 doi:10.20944/preprints202211.0041.v1

Article

Aging-Related Decline in Phonated and Whispered Speech Perception Not Compensated for by Increased Duration and Intensity: Evidence from Mandarin-Speaking Adult Listeners

Min Xu[1], Jing Shao[2*], Boquan Liu[3], Lan Wang[4], Hongwei Ding[5] and Yang Zhang[6*]

[1]Institute of Corpus Studies and Applications, Shanghai International Studies University, Shanghai, China
[2]Department of English Language and Literature, Hong Kong Baptist University, Hong Kong SAR, China
[3]School of Humanities, Shanghai Jiao Tong University, Shanghai, China
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[5]Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
[6]Department of Speech-Language-Hearing Sciences & Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, USA
* Correspondence: jingshao@hkbu.edu.hk(J.S.); zhanglab@umn.edu(Y.Z.)

Abstract

Purpose: This study aimed to examine how aging and modifications of critical acoustic parameters may affect the perception of whispered speech as a degraded signal.

Method: Forty Mandarin-speaking adults were included in the study. Part 1 of the study compared the perception of Mandarin lexical tones, vowels, and syllables in older and younger adults in whispered vs. phonated speech conditions. Parts 2 and 3 further examined how modification of duration and intensity cues contributed to the perceptual outcomes.

Results: Perception of whispered tones was compromised in older and younger adults. Older adults identified lexical tones less accurately than their younger counterparts, particularly for phonated T2, T3 and whispered T3. Aging also negatively affected the vowel identification of /i, u/ in the whispered condition. Syllable-level accuracy was largely dependent on the accuracy of lexical tones and vowels. Furthermore, reduced duration led to the decreased accuracy of phonated T3 and whispered T2, T3 but increased accuracy of phonated T4. Reduced intensity lowered the recognition accuracy for phonated vowels /i, ɤ, o, y/ in older adults and /i, u/ in younger adults, and it also lowered the accuracy of whispered vowels /a, ɤ/ in older adults. Contrary to our expectation, increased duration and intensity did not improve older adults' speech perception in either phonated or whispered conditions.

Conclusion: The results suggest that aging adversely affected speech perception in both phonated and whispered conditions with more challenges in identifying whispered speech for older adults. While older adults' diminished performance may be potentially due to problems with processing the degraded temporal and spectral information of the target speech sounds, it cannot be simply compensated for by increasing the duration and intensity of the target sounds beyond the audible level.

Keywords: older adults; whispered speech; lexical tone; vowel; duration; intensity.

1. Introduction

Aging has a negative impact on speech comprehension, especially for those who suffer from age-related hearing loss. Research studies have documented substantial age-related decline in perceiving speech segments (Gordon et al., 2006), words (Pichora-Fuller, 2008), and sentences (Helfer & Freyman, 2008). This can be attributed to the deterioration of peripheral, central-auditory, and cognitive systems (Humes & Dubno, 2010) with specific contributions from temporal and spectral processing deficits (Bidelman et al., 2014; Gordon et al., 2011; Gordon & Fitzgibbons, 1993; Kolodziejczyk & Szelag, 2008; Smith et al., 2012).

As the literature primarily focused on English-speaking participants, it remains unclear whether specific difficulties could arise from the unique linguistic structures and features in different languages. For the present study, our target language was Mandarin Chinese. Unlike English, Mandarin Chinese is a tonal language with four lexical tones, T1 (/mā/, "mother"), T2 (/má/, "hemp"), T3 (/mǎ/, "horse"), and T4 (/mà/, "to scold"), whose acoustic cues are mainly characterized by the fundamental frequency (F0), duration and intensity. Acoustically, T1 and T4 are realized as high pitch contours and falling pitch contours with high onsets. T2 and T3 are realized as rising and low rising pitch contours with low pitch at mid position. The amplitude contour corresponds well with the pitch contour. In terms of duration,

 doi:10.20944/preprints202211.0041.v1

T3 was the longest, followed by T2, T1 and T4 (Ho, 1976). It is generally accepted that F0 is the most salient cue but duration and amplitude contour are also contributed to lexical tone perception (Blicher et al., 1990; Fu & Zeng, 2000; Whalen & Xu, 1992). The vowel system in Mandarin is also different from English in that it has only six monophthongs (/a/, /ɤ/, /i/, /o/, /u/, /y/), which may pose less perceptual challenge for the elderly.

To date, studies about Mandarin speech perception in the aging population have mainly focused on lexical tone processing and reported mixed findings. Regarding the lexical tone perception in phonated speech, Feng et al. (2019, 2020) did not find that older adults with normal cognitive abilities had declined categorical perception of T1-T2, whereas Wang, Yang, and Liu (2017) reported the age group differences using the same paradigm. Moreover, Wang, Yang, and Liu (2017) found longer duration (100 ms, 200 ms, and 400 ms) facilitated categorical perception of T1-T2 in older adults but not in younger adults. Apart from the examination of aging effects on categorical perception of lexical tones, researchers also explored the aging effects on Mandarin syllable perception in quiet (Yang et al., 2015) and noise conditions (Liu et al., 2021) or sentence intelligibility with flat contour (Jiang et al., 2017). The results showed that less ideal conditions adversely affected perceptual outcomes at both syllable and sentence levels.

One common challenge for the aging population is to understand speech in adverse listening conditions (Gordon et al., 2007; Gordon et al., 2014; Helfer & Wilber, 1990; Moore et al., 2014; Wong et al., 2009). For example, Fostick et al. (2013) found that older adults showed comparable word perception with younger adults in quiet conditions but performed poorly in time-compressed speech or listening-in-noise conditions. Whispered speech is a natural mode of communication that involves signal degradation with a lack of fundamental frequency (F0) and harmonic structure, which could result in perception difficulties for the elderly. To the best of our knowledge, it remains underexplored how aging affects the perception of whispered speech.

 doi:10.20944/preprints202211.0041.v1

Unlike phonated speech which is characterized by the quasiperiodic pulses of airflow through the opening and closing of vocal folds, whispered speech is sourced from aperiodic noise owing to the opening vocal folds (Solomon et al., 1989). Apart from the F0 loss, several other acoustic modifications also occur, including slower speaking rates as measured by lengthened vowels and consonants (Heeren & Heuven, 2009; Jovičić & Šarić, 2008; Schwartz, 1967; Sharf, 1964), lower energy (Ito et al., 2005; Jovičić & Šarić, 2008), and upward shifts in formant frequencies (Eklund & Traunmüller, 1997; Ito et al., 2005; Jovičić, 1998; Kallail & Emanuel, 1984a, 1984b; Li & Guo, 2012; Sharifzadeh et al., 2012).

A few studies have investigated the perception of lexical tones in whispered speech in younger adult listeners. Findings are mixed regarding whether the absence of critical cue, F0, affects the lexical tone perception in whispered speech. In an earlier report, Jensen (1958) observed that native listeners of Norwegian, Swedish, Slovenian, and Mandarin did not have a dramatic decline in lexical tone perception in their native language. For instance, the accuracy of Mandarin tone perception in whispered speech was up to 88%. However, this study included only two Mandarin listeners and required listeners to compare two lexical tones in each trial, and thus the reliability of the results is questionable. Reduced perception accuracy of whispered Mandarin lexical tone was reported by Jiao and Xu, (2019) and Gao (2002), with T3 and T4 maintaining relatively higher accuracy and the accuracy of T1 and T2 decreasing to the chance level. Another similar study also showed that young Thai listeners performed 21-46% lower in the whispered mode than in the phonated mode when identifying Thai lexical tones (Arthur, 1972).

In assessing the perception of whispered vowels, Kallail & Emanuel (1985) reported the perception accuracy of whispered vowels (60%) was lower than in the phonated mode (80%). This lower accuracy was explained by the unusual acoustic cues, such as declined intensity and lifted formants as well as listeners' unfamiliarity with whispered speech.

 doi:10.20944/preprints202211.0041.v1

It is often observed that people are asked to speak slower or louder by older adults in our daily life. A number of studies have demonstrated the relationship between signal enhancement in duration or intensity and speech intelligibility in the aging population. Many studies on the perception of time-compressed sentences found that older listeners had particular difficulties in recognizing speech presented at a fast speed (Gordon & Fitzgibbons, 1993; Vaughan & Letowski, 1997; Konkle et al., 1977; Peelle & Wingfield, 2005). For instance, Gordon and Fitzgibbons (1993) adopted four compression ratios at 30%, 40%, 50%, and 60%, and found that speech recognition performance in the elderly started declining from the 40% time-compressed ratio with a marked drop occurring in the 60% time-compressed condition. Similarly, manipulating the duration cue in phonated speech has also been shown to affect older adults' lexical tone perception (Wang, Yang, & Liu, 2017). However, it remains unclear how duration modification may affect whispered Mandarin lexical tone or vowel perception in the aging population.

Intensity or sound presentation level plays an important role in speech perception (Gordon, 1986; Nábělek et al., 1996; Price & Simon, 1984; Konkle et al., 1977), although it may not appear as critical in normal audible conditions as other cues such as the fundamental frequency for lexical tones and formants for vowels. Nábělek et al. (1996) found the high intensity transition may facilitate the perception of diphthongs in older adults with hearing loss, especially in noise and reverberation. In addition, Price and Simon (1984) found interactions between aging, temporal factors and intensity. Specifically, the change from 80 to 100 dB SPL resulted in younger adults needing a shorter silent closure duration to discriminate "rapid" from "rabid" than older adults, suggesting that older adults might benefit less from intensity increments than younger adults. Based on these findings, we hypothesized that increasing and decreasing stimulus intensity would have differential influences on the perceptual outcomes of phonated and whispered speech and that the effects would be more noticeable in the older adults than the younger adults.

 doi:10.20944/preprints202211.0041.v1

In the current study, we designed an experiment with three parts to examine the identification of phonated and whispered syllables in older and younger adults. Part 1 required listeners to identify the whole CV syllable consisting of a vowel and a lexical tone. We predicted that aging might adversely affect speech perception, especially in the whispered condition. In Parts 2 and 3, we respectively examined the effects of duration and intensity modifications on the perceptual outcomes. Listeners did the same task as in Part 1 with the duration- or intensity-manipulated stimuli. It was expected that the reduction of duration or intensity would exert a negative effect on speech perception for older listeners and the increase of duration or intensity could facilitate older adults' speech perception. The findings would enrich our understanding of whispered speech perception in older adults, and further shed light on how compensatory strategies such as duration and intensity exaggeration may affect their speech processing.

2. Method

2.1. Participants

Two subject groups participated in the current study. They were all native Mandarin speakers from the northern China. The older adult group contained 20 participants (16 females, 4 males) with an age range of 59-72 years (Mean = 64.05, SD = 4.36). The younger adult group included 20 individuals (8 females, 12 males), ranging from 21 to 32 years old (Mean = 24.20, SD = 2.86). The younger adults were students at Shenzhen Institute of Advanced Technology and were recruited by posted advertisements. The older adults were local residents recruited with the help of community managers.

All subjects were right-handed and had used computers in their daily life. The self-reports showed that they had no history of neurological conditions or ear diseases. All participants' cognitive conditions were screened by the Montreal Cognitive Assessment-Basic (MoCA-B) Chinese Version (https://www.mocatest.org/). The MoCA-B is a 15

 doi:10.20944/preprints202211.0041.v1

min examination that consists of 10 subtests: executive, memory, fluency, orientation, calculation, abstraction, delay recall, visual, naming, and attention. The total score of MoCA-B is 30, and a score less than 26 is defined as cognitive impairment. All participants obtained a score higher than 26 (Mean = 27.75, SD = 1.52).

Hearing sensitivity was assessed by air-conduction audiometry from 250 to 8000 Hz using a portable audiometer (GSI 18). Clinically normal hearing thresholds were defined as having a pure-tone average between 250 Hz and 4000 Hz better than 25 dB HL (Goossens et al., 2017). Younger subjects had thresholds under 25 dB HL between 250 Hz and 8000 Hz. Older subjects had thresholds under 25 dB HL between 250 and 4000 Hz, and their thresholds were below 40 dB at 8000 Hz. All listeners' threshold differences between right and left ears were smaller than 15 dB at each frequency. The audiograms of the younger and older participants are shown in Figure 1. All participants were paid 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.

2.2. Stimuli

The speech stimuli were 24 Mandarin monosyllables with the combinations of six vowels (/a/, /ɤ/, /i/, /o/, /u/, /y/) and four lexical tones (T1, T2, T3, T4). The syllables were recorded in a sound-attenuated laboratory using a headset microphone (Sennheiser PC8) and were sampled at a rate of 44100 Hz with a 16-bit resolution. The speaker was a female adult (aged 29) from northern China, who spoke standard Mandarin Chinese. Each syllable was repeated three times to ensure the target syllables were produced as accurately and naturally as possible. For whispered syllables, the speaker was instructed to produce them in a way that was similar to whispering in a public library. A clearest token was selected by the first author. Another five native listeners without prior training in linguistics then evaluated the

 doi:10.20944/preprints202211.0041.v1

chosen tokens. The recorded stimuli and a hard copy with 24 selected target stimuli were provided to evaluators. They were asked to evaluate the auditory stimuli with two criteria. The first and most important one was whether the monosyllables were pronounced correctly (both vowels and lexical tones). The second was whether the monosyllable sounded natural. The stimuli could be listened to repeatedly. The incorrect/unnatural stimuli were replaced by re-recorded tokens and were reevaluated by the same evaluators until they all met the criteria.

The phonated syllables had a mean duration of 530 ms (SD = 99 ms) at 57.70 dB (SD = 3.05). The whispered syllables were of a mean duration of 535 ms (SD = 58 ms) with mean intensity at 44.41 dB (SD = 4.50). In order to better investigate the effect of F0 absence and further examine the differential effects of duration and intensity modifications, the original stimuli were scaled to 500 ms and 60 dB as the baseline condition in Part 1 of the study. These baseline parameters were based on the estimate of mean duration and intensity of original stimuli, as well as consulting the results from a previous study (Jiao & Xu, 2019).

In Part 2, two other duration variants were introduced while keeping the intensity constant. The time-compressed condition was 30% of the baseline length, and the time-expanded condition was 150% of the baseline. The duration-modified stimuli were also normalized in mean intensity and presented at 60 dB SPL. In Part 3, we manipulated the intensity of the target syllables to produce two variants, 40 dB and 75 dB, with a fixed duration of 500 ms. These stimuli were presented at corresponding sound pressure levels, 40 dB SPL and 75 dB SPL. The intensity-reduced condition was 20 dB SPL lower than the baseline. According to our pilot study, 40 dB SPL was the minimum level at which all older adults could discern syllables in the phonated and whispered conditions. All stimuli manipulations were conducted with the Praat software. Duration was controlled with the Pitch Synchronous Overlap Add (PSOLA) technique (Boersma & Weenink, 2018). Stimuli were played through Sennheiser 280 headphones binaurally. The sound presentation level was measured by a Rion sound-level meter (Model NL-21) with a linear weighting band.

 doi:10.20944/preprints202211.0041.v1

2.3. Procedure

E-prime 3.0 was used to present stimuli and collect responses. The task was a 24-alternative forced-choice identification without feedback in the formal sessions. The syllables were presented in Pinyin forms /ā, á, ǎ, à, ē, é, ě, è, ī, í, ǐ, ì, ō, ó, ǒ, ò, ū, ú, ǔ, ù, ǖ, ǘ, ǚ, ǜ/. Auditory stimuli were presented via Sennheiser 280 headphones binaurally to listeners. Participants were presented with a syllable in each trial, and 24-response alternatives corresponding with each vowel-plus-tone were presented simultaneously on the computer screen. The participants were then required to choose which of the 24 syllables they had just heard by clicking the corresponding response on the screen as soon as possible.

There were a total of 10 presentation blocks, which were divided into three parts. Each block contained 24 trials, and each trial was repeated three times, giving a total of 72 trials. Part 1 consisted of two blocks (phonated speech and whispered speech) with the baseline stimuli (600 ms and 60 dB). Part 2 included four blocks with duration-modified stimuli (phonated speech and whispered speech: 150 ms, 60 dB SPL; 750 ms, 60 dB SPL). Part 3 also had four blocks with intensity-modified stimuli (phonated speech and whispered speech: 500 ms, 40 dB SPL; 500 ms, 75 dB SPL). The order of blocks was counterbalanced within three parts. Each block cost 6-10 minutes for older adults, the whole experiment lasted approximately 1.5 h. Short breaks were available between conditions and after every 40 trials within each condition.

Before the formal test session, the participants were given sufficient time to familiarize themselves with the stimuli and experimental procedures. First, the participants were asked to read out monosyllables printed on hard copies to ensure they could recognize these monosyllables correctly. Then, practice sessions which followed the same procedure as the formal experiment were conducted to make participants familiarized with the procedure. The stimuli used in the practice sessions were produced by a different female native Mandarin speaker. Feedback for each trial was given in two practice conditions. The feedback information included: (1) accuracy: if the response made by the participant was

 doi:10.20944/preprints202211.0041.v1

correct or incorrect and (2) the correct response. Participants whose accuracy in the phonated condition fell below 50% had to repeat the practice again. If their accuracy was still below 50%, the participants were not invited to take the formal test. A total of six older adults were excluded from the familiarization process. The 24 response buttons corresponding to each vowel-plus-tone (i.e., ā referred to the vowel /a/ with tone 1) were displayed on a computer monitor with a layout of six vowels (rows) by four lexical tones (columns).

2.4. Data Analysis

The response for each trial was coded as 1 or 0 (correct or incorrect) according to the target information dimension of lexical tone, vowel, and syllable (tone-plus-vowel) to calculate accuracy rates for lexical tone, vowel, and syllable identification. For example, for the stimulus /a1/, a response of /a2/ was considered incorrect for syllable identification and lexical tone identification, but correct for vowel identification, and a response of /ɤ1/ was considered incorrect for syllable identification and vowel identification, but correct for lexical tone identification. The chance level was at 0.25 for lexical tone identification, 0.17 for vowel identification, and 0.04 for syllable identification. Trials with response times beyond the 500 – 7500 ms range were excluded due to inadvertent key press or attentional lapse. In total, 388 trials were excluded, accounting for 6.7% of the data.

All statistical analyses were performed with R (Version 4.0.5). A series of generalized linear mixed-effects (GLMM) models were constructed with the package of *lme4* package (Bates et al., 2015) to elucidate the effects of aging on Chinese syllable identification in phonated and whispered conditions and whether the modifications of duration and intensity have additional effects on perception. As the duration or intensity were manipulated independently while another factor remained the same, we divided the statistical analyses into three parts rather than analyzing all factors together. In the first part, only the baseline data (500 ms, 60 dB SPL) was analyzed. In the second part, the data of the duration-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

modified stimuli (150 ms and 750 ms, 60 dB SPL) were compared with the baseline. The third part compared the intensity-modified stimuli (40 dB SPL and 75 dB SPL, 500 ms) with the baseline. The accuracies of lexical tone, vowel, and syllable identification were analyzed with separate models in each part.

Group (older group, younger group), condition (whispered speech, phonated speech), tone (T1, T2, T3, T4), vowel (/a/, /ɤ/, /i/, /o/, /u/, /y/), syllable (a1, a2, a3, a4, ɤ1, ɤ2, ɤ3, ɤ4, i1, i2, i3, i4, o1, o2, o3, o4, u1, u2, u3, u4, y1, y2, y3, y4), were independent variables which might affect the identification accuracy independently or jointly. Considering that models with too many parameters and too high performance can be a sign of overfitting, the variance inflation factor (VIF) calculated by the function *check collinearity* in *performance* packages (Lüdecke et al., 2021) was used to assess and exclude high-correlated independent variables. A full model without interactions was first constructed. If variables were highly correlated (VIF > 10), one of the variables were removed from the full model and the collinearity of the updated model was retested. The procedure repeated until the VIF of independent variables was close to 1. For lexical tone and vowel accuracy analyses, the independent variable syllable was excluded, and for syllable accuracy analyses, the independent variables tone and vowel were excluded.

For the analyses of lexical tones in Part 1 (baseline), GLMM models were fitted with group, condition, tone, and their interactions as fixed effects, as well as vowel item and subject as random effects. The tone was included as a random effect for analyzing vowel identification accuracy. In analyzing results from Parts 2 and 3 (duration and intensity modifications), duration or intensity and their interactions with other independent variables were added to GLMM models as fixed effects. A full model and a set of reduced models by excluding fixed or random factors were defined. The best-fit models were selected using the function *compare performance* in *performance* packages in R. When there were significant interactions, Bonferroni post hoc tests were applied for pairwise comparisons using the *emmeans* package (Lenth, 2018).

 doi:10.20944/preprints202211.0041.v1

3. Results

3.1. Baseline Condition Results

For lexical tone identification, there were significant main effects of condition ($\chi^2(1) = 538.365$, $p < .001$), tone ($\chi^2(3) = 98.290$, $p < .001$). The interaction effects of group × condition ($\chi^2(2) = 82.008$, $p < .001$), group × tone ($\chi^2(3) = 121.896$, $p < .001$), and condition × tone ($\chi^2(3) = 61.685$, $p < .001$) were also significant, and no other effects were significant. Post hoc tests confirmed our hypothesis that aging had negatively effects on lexical tone perception. Older adults identified all four lexical tones less accurately than younger adults (*ps* < .001). As demonstrated in Figure 2(a) and Table 1, identifying phonated T2 and T3 and whispered T3 was more difficult for older adults and they showed more confusion between these two lexical tones. Moreover, relative to the phonated speech, the identification of the four Mandarin lexical tones became more difficult in the whispered condition (*ps* < .001). Accuracy of the four lexical tones ranked from high to low was T3>T4>T1>T2 in whispered condition as shown in Figure 2(a). Older adults had more difficulty identifying whispered T3 compared with younger adults.

For vowel identification, there were significant main effects vowel ($\chi^2(5) = 32.512$, $p < .001$), as well as interaction effects of group × vowel ($\chi^2(5) = 11.401$, $p < .05$) and condition × vowel ($\chi^2(5) = 16.809$, $p < .01$). Post hoc analyses revealed that older adults made more errors in recognizing vowels /i, u/ than younger adults (*ps* < .05). As shown in Table 2 and Figure 2(b), these errors were largely from the whispered condition, and older adults particularly had difficulties in distinguishing the /o/ and /u/ pair as well as the /i/ and /y/ pair. Whispering lowered the accuracy of /i, u, y/, especially for older adults as shown in Table 2 and Figure 2(b).

For syllable identification, there were significant main effects of condition ($\chi^2(1) = 827.086$, $p < .001$) and syllable ($\chi^2(23) = 168.883$, $p < .001$). The interaction effects of group × condition ($\chi^2(2) = 54.645$, $p < .001$), group × syllable ($\chi^2(23)$

 doi:10.20944/preprints202211.0041.v1

= 144.078, $p < .001$), condition × syllable ($\chi^2(23) = 116.996$, $p < .001$) were also significant. Post hoc analyses revealed that older adults perceived /a3, ɤ2, ɤ3, ɤ4, i2, i3, o3, o4, u1, u3, y3/ less accurate than younger adults ($ps < .05$) in both phonated and whispered conditions as shown in Figure 2(c). The lowered accuracy for these syllables was mainly related to the inability to recognize the lexical tones and vowels in the elderly.

3.2. Results for duration-modified stimuli

For lexical tone identification, there were significant main effects of group (P (phonated condition): $\chi^2(1) = 32.628$, $p < .001$; W (whispered condition): $\chi^2(1) = 13.445$, $p < .001$), duration (P: $\chi^2(2) = 6.713$, $p < .05$; W: $\chi^2(2) = 129.496$, $p < .001$), tone (P: $\chi^2(3) = 84.948$, $p < .001$; W: $\chi^2(3) = 369.791$, $p < .001$). The interaction effects of group × tone (P: $\chi^2(3) = 25.885$, $p < .001$; W: $\chi^2(3) = 303.234$, $p < .001$), duration × tone (P: $\chi^2(6) = 125.585$, $p < .001$; W: $\chi^2(6) = 262.354$, $p < .001$) were also significant. Post hoc tests indicated that time compression from baseline to 150 ms lowered the accuracy of T3 but increased the accuracy of T4 in the phonated condition and it negatively affected T2 and T3 ($ps < .001$) in the whispered condition. Older adults' identifications of T3 were more affected than younger adults as demonstrated in Figure 3. Expanding time from the 500 ms baseline to 750 ms did not facilitate the lexical tone identification as we expected; instead, it lowered the identification accuracy of T2 and T4 ($ps < .05$) in the whispered condition.

For vowel identification (Figure 4), there were main effects of duration (W: $\chi^2(2) = 20.805$, $p < .001$), vowel (P: $\chi^2(5) = 142.271$, $p < .001$; W: $\chi^2(5) = 230.143$, $p < .001$). The interaction effect of group × vowel (P: $\chi^2(5) = 14.829$, $p < .05$; W: $\chi^2(5) = 30.281$, $p < .001$) was also significant. The post hoc comparison revealed that the accuracy of vowels in 150 ms was lower than in 500 ms. The lack of interaction between duration and group suggested that duration modification did not have more effects on older adults. The interaction between group and vowels confirmed that older adults identified most vowels as accurately as younger adults in phonated condition but less accurately in whispered condition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

For syllable identification, there were main effects of group (P: $\chi^2(1) = 26.010$, $p < .001$; W: $\chi^2(1) = 13.041$, $p < .001$), duration (P: $\chi^2(2) = 16.553$, $p < .001$; W: $\chi^2(2) = 112.936$, $p < .001$), and syllable (P: $\chi^2(23) = 177.319$, $p < .001$; W: $\chi^2(23) = 728.245$, $p < .001$). The interaction effects of group × syllable (P: $\chi^2(23) = 136.478$, $p < .001$; W: $\chi^2(23) = 323.466$, $p < .001$), and duration × syllable (P: $\chi^2(46) = 154.985$, $p < .001$; W: $\chi^2(46) = 336.579$, $p < .001$) were also significant. Post hoc tests showed that time compression negatively affected /a3, ɤ3, i3, u3, y3/ in phonated condition, and /a2, a3, a4, ɤ3, ɤ4, i3, o3, u3, y3/ in whispered condition. Expanding time only increased the accuracy of /ɤ3/ for older adults, and other syllables were not affected. On the contrary, it lowered the accuracy of whispered /a4, y4/ for two groups as shown in Figure 5.

3.3. Results for intensity-modified stimuli

For lexical tone identification (Figure 6), there were main effects of group (P: $\chi^2(1) = 22.591$, $p < .001$; W: $\chi^2(1) = 13.431$, $p < .001$), intensity (P: $\chi^2(2) = 6.549$, $p < .05$; W: $\chi^2(2) = 13.107$, $p < .01$) and tone (P: $\chi^2(3) = 47.735$, $p < .001$; W: $\chi^2(3) = 592.115$, $p < .001$). The interaction effects of group × tone (P: $\chi^2(3) = 29.113$, $p < .001$: W: $\chi^2(3) = 304.993$, $p < .001$), intensity × tone (W: $\chi^2(6) = 14.062$, $p < .05$), and group × intensity × tone (W: $\chi^2(6) = 23.680$, $p < .001$) were also significant. The post hoc analyses showed that the reduced intensity did not affect the identification accuracy of lexical tones in the phonated condition but it lowered the accuracy of T4 for older adults in the whispered condition ($p < .01$). However, increasing the intensity to 75 dB SPL did not result in higher accuracy of identification ($ps > .05$).

For vowel identification, there were significant main effects of intensity (W: $\chi^2(2) = 51.724$, $p < .05$) and vowel (P: $\chi^2(5) = 88.663$, $p < .001$; W: $\chi^2(5) = 300.804$, $p < .001$). The interaction effects of group × vowel (W: $\chi^2(5) = 105.127$, $p < .001$), vowel × intensity (P: $\chi^2(10) = 25.022$, $p < .01$; W: $\chi^2(10) = 63.864$, $p < .001$), group × intensity (W: $\chi^2(3) = 68.687$, $p < .001$) and group × intensity × vowel (P; $\chi^2(10) = 21.098$, $p < .05$) were also significant. Post hoc analyses showed that intensity reduction lowered the accuracy of /ɤ, i, o, y/ for older adults and /i, u/ for younger adults in phonated condition. While

 doi:10.20944/preprints202211.0041.v1

in whispered condition, vowels /a, ɤ/ were most sensitive to intensity reduction, particularly in the older adults. Similar to the duration expansion, the increase of intensity from 60 dB SPL to 75 dB SPL brought no significant increment to the accuracy of vowel identifications for two groups in both phonated and whispered condition.

For syllable identification (Figure 8), there were significant main effects of group (P: $\chi^2(1) = 23.662$, $p < .001$; W: $\chi^2(1) = 9.445$, $p < .01$), intensity (P: $\chi^2(2) = 49.095$, $p < .001$; W: $\chi^2(2) = 21.769$, $p < .001$), and syllable (P: $\chi^2(23) = 209.612$, $p < .001$; W: $\chi^2(23) = 927.946$, $p < .001$). The interaction effects of group × intensity (P: $\chi^2(2) = 6.424$, $p < .05$), group × syllable (W: $\chi^2(23) = 321.919$, $p < .001$), and intensity × syllable (P: $\chi^2(46) = 85.235$, $p < .05$) were also significant. Post hoc analyses showed that reducing intensity to 40 dB detrimentally affected the recognitions of /ɤ4, i1, i2, i3, i4, o2, o4, u1, y3/ in the phonated condition and syllable identifications in the whispered condition. The 15 dB increment of intensity did not improve syllable identification accuracy for any group in either condition.

4. General Discussion

4.1. Aging effects on phonated and whispered lexical tone perception

Our findings revealed that older adults identified lexical tones, especially for T2 and T3, less accurately than their younger counterparts. This pattern is consistent with the previous findings (Yang et al., 2015; Liu et al., 2021). Although Feng et al. (2019) failed to find the decline in categorical perception of T1-T2 in older adults with normal cognition, other researchers reported the opposite results (Wang, Yang, & Liu, 2017; Wang, Yang, Zhang, et al., 2017) with the older adults showing inferior categorical perceptions of T2-T3 and T1-T2 but comparable performance of T1-T4. These data on older adults' lexical tone perception in the phonated condition suggest that T2 and T3 are most difficult to recognize for the elderly, which can be attributed to older adults' deteriorated temporal processing ability and the characteristics of four Mandarin lexical tones.

 doi:10.20944/preprints202211.0041.v1

Existing studies have demonstrated that aging adversely affects temporal processing, even for those with normal hearing (Gordon & Fitzgibbons, 1993; Strouse et al., 1998; Gordon et al., 2006). The deficit was also confirmed by a growing number of neural studies using brainstem response (ABR), frequency-following response (FFR), cortical auditory-evoked potential (CAEP), and cortical envelope tracking technology (Anderson et al., 2012; Burkard & Sims, 2001; Konrad-Martin et al., 2012; Tremblay et al., 2003). A large body of work has established that the acoustic features of T2 and T3 were most complex among four lexical tones (Ho, 1976; Howie, 1976). T1 and T4 are realized by unchanged level and falling F0 contours, while T2 and T3 are realized by rising and falling-rising F0 contours in the phonated condition. Previous studies have also suggested that the increased complexity of stimuli or tasks could exacerbate the temporal processing deficits (Gordon & Fitzgibbons, 1993; Grose et al., 2006). Therefore, the temporal processing deficit in the elderly combined with the most complex realization pitch tour of T2 and T3 might account for the low perception accuracy.

In the whispered condition, we found lexical tone identification was challenging for both older and younger adults owing to the loss of F0. However, both listener groups achieved high accuracy for the whispered T3 with the older adults showing lower identification accuracy of whispered T3 than the younger group. Previous studies on whispered Mandarin lexical tone perceptions indicated that pre-existing acoustic cues might become dominant in whispered lexical tone recognition (Jiao & Xu, 2019). This view was supported by several phenomena. For example, the secondary cues, duration and amplitude envelope, could be used to identify lexical tones when the primary F0 information was neutralized (Liu & Samuel, 2004). Moreover, amplitude contour could help distinguish T2, T3, and T4, even when the duration of the different tones was controlled (Whalen & Xu, 1992). Our results suggest that younger adults can efficiently use the secondary cues to recognize whispered T3, but this ability was degraded in the aging population. Since the duration was scaled to 500 ms in the first part of the experiment, the amplitude contour is most likely to be

 doi:10.20944/preprints202211.0041.v1

used as a perception cue. The most complex contour of T3 among the four tones makes it easier to recognize than other lexical tones when F0 is absent since the amplitude contour corresponds well with the pitch contour (Ho, 1976). Temporal processing deficits in older adults may impair their ability to efficiently use amplitude contours, resulting in their lower accuracy of whispered T3.

4.2. Aging effects on phonated and whispered vowel perception

Our results of vowel identification indicated that older adults identified vowels as well as younger adults in the phonated condition, which was also in accord with the observation in Yang et al. (2015) and Liu et al., (2021). We additionally found that older adults had lower accuracy in the whispered condition, especially for /i, u/. Confusion matrix data showed that they often confused /o-u/ and /i-y/. Given that identifying vowels is mainly dependent on the first two formants, these results may reflect the formant structure changes in whispered speech and older adults' reduced ability to process the altered formant information.

Previous studies on whispered vowels have reported the formants lift in whispering, especially for the first two formants (Kallail & Emanuel, 1984a, 1984b; Jovičić, 1998; Li & Guo, 2012). Therefore, recognizing whispered vowels might require listeners to remap the raised formants into typical vowel categories. It largely depends on the ability to process spectral information and differentiate fine spectral differences. Emerging studies have suggested the spectral processing deficit (Bidelman et al., 2014; Dorman & Lindholm, 1985; Vongpaisal & Pichora-fuller, 2007). For example, the behavioral results in Bidelman et al. (2014) revealed that older adults had a slower and less consistent classification of vowels than younger adults when presented with a vowel continuum from /u/ to /a/. Neural results also revealed that the brainstem-level encoding was reduced and cortical-level speech-evoked responses were increased but delayed in older adults during vowel categorical perception. Among the six vowels, older adults had more difficulty in

 doi:10.20944/preprints202211.0041.v1

recognizing /i, o, u/. Liu et al. (2021) also found vowels /i, u, y/ (/o/ was not included in their study) were more detrimentally affected than /a, ɤ/ in noise condition. These results may reflect the acoustic similarity of the first two formant frequencies for /o-u/ and /i-y/.

4.3. Effects of duration and intensity modifications in younger and older adults

Part 2 and Part 3 of the experiment were designed to investigate whether variation in duration and intensity may have differential effects on phonated and whispered speech for two groups. It was predicted that reduced duration and intensity would adversely affect speech perception, while increased duration and intensity would facilitate the identification of lexical tones, vowels, and syllables. All these effects would be more apparent in the whispered speech of older adults. These predictions were partially confirmed. Older adults had difficulty recognizing phonated T3, whispered T2 and T3 in the time-compressed condition. They also had difficulty recognizing whispered T4 in the intensity-reduced condition. Lowering intensity negatively affected recognizing phonated vowels /ɤ, i, o, y/ for older adults and /i, u/ for younger adults as well as whispered vowels /a, ɤ/ for older adutls. Contrary to our expectation, time compression increased the accuracy of T4 in phonated condition and the expansion of stimuli brought limited benefits for speech perception for two groups.

Pitch, duration, and intensity are three major factors that have varying degrees of influence on natural speech intelligibility. Interactions between these cues have been demonstrated in a large number of studies (Best et al., 1981; Stevens & Klatt, 1974). For example, although pitch height and contour were the most relevant cues in lexical tone perception, duration also played an influential role (Whalen & Xu,1992; Blicher et al., 1990; Lin & REPP, 1989). These findings underlay our hypotheses that the modification of duration and intensity would affect the perceptions of syllables, vowels, and lexical tones, particularly in whisper speech with F0 loss.

 doi:10.20944/preprints202211.0041.v1

We found the duration reduction negatively affected older adults' perception of phonated T3 and whispered T2, T3. Moreover, the identification accuracy of T4 decreased in the lowered intensity condition but increased in the time-compressed condition. Numerous studies have shown that the duration of lexical tones in Mandarin followed the order of T3>T2>T1>T4 in both phonated and whispered conditions, and the intensity patterns followed the order of T4>T1>T2>T3 in the whispered condition (Fu & Zeng, 2000; Gao, 2002; Li & Guo, 2012; Jiao & Xu, 2016; Liu & Samuel, 2004). Liu & Samuel (2004) also found that in whispered speech, the perception of T3 was correlated with duration positively, but the perception of T1 and T4 correlated with duration negatively. The asymmetric duration and intensity features of different lexical tones might result in older adults' lower accuracy of T2 and T3 and higher accuracy of T4 in short duration because the secondary cues, duration and amplitude contour, could be used as efficient cues for lexical tone identification.

Lowering stimulus intensity affected the identifications of phonated /ɤ, i, o, y/ in the older adults and /i, u/ in younger adults. In addition, it lowered the accuracy of whispered /a, ɤ/ in older adults. Vowel identification is largely based on the first two formants, and a lower intensity or presentation level might make it difficult for listeners, especially for older listeners, to recognize formants accurately. Our findings here are consistent with earlier studies. For instance, diphthongs with high-intensity transition were easier to recognize for older adults with hearing loss, especially in noise and reverberation (Nábělek et al., 1996). Konkle et al., (1977) also found that speech intelligibility decreased as the sensation level decreased from 40 dB SL to 24 SL, especially for those time-compressed stimuli.

Contrary to our expectation, signal enhancement in terms of increasing the duration and intensity of the target stimuli did not bring benefits to lexical tone or vowel perception. The results of intensity modification are in line with Plyler & Hedrick (2002), who found increasing the presentation level from 62 dB SPL to 92 dB SPL did not result in the similar stop consonant perception for impaired hearing listeners. As for the duration, it could boil down to the issue of

 doi:10.20944/preprints202211.0041.v1

the optimal range of duration for lexical tone perception. In Gao (2002), the duration for lexical tones ranged from 200 ms to 400 ms in two articulatory modes when produced in isolation. But the duration was shortened to 150 ms in sentence-medial position. The mean duration of lexical tone in the middle of sentences was around 250 ms (Howie,1976). These findings suggest that the most sensitive duration range of lexical tones might be around 150 ms to 500 ms in natural speech and may explain why lengthening duration to 750 ms has little benefits. It is worth noting that the amounts of duration and intensity increments (150 ms and 15 dB) in our experiment were less than the reduction (350 ms and 20 dB). So we could not rule out the possibility that larger increments could bring noticeable benefits.

4.4. Limitations and Future Direction

Some limitations need to be considered when interpreting the results of the current study. First, it is difficult to tease apart aging from hearing loss (and hidden hearing loss that is not shown in an audiogram), although most older adults in our study have normal hearing at all octave intervals between 250 and 8000 Hz and only five older adults have moderate hearing loss at 8000 Hz. Numerous studies reported the independent effect of hearing loss on speech perception (Gordon & Fitzgibbons, 1993; Phillips et al., 2000;Rooij & Plomp, 1991). Further studies are needed with more advanced methods to explore the effects of hearing loss on speech perception. Second, aging-related cognitive decline such as attention and working memory and older adults' unfamiliarity with computer operation could lead to fatigue, which could have an effect on the results. Third, this study used a very limited set of intensity and duration conditions. More levels of duration and intensity between 150 ms and 500 ms as well as between 40 dB SPL and 60 dB SPL with a larger sample size are desirable in this regard. Finally, we measured the sound pressure level by a calibrated sound-level meter (Rion: Model NL-21) with a linear weighting band and ensured the relative differences for sound intensity differences in Part 3 are fixed and the sound levels for Part 1 and Part 2 are kept the same.

 doi:10.20944/preprints202211.0041.v1

However, we did not further calibrate the sound pressure level using ear simulator or coupler. Future studies should calibrate the sound pressure level to achieve a more accurate SPL.

5. Conclusion

The current study investigated the perception of phonated and whispered speech in younger and older adults and explored the influences of duration and intensity modification on speech perception in the two groups. The results revealed that whispered lexical tone identification became more challenging for both listener groups owing to the loss of F0. Older adults showed lower accuracy of lexical tone perception, especially for the phonated T2, T3 and whispered T3, than younger adults. They also identified vowels /i, u/ less accurately than younger adults, mainly in the whispered condition. Reducing stimulus duration led to lower identification accuracy of phonated T3 and whispered T2 and T3 but increased accuracy of phonated T4 for both listener groups. Older adults' accuracy of phonated T3 were more affected than younger adults. Reduced intensity also had adverse impacts on /ɤ, o, i, y/ for older adults and /i, u/ for younger adults in phonated condition as well as /a, ɤ/ for older adults in whispered condition. Increased duration and intensity in the current study did not help older adults to improve the speech perception in either articulatory mode. These findings highlight how speech perception of the older listeners is impacted for different speech sounds under adverse conditions, which has important implications for auditory rehabilitation of the target speech sounds and the development of the whispered speech intelligibility test for assessing age-related hearing loss.

Acknowledgments

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

 doi:10.20944/preprints202211.0041.v1

(162646). YZ received additional support from University of Minnesota's Brain Imaging Grant and Grand Challenges Exploratory Grant for international collaboration.

Data availability statement

This manuscript qualifies for an open Data badge. The data that support the findings of this study have been made available for public access at https://osf.io/ezxgd/.

References

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, *32*(41), 14156–14164. https://doi.org/10.1523/JNEUROSCI.2176-12.2012

Arthur S. Abramson; (1972). Tonal experiments with whispered THAI. *In Papers in Linguistics and Phonetics to the Memory of Pierre Delattre.De Gruyter Mouton.*, 31–44.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67, 1–48.

Best, C. T., Morrongiello, B., & Robson, R. (1981). *Perceptual equivalence of acoustic cues in speech and nonspeech perception*. 191–211.

Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging*, *35*(11), 2526–2540. https://doi.org/10.1016/j.neurobiolaging.2014.05.006

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37–49. https://doi.org/10.1016/s0095-4470(19)30357-2

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer [Computer program]*. 2018.

Burkard, R. F., & Sims, D. (2001). The Human Auditory Brain-stem Response to High Click Rates. *American Journal of Audiology*, *11*(1), 12–12. https://doi.org/10.1044/1059-0889(2002/er01)

Dorman, M. F., & Lindholm, J. (1985). Phonetic identification by elderly normal and hearing-impaired listeners Phonetic identification by elderly normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *77*(2), 664–670. https://doi.org/10.1121/1.391885

Eklund, I., & Traunmüller, H. (1997). Comparative Study of Male and Female Whispered and Phonated Versions of the Long Vowels of Swedish. *Phonetica*, *54*(1), 1–21. https://doi.org/10.1159/000262207

Feng, Y., Meng, Y., & 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*, 909–913.

Feng, Y., Peng, G., & Wang, W. S. Y. (2020). Age-related differences of tone perception in Mandarin-speaking seniors. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020-Octob*, 1629–1633. https://doi.org/10.21437/Interspeech.2020-2194

Fostick, L., Ben-Artzi, E., & Babkoff, H. (2013). Aging and speech perception: Beyond hearing threshold and cognitive ability. *Journal of Basic and Clinical Physiology and Pharmacology*, *24*(3), 175–183. https://doi.org/10.1515/jbcpp-2013-0048

Fu, Q.-J., & Zeng, F.-G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, *5*(1), 45–57.

Gao, M. (2002). *Tones in Whispered Chinese: Articulatory Features and Perceptual Cues*.

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2017). Masked speech perception across the adult lifespan: Impact of age and hearing impairment. *Hearing Research*, *344*, 109–124. https://doi.org/10.1016/j.heares.2016.11.004

Gordon-Salant, S. (1986). Recognition of natural and time/intensity altered cvs by young and elderly subjects with normal hearing. *Journal of the Acoustical Society of America*, *80*(6), 1599–1607. https://doi.org/10.1121/1.394324

Gordon-salant, S., Fitzgibbons, P. J., & Friedman, S. A. (2007). Recognition of Time-Compressed and Natural Speech With Selective and Elderly Listeners. *Journal of Speech, Language, and Hearing Research*, *50*(October), 1181–1194.

Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. H. (2011). Auditory temporal processing and aging: implications for speech understanding of older people. *Audiology Research*, *1*(1S), 9–15. https://doi.org/10.4081/audiores.2011.e4

Gordon-Salant, S., Zion, D. J., & Espy-Wilson, C. (2014). Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners. *The Journal of the Acoustical Society of America*, *136*(4), EL268–EL274. https://doi.org/10.1121/1.4895014

Gordon, Salant, S., & J.Fitzgibbons, P. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech, Languge,and Hearing Research*, *36*(6), 1276–1285.

Gordon Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *The Journal of the Acoustical Society of America*, *119*(4), 2455–2466. https://doi.org/10.1121/1.2171527

Grose, J. H., Hall, J. W., & Buss, E. (2006). Temporal processing deficits in the pre-senescent auditory system. *The Journal of the Acoustical Society of America*, *119*(4), 2305–2315. https://doi.org/10.1121/1.2172169

Heeren, W., & Heuven, V. J. Van. (2009). Perception and production of boundary tones in whispered Dutch. *Proceedings of Interspeech*, 2411–2414. https://doi.org/10.21437/interspeech.2009-302

Helfer, K. S., & Freyman, R. L. (2008). Aging and Speech-on-Speech Masking. *Ear Hear*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Helfer, K.S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal of Speech and Hearing Research*, *33*(1), 149–155. https://doi.org/10.1044/jshr.3301.149

Ho, A. T. (1976). The acoustic variation of mandarin tones. In *Phonetica* (Vol. 33, Issue 5, pp. 353–367). https://doi.org/10.1159/000259792

Howie, J. M. (1976). Acoustical Studies of Mandarin Vowels and Tone. *Cambridge University Press*.

Humes, L. E., & Dubno, J. R. (2010). Factors Affecting Speech Understanding in Older Adults. In *The Aging Auditory System* (pp. 211–257). https://doi.org/10.1007/978-1-4419-0993-0_8

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication*, *45*(2), 139–152. https://doi.org/10.1016/j.specom.2003.10.005

Jensen, M. K. (1958). Recognition of Word Tones in Whispered Speech. *WORD*, *14*(2–3), 187–196. https://doi.org/10.1080/00437956.1958.11659663

Jiang, W., Li, Y., Shu, H., Zhang, L., & Zhang, Y. (2017). Use of semantic context and F 0 contours by older listeners during Mandarin speech recognition in quiet and single-talker interference conditions. *The Journal of the Acoustical Society of America*, *141*(4), EL338–EL344. https://doi.org/10.1121/1.4979565

Jiao, L., & Xu, Y. (2016). Interactions of Tone and Intonation in Whispered Mandarin. Speech Prosody, 94–98.

Jiao, L., & Xu, Y. (2019). Whispered Mandarin has no production-enhanced cues for tone and intonation. *Lingua*, *218*, 24–37. https://doi.org/10.1016/j.lingua.2018.01.004

Jovičić, S. T. (1998). Formant feature differences between whispered and voiced sustained vowels. *Acta Acustica United with Acustica*, *84*(4), 739–743.

Jovičić, S. T., & Šarić, Z. (2008). Acoustic Analysis of Consonants in Whispered Speech. *Journal of Voice*, *22*(3), 263–274.

Kallail, K. J., & Emanuel, F. W. (1984a). An acoustic comparison of isolated whispered and phonated vowel samples produced by adult male subjects. *Journal of Phonetics*, *12*(2), 175–186. https://doi.org/10.1016/s0095-4470(19)30864-2

Kallail, K. J., & Emanuel, F. W. (1984b). Formant-frequency differences between isolated whispered and phonated vowel samples produced by adult female subjects. *Journal of Speech and Hearing Research*, *27*(2), 245–251.

Kallail, K. J., & Emanuel, F. W. (1985). The identifiability of isolated whispered and phonated vowel samples. Journal of Phonetics, 13(1), 11–17. https://doi.org/10.1016/S0095-4470(19)30722-3

Kolodziejczyk, I., & Szelag, E. (2008). Auditory perception of temporal order in centenarians in comparison with young and elderly subjects. *Acta Neurobiologiae Experimentalis*, *68*(3), 373–381.

Konkle, D. F., Beasley, D. S., & Bess, F. H. (1977). Intelligibility of time altered speech in relation to chronological aging. *Journal of Speech and Hearing Research*, *20*(1), 108–115. https://doi.org/10.1044/jshr.2001.108

Konrad-Martin, D., Marilyn F., D., Garnett, M., Susan, G., Daniel, M., Stephen A., F., & Donald F., A. (2012). Age-Related Changes in the Auditory Brainstem Response. *J Am Acad Audiol.*, *23*(1), 18–75. https://doi.org/10.3766/jaaa.23.1.3.Age-Related

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means. R package version 1.2.3. Retrieved from https://CRAN.R-project.org/package=emmeans

Li, B., & Guo, Y. (2012). Mandarin tone contrast in whisper. *Third International Symposium on Tonal Aspects of Languages*, 10–13. http://20.210-193-52.unknown.qala.com.sg/archive/tal_2012/papers/tl12_P1-10.pdf

Lin, H., & REPP, B. H. (1989). Cues to the perception of Tanwanese tones. *Language and Speech*, *32*(1), 25–44.

Liu, C., Xu, C., Wang, Y., Xu, L., Zhang, H., & Yang, X. (2021). Aging Effect on Mandarin Chinese Vowel and Tone Identification in Six-Talker Babble. *American Journal of Audiology*, *30*(3), 616-630.

Liu, S., & Samuel, A. G. (2004). Perception of mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*(2), 109–138. https://doi.org/10.1177/00238309040470020101

Lüdecke et al., (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139

Moore, D. R., Edmondson-Jones, M., Dawes, P., Fortnum, H., McCormack, A., Pierzycki, R. H., & Munro, K. J. (2014). Relation between speech-in-noise threshold, hearing loss and cognition from 40-69 years of age. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0107720

Nábělek, A. K., Ovchinnikov, A., Czyzewski, Z., & Crowley, H. J. (1996). Cues for perception of synthetic and natural diphthongs in either noise or reverberation. *The Journal of the Acoustical Society of America*, *99*(3), 1742–1753. https://doi.org/10.1121/1.415238

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

Phillips, S. L., Gordon-salant, S., Fitzgibbons, P. J., & Yeni-komshian, G. (2000). Frequency and Temporal resolution in elderly listerners with good and poor word recognition. *Journal of Speech,Language, and Hearing Research*, *43*, 217–228.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology*, *47*(SUPPL. 2), 72–82. https://doi.org/10.1080/14992020802307404

Plyler, P. N., & Hedrick, M. S. (2002). Effects of stimulus presentation level on stop consonant identification in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*, *13*(3), 154–159. https://doi.org/10.1055/s-0040-1715957

Price, P. J., & Simon, H. J. (1984). Perception of temporal differences in speech by “normal-Hearing” adults: Effects of age and intensity. *Journal of the Acoustical Society of America*, *76*(2), 405–410. https://doi.org/10.1121/1.391581

 doi:10.20944/preprints202211.0041.v1

Rooij, J. C. G. M. V., & Plomp, R. (1991). Auditive and cognitive factors in speech perception by elderly listeners. *Acta Oto-Laryngologica*, *111*(S476), 177–181. https://doi.org/10.3109/00016489109127275

Schwartz, M. F. (1967). Syllable duration in oral and whispered reading. *The Journal of the Acoustical Society of America*, *41*, 1367–1369.

Sharf, D. J. (1964). Vowel Duration in Whispered and in Normal Speech. *Language and Speech*, *7*(2), 89–97. https://doi.org/10.1177/002383096400700204

Sharifzadeh, H. R., Mcloughlin, I. V, Russell, M. J., & Kingdom, U. (2012). A Comprehensive Vowel Space for Whispered Speech. *Journal of Voice*, *26*(2), e49–e56. https://doi.org/10.1016/j.jvoice.2010.12.002

Smith, S. L., Pichora-Fuller, M. K., Wilson, R. H., & MacDonald, E. N. (2012). Word recognition for temporally and spectrally distorted materials: The effects of age and hearing loss. *Ear and Hearing*, *33*(3), 349–366. https://doi.org/10.1097/AUD.0b013e318242571c

Solomon, N. P., McCall, G. N., Trosset, M. W., & Gray, W. C. (1989). Laryngeal configuration and constriction during two types of whispering. *Journal of Speech and Hearing Research*, *32*(1), 161–174. https://doi.org/10.1044/jshr.3201.161

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, *55*(3), 653–659. https://doi.org/10.1121/1.1914578

Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America*, *104*(4), 2385–2399. https://doi.org/10.1121/1.423748

Tremblay, K. L., Piskosz, M., & Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clinical Neurophysiology*, *114*(7), 1332–1343. https://doi.org/10.1016/S1388-2457(03)00114-7

Vaughan, N. E., & Letowski, T. (1997). Effects of age, speech rate, and type of test on temporal auditory processing. *Journal of Speech, Language, and Hearing Research*, *40*(5), 1192–1200. https://doi.org/10.1044/jslhr.4005.1192

Vongpaisal, T., & Pichora-fuller, M. K. (2007). Effect of Age on F0 Difference Limen and Concurrent Vowel Identification. *Journal of Speech, Language, and Hearing Research*, *50*, 1139–1156.

Wang, Y., Yang, X., & Liu, C. (2017). Categorical perception of mandarin chinese tones 1–2 and tones 1–4: Effects of aging and signal duration. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3667–3677. https://doi.org/10.1044/2017_JSLHR-H-17-0061

Wang, Y., Yang, X., Zhang, H., & Xu, L. (2017). Aging effect on categorical perception of Mandarin tones 2 and 3 and thresholds of pitch contour discrimination. *American Journal of Speech-Language Pathology*, *25*(October), 1–15. https://doi.org/10.1044/2016

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25–47. https://doi.org/10.1159/000261901

Wong, P. C. M., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693–703. https://doi.org/10.1016/j.neuropsychologia.2008.11.032

Yang, X., Wang, Y., Xu, L., Zhang, H., Xu, C., & Liu, C. (2015). Aging effect on Mandarin Chinese vowel and tone identification. *The Journal of the Acoustical Society of America*, *138*(4), EL411–EL416. https://doi.org/10.1121/1.4933234

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

Figure Captions

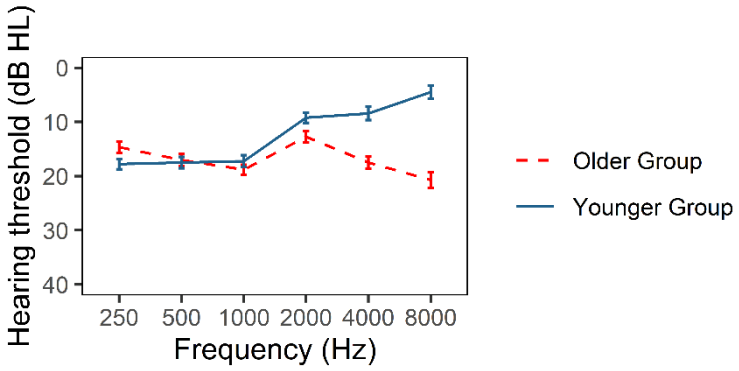

Figure 1. Pure tone thresholds and standard errors at 250, 500, 1000, 2000, 4000 and 8000 Hz for the two groups.

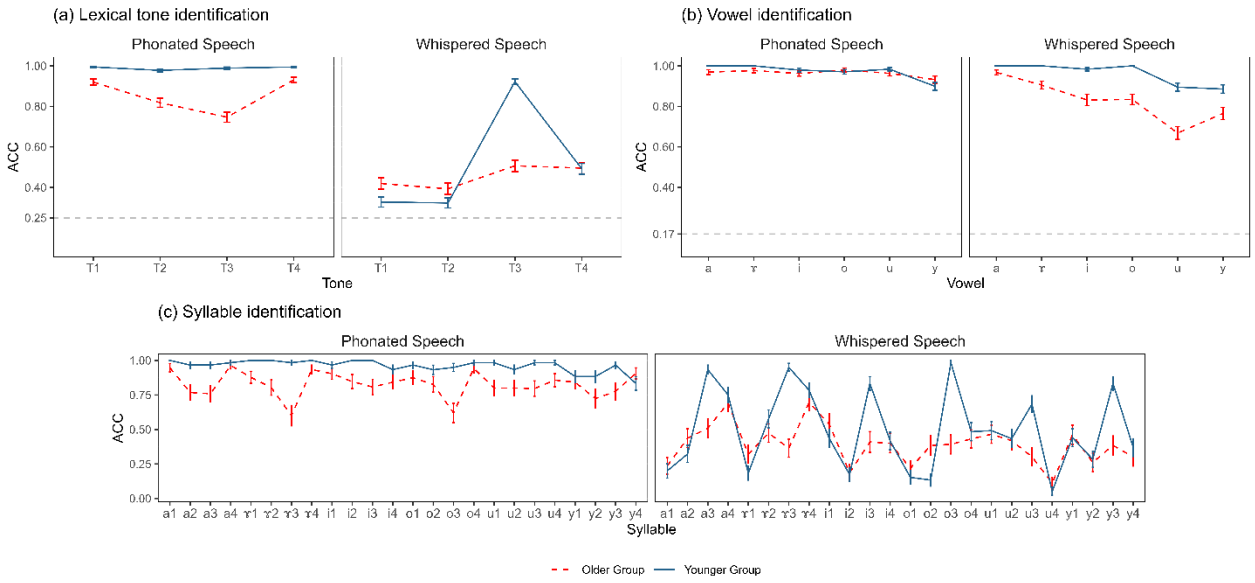

Figure 2. Results of the (a) lexical tone Identification accuracy, (b) vowel identification accuracy, and (c) syllable identification accuracy in two articulatory speech (phonated speech and whispered speech) for two listener groups (older group and younger group). For each identification task, the results in phonated condition are displayed on the left panel, and the results in whispered condition are displayed on the right panel. The gray lines represent the chance level of the identification accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

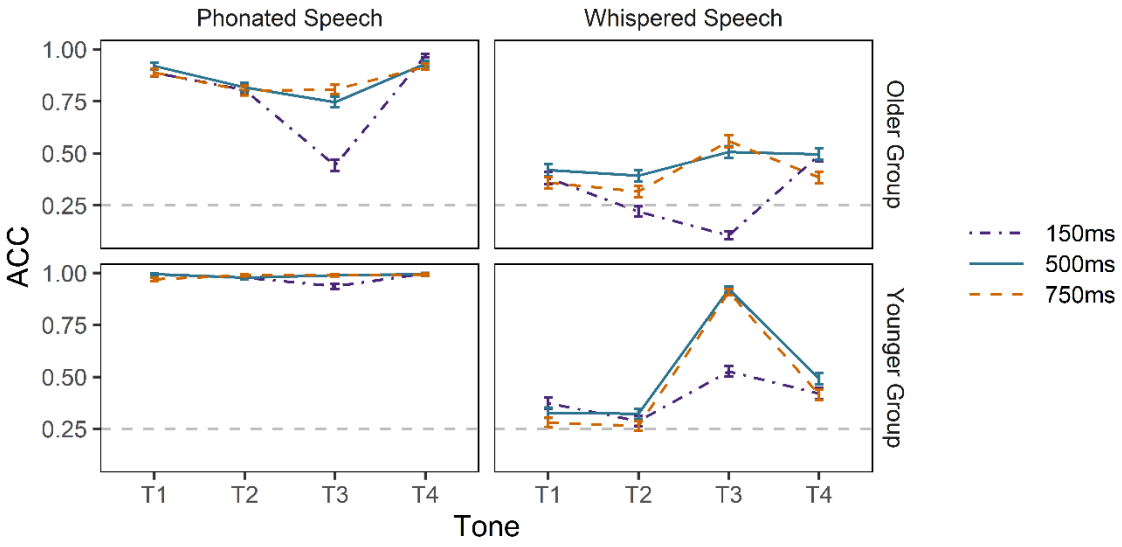

Figure 3. Lexical tone identification accuracy in phonated and whispered condition in three duration conditions.

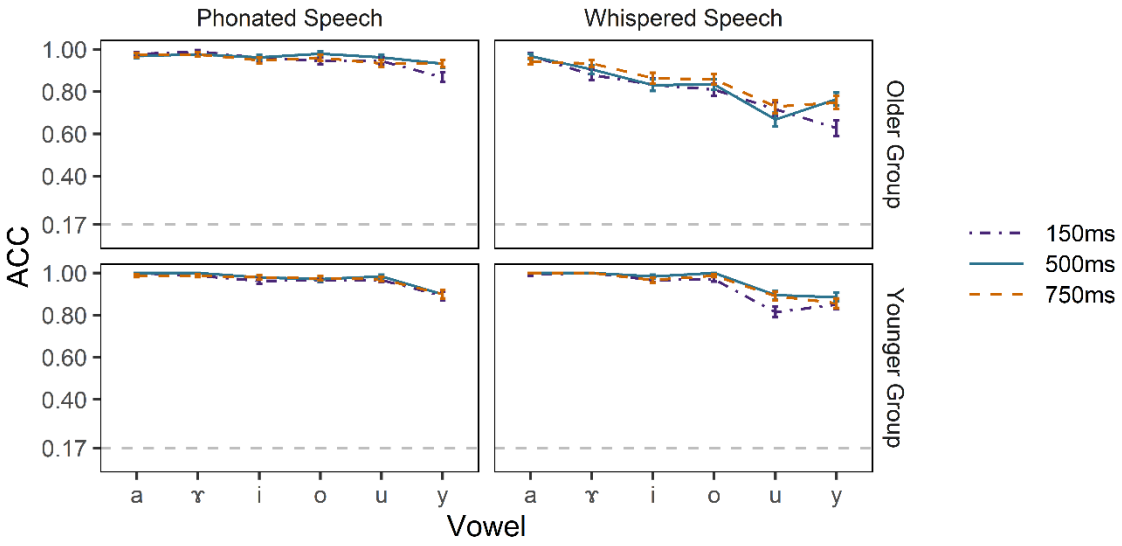

Figure 4. Vowel identification accuracy in phonated and whispered condition in three duration conditions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

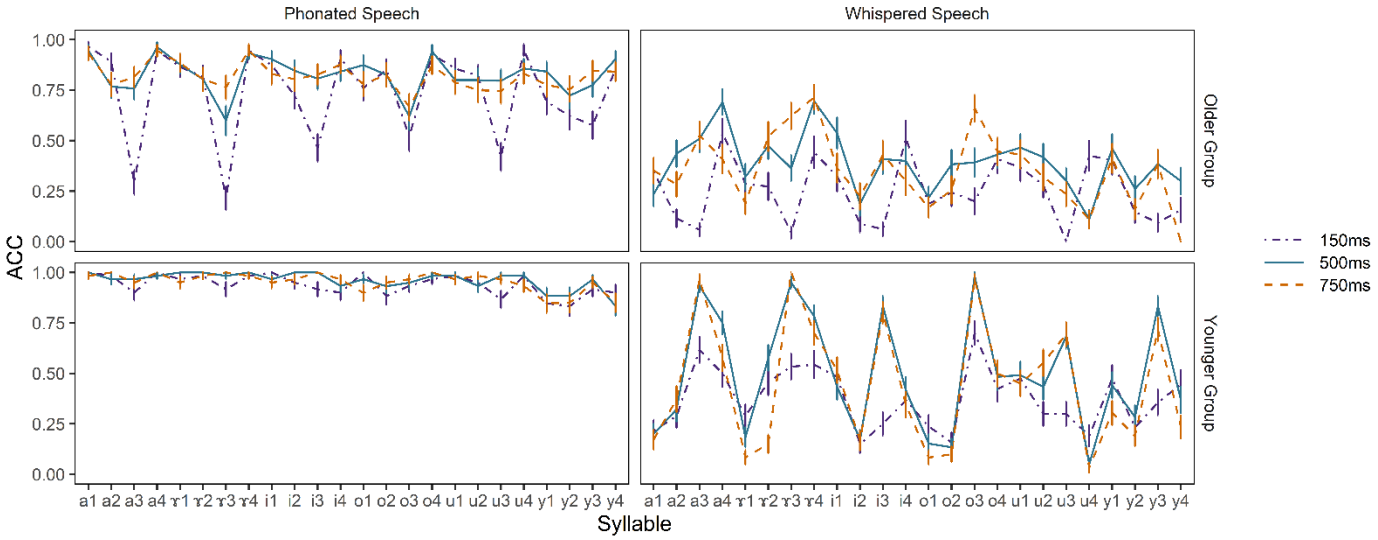

Figure 5. Syllable identification accuracy in phonated and whispered condition in three duration conditions.

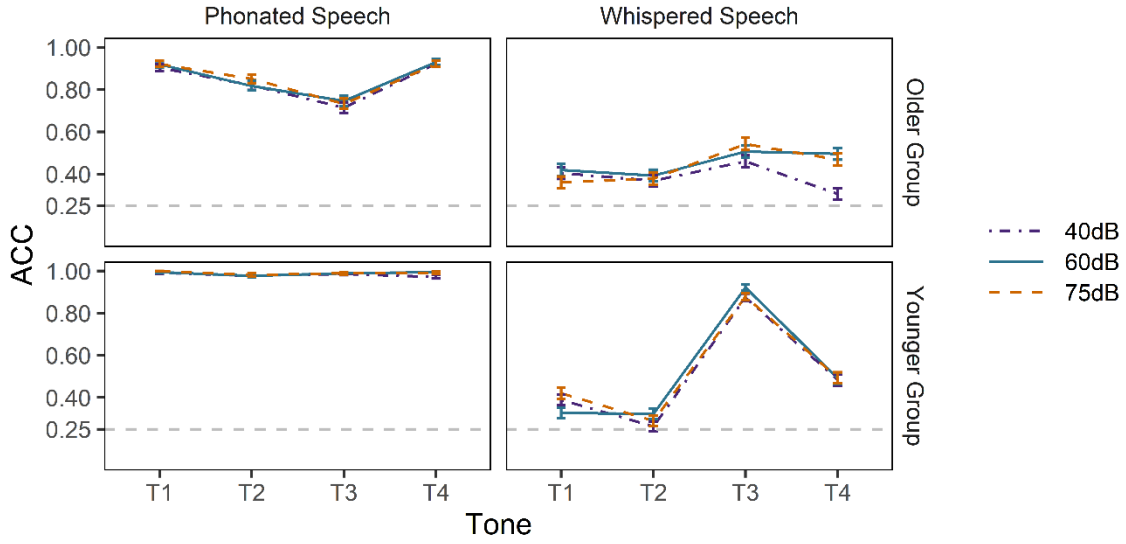

Figure 6. Lexical Tone identification accuracy in phonated and whispered condition in three intensity conditions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2022 doi:10.20944/preprints202211.0041.v1

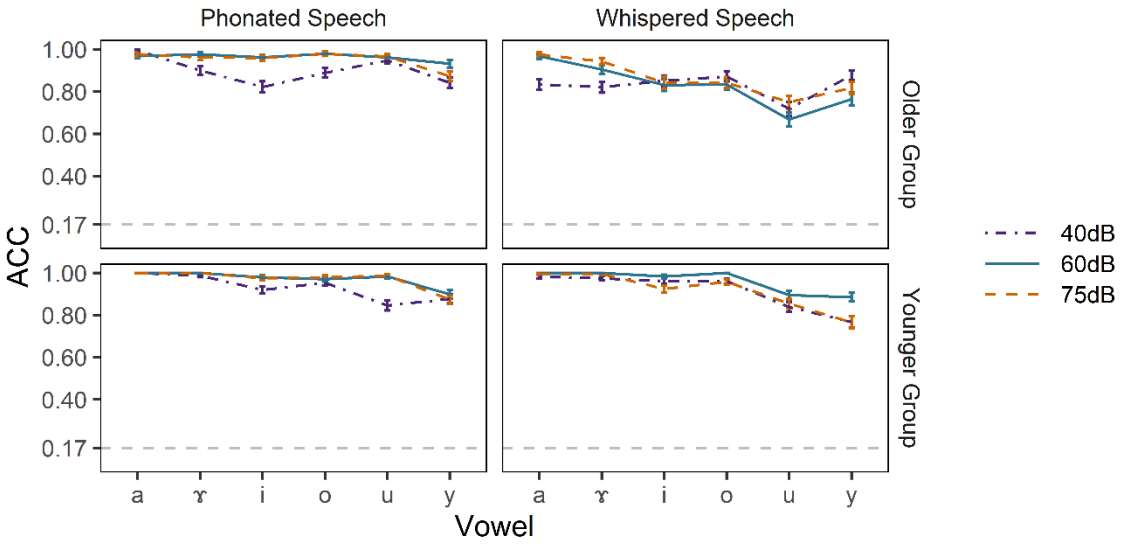

Figure 7. Vowel identification accuracy in phonated and whispered condition in three intensity conditions.

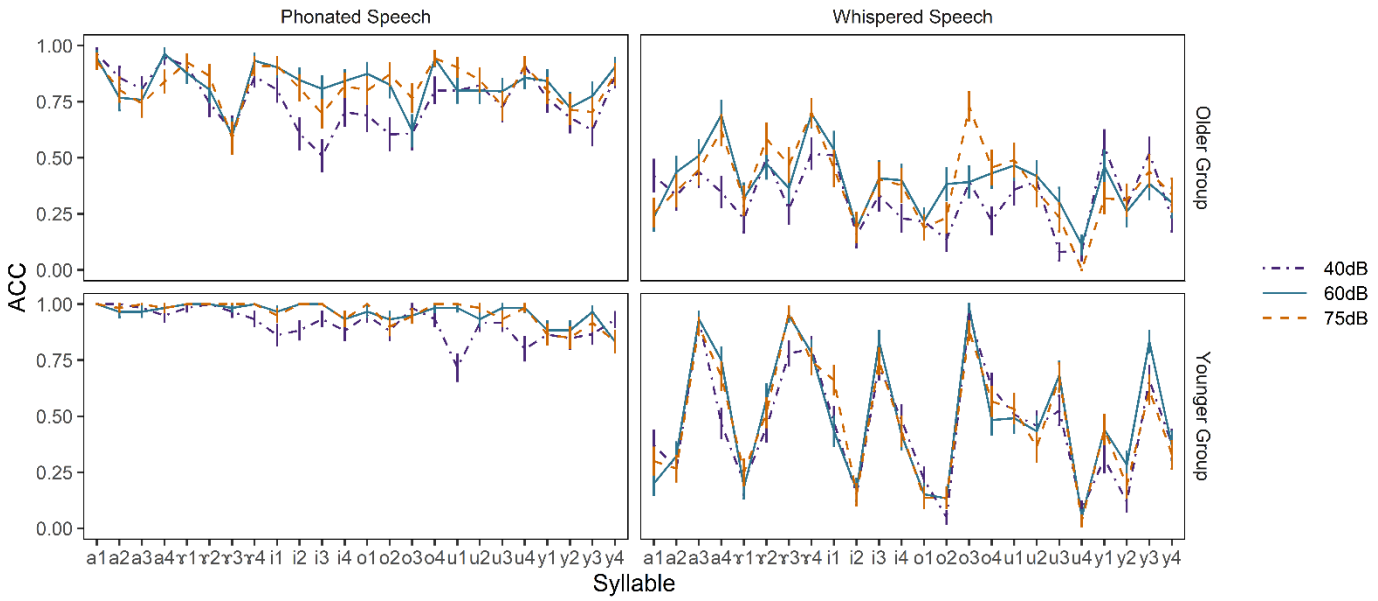

Figure 8. Syllable identification accuracy in phonated and whispered condition in three intensity conditions.