Age-related differences of Mandarin tone and consonant aspiration perception in babble noise

Yaru Meng,[a,e,*] Fei Chen,[b] Yan Feng,[c,d] Gang Peng,[d,e] Wei Zheng[a]

[a] Department of Chinese Language and Literature, East China Normal University, Shanghai, China

[b] School of Foreign Languages, Hunan University, Changsha, China

[c] School of Foreign Studies, Nanjing University of Science and Technology, Nanjing, China

[d] Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China

[e] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

***Corresponding author**: Department of Chinese Language and Literature, East China Normal University, Shanghai, China.

E-mail address: yaru_meng@163.com

Conflict of Interest: The authors have declared that no competing interests existed at the time of publication.

Purpose: This study investigated the categorical perception of Mandarin tones and consonant aspiration contrasts in babble noise among adults and adolescents aged 12-14 years, and explored the association between working memory and categorical perception.

Method: Twenty-four adults and twenty adolescents with Mandarin as their native language were recruited. Their performances of phonemic identification and discrimination in babble noise and quiet conditions, digit span tasks, and nonword repetition were assessed.

Results: Results indicated that, firstly, in the noise condition, both adults and adolescents showed wider boundary widths and lower between-category accuracies when perceiving aspiration of consonants than in the quiet condition, and the categorical boundary of tone perception in adolescents showed a transitional tendency towards Tone 1. Secondly, discrimination of consonant aspiration in adolescents needed to be further developed. Lastly, the accuracy of nonword repetition in adolescents was lower than that in adults, and adults with better auditory verbal working memory had better performance on tone perception.

Conclusions: Our results provided evidence that tone perception is acquired easier than consonant aspiration perception, and tone perception is more robust and less susceptible to noise interference. Categorical perception performance relates to the capacity and utilization of auditory verbal working memory in some ways.

Introduction

Categorical perception (CP) of speech refers to the ability that people could map continuous and infinite sounds into finite phonemic categories that the specific language allows. Previous studies have well-documented the characteristics of CP: a steep rise or fall in identification curve near the boundary between different phonemic categories (Liberman et al., 1957), and better discrimination ability of between-category phonemes than that of within-category ones (Xu et al., 2006).

CP of Mandarin lexical tones and aspiration of consonants

Earlier studies on CP have focused on segmental features in Indo-European languages (e.g., Liberman et al., 1957, 1961), and recent studies have paid attention to suprasegmental features, for example, lexical tone (e.g., Peng et al., 2010; Xu et al., 2006). Mandarin is a tone language that typically uses fundamental frequency (F0) contour to provide contrasts in word meaning. The function of tones to distinguish lexical meanings is analogous to segments such as voiced/voiceless consonants, except that the tone can be carried by more than one segment. Mandarin Chinese phonemically distinguishes four citation tones with different pitch contours: high-level tone (Tone 1), mid-rising tone (Tone 2), low-falling-rising tone (Tone 3), and high-falling tone (Tone 4). The same syllable expresses different meanings depending upon the tones. For example, the meaning of Mandarin /pa/ with Tone 1 is "八 eight", with Tone 2 means "拔 to pull", with Tones 3 and 4 are "把 to hold" and "爸 father", respectively. The importance of F0 as a primary cue to the perception of Mandarin tones is supported by a variety of studies (e.g., Gandour, 1978; Wang, 1967). For example, Massaro, Tseng, and Cohen (1985) reported that both F0 height and F0 contour were important in tone perception. Many studies have shown that the perception of Mandarin lexical tones is typically categorical and is influenced by language experience (cross-language or crossdialect) (Peng et al., 2010; Wang, 1976; Xu et al., 2006). For example, Wang (1976) investigated CP of level tone and rising tone among native-Mandarin and native-English speakers, and found two different types of perceptual boundary (i.e., psychophysical boundary vs. linguistic boundary). For native-English listeners who lack tonal language experience, they were more likely to distinguish the "real level tone" and the rest according to their psychophysical experience. Therefore, their boundary position of tone perception was closer to the level tone. While for native-Mandarin speakers with tonal language experience, they could distinguish tones according to their mental linguistic categories.

Like lexical tone, consonant aspiration is also a distinctive feature in Mandarin. In English, voiced and voiceless consonants are distinct because of voice onset time (VOT) (Lisker & Abramson, 1964), and the perception of English consonants varying in VOT was essentially categorical (Liberman et al., 1961). The perceptual characteristics of the VOT have been documented to be related to specific languages (Lisker & Abramson, 1970). Different from English, the temporal cue of VOT is utilized to identify the aspiration of consonants, which is an important distinctive feature in Mandarin. There are six pairs of unaspirated and aspirated consonants that can distinguish lexical meanings. For example, the meaning of Mandarin /pa/ with Tone 1 is "八 eight", while the aspirated one /p^ha/ with Tone 1 means "趴 grovel". Numerous studies have shown that VOT is an important perceptual cue for the distinction between aspirated and unaspirated consonants such as /p/ and /p^h/ (Lisker & Abramson, 1964, 1967). Many investigations showed that the perception pattern of VOT in Mandarin speakers is highly categorical (Wang & Shangguan, 2004; Xi et al., 2009). Given that lexical tones and aspiration of stops were two representative distinctive features of Mandarin and were perceived categorically, they were chosen in this study to investigate the age-related differences of CP of Mandarin segments and suprasegments in babble noise.

Effects of babble noise on CP

Speech perception is always accompanied by background noise in daily life, especially speech noise, instead of in acoustically ideal conditions in the lab. A considerable number of studies have shown noise-associated deterioration in speech recognition for adult listeners, especially under the babble noise which contains phonetic, linguistic, and semantic cues. Babble noise disrupts the CP of speech from spectral energy and language information, and distracts attention and working memory (WM) needed for CP of target speech as well (Brungart et al., 2001; Schneider et al., 2007). Therefore, investigating the CP in babble noise could help to reveal the development of CP in real life, and the utilization of WM involved. In view of the sparse literature about the effect of babble noise on the perception of Mandarin tones and consonants, the first goal of this study was to examine how the babble noise influenced listeners' CP of Mandarin tones and aspiration of consonants involving both identification and discrimination functions.

Different noise types (e.g., white noise and babble noise) with diverse masking types modulate the difficulty of speech perception (Brungart et al., 2001; Lecumberri & Cooke, 2006). Nonspeech noise (e.g., white noise) mainly masks the energy in the same critical bands and produces disturbance at the auditory periphery. For example, Turner et al. (1992) investigated the detection and recognition of stop consonants in white noise, and found recognition was largely determined by the audibility of speech in normal-hearing listeners. And results also indicated the strong effect of noise on consonants recognition, which only achieved above-chance accuracy (16.67%) at the SNR of -5 to 0 dB. Besides masking energy, speech noise (e.g., babble noise) interferes with the processing of the target signal at phonetic and/or semantic levels because of the similar-sound-pattern distracter. Especially, babble noise which contains phonetic, linguistic, and semantic cues has comparative interference with speech perception in both auditory and cognitive aspects. The effects of babble noise on the recognition of lexical tones and consonants had been reported in several studies. For example, Dees et al. (2007) investigated the recognition of tones carried by different vowels in white noise with signal-to-noise ratios (SNR) of -15 dB and -20 dB and babble noise with the SNR of -5 dB and -20 dB among native Mandarin speakers and English speakers without exposure to tonal languages. Results showed that firstly both groups performed better in the condition with higher SNR. In addition, native Mandarin speakers overall performed better than English speakers in white noise conditions, but Mandarin-speaking listeners were more likely to be affected by babble noise. Moreover, the interference of babble noise was greater than white noise in both groups. These findings indicated that babble noise had a stronger interference effect than white noise on listeners' perception of lexical tones by masking information transmission. Similar results were also found in the studies about consonants where the recognition accuracy of 57 CV syllables decreased with the gradual reduction of SNR by 12 dB, 6 dB, and 0 dB, and was below 30% in babble noise at the 0dB SNR (Gordon-Salant, 1985). Listeners also were found to suffer more in babble noise than in white noise and to compete with speech noise when identifying the consonant of VCV combinations (Lecumberri & Cooke, 2006). Since previous studies have generally suggested that babble noise impedes the recognition of tones and consonants because of information masking, it is hypothesized that babble noise could impair the CP of Mandarin tones and consonant aspirations by masking F0 and VOT information. There would be a shallower identification slope and poorer discrimination performance across categories in the babble noise condition than in the quiet condition.

Perceptual development of lexical tones and consonants

The development of CP has been extensively studied, which indicated that CP of tones and aspiration of consonants were found less categorical in children compared to adults (Chen et al., 2017; Feng, 2022; Hazan & Barrett, 2000; Ma et al., 2021; Xi et al., 2009). Previous studies suggested that speech perception developed significantly between childhood and later adolescence, and matured during the period from 12 to early 20s of age (Feng, 2022; Hazan & Barrett, 2000; Medina et al., 2010; Westerhausen et al., 2015). Considering that CP of speech develops and possibly reaches maturity during adolescence, it is important to understand the trajectory of development in adolescents. The second goal of the current study was to investigate the age-related differences in the CP of Mandarin tones and aspiration of consonants between adults and adolescents, and particularly, how age interacted with babble noise.

Previous literature has shown that it takes newborns six to twelve months to primarily form the concept of the phonological category with the increased native language experience after birth (Kuhl et al., 2006). And speech perception ability is continuously developed till reaching a peak in the early 20s (Westerhausen et al., 2015). Simon and Fourcin (1978) explored the perception of VOT and the first-formant transition of synthetic speech-like consonants in French-speaking and English-speaking children. Results showed that there were three stages in the development of children's labeling behavior: scattered labeling, progressive labeling, and categorical labeling. Categorical labeling only becomes the dominant pattern at the ages of five to six for the English-speaking children, and one or two years later for the French-speaking children. The study of Hazan and Barrett (2000) further revealed that 12-year-old children still do not acquire the same abilities of CP to identify consonants as adults by investigating the identification of continua of /g/-/k/, /d/-/g/, /s/-/z/, and /s/-/ʃ/. Results showed that the slope of the identification curve increased significantly with age between six and 12 years. Medina et al. (2010) found the steepness of the identification function increases significantly between 9-year-old children and 17-year-old adolescents, but not between 17-year-old adolescents and adults, by investigating the identification and discrimination of /də/-/tə/ continuum in French.

In terms of the CP of Mandarin phonemic categories, Xi et al. (2009) examined the development of CP by testing the behavioral identification performance of 51 children aged five to seven years. Results showed that children's perception of tones and aspiration of consonants were categorical, but the perceptual developmental trajectories of the two phonological categories were different. Children aged six years had obtained an adult-like ability to identify tones, while consonant aspiration perception among children aged seven years was still worse than adults. Similar results involving both identification and discrimination were found in Ma et al. (2021), where the identification and discrimination reached the adult-like level at age six and five respectively for lexical tones, but still not for aspiration of stops at age six. Feng (2022) further expanded the age range to four to 14 years old. Results showed that six-year-old children could perceive tones categorically as adults, and the identification of aspiration in ten-year-old children had reached the adult-like level, but their ability to discriminate aspiration became mature at the age of 14. The different perceptual developmental trajectories between lexical tones and aspiration of consonants were consistent with the phonological saliency hypothesis (Zhu & Dodd, 2000), which proposed that phonological saliency determined the acquisition order of syllable components. The degree of phonological saliency was determined by the combination of the status of the component in the syllable structure, the capacity in differentiating lexical meaning, and the number of permissible choices within the component. Therefore, tones had a higher saliency and were acquired earlier than the aspiration of consonants, because of its necessity for forming syllables and fewer alternative phonemes in Mandarin.

The literature above documented the effect of babble noise on adult listeners' perception of lexical tones and consonant contrasts, and tried to uncover the perceptual development of lexical tones and consonants. However, few studies investigated the perceptual performance of lexical tones and consonants among adolescents over the age of 12. Moreover, the effects of noise on the CP of tones and consonants in adolescents remain unclear. In consequence, it remains unknown about the age-related differences in CP of lexical tones and aspiration of consonants between adults and adolescents, and the interaction effect of age and noise. According to the results of previous studies, the development of categorical precision (slope of the identification curve and magnitude of the discrimination peak) of VOT perception is completed sometime during the period from 12 to 17 years of age (Feng, 2022; Hazan & Barrett, 2000; Medina et al., 2010), and the six-year-old children could perceive tones categorically as adults in the quiet condition (Chen, 2017; Feng, 2022). Therefore, the adolescents aged 12-14 were chosen in this study to investigate the age-related differences of the CP of Mandarin tones and aspiration of consonants in babble noise, which may help to better understand the real development of CP of lexical tones and consonant aspiration in real life. In the quiet condition, adolescents aged 12 years had obtained an adult-like ability to identify lexical tones, while their consonant perception was still in development. And cross-linguistic studies of noise suggest that lacking experience of the native language may lead to less robustness in phonemic category learning, which is susceptible to noise (Lecumberri & Cooke, 2006; Xu et al., 2018). On this account, it is assumed that, for adolescents aged 12-14 years, CP of Mandarin tones was comparable with adults, while CP of consonant aspiration was poorer than that of adults and the gap would become larger in the babble noise condition.

The role of working memory in CP

Considering that babble noise produces the energetic and informational interference, and attention and WM distractions in the processing of CP (Brungart et al., 2001; Schneider et al., 2007), the third purpose of this study was to examine the difference in working memory capacity (WMC) between adults and adolescents, and assess the relationship between speech perception in the quiet or noise conditions and components of auditory verbal working memory (AVWM).

WM is considered to support language acquisition and subsequent processing via the maintenance, storage, and processing of ambient language sounds (Baddeley, 1992, 2000; Gathercole & Baddeley, 1989). Previous studies suggest that cognitive development may represent a cascade in which age-related increases in processing speed mediate most of the developmental increases in WMC (Fry & Hale, 1996; 2000). However, the relationship between developmental differences in WMC and speech perception, an important high-level cognitive ability, remains unclear and needs further investigation. In addition, it has been well documented that WM is always deployed when speech is presented against a noisy background, especially when the masker consists of speech (Koelewijn et al., 2012; Zekveld et al., 2013). WM was thought to compensate for the ambiguity as a result of degraded phonological representations, and help predictive coding before or during the early portions of a spoken stimulus in adverse listening conditions (Rönnberg et al., 2011; 2013). Previous studies based on the perception of sentences and words suggested that listeners with larger WMC tended to have a better perceptual performance in noise (e.g. Ingvalson et al., 2015; Millman & Mattys, 2017). For example, Millman and Mattys (2017) found that listeners with greater phonological WMC were better able to correctly identify sentences in modulated noise backgrounds. Previous research has focused on adults, however, less is known about the underage group. Moreover, how WM contributes to CP in noise remains unclear.

WMC has been assessed with various auditory and visual tests in previous studies, often pertaining to different components of WM. Standardized tests of AVWM include forward digit span task (FDS), backward digit span task (BDS), and non-word repetition (NWR), which were considered in terms of a gradient of processing demands (Daneman & Merikle, 1996). FDS requires minimum processing demand for temporarily maintaining verbal information, whereas BDS requires more executive control for manipulation of the digit sequence (Alloway et al., 2004; Gathercole et al., 2004). In the WM models, short-term memory generally refers to maintaining information temporarily, while complex WM involves both information storage and processing (Baddeley, 2000; Zebib et al., 2019). Therefore, FDS and BDS assessed short-term memory (temporarily storing information), and complex working memory (storage and manipulation by the executive control), respectively (Hamann, 2017). And NWR, with the maximum processing demand, measures more than the phonological component of WM and, to some extent, long-term memory (Gathercole, 1995; Gathercole & Adams, 1994). Recently, NWR was considered as an indicator of sub-lexical knowledge, which was the representation of ‘chunks’ of phonemes in the long-term memory (Jones et al., 2014; Szewczyk, 2018).

FDS, BDS, and NWR have been widely used in speech recognition, such as words and sentence recognition, however, any relationship between different sub-components of AVWM and CP remains ill-defined. Xu et al. (2006) proposed a multistore model of CP consisting of unanalyzed and analyzed sensory memory, short-term and long-term categorical memory, which are inputs for decision-making. Information is encoded hierarchically but short-term categorical memory and analyzed sensory memory can be processed in parallel. However, the measurement corresponding to each memory component in this model has not emerged. The multistore model of CP could correspond to the components of generalized WM: short-term memory (sensory and short-term categorical information storage), long-term memory（long-term categorical representations), and complex memory（information encoding and decision-making). Therefore, the FDS, BDS, and NWR, which were frequently used to measure short-term memory, complex memory, and long-term memory respectively, were chosen in this study to clarify the role of sub-components of AVWM in CP of adolescents and adults in the quiet and babble noise conditions via these different tests.

Taken together, the current work attempted to address the following three research questions: (a) how does babble noise affect adult and adolescent listeners' CP of Mandarin tones and aspiration of consonants? (b) how does age interact with noise to affect CP of Mandarin tones and aspiration of consonants? and (c) whether there is a difference in WMC between adults and adolescents? and whether adult and adolescent listeners' CP of Mandarin tones and aspiration of consonants in the quiet or babble noise conditions were associated with their AVWM?

Method

Participants

Twenty-four adults (13 males, 21-33 years old, mean = 25.1, *SD* = 3.01) and twenty adolescents (14 males, 12-14 years old, mean = 13.5, *SD* = 0.79) with Mandarin as their native language were recruited from northern China for this study. All participants were right-handed and had no formal musical experience or a history of speech or hearing impairment. Informed consent approved by the Behavioral Research

Ethics Committee of Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences was obtained from all adult participants and adolescents' parents.

Materials

Tone 1 - Tone 2 pair had been investigated in many studies, and was well-documented to be perceived categorically in native listeners (e.g., Wang, 1976; Xi et al., 2009). Therefore, Tone 1 - Tone 2 pair was chosen for the investigation of lexical tone perception in the current study. Speech samples, including the Mandarin syllable /pa/ with Tone 1, the syllable /pa/ with Tone 2, and the syllable /pʰa/ with Tone 1, were recorded by a female native Mandarin speaker using Praat (Boersma & Weenink, 2019), with a sampling rate of 44,100 Hz and 16-bit resolution. The 7-stimuli continuum of tones from Tone 2 (mid-rising tone) to Tone 1 (high-level tone) and the 7-stimuli continuum of consonant aspiration from /pa/ to /pʰa/ with Tone 1 were synthesized using TANDEM-STRAIGHT software (Kawahara et al., 2009). For example, the tone continuum was generated based on the natural Tone 1 and Tone 2 sounds, after F0 extraction and editing, aperiodicity and STRAIGHT spectrum extraction, assigning temporal and frequency anchors, and morphing rate manipulation. The step sizes of tone and aspiration continua were 12 Hz and 8 ms, respectively. Schematic diagrams of the pitch contours in the tone continuum and VOT in the aspiration continuum are shown in Figures 1 (A) and (B).

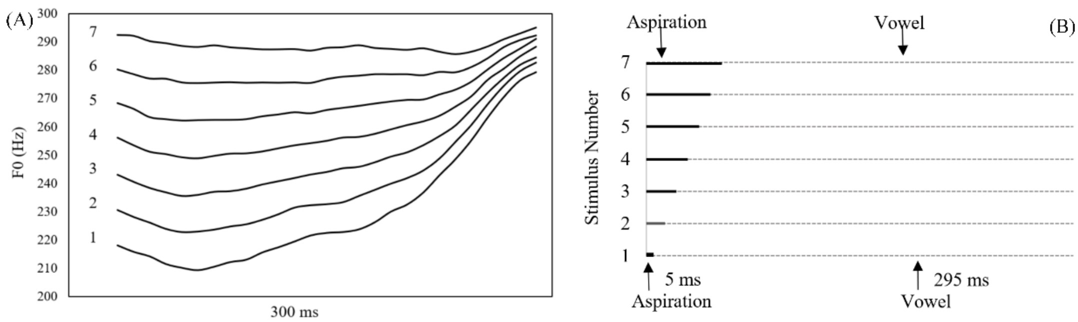

Figure 1. (A) Schematic diagram of pitch contours in tone continuum; (B) Schematic diagram of VOT patterns in aspiration continuum.

Each stimulus of the noisy continua was resynthesized by mixing the corresponding clean speech stimulus with a frozen piece of babble noise with identical duration and SPL, resulting in an SNR of 0 dB. The babble noise was generated from the recording of six talkers' (3 males) simultaneous speech. They were asked to read some sentences matched in length and composed of common words in Chinese, which were selected from some books, journals, and newspapers. As the duration of babble noise is as short as 300 ms, it was hardly recognizable to listeners. To make the loudness and duration of all stimuli comparable, each stimulus was fixed at 70 dB SPL by modulating the root-mean-square amplitude of each sound, and was adjusted to 300 ms through PSOLA in Praat. Before the experiment, the volume was adjusted to a comfortable level for each participant. There are two reasons for the selection of 0 dB SNR in this study. Firstly, previous studies have shown that noise at 0 dB SNR had significant interference on speech perception (Gordon-Salant, 1985; Turner et al., 1992). Meanwhile, it is reported that the CP of consonants in 12-year-olds was still immature (Hazan and Barrett, 2000). Therefore, the teen-friendly SNR at 0 dB was selected in this study.

Tasks to measure different sub-components of AVWM included FDS, BDS, and NWR. There were two trials at each length, from two numbers (e.g., 3-7) to eight numbers (e.g., 4-6-2-4-7-9-0-2-1) for both FDS and BDS tasks. None of the lists contain easily memorable sequences (e.g., 1-2-3). There was a nonword list consisting of 20 trials at the length of one word (e.g., rài in Pinyin), two words (e.g., bōng nuá), and three words (e.g., rěi fáo shōng). All of the nonwords were nonsense syllables which were phonotactically illegal in Mandarin.

Procedure

The CP test included the identification task and the discrimination task. Both tasks were performed on a laptop through E-prime 2.0 (http://www.pstnet.com/eprime.cfm). The stimuli were presented to participants randomly through a headphone (Sennheiser HD280 Pro). Participants were asked to press key '1' on the keyboard when they heard /pa/ with Tone 1 and press key '2' when they heard /pa/ with Tone 2, or press key '1' when they heard /pa/ and press key '2' when they heard /p^{h}a/ in the tone or consonant aspiration identification tasks, respectively. Each stimulus was repeated five times.

The AX paradigm was used in the discrimination task. The speech stimuli were presented in pairs with a 500 ms inter-stimulus interval. Participants were asked to judge whether the two speech stimuli they heard were the same or not by pressing '1' for the same and pressing '2' for different on the keyboard. There were 10 stimuli pairs in the discrimination task, including six pairs which consisted of two different stimuli separated by two steps in either forward (1-3, 3-5, 5-7,) or reverse order (3-1, 5-3, 7-5), and four pairs which consisted of four stimuli each paired with itself (1-1, 3-3, 5-5, 7-7). Each stimuli pair was repeated four times and released in random order.

For AVWM tests, the participants heard a string of numbers and were asked to repeat it forward or backward immediately in digit span tasks. The testing session began with span lists of two numbers in length and only increased in length when at least one of the two sequences was correctly recalled. Besides, participants were asked to repeat the nonwords immediately in the NWR task. All of the nonwords were required to be recalled. The responses of listeners were verbal and the whole process was recorded by Praat for further data analysis.

There was a practice session before all tasks, making sure that participants understood the tasks and were familiar with the procedure. The order of four continua (i.e., tone and aspiration continua in the noise or quiet conditions) was counterbalanced across all participants. The experiment was completed in a quiet lab.

Data analysis

In the identification task, the identification score of each particular stimulus was defined as the average percentage of responses as Tone 1. The boundary position and boundary width of each participant were calculated using Probit analyses (Finney, 1971). The boundary position was defined as the 50th percentile and the boundary width was defined as the linear distance between the 25th and 75th percentiles. The narrow boundary width refers to a clear distinction between phonological categories.

In the discrimination task, the discrimination accuracy of each comparison unit was obtained by the formula: $P = P\ (\text{'S'}|S) \times P\ (S) + P\ (\text{'D'}|D) \times P\ (D)$, where P (S) represents the percentage of the same stimuli pairs (e.g., 1-1, 3-3) and P (D) represents the percentage of different stimuli pairs (e.g., 1-3, 3-1). P (‘S’|S) and P (‘D’|D) represent the percentage of ‘the same’ and ‘different’ responses to the same and different stimuli pairs, respectively (Xu et al, 2006). Between-category discrimination accuracy was defined as the mean value of discrimination accuracy between two stimuli pairs that stretch over the boundary position, and discrimination accuracy of within-category was defined as the mean value of discrimination accuracy of residual stimulus pairs.

In the digit span tasks, the score was defined as the repetition accuracy of digit combinations. Accuracies of FDS and BDS were calculated separately. In the NWR tasks, in view of the structure of Chinese syllables composed of initial, rhymes, and lexical tone, the score was defined as the mean of repetition accuracy of tones, initial consonants, and rhymes of nonwords.

Results

Identification and discrimination of tones

The identification curves of tones among adults and adolescents in both quiet and noise conditions are shown in Figure 2 (A). The between- and within-category accuracies of tone discrimination in the two groups are shown in Figure 2 (B). The boundary position and width are shown in Table 1.

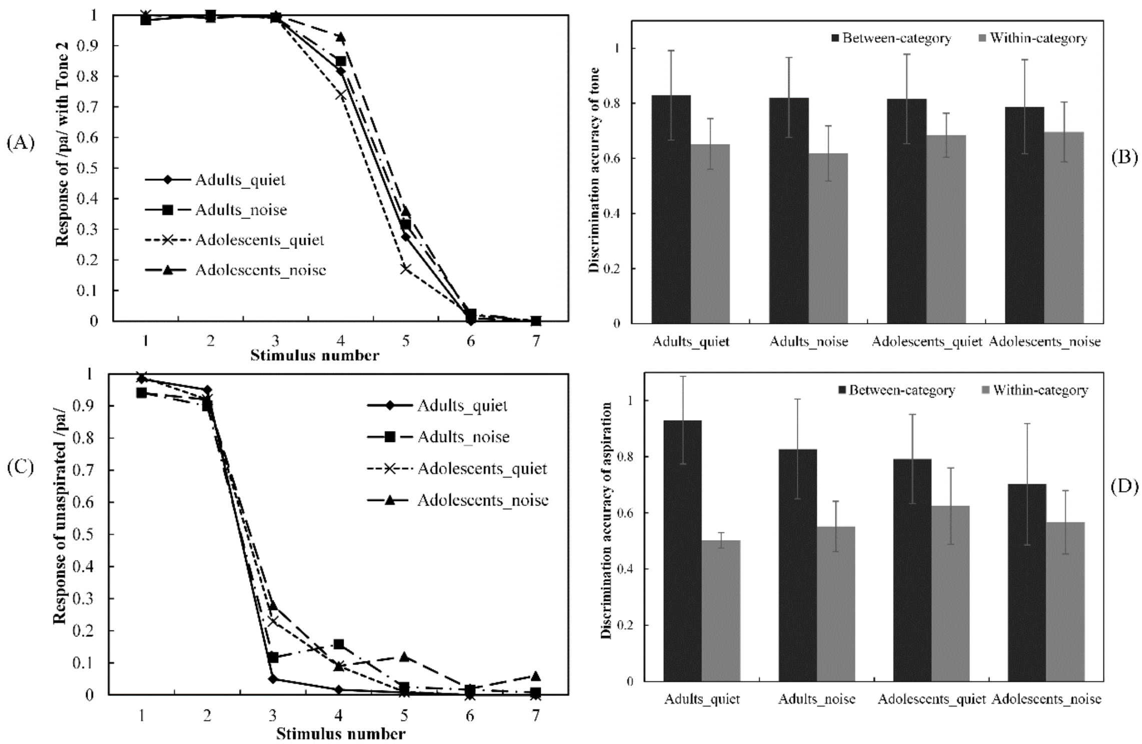

Figure 2. (A) Identification curves of tone perception; (B) Between-category and within-category accuracy of tone discrimination; (C) Identification curves of aspiration perception; (D) Between-category and within-category accuracy of aspiration discrimination. Error bars = ±1 SD.

Table 1. *Boundary position and width of tone and aspiration identification.*

	Tone identification				Aspiration identification			
	Adults		Adolescents		Adults		Adolescents	
	Quiet	Noise	Quiet	Noise	Quiet	Noise	Quiet	Noise
Boundary position (SD)	4.80 (0.50)	4.70 (0.56)	4.42 (0.43)	4.81 (0.44)	2.49 (0.14)	2.66 (0.59)	2.75 (0.33)	2.90 (0.92)
Boundary width (SD)	0.51 (0.04)	0.64 (0.32)	0.63 (0.33)	0.61 (0.34)	0.46 (0.08)	1.00 (0.66)	0.84 (0.38)	1.61 (1.59)

Results of 2 (*group*) × 2 (*condition*) repeated measures ANOVA showed a marginal significant main effect of *condition* (F (1,42) = 4.02, p = .052) and a significant interaction effect between *group* and *condition* (F (1,42) = 4.37, p = .043) on boundary position. Simple main effect analysis indicated that there was no significant difference in the adult group between the quiet and noise conditions (p = .95), while the boundary position of adolescents in the babble noise condition significantly moved to Tone 1 (p = .008), indicating that more sound stimuli were perceived as Tone 2 under the interference of noise. The main effect of *group* on boundary position was not significant (p = .074). There were neither significant main effects of *condition* (p = .76) and *group* (p = .51), nor an interaction effect on boundary width (p = .54), which indicated that there was no significant difference in the boundary width of tone perception between adults and adolescents, and babble noise at an SNR of 0 dB exerted no significant influence on the boundary width in both groups.

For tone discrimination, the results of 2 (*group*) × 2 (*condition*) × 2 (*category type*) repeated measures ANOVA showed a significant main effect of *category type* (F (1, 42) = 55.46, p < .001), and a significant interaction effect between *group* and *category type* (F (1, 42) = 4.37, p = .043). Simple main effect analysis showed that there was no significant difference in the accuracy of between-category discrimination between adults and adolescents (p = .47), while within-category discrimination accuracy of adolescents (mean = 0.69) was significantly higher than that of adults (mean = 0.63) (p = .011). There were neither significant main effect of *condition* (p = .36) and *group* (p

= .47), two-way interaction effect of *condition* and *group* (p = .69), *condition* and *category type* (p = .87), nor three-way interaction effect of *condition*, *group* and *category type* (p = .45).

Identification and discrimination of consonant aspiration

Figure 2 (C) shows the identification curves of consonant aspiration among adults and adolescents in both quiet and noise conditions. Figure 2 (D) shows the between- and within-category discrimination accuracies among the two groups in both quiet and noise conditions. The boundary position and width are shown in Table 1.

Results of 2 (*group*) × 2 (*condition*) repeated measures ANOVA showed that there was no significant main effect of *group* (p = .22) and *condition* (p = .89), nor the interaction effect on boundary position of aspiration identification (p = .07). But there was a significant main effect of *condition* on boundary width (F (1,42) = 11.23, p = .002). There was no significant main effect of *group* (p = .14), nor the interaction effect (p = .61) on the boundary width of aspiration identification. These indicated that there was no significant difference in boundary position and width between adults and adolescents, but babble noise had a significant negative effect on both groups, reflected by a wider boundary width of aspiration identification.

For aspiration discrimination, the results of 2 (*group*) × 2 (*condition*) × 2 (*category type*) repeated measures ANOVA revealed a significant main effect of *condition* (F (1,42) = 11.19, p = .002) and *category type* (F (1,42) = 70.12, p < .001), and a significant interaction between *category type* and *group* (F (1,42) = 11.08, p = .002). Simple main effect analysis indicated that between-category discrimination accuracy in adults was significantly higher than that in adolescents (p = .006), while within-category discrimination accuracy in adults was significantly lower than that in adolescents (p =

.004). In addition, there was a significant interaction between *category type* and *condition* (F (1,42) = 5.56, p = .023). Simple main effect analysis indicated that the discrimination accuracy of between-category phonemes in the quiet condition was significantly higher than that in the noise condition (p = .002), while within-category discrimination accuracy in both conditions were not significantly different (p = .83). There were neither significant main effect of *group* (p = .11), two-way interaction effect of *condition* and *group* (p = .12), nor three-way interaction effect of *condition*, *group* and *category type* (p = .13).

Digit span and nonword repetition

The distributions of accuracies of FDS, BDS, and NWR tasks of adults and adolescents are presented in Figure 3 (A). Independent t-test indicated that there was a significant difference between the nonword repetition accuracy of adults and adolescents (t (35) = 2.848, p = .007), while the difference of FDS and BDS between the two groups did not reach significance (ps > .05).

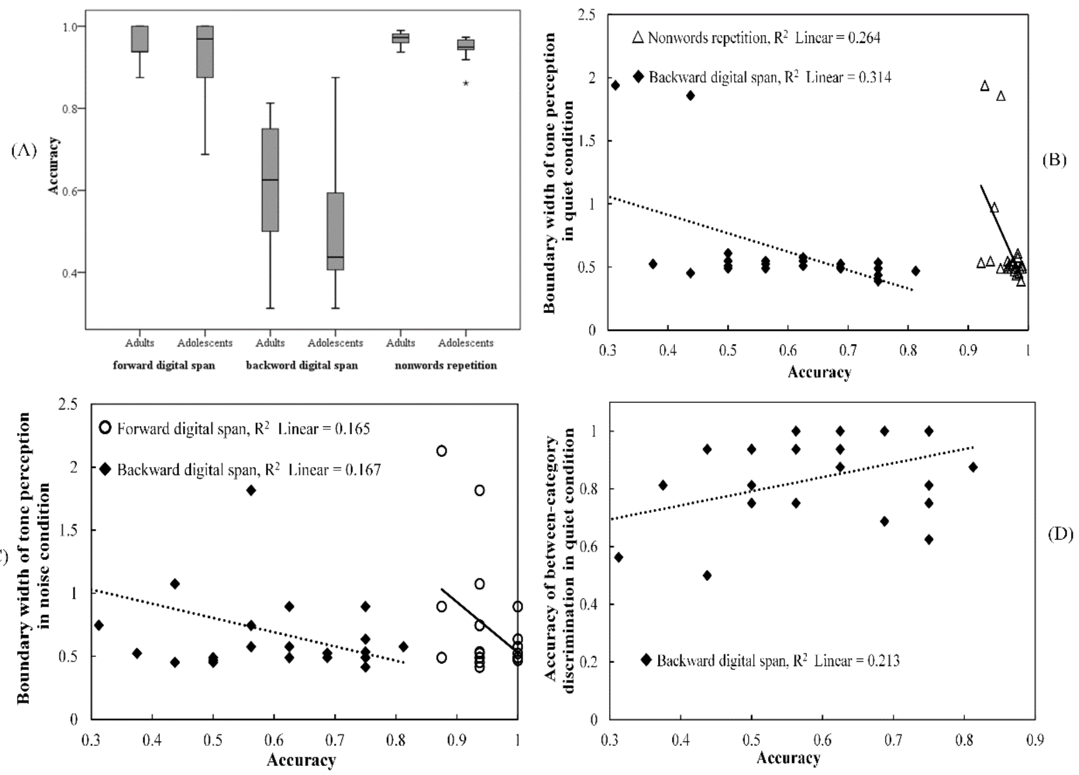

Figure 3. (A) Recall accuracy of AVWM tasks; (B) Correlation between the boundary width of tone perception in the quiet condition and recall accuracy of AVWM tasks in adults; (C) Correlation between the boundary width of tone perception in the noise condition and recall accuracy of AVWM tasks in adults; and (D) Correlation between the accuracy of between-category tone discrimination in the quiet condition and recall accuracy of AVWM tasks in adults.

Pearson correlation analysis indicated significant correlations between accuracies of AVWM tasks and CP of tones among adults. There were negative correlations between boundary width of tone perception in the quiet condition and accuracy of BDS task (r = -.560, p = .004), or accuracy of NWR (r = -.514, p = .010). There were also significant negative correlations between boundary width of tone perception in the noise condition and accuracy of FDS task (r = -.406, p = .049), or accuracy of BDS task (r = -.409, p = .047). Besides, there was a significant positive correlation between the accuracy of the BDS task and the accuracy of between-category discrimination of tone perception in the quiet condition (r = .462, p = .023). The corresponding scatterplots are shown in Figure 3 (B), (C) and (D), respectively. However, there was no significant correlation either between accuracies of AVWM tasks and CP of tones and aspiration of consonants among adolescents, or between accuracies of AVWM tasks and CP of consonant aspiration among adults. The correlation results between the accuracy of cognitive tests and CP performance are shown in Table 2.

Table 2. *The correlation between accuracy of cognitive tests and CP performance.*

Cognitive tests			FDS	BDS	NWR
Tone perception of Adults	Quiet	Boundary position	*r* = -.084	*r* = -.079	*r* = -.090
		Boundary width	*r* = -.239	*r* = -.560**	*r* = -.514*
		Accuracy of between-category discrimination	*r* = .331	*r* = .462*	*r* = .376
		Accuracy of within-category discrimination	*r* = -.348	*r* = .020	*r* = .232
	Noise	Boundary position	*r* = .310	*r* = .037	*r* = -.286
		Boundary width	*r* = -.406*	*r* = -.409*	*r* = -.194
		Accuracy of between-category discrimination	*r* = -.012	*r* = .313	*r* = .124
		Accuracy of within-category discrimination	*r* = -.152	*r* = .017	*r* = -.050
Tone perception of Adolescents	Quiet	Boundary position	*r* = -.393	*r* = -.367	*r* = .163
		Boundary width	*r* = -.420	*r* = -.048	*r* = -.295
		Accuracy of between-category discrimination	*r* = .265	*r* = .262	*r* = .139
		Accuracy of within-category discrimination	*r* = .209	*r* = .202	*r* = -.164
	Noise	Boundary position	*r* = -.195	*r* = .262	*r* = .271
		Boundary width	*r* = .070	*r* = .027	*r* = -.054
		Accuracy of between-category discrimination	*r* = -.007	*r* = .050	*r* = -.227
		Accuracy of within-category discrimination	*r* = .097	*r* = .112	*r* = .010
Aspiration perception of Adults	Quiet	Boundary position	*r* = -.264	*r* = .004	*r* = -.050
		Boundary width	*r* = -.082	*r* = -.047	*r* = .000
		Accuracy of between-category discrimination	*r* = .022	*r* = -.077	*r* = -.192
		Accuracy of within-category discrimination	*r* = -.113	*r* = -.055	*r* = .104
	Noise	Boundary position	*r* = -.080	*r* = .122	*r* = -.070
		Boundary width	*r* = .056	*r* = -.117	*r* = .135
		Accuracy of between-category discrimination	*r* = .275	*r* = .291	*r* = .196
		Accuracy of within-category discrimination	*r* = -.324	*r* = .012	*r* = -.196
Aspiration	Quiet	Boundary position	*r* = .061	*r* = .069	*r* = -.255

perception of Adolescents		Boundary width	r = .181	r = .260	r = -.007
		Accuracy of between-category discrimination	r = -.031	r = .008	r = .312
		Accuracy of within-category discrimination	r = .201	r = .238	r = -.257
	Noise	Boundary position	r = .370	r = .060	r = -.204
		Boundary width	r = -.388	r = -.110	r = .047
		Accuracy of between-category discrimination	r = .194	r = -.237	r = .303
		Accuracy of within-category discrimination	r = .162	r = -.070	r = .084

Note. CP = categorical perception, FDS = forward digit span task, BDS = backward digit span task, NWR = non-word repetition, * $0.01 < p < 0.05$, ** $p < 0.01$.

Discussion

In this study, we investigated the CP of Mandarin tones and consonant aspiration contrasts in babble noise among adults and adolescents aged 12-14 years, and explored the association between WM and CP. Results indicated that, firstly, in the noise condition, both adults and adolescents showed wider boundary widths and lower between-category accuracies when perceiving aspiration of consonants, and the categorical boundary of tone perception in adolescents showed a transitional tendency towards Tone 1. Secondly, discrimination of consonant aspiration in adolescents needed to be further developed. Lastly, the accuracy of NWR in adolescents was lower than that in adults, and adults with better AVWM had better performance on tone perception. These results were discussed in detail as below.

Effect of babble noise on the CP of Mandarin tones and aspiration of consonants

In this study, we found a negative effect of babble noise on certain aspects of CP of Mandarin tones and aspiration of consonants, which seemed to vary among different groups. Babble noise weakened the capacity of adults and adolescents to perceive aspiration of consonants, reflected by wider boundary widths and lower between-category accuracies, while babble noise influenced tone perception only in adolescents who showed a boundary moving towards Tone 1 in the noise condition.

For tone perception, the obtained boundary position in adolescents moved towards the ending of Tone 1 in the babble noise condition, while adults did not show the same tendency. Babble noise interferes with the CP of adolescents from spectral energy and language information, and distracts attention and working memory needed for CP of target speech as well (Brungart et al., 2001; Schneider et al., 2007). The generalized pulse-skipping hypothesis (Chiu et al., 2019; Feng et al., 2021) suggests that divided attention and WM resources could cause a loss of pulses in auditory inputs, and further reduce listeners' sensitivity to F0 information. Therefore, for adolescents with immature WM, they could not maintain a robust boundary position of tone perception in the noise condition. On the other hand, the accuracy of NWR, which is an indicator of sub-lexical knowledge, was lower in adolescents than that in adults. Therefore, it's also possible that adolescents performed a psychophysical boundary trend in the noise condition, for lacking abundant sub-lexical knowledge like adults. Otherwise, the babble noise exerted no significant effect on either adolescents' CP of tones in terms of boundary width and discrimination accuracy or overall CP performance in adults. Tone perception has been found to have robust noise resistance (Qi et al., 2017; Wang & Xu, 2020). According to Wang and Xu (2020), for example, the accuracy of tone recognition was approximately 80% at -6 dB SNR in eight-talker babbles. Therefore, it is likely that the SNR was not low enough to make a significant difference in CP of tones between the quiet condition and the noise condition at 0 dB SNR in the current study.

Unlike the only adolescents' moving boundary position of the CP of tones affected by noise, babble noise impoverished both groups' ability to perceive aspiration of consonants, reflected by wider boundary width and lower accuracy of between-category discrimination. Firstly, the duration of tones is longer (e.g., 300 ms in this study) than that of aspiration (e.g., 5 - 53 ms in this study). That is to say, aspiration changes faster and has lower clarity, therefore, perception of consonant aspiration was more difficult than that of tones (Bao & Lin, 2014). Moreover, F0 was noise-resistant relatively, while aspiration itself could be regarded as a kind of noise and the transient noise burst is an essential cue for perceiving VOT, which is easily masked by babble noise. Therefore, under the masking of babble noise, people perceived aspiration of consonants in a more adverse condition than tones, and were unable to obtain enough acoustic information.

Perceptual development of Mandarin lexical tones and aspiration of consonants

For the development of CP, Mandarin-speaking adolescents' perception of lexical tones had reached maturity basically in terms of boundary position, boundary width, and between-category discrimination, except for the moving boundary towards Tone 1 in the noise condition. And there was a significantly higher within-category accuracy of tones in adolescents than adults. Xu et al. (2006) found that English listeners showed better within-category discrimination sensitivity of pitch than Chinese, and pointed out that the peakedness of discrimination was a result of both an enhancement of between-category discrimination and a reduction of within-category discrimination. Therefore, it was revealed that tone discrimination of adolescents was still in the process of further development, for their higher within-category discrimination accuracy than adults.

The identification of consonant aspiration in adolescents had reached the adult level. However, the between-category accuracy was significantly lower in adolescents than adults, while the within-category accuracy was significantly higher in adolescents than adults. The results indicated that both between- and within-category discrimination skills still needed to be further developed, which was consistent with previous studies about the development of the CP of consonants in 10-year-old children and 12-year-old adolescents (e.g., Feng, 2022; Hazan & Barrett, 2000).

Comparing the perceptual results of tones and consonant aspiration, it could be found that CP of tones developed easier than that of consonant aspiration, which was also consistent with previous research. Feng (2022), for example, reported that the tone perception ability of six-year-old children had not been significantly different from that of adults, while the aspiration discrimination ability of ten-year-old children had not been maturely developed. According to the acoustic characteristics of phonemes, as mentioned above, compared with the tones, consonant aspiration shows a faster change and lower resolution, therefore, its perception was more difficult than tones. Tone perception tends to be easier to be acquired than aspiration perception for its clearer acoustic pattern. Furthermore, according to the phonological saliency hypothesis (Zhu & Dodd, 2000), the tone had a higher saliency than the aspiration of consonants in Mandarin, because it is compulsory for every syllable and there are fewer alternative phoneme choices than aspiration. Therefore, lexical tones could be acquired easier than the aspiration of consonants as a result of the higher saliency of tones.

Relationship between working memory and CP

A significant difference between the NWR accuracy of adults and adolescents was found in AVWM tasks, which might be the result of the gap in the phonological component of WM and language experience between the two groups. Because NWR was well-documented to be an indicator of sub-lexical knowledge, which was the representation of ‘chunks’ of phonemes in long-term memory (Szewczyk et al., 2018). It also might provide evidence that tone perception of native Mandarin adolescents perhaps perform a psychophysical boundary trend in the noise condition, for lacking abundant sub-lexical knowledge like adults. In addition, NWR has the maximum processing demand in AVWM tasks (Daneman & Merikle, 1996), therefore, the difference in NWR accuracy revealed that the processing capacity of adolescents might be lower than that of adults. Other than NWR, there was no significant difference in FDS and BDS accuracy between adults and adolescents. There might be a ceiling effect for both groups, because of the less processing demand of FDS and BDS than NWR.

For the correlations between CP and AVWM, it seemed that the roles of different sub-components of AVWM varied in CP of tones and aspiration of consonants among adults and adolescents in different conditions. Firstly, there was no significant correlation between any sub-components of AVWM and CP in adolescents. The processing capacity gap between adolescents and adults reflected by the NWR performance difference may suggest the immaturity of adolescents in the utilization of AVWM in speech perception. Secondly, with respect to adults, the correlation between sub-components of AVWM and CP of consonant aspiration was not significant, while accuracies of AVWM tasks were positively associated with performance in tone perception in some ways. According to the cue-duration hypothesis in Fujisaki and Kawashima (1970), the duration of the critical information of consonants was majorly responsible for the inferior auditory short-term memory, which stores information for subsequent speech processing. The acoustic cues (VOT) that distinguish aspiration of consonants are relatively short in duration and presumably cannot be stored well in memory, while the acoustic cues (F0) distinguish tones throughout the entire duration of the stimulus.

Thirdly, for CP of tones in adults, sub-components of AVWM contributed to its identification and discrimination in different conditions differentially. The significant correlations between the accuracy of BDS and performance of identifying tones in both conditions, and performance of discriminating tones in the quiet condition indicated that identification in both quiet and noise conditions and discrimination in the quiet condition took advantage of complex working memory, especially the executive component of AVWM (as assessed by BDS), which was necessary for encoding information in different domains (Baddeley, 2000). Besides, those with better sub-lexical knowledge (as assessed by NWR) and short-term memory (as assessed by FDS) performed better in the quiet and noise conditions respectively. It is possible that speech perception in the noise condition also needed to utilize short-term memory to store extra information of babble noise, and noise masking made it more difficult to extract categorical features and to link to long-term memory.

Limitation

Limit effect of babble noise on tone perception in this study may result from the 0 dB SNR. Therefore, CP, especially that of tone in more challenging babble with lower SNR and more talkers needs to be investigated in future studies, to find more interaction between noise and different age groups. In addition, adolescents over the age of 14 should also be considered in further research on the development of categorical speech perception, especially CP of consonants, since the consonant discrimination in adolescents aged 12-14 was still immature.

Conclusion

In this study, we examined CP of Mandarin lexical tones and aspiration of consonants in babble noise among native young adults and adolescents, and further explored the association between different sub-components of AVWM and CP. Results revealed that babble noise weakened the ability of adults and adolescents to perceive aspiration of consonants with a wider boundary and lower between-category accuracy, and influenced tone perception in adolescents who showed a moving boundary towards Tone 1 in the noise condition. Our results also provided evidence that the CP of consonant aspiration is acquired more difficultly than that of tones. Moreover, CP performance relates to the capacity and utilization of AVWM in some ways. Better AVWM might lead to better performance of tone perception in adults. Tone identification of adults in both quiet and noise conditions and discrimination in the quiet condition took advantage of complex working memory, especially the executive component of AVWM. And better sub-lexical knowledge and short-term memory contribute to better performances of tone identification in quiet and noise conditions, respectively.

Acknowledgments

This work was supported by grants from the National Social Science Fund of China (18ZDA296) awarded to Wei Zheng and the National Natural Science Foundation of China (11974374) awarded to Gang Peng.

References

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A-M. (2004). A structural analysis of working memory and related cognitive skills in young children.

Journal of Experimental Child Psychology, *87*(2), 85-106. https://doi.org/10.1016/j.jecp.2003.10.002

Baddeley, A. (1992). Working memory. *Science, 255*(5004)*,* 556-559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bao, H. Q, & Lin, M. C. (2014). *Summary of Experimental Phonetics (revised Ed.).* Peking University Press.

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer (Version 6.0.49)*. http://www.praat.org

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America, 109*(3), 2527-2538. https://doi.org/10.1121/1.1408946

Chiu, F., Rakusen, L. L., & Mattys, S. L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f0 for a synthesized vowel. *The Journal of the Acoustical Society of America, 146*(2), 1077–1084. https://doi.org/10.1121/1.5120404

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language, 44*(6), 1413-1434. https://doi.org/10.1017/S0305000916000581

Dees, T. M., Bradlow, A. R., Dhar, S., & Wong, P. C. M. (2007). Effects of noise on lexical tone perception by native and non-native listeners. In *Proceedings of the*

16th International Congress of Phonetic Sciences (pp. 817-820). Saarbrucken, Germany.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433. https://doi.org/10.3758/BF03214546

Feng, Y. (2022). *Categorical speech perception across the lifespan* (Doctoral dissertation). The Hong Kong Polytechnic University. https://theses.lib.polyu.edu.hk/handle/200/11620

Feng, Y., Meng, Y. R., Li, H. F., & Peng, G. (2021). Effects of cognitive load on the categorical perception of Mandarin tones. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3794-3802. https://doi.org/10.1044/2021_JSLHR-20-00695

Finney, D. J. (1971). *Probit Analysis (3rd Ed.).* Cambridge University Press. https://doi.org/10.1080/00401706.1973.10489027

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7*, 237–241. http://doi.org/10.1111/j.1467-9280.1996.tb00366.x

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological Psychology, 54*(1-3), 1–34. http://doi.org/10.1016/S0301-0511(00)00051-X

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Vol.29*, 207-214. Tokyo, Japan.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or lexical knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83-94. http://doi.org/10.3758/BF03210559

Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*(5), 672-388. http://doi.org/10.1006/jmla.1994.1032

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200–213. http://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177-190. https://doi.org/10.1037/0012-1649.40.2.177

Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41–76). Academic Press. https://doi.org/10.1016/B978-0-12-267350-4.50007-8

Gordon-Salant, S. (1985). Some perceptual properties of consonants in multi-talker babble. *Perception & Psychophysics 38*(1), 81-90. https://doi.org/10.3758/BF03202928

Hamann, C. (2017). Phonological working memory and language development: What are the measures and what do they measure? *Applied Psycholinguistics, 38*(06), 1313-1321. https://doi.org/10.1017/S0142716417000261

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics, 28*(4), 377-396. https://doi.org/10.1006/jpho.2000.0121

Ingvalson, E. M., Dhar, S., Wong, P., & Liu, H. (2015). Working memory training to improve speech perception in noise across languages. *The Journal of the Acoustical Society of America, 137*(6), 3477-86. http://doi.org/10.1121/1.4921601

Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., & Pine, J. M. (2014). Why computational models are better than verbal theories: the case of nonword repetition. *Developmental Science*, *17*(2), 298–310. http://doi.org/10.1111/desc.12111

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In *Proceedings of Annual Summit and Conference on Asia-Pacific Signal and Information Processing Association* (pp. 111-121). Sapporo, Japan. http://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/39651

Koelewijn, T., Zekveld, A. A., Festen, J. M., Rönnberg, J., & Kramer, S. E. (2012). Processing load induced by informational masking is related to linguistic abilities. *International Journal of Otolaryngology, 2012*(3), 1-11. https://doi.org/10.1155/2012/865731

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America, 119*(4), 2445-2454. https://doi.org/10.1121/1.2180210

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368. https://doi.org/10.1037/h0044417

Liberman, A. M., Eimas, P. D., Harris, K. S., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4*(4), 175-195. https://doi.org/10.1177/002383096100400401

Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L. & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10*(1), 1–28. https://doi.org/10.1177/002383096701000101

Lisker, L. & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences* (pp. 563–567). Prague: Academia.

Ma, J. Z., Zhu, J. Q., Yang, Y. X., & Chen, F. (2021). The development of categorical perception of segments and suprasegments in Mandarin-speaking preschoolers. *Frontiers in Psychology, 12*, 2427-2544. https://doi.org/10.3389/fpsyg.2021.693366

Massaro, D. W., Cohen, M. M., & Tseng, C. C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*(2), 267–289. https://doi.org/ 10.2307/23767517

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in french: comparing adults, adolescents, and children. *Journal of Phonetics, 38*(4), 493-503. https://doi.org/10.1016/j.wocn.2010.06.002

Millman, R. E., & Mattys, S. L. (2017). Auditory verbal working memory as a predictor of speech perception in modulated maskers in listeners with normal hearing. *Journal of Speech, Language, and Hearing Research, 60*(5), 1236-1245. https://doi.org/10.1044/2017_JSLHR-S-16-0105

Peng, G., Zhang, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624. https://doi.org/ 10.1016/j.wocn.2010.09.003

Qi, B., Mao, Y., Liu, J., Liu, B., & Xu, L. (2017). Relative contributions of acoustic temporal fine structure and envelope cues for lexical tone perception in noise. *The Journal of the Acoustical Society of America, 141*(5), 3022. https://doi.org/10.1121/1.4982247

Rönnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternäng, O., Wahlin, Å., & Nilsson, L-G. (2011). Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory. *Journal of Speech, Language, and Hearing Research*, *54*(2), 705-726. https://doi.org/10.1044/1092-4388(2010/09-0088)

Rönnberg, J., Lunner, T., Zekveld, A. A, Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*(31), 1-17. https://doi.org/10.3389/fnsys.2013.00031

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology, 18*(7), 559-572. https://doi.org/10.3766/jaaa.18.7.4

Simon, C. & Fourcin, A. (1978). Cross-language study of speech-pattern learning. *The Journal of the Acoustical Society of America, 63*(3), 925-935. https://doi.org/10.1121/1.381772

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition, 179*, 23-26. https://doi.org/10.31219/osf.io/vq69e

Turner, C. W., Fabry, D. A., Barrett, S., & Horwitz, A. R. (1992). Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research, 35*(4), 942–949. https://doi.org/10.1044/jshr.3504.942

Wang, Y. J., & Shangguan, X. N. (2004). How Japanese learners of Chinese process the aspirated and unaspirated consonants in standard Chinese. *Chinese Teaching in the World*, *3*, 54–66. https://doi.org/10.13724/j.cnki.ctiw.2004.03.008

Wang, W. S-Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*(2), 93–105. https://doi.org/10.1086/464946

Wang, W. S-Y. (1976). Language change. *Annals of the New York Academy of Sciences, 208*, 61-72. https://doi.org/10.1111/j.1749-6632.1976.tb25472.x

Wang, X. H., & Xu, L. (2020). Mandarin tone perception in multiple-talker babbles and speech-shaped noise. *The Journal of the Acoustical Society of America, 147*(4), EL307-EL313. https://doi.org/10.1121/10.0001002

Westerhausen, R., Bless, J. J., Passow, S., Kompus, K., & Hugdahl, K. (2015) Cognitive control of speech perception across the lifespan: A large-scale cross-sectional dichotic listening study. *Developmental Psychology, 51*(6), 806-815. https://doi.org/10.1037/dev0000014

Xi, J., Jiang, W., Zhang, L. J., & Shu, H. (2009). Categorical perception of VOT and lexical tones in Chinese and the developmental course. *Acta Psychologica Sinica, 41*(7), 572-579. https://doi.org/10.3724/SP.J.1041.2009.00572

Xu, C, Hui, Z., Wang, Y., & Ding, H. (2018). Informational masking of six-talker babble on mandarin Chinese vowel and tone identification: comparison between native Chinese and Korean listeners. *The Journal of the Acoustical Society of America, 140*(4), 3337. https://doi.org/10.1121/1.4970645

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063-1074. https://doi.org/10.1121/1.2213572

Zebib, R., Tuller, L., Hamann, C., Ibrahim, L. A., & Prévost, P. (2019). Syntactic complexity and verbal working memory in bilingual children with and without developmental language disorder. *First Language, 40*(4), https://doi.org/10.1177/0142723719888372.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Ronnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225-2234. https://doi.org/10.1121/1.4817926

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*(1), 3-42. https://doi.org/10.1017/S030500099900402X