability to decode emotional prosody: Primary or secondary phenomenon? *Cogn. Emot.* 21, 1435–1454.

[7] Sen, A., Isaacowitz, D. & Schirmer, A. 2018. Age differences in vocal emotion perception: on the role of speaker age and listener sex. *Cogn. Emot.* 32, 1189–1204.

[8] Martzoukou, M., Nasios, G., Kosmidis, M. H. & Papadopoulou, D. 2022. Aging and the Perception of Affective and Linguistic Prosody. *J. Psycholinguist. Res.*.

[9] Dupuis, K., & Pichora-Fuller, M. K. 2015. Aging affects identification of vocal emotions in semantically neutral sentences. *J. Speech, Lang. Hear. Res.* 58, 1061–1076.

[10] Mitchell, R. L. C., Kingston, R. A. & Barbosa Bouças, S. L. 2011. The Specificity of Age-Related Decline in Interpretation of Emotion Cues From Prosody. *Psychol. Aging*. 26, 406–414.

[11] Grunwald, I. S. *et al.* 1999. The effects of age and gender on the perception of lexical emotion. *Appl. Neuropsychol.* 6, 226–238.

[12] Isaacowitz, D. M. *et al.* 2007. Age differences in recognition of emotion in lexical stimuli and facial expressions. *Psychol. Aging*. 22, 147–159.

[13] Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H. H. M. & Shakuf, V. 2019. Age-related differences in the perception of emotion in spoken language: The relative roles of prosody and semantics. *J. Speech, Lang. Hear. Res.* 62, 1188–1202.

[14] Kim, D., Sim, H. S. & Lee, Y. 2022. Age-Related Differences in the Perception of Emotion in Emotional Speech: The Effects of Semantics and Prosody. *Audiol. Speech Res.* 18, 48–59.

[15] Patterson, R. D., & Weber, L. 1982. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.* 76, 1788-1803.

[16] Strouse, A., Ashmead, D. H., Ohde, R. N. & Grantham, D. W. 1998. Temporal processing in the aging auditory system. *J. Acoust. Soc. Am.* 104, 2385–2399.

[17] Xu, M., Shao, J. & Wang, L. 2021. Effects of aging and age-related hearing loss on talker discrimination. *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*. 2917–2921.

[18] Lena, L., Kreifelts, B. & Wildgruber, D. 2012. Age-related decrease in recognition of emotional facial and prosodic expressions. *Emotion*. 12, 529–539.

[19] Picou, E. M. 2016. How hearing loss and age affect emotional responses to nonspeech sounds. *J. Speech, Lang. Hear. Res.* 59, 1233–1246.

[20] Christensen, J. A., Sis, J.,Kulkarni, A. M., Chatterjee, M., Prostheses, A. & National, B. T. 2020. Effects of age and hearing loss on the recognition of emotions in speech. *Ear and hearing.* 40, 1069–1083.

[21] Liu, P., Lu, Q., Zhang, Z., Tang, J. & Han, B. 2021. Age-Related Differences in Affective Norms for Chinese Words (AANC). *Front. Psychol.* 12. 1–17.

[22] Lin, Y., Ding, H. & Zhangb, Y. 2020. Prosody Dominates Over Semantics in Emotion Word Processing: Evidence From Cross-Channel and Cross-Modal Stroop Effects. 63, 896–912.

[23] Lin, Y., & Ding, H. 2020. Effects of communication channels and actor ' s gender on emotion identification by native Mandarin speakers. *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*. 3151–3155.

[24] Studebaker, G. A., 1985. A “rationalized” transform arcsine. *J. Speech, Lang. Hear. Res.* 28, 455–462.

[25] Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. 2017. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26.

[26] Lenth, R.,Singmann, H.,Love, J.,Buerkner, P. & Herve, M. 2018. Emmeans: Estimated marginal means, aka least-squares means. *R Packag. version*. 1.

[27] Frisina, R. D. 2001. Possible neurochemical and neuroanatomical bases of age-related hearing loss - Presbycusis. *Semin. Hear.* 22, 213–225.

[28] Tremblay, K. L., Piskosz, M. & Souza, P. 2003. Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin. Neurophysiol.* 114, 1332–1343.

[29] Husain, F. T., Carpenter-Thompson, J. R. & Schmidt, S. A. 2014. The effect of mild-to-moderate hearing loss on auditory and emotion processing networks. *Front. Syst. Neurosci.* 8, 1–13.

[30] Zhang, S., Ching, P. C. & Kong, F. 2006. Acoustic analysis of emotional speech in Mandarin Chinese, in *International symposium on chinese spoken language processing*. 57–66.

[31] Yuan, J., Shen, L. & Chen, F. 2002. The acoustic realization of anger, fear, joy and sadness in Chinese. *7th Int. Conf. Spok. Lang. Process. ICSLP*. 2025–2028.

[32] Davis, M. H., & Johnsrude, I. S. 2007. Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147.

[33] Lesicko, A. M. H., & Llano, D. A. 2017. Impact of peripheral hearing loss on top-down auditory processing. *Hear. Res.* 343, 4–13.

[34] Pals, C., Sarampalis, A., van Rijn, H. & Başkent, D. 2015. Validation of a simple response-time measure of listening effort. *J. Acoust. Soc. Am.* 138, EL187–EL192.