



OPEN ACCESS

EDITED BY

Xiaowei Zhao,
Emmanuel College, United States

REVIEWED BY

Mingshuang Li,
California State University, Northridge,
United States
Junru Wu,
East China Normal University, China

*CORRESPONDENCE

Jing Shao
jingshao@hkbu.edu.hk

SPECIALTY SECTION

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

RECEIVED 06 June 2022

ACCEPTED 21 July 2022

PUBLISHED 12 August 2022

CITATION

Xu M, Shao J, Ding H and Wang L
(2022) The effect of aging on
identification of Mandarin consonants
in normal and whisper registers.
Front. Psychol. 13:962242.
doi: 10.3389/fpsyg.2022.962242

COPYRIGHT

© 2022 Xu, Shao, Ding and Wang. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction
in other forums is permitted, provided
the original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

The effect of aging on identification of Mandarin consonants in normal and whisper registers

Min Xu¹, Jing Shao^{2*}, Hongwei Ding³ and Lan Wang⁴

¹Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China, ²Department of English Language and Literature, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China, ³Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China, ⁴Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Consonant perception in older adults has been widely explored in recent years. However, how aging affects the identification of Mandarin consonants, especially in whispered condition, are under studied. Mandarin consonants and whispering speech have unique features, which may result in different identification difficulties for older adults. The current study investigated older adults' identification of Mandarin consonants in phonated and whispered conditions in comparison with the performance of younger adults. It was found that in phonated condition, older adults showed the lowest accuracy for affricatives and fricatives owing to their insensitivity to high-frequency information. The lower accuracy of affricatives and plosives was largely attributed to the difficulty in recognizing articulatory places. Identifying aspirated plosives was much more difficult than unaspirated counterparts for older adults. In whispered condition, the recognition of voiced consonants and aspirated-unaspirated distinction became challenging, especially for older adults. Contrary to the expectation, some consonants became easier to be recognized in the whispered condition, i.e., /p^h, t_ɕ^h, x/. These findings enrich our understanding of how aging affects consonant identification in different languages and less ideal conditions. It also suggests that the listener's ability, language uniqueness, and characteristics of distorted speech should be all taken into consideration when investigating speech perception in adverse conditions.

KEYWORDS

Mandarin, speech perception, consonant, aging, whispering

Introduction

The deteriorated abilities in older adults

Aging is accompanied by various declined abilities, such as deteriorated memory and lower capacity for attention and inhibition. Apart from that, one of the most common phenomena in older adults is age-related hearing loss. It has been well-documented that the aging ear is associated with the loss of sensory hair

cells (Schuknecht, 1974; Gates and Mills, 2005), which can amplify the traveling wave vibration in the auditory system. The reduced amplification led to high-frequency information decoded in the cochlear base being more likely to be affected than low-frequency information decoded in the cochlear apex (Cooper and Rhode, 1997; Robles and Ruggero, 2001). The insensitivity to high-frequency information in the aging population adversely affected their perception of certain consonants, including fricatives and affricatives. For example, Scharenborg et al. (2015) found that older adults were less sensitive to acoustic differences between fricatives /f/ and /s/ in categorical perception and hearing loss further lowered the sensitivity.

Another well-explored issue with older adults is their poor ability to process temporal information. The evidence was from psychoacoustic experiments, categorical perception using stimuli with varied temporal cues, and temporally distorted speech perception, such as time-compressed, reverberated, and noise-masked speech. The psychoacoustic experiments mainly revealed this deficit through the threshold examination of temporal order discrimination/identification and gap detection (Fitzgibbons and Gordon-Salant, 1998; Grose et al., 2006). With regard to speech perception, most studies modified the duration of VOT, closure duration, and formant transition. It was found that older adults showed inferior ability in temporal processing in speech perception (Ginzel et al., 1982; Gordon Salant et al., 2006). The temporally distorted speech perception was usually related to age-related hearing loss, which suggested that hearing loss played an influential role in temporal information processing (Rooij and Plomp, 1991). However, there were other studies showing that aging alone may also contribute to temporal processing deficits, independent from hearing loss (Fitzgibbons and Gordon-Salant, 1998; Strouse et al., 1998). This temporal processing deficit was also proved by neural evidence (see e.g., Walton, 2010 for a review). Based on these existing findings, it is likely that older adults' temporal processing deficits may negatively affect Mandarin consonant perception.

Apart from the deteriorated abilities mentioned above, aging is also associated with other inferior abilities, such as spectral processing deficit and insensitivity to pitch and intensity, which may also hinder speech perception (He et al., 1998, 2007; Harris et al., 2007; Souza et al., 2011; Bidelman et al., 2014). For example, He et al. (1998) and Harris et al. (2007) reported that older adults had higher thresholds for frequency and intensity than younger adults, although two groups had closely matched audiograms. Older adults also classified vowel continuum more variably which indicated they utilized spectral information less effectively (Bidelman et al., 2014).

Consonant perception in older adults

In the past few years, a large number of studies have confirmed the adverse influence of aging on consonant

perception (Gelfand et al., 1985; Ohde and Abou-Khalil, 2001; Tremblay et al., 2003; Scharenborg et al., 2015; Kalaiah et al., 2016; Fogerty et al., 2017; Rishiq et al., 2020). Some studies employed synthesized continua varied in the transition of second formant, VOT or frication duration to examine the categorical perception of consonants, such as plosives (/b, d, g/; /p, t, k/) and fricatives (/f, s/). This method is thought to reveal older adults' specific abilities in consonant perception. For example, Harkrider et al. (2005) and Tremblay et al. (2003) revealed older adults' temporal processing deficit and their inferior ability to process dynamic spectral information in categorical perception experiments. On the other hand, some studies examined the identification of a larger scope of consonants using natural stimuli. This line of research may better reflect older adults' difficulty in perceiving consonants in their daily communications. For example, identifying places of articulation was most difficult, followed by manners of articulation and voicing (Xia et al., 2018). Furthermore, consonants with nasality, glide/liquid, and sibilance quality were easier to be recognized than stop, frication, and place information (Gelfand et al., 1985).

The above-mentioned studies mostly focused on English. Existing works on the aging effects on phoneme perception in Mandarin mainly focus on the lexical tones and vowels (Yang et al., 2015; Wang et al., 2017a,b; Feng et al., 2019, 2021; Liu et al., 2021), while consonant recognition was understudied. There was only one study that examined the perception of 23 Mandarin consonants [including approximants/lateral approximants (A/LA)] in older adults (Zhao et al., 2005). It found that the error rate of high-frequency consonants (/p^h, t^h, f, tɕ^h, tɕ^h, s, ʂ/), which have the largest amplitude in the spectrum over 3,000 Hz, was significantly higher than other consonants in older adults.

Mandarin consonants can be grouped into plosives (/p, p^h, t, t^h, k, k^h/), fricatives (/ç, s, ʂ, f, x/), affricatives (/tɕ, tɕ^h, ts, ts^h, tʂ, tʂ^h/), nasals (/m, n/), and A/LA (/l, z/) according to the manner of articulation. In terms of the place of articulation, these consonants can be categorized into bilabials/labiodental (B/LD: /p, p^h, m, f/), alveolars (/t, t^h, n, l, ts, ts^h, s/), palatals (/tɕ, tɕ^h, ç/), velars (/k, k^h, x/) and retroflexes (/tʂ, tʂ^h, ʂ/). These consonants share some similarities with English but also have unique features. Unlike English which has voicing distinction in plosives and fricatives/affricatives, all these consonants in Mandarin are voiceless, and the plosives and affricatives have aspirated distinction. Besides, there are no consonant clusters in Mandarin and all consonants can only appear in the initial position, except for /n/. Moreover, it has been reported that the three palatal phonemes in Mandarin do not exist in English and are less common in other languages (Ladefoged and Maddieson, 1996). Another worth noting difference is that the three fricatives (/ç, s, ʂ/) share the same articulatory places as affricatives (/tɕ, tɕ^h, ts, ts^h, tʂ, tʂ^h/), which may increase the perceptual difficulty due to the increased competition in a similar range.

From the acoustic perspective, voice onset time (VOT), formant transitions, and spectral moments were the most effective cues for recognizing consonants in most languages (Lieberman et al., 1967; Forrest et al., 1988; Mitani et al., 2006; Li and Gu, 2015), with the values varied in different languages. For example, the VOT differences between voiced and voiceless consonants in English are around 20–30 ms, whereas the differences between aspirated and unaspirated consonants in Mandarin could be as large as 60 ms. It is reasonable to speculate that the larger VOT difference might facilitate the perception of the distinction of aspiration. Together, Mandarin consonants have many features that set them apart from other languages, both phonologically and acoustically. However, how aging affects consonant perception is still largely unknown.

Older adults' consonant perception in less ideal conditions

There has been a large number of studies focused on older adults' speech perception in less ideal conditions, including noise, reverberation, time-compression, foreign-accent, or mixed degraded speech (Helfer and Huntley, 1991; Helfer, 1994; Gordon-Salant and Fitzgibbons, 2004; Ferguson et al., 2010; Alwan et al., 2011; Reinhart et al., 2016). These studies did report their worse performance than younger adults, but it also varied with conditions. For instance, reverberation and noise influenced consonant perception differently. Plosives were easier to understand in noise than in the reverberation plus noise condition and reverberation exerted more influence on the syllable-final consonants than syllable-initial consonants. It indicated that the underlying mechanisms for speech perception in these conditions might be different (Helfer, 1994). Reverberation and noise affected speech perception mainly through masking the temporal envelopes of original signals rather than changing any acoustic features of original speech (Fogerty et al., 2010), while time-compression and foreign accents mainly distort information such as the temporal cues inherent to natural speech.

Whispering is also a degraded speech condition. It might be difficult for older adults, but as far as we know, how whispering affects older adults' speech perception has yet to be explored. Whispered speech is characterized by a turbulent flow created by exhaled air passing through the opened glottis (Sundberg, 1994; Matsuda and Hideki, 1999). Recognizing whispered speech requires decoding these noisy and weak signals, which are therefore more demanding. The most prominent change in whispered speech is that the fundamental frequency is not available owing to the absence of vocal vibration, which resulted in a dramatic decrease in lexical tone perception (Gao, 2002; Jiao and Xu, 2019), intonation identification (Meyer-Eppler, 1956; Heeren and van Heuven, 2009; Heeren and Van Heuven, 2014;

Zygis et al., 2017) and talker recognition (Lass et al., 1976; Fan and Hansen, 2011; Smith et al., 2017). Whispering also brings a shift of formants, especially for the first two formants, which leads to the perceptual difficulty for vowels (Kallail and Emanuel, 1984, 1985; Jovičić, 1998; Jiao and Xu, 2019).

Previous studies also demonstrated that whispering changed the acoustic features of consonants (Schwartz, 1972; Parnell et al., 1977; Jovičić and Šarić, 2008). It was found that the duration of whispered consonants was prolonged by about 10% on average and the prolongation of voiced consonants was greater than that of unvoiced. In the intensity domain, voiced consonants, that were nasals and A/LA, had reduced intensity over 25 dB but the intensity for unvoiced consonants remained unchanged. These acoustic changes brought perceptual difficulties, especially for the voiced-voiced distinction and the aspirated-unaspirated distinction (Dannenbring, 1980; Tartter, 1989; Mills, 2003; Meynadier et al., 2013). Our acoustic data (unpublished) on Mandarin consonants showed whispering led to an increased duration of unaspirated consonants but a decreased duration of their aspirated counterparts which narrowed the duration difference between them in whispered mode. Similarly, whispering lowered consonants' intensity, but increased the consonant-vowel intensity ratio (CVR), except for nasals. As a result of these changes in the acoustic properties, whispering may lead to heightened perceptual difficulties in Mandarin.

The current study

As we mentioned above, a large number of studies have proved that aging was accompanied by high-frequency insensibility, temporal and spectral processing deficits, as well as higher thresholds of intensity and frequency, which may detrimentally affect consonant recognition. Whispering distorts the original speech owing to the lack of vocal vibration, which is different from the masked effects of noise or reverberation. Examining the effects of aging and whispering on Mandarin consonant identification might enrich our knowledge of speech perception in older adults, especially in challenging conditions. However, this question has yet to be investigated. To this end, we conducted an identification experiment with Mandarin consonants as stimuli. Older and younger listeners were required to recognize the 21 Mandarin consonants in phonated and whispered conditions. Consonants were mainly grouped by place and manner of articulation in statistical analyses. We hypothesized that older adults might show lower accuracy than younger adults, even in the phonated condition. Among the consonants, recognizing affricatives and fricatives that are mostly dependent on decoding high-frequency information might be most difficult for the elderly. In contrast, the identification of nasals and semivowels, which are based on formant information might be less affected. In terms of the

aspirated-unaspirated distinction, it might be difficult as well considering the temporal processing deficit in older adults, but it is also possible that the difficulty might be compensated by the longer duration differences. In relation to whispering, we speculated that the identification of voiced consonants (nasals/A/LA), and aspirated-unaspirated distinction might be negatively affected in both age groups, but older adults will be more vulnerable to this condition.

Methods

Participants

Two groups of listeners, representing older and younger adults were recruited to participate in the current study. The older listener group consisted of 20 participants (16 females, 4 males) with ages ranging from 60 to 72 (Mean = 64.05, SD = 4.36). They were recruited in Shenzhen with the assistance of community managers. The younger listener group contained 20 students (8 females, 12 males) with an age range of 21–32 years old (Mean = 24.20, SD = 2.86). They were selected from the subject pool at Shenzhen Institute of Advanced Technology. Although the participants were recruited in Shenzhen, all the participants (younger and older adults) were from the northern China and spoke standard Mandarin. Furthermore, we have asked the participants to read the syllables on the hard copies before the formal experiment, which allowed us to make sure that all participants were able to distinguish the most confusing pairs, such as /ts-tʂ/, /ts^h-tʂ^h/, /s-ʂ/, and /n-l/. Due to the careful screening the dialect background of the participants, we assumed that the influence of sociolinguistic factors might have been kept at a lesser extent. Nobody reported neurological history or ear diseases. All participants were compensated for their participation. The experimental procedures were approved by the Human Subjects Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. Informed written consent was obtained from participants in compliance with the experiment protocols.

A portable audiometer (GSI 18) was used to assess participants' hearing condition by air-conduction audiometry from 250 to 8,000 Hz. Given that the screening test was conducted in a quiet room rather than in a soundproof laboratory or a hospital, the average pure tone threshold lower than 25 dB rather than 20 dB was chosen to define as clinically hearing normal (Brandy, 2002). All pure-tone thresholds of younger adults were lower than 25 dB at all octave intervals between 250 Hz and 8,000 Hz, and their binaural threshold differences were lower than 15 dB at each frequency. Ten of the older adults had matched audiograms with younger adults. Their pure-tone thresholds were lower than 25 dB at all octave intervals between 250 and 8,000 Hz, and binaural threshold differences were lower than 15 dB at each frequency. The rest

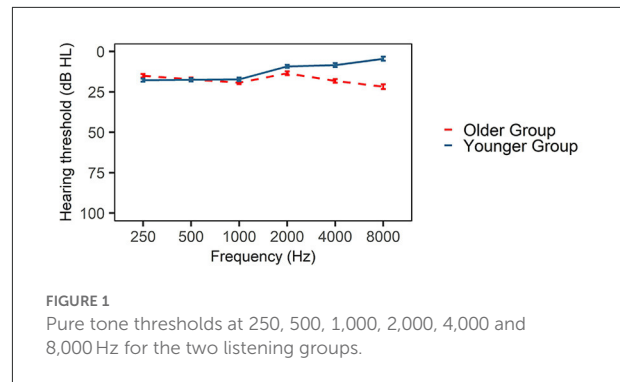


FIGURE 1
Pure tone thresholds at 250, 500, 1,000, 2,000, 4,000 and 8,000 Hz for the two listening groups.

10 of older adults had normal hearing below 4,000 Hz but mild hearing loss (<40 dB) at 8,000 Hz. The audiograms of two groups of listeners are shown in Figure 1. The hearing test took 10–15 min for each participant.

The cognitive abilities were tested with the Chinese version of Montreal Cognitive Assessment-Basic (MoCA-B) (<https://www.mocatest.org/>), which can be used for individuals with poor educational status. The test consists of 10 subtests: executive, memory, fluency, orientation, calculation, abstraction, delay recall, visual, naming, and attention. The scores of the older adults were all above 26 (Mean = 27.75, SD = 1.52), which indicate normal cognitive functioning. The cognitive screening lasted around 15 min.

Stimuli

Mandarin consonants /p, p^h, t, t^h, k, k^h, tʂ, tʂ^h, ʂ, ts, ts^h, s, tʂ, tʂ^h, ʂ, f, x, m, n, l, z/ were selected as target stimuli, and /a, i, u/ were used as the following vowels to ensure the diversity of the phonetic contexts. According to the classifications of Mandarin vowels, /a/ is an open vowel with the lowest tongue position, and /i/, /u/ are close vowels with the highest tongue position. The difference between /i/ and /u/ is the vowel backness with /i/ being a front vowel, while /u/ being a back vowel. The carrier lexical tones were high tone (Tone 1: 55). However, not all combinations of consonants, vowels, and lexical tones were legal in Mandarin. For example, the combinations of /tʂa55, tʂ^ha55, ʂa55, tʂu55, tʂ^hu55, ʂu55, zu55, ki55, k^hi55, xi55, fi55, zi55/ do not exist in Mandarin. To rule out the possibility that the pseudoword might elicit inferior performance in older adults, these illegal syllables were replaced by most similar syllables which are real words in Mandarin: /tʂia55, tʂ^hia55, ʂia55, ʂa^u51 (Tone 4), tʂio^u55, tʂ^hio^u55, ʂio^u55, zu51, kei214 (Tone 3), k^hei55, xei55, fei55, zi51/.

The original syllables were recorded with a female, speaking standard Mandarin, in a sound-attenuated laboratory. The target syllables were embedded in a sentence presented in both Chinese character and Pinyin through Powerpoint, for example,

“这个音节是*aba*. zhè ge yīn jié shì *aba*” (This syllable is *aba*). The speaker was given enough time to get familiar with the pronunciation of all stimuli. Each token was repeated at least three times to ensure the intelligibility of target syllables, giving rise to 6 blocks in total. Each block contained 21 target consonants. Phonated and whispered blocks were recorded alternately. For whispered sentences, the speaker was instructed to mimic whispering in a public library.

The first author selected the clearest syllable by observing the spectrogram and judging the intelligibility. Five younger listeners participated in a pilot study that had similar procedures to the formal study. The tokens with accuracy lower than 70% were replaced. Reevaluation was conducted until the tokens reached high intelligibility in younger adults. The target syllables were recorded and segmented at 44,100 Hz and 24 bits. Acoustic analyses were conducted for the selected phonated syllables (Mean duration = 621 ms, SD = 113 ms; Mean intensity = 57.22 dB, SD = 6.13 dB) and whispered syllables (Mean duration = 588 ms, SD = 111 ms; Mean intensity = 51.04 dB, SD = 7.05 dB). The selected syllables were then scaled to a mean duration of 600 ms at 60 dB by Praat. All the stimuli were presented at 60 dB sound pressure level (SPL) measured by sound pressure meter (Rion NL-21).

Procedure

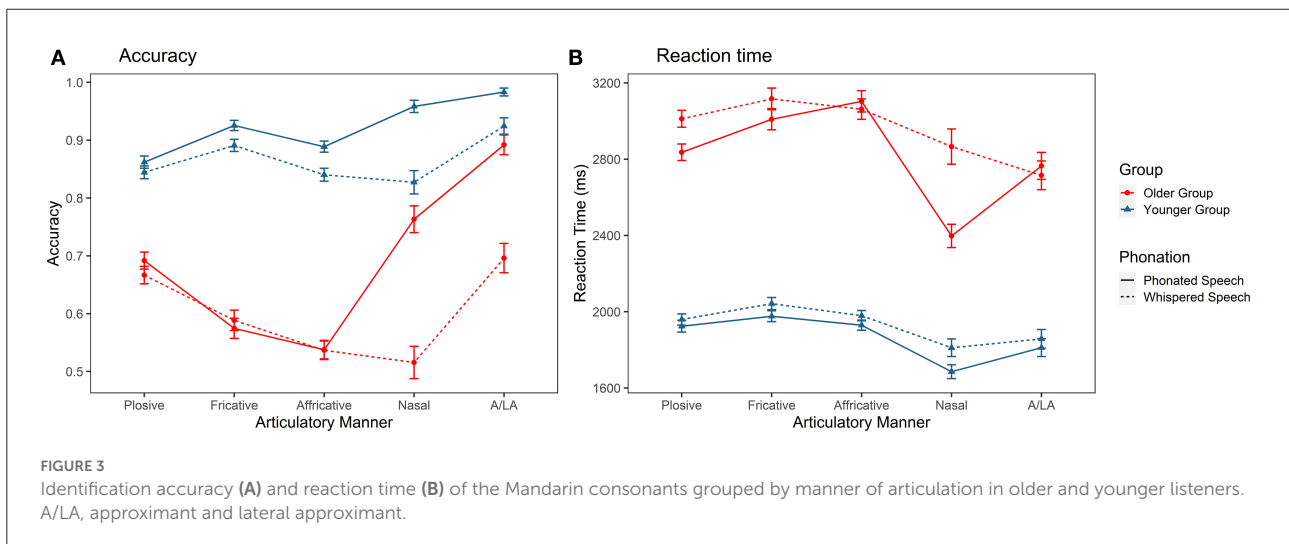
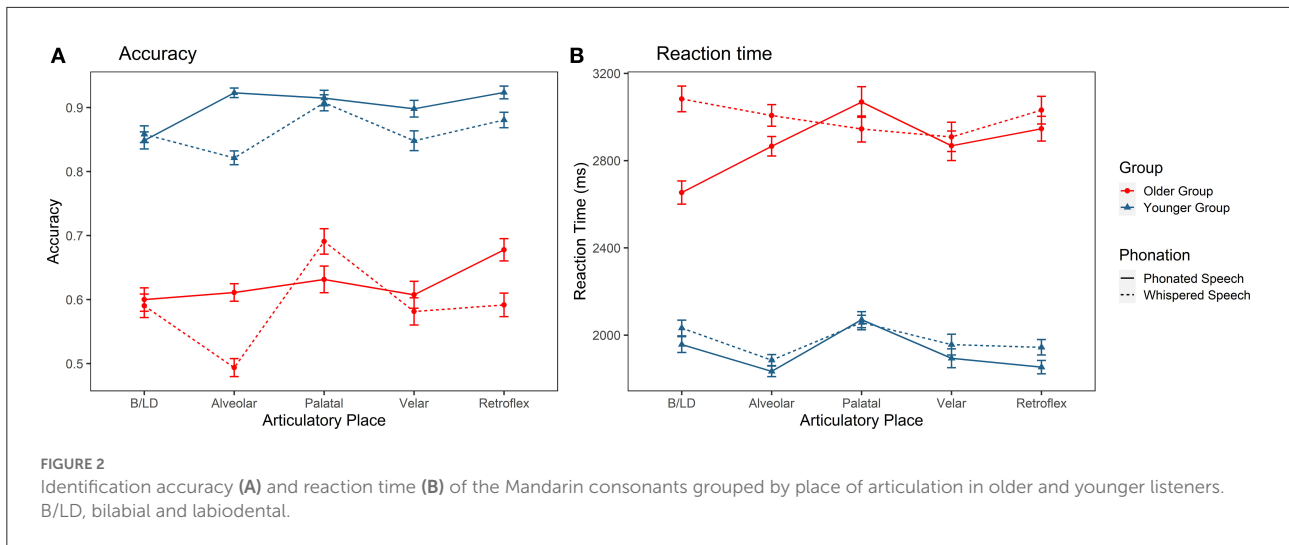
The procedure was a 21-alternative forced-choice identification task. Psychological software E-prime 3.0 was used to present stimuli and collect data. The response syllables were displayed on a computer monitor in Pinyin forms in six rows and four columns. The syllables in each row shared a similar place of articulation, for example, the first row contained bilabials and labiodental, the second row contained alveolars, and the third row contained velars. The auditory stimuli were randomly presented to listeners *via* Sennheiser 280 headphones binaurally. Participants were asked to identify the consonant by selecting one of the choices presented simultaneously on the screen as correctly and quickly as possible. Whispered and phonated stimuli were presented in separate blocks. There were six blocks that varied in articulatory modes and vowels. Each block consisted of 21 trials, each of which was repeated three times. The order of the blocks was counterbalanced. Participants could take a break after each block or after 30 trials within a block.

Listeners, especially older adults, were provided enough time to familiarize themselves with the stimuli and procedures. This was to ensure that consonant identification was affected by aging and articulatory conditions rather than other factors. In the familiarization stage, as has been described briefly above, participants were first asked to read out the 63 syllables (the combinations of 21 consonants and 3 vowels) printed on

the hard copies. Participants who could not read consonants correctly had to be excluded. Note that to minimize the influence of Pinyin proficiency on the perceptual identification results, the syllables were presented in Pinyin format, which were identical to the materials used in the formal test. Only those who can recognize Pinyin accurately were invited to the formal experiments. Secondly, listeners were required to complete two practice blocks (a phonated condition and a whispered condition) with feedback given for each trial, including accuracy and the correct answer. The auditory stimuli used in practice sessions were recorded by another female speaker. The participants were asked to redo the practice if their scores were lower than 50% in the phonated block. They would not be invited to the formal test if their accuracy was still lower than 50%. Nine older adults were excluded during this process.

Data analysis

Identification accuracy and reaction time (RT) were analyzed with a series of generalized linear mixed-effects models (for accuracy) and linear mixed-effects models (for RT). For accuracy analyses, trials with response times beyond the 500–7,500 ms were excluded as a result of accidental key pressing or lapses in attention. A total of 819 trials were excluded, accounting for 5.4% of the data. For RT, incorrect responses (26.11%) and responses over 2 SDs from the mean (3.88%) were excluded. The original data were then log-transformed for statistical analyses. The statistical analysis aims to elucidate consonant identification in two listener groups and in different places and manners of articulation. Generalized linear mixed-effects models were first fitted with group (younger adults, older adults), condition (phonated speech, whispered speech), and place (B/LD, alveolar, palatal, velar, retroflex), and their interactions as fixed effects, and subject was included as a random effect. Additional generalized mixed-effects models were then fitted with group, condition, and manner (plosive, fricative, affricative, nasal, A/LA), and their interactions as fixed effects, and with subject as a random effect. Further, more generalized linear mixed-effects models were fitted to investigate the identification differences within plosives, fricatives, and affricatives. Finally, to rule out the influence of high-frequency hearing loss (at 8,000 Hz), a series of generalized linear mixed-effects models were conducted to test whether there are differences between two older adults with normal hearing and mild high-frequency hearing loss. When there were significant main effects or interactions, Bonferroni *post-hoc* tests were applied for pairwise comparisons. The data analyses were performed with R (Version 4.0.5), using the *lme4* package (Bates et al., 2015) and *emmeans* package (Lenth, 2018).

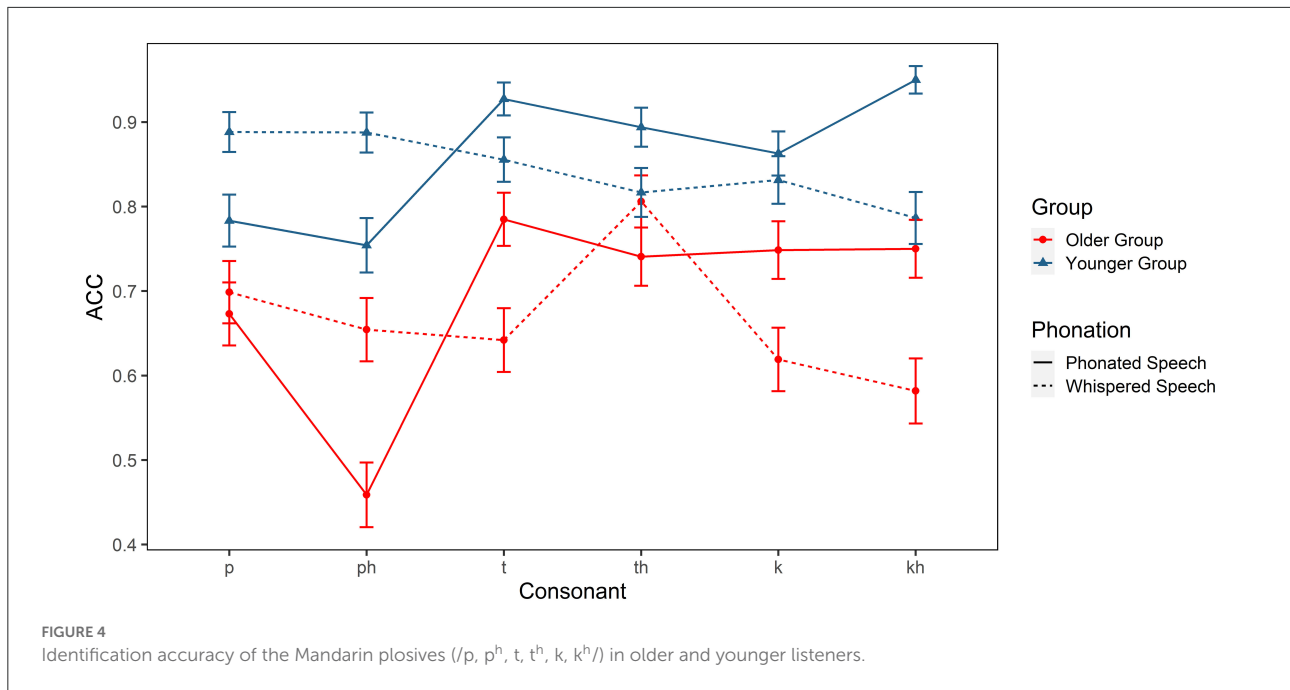


Results

As shown in Figure 2A, the identification accuracy of consonants with different places of articulation differs among the two listener groups and articulatory modes. There were significant main effects of group [$\chi^2_{(1)} = 73.909, p < 0.001$], condition [$\chi^2_{(1)} = 46.123, p < 0.001$], and place [$\chi^2_{(4)} = 63.639, p < 0.001$]. The interaction effects of group \times place [$\chi^2_{(4)} = 4.951, p < 0.01$], condition \times group [$\chi^2_{(1)} = 7.016, p < 0.001$], and condition \times place [$\chi^2_{(4)} = 60.494, p < 0.01$] were also significant. Three-way interaction was not significant ($p > 0.05$). *Post-hoc* tests indicated that older listeners identified B/LD, alveolars, palatals, velars, and retroflexes less accurately than younger adults ($ps < 0.001$). Furthermore, whispering lowered the accuracy of alveolars, retroflexes and velars ($ps < 0.05$). Regarding RTs (Figure 2B), there were significant main

effects of group [$\chi^2_{(1)} = 44.124, p < 0.001$], condition [$\chi^2_{(1)} = 33.441, p < 0.001$] and place [$\chi^2_{(4)} = 61.944, p < 0.001$]. The interaction effects of group \times place [$\chi^2_{(3)} = 44.300, p < 0.001$], and group \times condition \times place [$\chi^2_{(9)} = 57.368, p < 0.001$] were also significant. The *post-hoc* tests showed that older adults employed longer time in recognizing all consonants than younger adults under both phonated and whispered conditions. Among different consonants, the RT used in recognizing whispered labials was much longer than other types of consonants for older adults.

Figure 3A demonstrates the identification accuracy of consonants with different manners of articulation in the two listener groups and articulatory modes. There were significant main effects of group [$\chi^2_{(1)} = 70.584, p < 0.001$], condition [$\chi^2_{(1)} = 41.720, p < 0.001$], and manner [$\chi^2_{(4)} = 156.307, p < 0.001$]. The interaction effects of group \times manner [$\chi^2_{(4)} =$

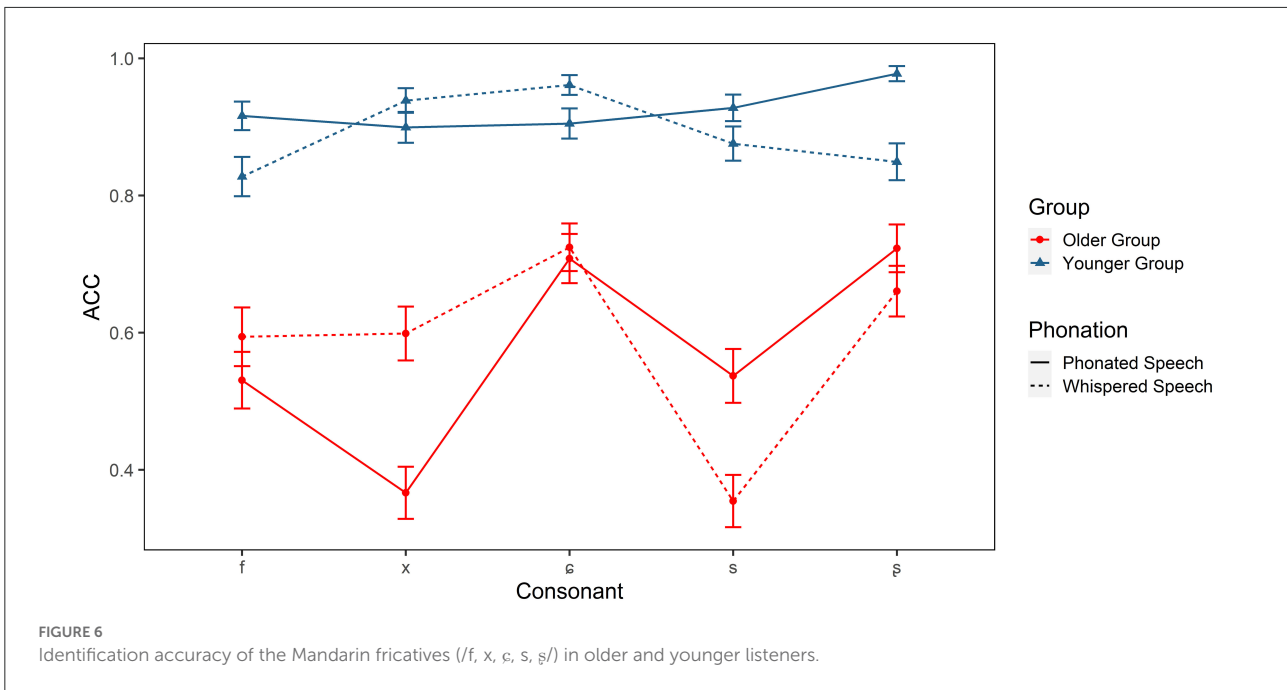
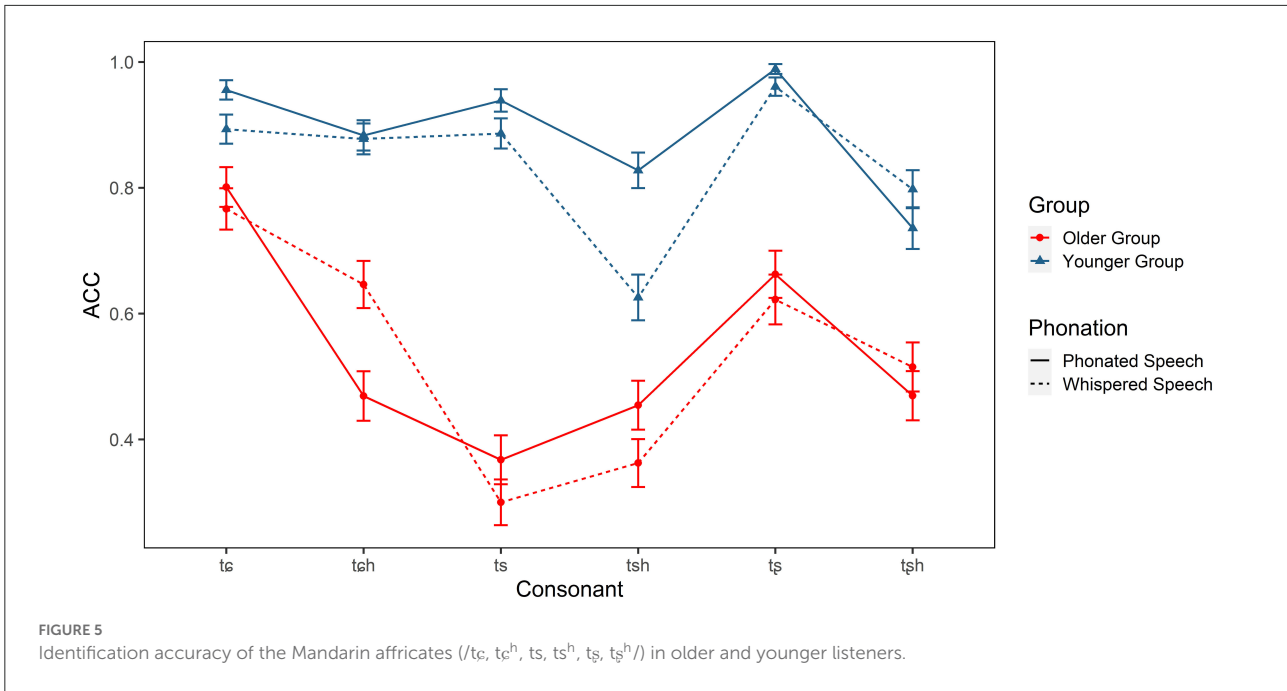


76.300, $p < 0.001$], group \times condition [$\chi^2_{(1)} = 9.655, p < 0.001$], and condition \times manner [$\chi^2_{(4)} = 94.525, p < 0.001$] were also significant. *Post-hoc* tests revealed that older adults had lower recognition accuracy than younger adults in two conditions ($ps < 0.001$). In addition, the *post-hoc* tests also indicated that two listener groups showed different perceptual patterns. For younger adults, the accuracy for A/LA was highest among all types of consonants ($ps < 0.001$) and the accuracy for nasals was higher than plosives and affricatives ($ps < 0.05$). As for fricatives and affricatives, affricatives seemed more difficult than fricatives ($p < 0.001$). For older adults, similar to younger adults, the accuracy for A/LA was highest ($ps < 0.001$). The accuracy for nasals was also higher than fricatives and affricatives ($ps < 0.05$). As for fricatives and affricatives, affricatives were also more challenging than fricatives ($p < 0.001$). The most notable difference between the two listener groups was that older adults performed less accurately on the identification of affricatives and fricatives than other types of consonants ($ps < 0.001$). In terms of the effect of whispering, it lowered the accuracies of nasals and A/LA for older adults ($ps < 0.001$). Affricative identification was also negatively affected ($ps < 0.001$) in addition to nasals and A/LA for younger adults. In terms of RTs (Figure 3B), there was a main effect of group [$\chi^2_{(1)} = 44.124, p < 0.001$], with the RT longer in the older listener group. There was also a main effect of condition [$\chi^2_{(1)} = 33.441, p < 0.001$], in that RT used in recognizing whispered consonants was significantly longer. Lastly, there was a main effect of manner of articulation [$\chi^2_{(4)} = 200.310, p < 0.001$], showing that the RT used in recognizing nasals was shortest. No other effects were significant.

In order to examine the perceptual patterns in a more detailed manner, linear mixed effects models were further constructed for plosives, affricatives, and fricatives separately. Group, condition, and consonant (plosives: /p, p^h, t, t^h, k, k^h/; affricatives: /tʃ, tʃ^h, ts, ts^h, tʂ, tʂ^h/; fricatives: /f, x, ç, s, ʃ/), and their interactions were treated as fixed effects and subject was treated as a random effect. Figures 4–6 show the results.

For plosives, there were significant main effects of group [$\chi^2_{(1)} = 21.551, p < 0.001$] and consonant [$\chi^2_{(5)} = 39.504, p < 0.001$]. The interaction effects of group \times consonant [$\chi^2_{(5)} = 14.447, p < 0.05$], condition \times consonant [$\chi^2_{(5)} = 75.022, p < 0.05$], and group \times condition \times consonant [$\chi^2_{(5)} = 13.253, p < 0.05$] were also significant. *Post hoc* analyses showed that the accuracy for /p^h, k^h/ was significantly lower in older adults ($ps < 0.05$) but the accuracy for /p, t, k/ was comparable between the two groups ($ps > 0.05$) in phoned condition. In whispered condition, older adults identify /t, k, p^h/ less accurately than younger adults ($ps < 0.001$). As for the effect of whispering, it lowered the identification accuracy of /k^h/ in two groups ($ps < 0.05$) but it facilitated the recognition of /p^h/ for both groups ($ps < 0.05$) which was contrary to our expectations.

For affricatives, there were significant main effects of group [$\chi^2_{(1)} = 71.827, p < 0.001$] and consonant [$\chi^2_{(5)} = 213.502, p < 0.001$]. The interaction effects of group \times condition [$\chi^2_{(1)} = 7.734, p < 0.01$], group \times consonant [$\chi^2_{(5)} = 70.927, p < 0.001$], and condition \times consonant [$\chi^2_{(5)} = 34.056, p < 0.001$] were also significant. *Post-hoc* tests indicated that older adults identified all affricatives less accurately than younger adults in both conditions ($ps < 0.001$). Moreover, whispering resulted in



a lower accuracy score of /ts, tsʰ, tɕ/, ($p < 0.05$), but a higher score of /tɕʰ/.

For fricatives, there were significant main effects of group [$\chi^2_{(1)} = 70.886, p < 0.001$] and consonant [$\chi^2_{(4)} = 60.053, p < 0.001$]. The interaction effects of group \times condition [$\chi^2_{(1)} = 4.651, p < 0.05$], group \times consonant [$\chi^2_{(4)} = 23.084, p < 0.001$], condition \times consonant [$\chi^2_{(4)} = 44.665, p < 0.001$], and

group \times condition \times consonant [$\chi^2_{(4)} = 16.520, p < 0.01$] were also significant. *Post-hoc* analyses showed that older adults had difficulties in recognizing all fricatives in two conditions ($ps < 0.01$). Fricatives /f, x, s/ were more challenging for older adults ($ps < 0.01$) compared with /ɕ, ʃ/ in the phonated condition, while /s/ was most difficult to be identified in whispered condition ($ps < 0.001$). Whispering especially lowered the

TABLE 1 Most frequently confused pairs and their error percentage for two groups in two articulatory modes.

Phonated speech				Whispered speech					
Younger group		Older group		Younger group		Older group			
p-t	20%	p-t	12.78%	t-k	10.56%	p-t	10%	p-p ^h	14.44%
p ^h -t ^h	20%	p ^h -t ^h	36%	t ^h -k ^h	8.89%	p ^h -p	17.50%	p ^h -t ^h	11.67%
t _s ^h -t _s ^h	18.33%	k ^h -t ^h	16.67%	k-k ^h	14.44%	t-p	10%	t-t ^h	11.67%
ts-s	11.11%	x-t ^h	15.56%	k ^h -k	18.89%	k-k ^h	31.11%		
t _c ^h -c	10%	x-k ^h	13.89%	ts-t _s	8.33%	k ^h -k	31.67%		
		x-p ^h	11.67%	ts ^h -s	23.89%	t _c -t _c ^h	10.56%		
		t _c ^h -c	39%	ts ^h -ts	8.33%	t _c ^h -t _c	13.89%	t _c ^h -c	14.44%
		c-t _c	20.56%	t _s ^h -t _s	15.56%	c-t _c ^h	8.33%		
		ts-t _c	25.56%	c-z _c	11.67%	ts-t _s	37.22%		
		ts-z _c	13.89%	m-n	8.33%	ts ^h -t _s ^h	21.11%	ts ^h -s	18.89%
		ts ^h -t _s ^h	24.24%	z _c - _s	7.22%	s- _s	19.44%	s-t _s ^h	11.67%
		s- _s	25.56%			t _s -ts	22.22%	t _s -t _s ^h	11.11%
		t _s -ts	23.89%			t _s ^h -t _s	20%	t _s ^h -ts	16.67%
		t _s ^h -ts ^h	18.33%			_s -s	23.89%		
		_s -s	15.56%			f-p ^h	11.67%		
		n-m	21.11%			m-n	10%		
						n-m	10%	n-z _c	10%
						l-n	10%		
						z _c - _s	21.67%		

accuracy of /s/ for younger adults and /x/ for older adults. However, the accuracy of /x/ was increased in the whispered condition in older adults.

Lastly, to rule out the possibility that hearing loss may also have played a role in the results, additional generalized linear mixed-effects models were fitted. It revealed that there were no significant differences between the two groups of older adults (with normal hearing and mild hearing loss in high-frequency) for all performances mentioned above (*ps* > 0.05).

We also analyzed the perceptual error patterns. Table 1 demonstrates the most frequent confusion pairs and their error percentage for younger and older adults in two articulatory modes. In phonated condition, younger adults' confusion was within bilabial plosives (/p-t/, /p^h-t^h/), aspirated affricatives (/t_s^h-t_s^h/), as well as affricatives and fricatives with the same place of articulation (/ts-s/, /t_c^h-c/). Compared with younger adults, older adults demonstrated more confusing pairs. They were more confused about aspirated bilabial plosives /p^h-t^h/. In addition, pairs of affricatives and fricatives distinguished by high-frequency information seemed confusing to them too. For example, the most frequent confusion pairs of /ts-t_s/, /s-_s/, /ts^h-t_s^h/ and /t_c^h-c/ were observed in older adults. In whispered condition, both listener groups exhibited more confusion pairs. Differentiating aspirated and unaspirated consonants, such as the pair /p-p^h/, /t-t^h/, /k-k^h/, /t_c-t_c^h/, /t_s-t_s^h/ was challenging for older adults, implying that processing temporal cues was more difficult in whispered condition for them. In addition,

consonants with high-frequency information were also difficult to distinguish for older adults too, especially for /ts, t_s^h, s, t_s, t_s^h, _s/. Lastly, distinguishing voiced consonants, including nasals (/m-n/) and A/LA (/l-z/) became more challenging for both groups.

Discussion

There are many potential factors contributing to consonant identification, including listeners' speech processing ability and cognition ability, acoustic characteristics and the phonological structure of consonants, whether it is presented in a degraded mode, and the difficulty of tasks etc. The current study aims to investigate the effects of aging and whispering on Mandarin consonant identification. Our hypothesis was that cognitive decline and aging-related speech processing deficits, such as the insensitivity to high-frequency information and the deteriorating temporal and spectral processing ability, would adversely affect older adults' consonant identification in normal condition. Moreover, whispering would further worsen the consonant identification in older adults as the acoustic differences became less prominent in whispered speech. It is noteworthy that in the current study, the intensity of phonated and whispered syllables was scaled to the same level (60 dB) to ensure speech audibility, which might reduce the intensity differences between phonated and whispered consonants.

task. This finding of age-related processing speed decline is also consistent with previous studies showing that older adults are associated with general slower speed in both speech and cognitive tasks (Dey and Sommers, 2015).

The effect of whispering on consonant perception

Findings of the current study revealed that the identification of nasals and A/LA was compromised in whispered condition in both groups, especially for older adults, whereas the accuracy of voiceless consonants was less affected. It was also found that RT in the whispered condition was significantly longer for most consonants. The absence of vocal fold vibration in whispering turns voiced consonants to voiceless. The turbulent flow in whispering also distorted the resonance which results in degraded formant information. Furthermore, whispering lowered the intensity of consonants (Jovičić and Šarić, 2008). Several studies have demonstrated that reduced-intensity compromised speech perception (Gordon-Salant, 1986; Nábělek et al., 1996; Kennedy et al., 1998), especially in less ideal conditions. For example, Nábělek et al. (1996) found that the reduction of intensity ranging between 1 and 15 dB affected the diphthong perceptions, particularly for listeners with hearing loss and in noise and reverberation conditions. We speculated that the distorted formants, low intensity collectively lead to the difficulty of voiced consonant recognition and longer RT in whispered speech. Older adults' poor performance further indicated that they may have difficulty processing degraded signals, especially the voiced information, at lower intensity.

Analyses of the matrices showed that the error pattern in the two articulatory modes was different, particularly for plosives and affricatives. In phonated condition, most of the difficulties were differentiating articulatory places, while in whispered condition, distinguishing aspirated from unaspirated consonants became also challenging, especially for older adults. Our acoustic data (unpublished) showed that whispering led to the decreased frication duration of aspirated consonants and increased frication duration of unaspirated consonants, which narrowed the duration differences between aspirated and unaspirated consonants. It is possible that this reduced duration difference, together with temporal processing deficit in older adults, leads to more errors.

Contrary to our expectation, whispering increased the identification accuracy of /p^h/ for both groups, and /x, tɕ^h/ for older adults. As mentioned above, /x/ was one of the most difficult fricatives for older adults in phonated condition, owing to the lower intensity and homogeneous energy distribution. The increased intensity in whispered condition and intensity manipulation of stimuli may also have played a role here.

Whispering is one of the challenging listening conditions in our daily life. Whether it influences speech perception differently compared to other degraded speech is less studied in the literature. In a previous study, Helfer (1994) investigated the English consonants perception (/b, d, g, p, t, k, f, v, s, z, θ, ð, m, n/) in three less ideal conditions, that is noise, reverberation, and noise plus reverberation conditions. Comparing our findings with it, we found that whispering led to different perceptual error patterns. Firstly, whispered speech detrimentally affected nasals and A/LA while the accuracy of nasals was well-maintained in all the three degraded conditions in the previous study. Secondly, whispering resulted in new error patterns while the error patterns in other three less ideal conditions were similar to the normal condition. For example, new error patterns (/k-k^h/, /ts-ts^h/, /tʃ-tʃ^h/) emerged in whispered condition only. On the contrary, distinguishing articulatory places (/b-g/, /p-t-k/) was most difficult in both normal and other three degraded conditions. The possible reason is that speech perception difficulties in noise and reverberation are caused by these additional interferences (i.e., noise) to target phonemes, while the difficulties in whispering are attributed to the changed quality of targets. Lastly, the previous study found that some fricatives with weaker energy were more likely to be masked by noise while relatively resistant to reverberation. In contrast, our study found that certain fricative, i.e., /x/, became easier in whispering. These results suggested that the underlying mechanism under whispered speech perception might be different from other degraded conditions, owing to its special phonation nature.

Limitations

In the current study, to make the two conditions comparable, the phonated and whispered syllables were scaled to having the same duration and intensity. Some of their original/natural cues and the difference between the two articulatory modes may have been changed due to these manipulations. Stimuli without any manipulation could be used to retest the effects of whispering on consonant identification in future studies. Another question is that tokens from only one female speaker were used as stimuli. Previous studies of consonant perception showed response to the same phoneme could be variable due to various factors, such as acoustic differences in speech, within- and across- speakers/listeners (Cutler et al., 2004; Toscano and Allen, 2014; Zaar and Dau, 2015). Further studies may increase the stimulus complexity by including more talkers to better understand the effects of aging and whispering on consonant perception in Mandarin. Lastly, we have administered cognitive tests to exclude those with cognition impairment. We have also provided adequate practice sessions and did not set a response time limit in order

to reduce cognitive load. However, the 21-alternative forced-choice task might still be challenging for older adults. Therefore, declined cognitive ability could not be completely ruled out as one of the possible reasons older adults' lower accuracy and longer RT in consonant identification.

Conclusion

The current study aims to explore whether and how aging and whispering affect consonant identification. Considering older adults' cognition declining, deteriorated sensitivity to high-frequency information, poor ability to process temporal and spectral cues and other listening deficits documented in previous literature, we hypothesized that the consonant identification ability in the aging population might decline, even in the quiet condition. The results confirmed our hypothesis and revealed that the difficulty in different consonants was uneven. The recognition of affricatives and fricatives based on decoding high-frequency information is most challenging in the elderly population even for those with normal hearing condition. Identifying the place of articulation in affricatives and plosives was more difficult as well. We also hypothesized that whispering may further affect consonant identification, particularly for older adults. The results showed that the degraded formant information and lower intensity in whispering adversely affected the identification of voiced consonants (nasals and L/LA). Moreover, the duration differences between aspirated and unaspirated consonants decreased in the whispered condition which resulted in difficulties in distinguishing these two types of consonants. Contrary to our expectations, whispering facilitated the identification of certain consonants, such as /p^h, t^h, x/. Our study increased the understanding of how aging affects speech perception in Mandarin which has unique phonological and acoustic features. Findings on whispered speech perception also enriched our knowledge of speech perception under different less ideal conditions, which in turn may assist us in developing effective speech rehabilitation strategies for older adults.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

and accession number(s) can be found below: <https://osf.io/rhwqv/>.

Ethics statement

The studies involving human participants were reviewed and approved by Human Subjects Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MX and JS contributed to conception and design of the study, and prepared the manuscript. MX collected data and performed the statistical analysis. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by grants from the National Natural Science Foundation of China (NSFC: 11904381), the National Key R&D Program of China (2020YFC2004100), and the Start-up Grant from Hong Kong Baptist University (162646).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alwan, A., Jiang, J., and Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. *Speech Commun.* 53, 195–209. doi: 10.1016/j.specom.2010.09.001

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

- Bidelman, G. M., Villafuerte, J. W., Moreno, S., and Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol. Aging* 35, 2526–2540. doi: 10.1016/j.neurobiolaging.2014.05.006
- Brandy, W. T. (2002). "Speech audiometry," in *Handbook of Clinical Audiology*, ed J. Katz (Baltimore, MD: Lippincott Williams and Wilkins), 96–110.
- Cooper, N. P., and Rhode, W. S. (1997). Mechanical responses to two-tone distortion products in the apical and basal turns of the Mammalian Cochlea. *J. Neurophysiol.* 78, 261–270.
- Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *J. Acoust. Soc. Am.* 116, 3668–3678. doi: 10.1121/1.1810292
- Dannenbring, G. L. (1980). Perceptual discrimination of whispered phoneme pairs. *Percept. Motor Skills* 51(3 Pt 1), 979–985. doi: 10.2466/pms.1980.51.3.979
- Dey, A., and Sommers, M. S. (2015). Age-related differences in inhibitory control predict audiovisual speech perception. *Psychol. Aging* 30, 634. doi: 10.1037/pag0000033
- Dorman, M. F., Marton, K., Hannley, M. T., and Lindholm, J. M. (1985). Phonetic identification by elderly normal and hearing-impaired listeners. *J. Acoust. Soc. Am.* 77, 664–670. doi: 10.1121/1.391885
- Fan, X., and Hansen, J. H. L. (2011). Speaker identification within whispered speech audio streams. *IEEE Trans. Audio Speech Lang. Process.* 19, 1408–1421. doi: 10.1109/TASL.2010.2091631
- Feng, Y., Meng, Y., Li, H., and Peng, G. (2021). Effects of cognitive load on the categorical perception of mandarin tones. *J. Speech Lang. Hear. Res.* 64, 3794–3802. doi: 10.1044/2021_JSLHR-20-00695
- Feng, Y., Meng, Y., and Peng, G. (2019). "The categorical perception of mandarin tones in normal aging seniors and seniors with mild cognitive impairment," in *ICPhS 2019-19th International Congress of Phonetic Sciences* (Melbourne, VIC), 909–913.
- Ferguson, S. H., Jongman, A., Sereno, J. A., and Keum, K. A. (2010). Intelligibility of foreign-accented speech for older adults with and without hearing loss. *J. Am. Acad. Audiol.* 21, 153–162. doi: 10.3766/jaaa.21.3.3
- Fitzgibbons, P. J., and Gordon-Salant, S. (1998). Auditory temporal order perception in younger and older adults. *J. Speech Lang. Hear. Res.* 41, 1052–1060. doi: 10.1044/jslhr.4105.1052
- Fogerty, D., Bologna, W. J., Ahlstrom, J. B., and Dubno, J. R. (2017). Simultaneous and forward masking of vowels and stop consonants: effects of age, hearing loss, and spectral shaping. *J. Acoust. Soc. Am.* 141, 1133–1143. doi: 10.1121/1.4976082
- Fogerty, D., Humes, L. E., and Kewley-Port, D. (2010). Auditory temporal-order processing of vowel sequences by young and elderly listeners. *J. Acoust. Soc. Am.* 127, 2509–2520. doi: 10.1121/1.3316291
- Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. *J. Acoust. Soc. Am.* 84, 115–123.
- Gao, M. (2002). *Tones in Whispered Chinese: Articulatory Features and Perceptual Cues* (Doctoral dissertation). University of Victoria, Victoria, BC, Canada.
- Gates, G. A., and Mills, J. H. (2005). Presbycusis. *Lancet* 366, 1111–1120. doi: 10.1016/S0140-6736(05)67423-5
- Gelfand, S. A., Piper, N., and Silman, S. (1985). Consonant recognition in quiet as a function of aging among normal hearing subjects. *J. Acoust. Soc. Am.* 78, 1198–1206. doi: 10.1121/1.392888
- Ginzel, A., Pedersen, C. B., Spliid, P. E., and Andersen, E. (1982). The role of temporal factors in auditory perception of consonants and vowels: a study of different age groups. *Scand. Audiol.* 11, 93–100. doi: 10.3109/01050398209076205
- Gordon Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., and Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *J. Acoust. Soc. Am.* 119, 2455–2466. doi: 10.1121/1.2171527
- Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered cvs by young and elderly subjects with normal hearing. *J. Acoust. Soc. Am.* 80, 1599–1607. doi: 10.1121/1.394324
- Gordon-Salant, S., and Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *J. Acoust. Soc. Am.* 115, 1808–1817. doi: 10.1121/1.1645249
- Grose, J. H., Hall, J. W., and Buss, E. (2006). Temporal processing deficits in the pre-senescent auditory system. *J. Acoust. Soc. Am.* 119, 2305–2315. doi: 10.1121/1.2172169
- Harkrider, A. W., Plyler, P. N., and Hedrick, M. S. (2005). Effects of age and spectral shaping on perception and neural representation of stop consonant stimuli. *Clin. Neurophysiol.* 116, 2153–2164. doi: 10.1016/j.clinph.2005.05.016
- Harris, K. C., Mills, J. H., and Dubno, J. R. (2007). Electrophysiologic correlates of intensity discrimination in cortical evoked potentials of younger and older adults. *Hear. Res.* 228, 58–68. doi: 10.1016/j.heares.2007.01.021
- He, N., Dubno, J. R., and Mills, J. H. (1998). Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects. *J. Acoust. Soc. Am.* 103, 553–565. doi: 10.1121/1.421127
- He, N., Mills, J. H., and Dubno, J. R. (2007). Frequency modulation detection: effects of age, psychophysical method, and modulation waveform. *J. Acoust. Soc. Am.* 122, 467–477. doi: 10.1121/1.2741208
- Heeren, W., and Van Heuven, V. J. (2014). The interaction of lexical and phrasal prosody in whispered speech. *J. Acoust. Soc. Am.* 136, 3272–3289. doi: 10.1121/1.4901705
- Heeren, W. F. L., and van Heuven, V. J. (2009). "Perception and production of boundary tones in whispered Dutch," in *Proceedings of Interspeech* (Brighton: Interspeech), 2411–2414.
- Helfer, K. S. (1994). Binaural cues and consonant perception in reverberation and noise. *J. Speech Hear. Res.* 37, 429–438. doi: 10.1044/jshr.3702.429
- Helfer, K. S., and Huntley, R. A. (1991). Aging and consonant errors in reverberation and noise. *J. Acoust. Soc. Am.* 90, 1786–1796. doi: 10.1121/1.401659
- Jiao, L., and Xu, Y. (2019). Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua* 218, 24–37. doi: 10.1016/j.lingua.2018.01.004
- Jovićić, S. T. (1998). Formant feature differences between whispered and voiced sustained vowels. *Acta Acust. United Acust.* 84, 739–743.
- Jovićić, S. T., and Šarić, Z. (2008). Acoustic analysis of consonants in whispered speech. *J. Voice* 22, 263–274. doi: 10.1016/j.jvoice.2006.08.012
- Kalalah, M. K., Thomas, D., Bhat, J. S., and Ranjan, R. (2016). Perception of consonants in speech-shaped noise among young and middle-aged adults. *J. Int. Adv. Otol.* 12, 184–188. doi: 10.5152/iao.2016.2467
- Kallail, K. J., and Emanuel, F. W. (1984). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *J. Speech Hear. Res.* 27, 245–251.
- Kallail, K. J., and Emanuel, F. W. (1985). The identifiability of isolated whispered and phonated vowel samples. *J. Phonetics* 13, 11–17.
- Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (1998). Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners. *J. Acoust. Soc. Am.* 103, 1098–1114. doi: 10.1121/1.423108
- Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Malden, MA: Blackwell Publishing.
- Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J. Acoust. Soc. Am.* 59, 675–678. doi: 10.1121/1.380917
- Lenth, R. (2018). *Emmeans: Estimated marginal means, aka leastsquares means*. R Package. Available online at: <https://cran.r-project.org/web/packages/emmeans/emmeans.pdf>
- Li, S., and Gu, W. (2015). "Acoustic analysis of mandarin affricates acoustic analysis of mandarin affricates," in *Sixteenth Annual Conference of the International Speech Communication Association* (Dresden: Interspeech). doi: 10.21437/Interspeech.2015-387
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431. doi: 10.1037/h0020279
- Lieberman, A. M., Delattre, P., and Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* 65, 497–516.
- Liu, C., Xu, C., Wang, Y., Xu, L., Zhang, H., and Yang, X. (2021). Aging effect on mandarin chinese vowel and tone identification in six-talker babble. *Am. J. Audiol.* 30, 616–630. doi: 10.1044/2021_AJA-20-00139
- Matsuda, M., and Hideki, K. (1999). "Acoustic nature of the whisper," in *Sixth European Conference on Speech Communication and Technology* (Budapest), 133–136.
- Meyer-Eppler, W. (1956). Realization of prosodic features in whispered speech. *J. Acoust. Soc. Am.* 28, 760–760. doi: 10.1121/1.1905013
- Meynadier, Y., Dufour, S., and Gaydina, Y. (2013). "Duration as perceptual voicing cues in whisper," in *Proceedings of the 6th Phonetics and Phonology in Iberia Conference* (Lisbon).

- Mills, T. I. P. (2003). *Cues to Voicing Contrasts in Whispered Scottish Obstruents* (Master of Science) Edinburgh University, Edinburgh, United Kingdom.
- Mitani, S., Kitama, T., and Sato, Y. (2006). Voiceless affricate/fricative distinction by frication duration and amplitude rise slope. *J. Acoust. Soc. Am.* 120, 1600. doi: 10.1121/1.2221390
- Nábělek, A. K., Ovchinnikov, A., Czyzewski, Z., and Crowley, H. J. (1996). Cues for perception of synthetic and natural diphthongs in either noise or reverberation. *J. Acoust. Soc. Am.* 99, 1742–1753. doi: 10.1121/1.415238
- Ohde, R. N., and Abou-Khalil, R. (2001). Age differences for stop-consonant and vowel perception in adults. *J. Acoust. Soc. Am.* 110, 2156–2166. doi: 10.1121/1.1399047
- Parnell, M., Amerman, J. D., and Beverly, W. (1977). Closure and constriction duration for alveolar consonants during voiced and whispered speaking conditions. *J. Acoust. Soc. Am.* 61, 612–613. doi: 10.1121/1.381309
- Plyler, P. N., and Hedrick, M. S. (2002). Effects of stimulus presentation level on stop consonant identification in normal-hearing and hearing-impaired listeners. *J. Am. Acad. Audiol.* 13, 154–159. doi: 10.1055/s-0040-1715957
- Reinhart, P. N., Souza, P. E., Srinivasan, N. K., and Gallun, F. J. (2016). Effects of reverberation and compression on consonant identification in individuals with hearing impairment. *Ear Hear.* 37, 144. doi: 10.1097/AUD.0000000000000229
- Rishiq, D., Harkrider, A., Springer, C., and Hedrick, M. (2020). Effects of aging on the subcortical encoding of stop consonants. *Am. J. Audiol.* 29, 391–403. doi: 10.1044/2020_AJA-19-00044
- Robles, L., and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol. Rev.* 81, 1305–1352. doi: 10.1152/physrev.2001.81.3.1305
- Roosij, J. C. G. M. V., and Plomp, R. (1991). Auditive and cognitive factors in speech perception by elderly listeners. *Acta Oto Laryngol.* 111, 177–181. doi: 10.3109/00016489109127275
- Scharenborg, O., and Janse, E. (2012). “Hearing loss and the use of acoustic cues in phonetic categorisation of fricatives,” in *INTERSPEECH 2012: 13th Annual Conference of the International Speech Communication Association, vol 2*, 1458–1461. doi: 10.21437/Interspeech.2012-414
- Scharenborg, O., Weber, A., and Janse, E. (2015). Age and hearing loss and the use of acoustic cues in fricative categorization. *J. Acoust. Soc. Am.* 138, 1408–1417. doi: 10.1121/1.4927728
- Schuknecht, H. F. (1974). “Presbycusis,” in *Pathology of the Ear*, ed H. F. Schuknecht (Cambridge, MA: Harvard University Press).
- Schwartz, M. F. (1972). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *J. Acoust. Soc. Am.* 51, 2025–2029.
- Smith, I., Foulkes, P., and Sóskuthy, M. (2017). Speaker identification in whisper. *Letras de Hoje* 52, 5–14. doi: 10.15448/1984-7726.2017.1.26659
- Souza, P., Arehart, K., Miller, C. W., and Muralimanohar, R. K. (2011). Effects of age on f0 discrimination and intonation perception in simulated electric and electroacoustic hearing. *Ear Hear.* 32, 75–83. doi: 10.1097/AUD.0b013e3181eccfe9
- Stevens, K. N., and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.* 55, 653–659. doi: 10.1121/1.1914578
- Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (1998). Temporal processing in the aging auditory system. *J. Acoust. Soc. Am.* 104, 2385–2399. doi: 10.1121/1.423748
- Sundberg, J. (1994). Vocal fold vibration patterns and phonatory modes. *STL-QPSR* 35, 69–80.
- Tartter, V. C. (1989). What’s in a whisper? *J. Acoust. Soc. Am.* 86, 1678–1683. doi: 10.1121/1.398598
- Toscanoa, J. C., and Allen, J. B. (2014). Across- and within-consonant errors for isolated syllables in noise. *J. Speech Lang. Hear. Res.* 57, 2293–2307. doi: 10.1044/2014
- Tremblay, K. L., Piskosz, M., and Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin. Neurophysiol.* 114, 1332–1343. doi: 10.1016/S1388-2457(03)00114-7
- Tsui, I. Y. H., and Ciocca, V. (2000). Perception of aspiration and place of articulation of Cantonese initial stops by normal and sensorineural hearing-impaired listeners. *Int. J. Lang. Commun. Disord.* 35, 507–525. doi: 10.1080/136828200750001269
- Walton, J. P. (2010). Timing is everything: Temporal processing deficits in the aged auditory brainstem. *Hear. Res.* 264, 63–69. doi: 10.1016/j.heares.2010.03.002
- Wang, Y., Yang, X., and Liu, C. (2017a). Categorical perception of mandarin chinese tones 1–2 and tones 1–4: effects of aging and signal duration. *J. Speech Lang. Hear. Res.* 60, 3667–3677. doi: 10.1044/2017_JSLHR-H-17-0061
- Wang, Y., Yang, X., Zhang, H., Xu, L., Xu, C., and Liu, C. (2017b). Aging effect on categorical perception of Mandarin tones 2 and 3 and thresholds of pitch contour discrimination. *Am. J. Audiol.* 26, 18–26. doi: 10.1044/2016_AJA-16-0020
- Xia, J., Xu, B., Pentony, S., Xu, J., and Swaminathan, J. (2018). Effects of reverberation and noise on speech intelligibility in normal-hearing and aided hearing-impaired listeners. *J. Acoust. Soc. Am.* 143, 1523–1533. doi: 10.1121/1.5026788
- Yang, X., Wang, Y., Xu, L., Zhang, H., Xu, C., and Liu, C. (2015). Aging effect on Mandarin Chinese vowel and tone identification. *J. Acoust. Soc. Am.* 138, EL411–EL416. doi: 10.1121/1.4933234
- Zaar, J., and Dau, T. (2015). Sources of variability in consonant perception of normal-hearing listeners. *J. Acoust. Soc. Am.* 13, 1253–1267. doi: 10.1121/1.4928142
- Zeng, F. G., and Turner, C. W. (1990). Recognition of voiceless fricatives by normal and hearing-impaired subjects. *J. Speech Hear. Res.* 33, 440–449. doi: 10.1044/jshr.3303.440
- Zhao, Z., Á, P. P. R., and Yang, A. (2005). Enhancing hearing of computer commands for the aging Chinese population by filtering the initial consonant sounds. *Int. J. Ind. Ergon.* 35, 1133–1145. doi: 10.1016/j.ergon.2005.06.010
- Zygis, M., Pape, D., Koenig, L. L., Jaskula, M., and Jesus, L. M. T. (2017). Segmental cues to intonation of statements and polar questions in whispered, semi-whispered and normal speech modes. *J. Phonetics* 63, 53–74. doi: 10.1016/j.wocn.2017.04.001