The Role of Lexical Tone Information in the Recognition of Mandarin Sentences in Listeners with Hearing Aids

[Author names and affiliations]
Yuan Chen
Department of Special Education and Counselling, The Education University of Education, Hong Kong SAR, China.
Lena L. N. Wong
Clinical Hearing Sciences Laboratory, Division of Speech and Hearing Sciences, University of Hong Kong, Hong Kong SAR, China.
Jinyu Qian
Sonova China, Shanghai, China.
Volker Kuehnel
Sonova AG, Stäfa, Switzerland.
Solveig Christina Voss
Sonova China, Shanghai, China.
Fei Chen
Southern University of Science and Technology, Shenzhen, China.

[Corresponding author]
Yuan Chen
Department of Special Education and Counselling, The Education University of Education, Hong Kong SAR, China.

[Conflicts of interest]
The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

[Funding]
The work described in this article was partially supported by the Research Seed Fund 2018/2019 of the Department of Special Education and Counselling at the Education University of Hong Kong and by Phonak China.

Abstract

Objectives: Lexical tone information provides redundant cues for the recognition of Mandarin sentences in listeners with normal hearing in quiet conditions. The contribution of lexical tones to Mandarin sentence recognition in listeners with hearing aids (HAs) is unclear. This study aimed to remove lexical tone information and examine the effects on Mandarin sentence intelligibility in HA users. The second objective was to investigate the contribution of cognitive abilities (i.e., general cognitive ability, working memory, and attention) on Mandarin sentence perception when the presentation of lexical tone information was mismatched.

Design: A text-to-speech synthesis engine was used to manipulate Mandarin sentences into 3 test conditions: 1) a Normal Tone test condition, where no alterations were made to lexical tones within sentences; 2) a Flat Tone test condition, where lexical tones were all changed into Tone 1 (i.e., the flat tone); and 3) a Random Tone test condition, where each word in test sentences was randomly assigned 1 of 4 Mandarin lexical tones. The manipulated sentence signals were presented to 32 listeners with hearing aids in both quiet and noisy environments at an 8 dB signal-to-noise ratio.

Results: Speech intelligibility was reduced significantly (by approximately 40 percentage points) in the presence of mismatched lexical tone information in both quiet and noise. The difficulty in correctly identifying sentences with mismatched lexical tones among adults with hearing loss was significantly greater than that of adults with normal hearing. Cognitive function was not significantly related to a decline in speech recognition scores.

Conclusions: Contextual and other phonemic cues (i.e., consonants and vowels) are inadequate for HA users to perceive sentences with mismatched lexical tone contours in quiet or noise. Also, HA users with better cognitive function could not compensate for the loss of lexical tone

information. These results highlight the importance of accurately representing lexical tone information for Mandarin speakers using HAs.

INTRODUCTION

Pitch can express emotions and serve a variety of paralinguistic functions, such as marking emphasis and phrase boundaries in nontonal languages (e.g., English). In tonal languages (e.g., Mandarin), pitch also has lexical importance—it is used to distinguish words, analogous to consonants and vowels. Based on pitch pattern, the four Mandarin lexical tones can be characterized as follows: Tone 1 (flat and high), Tone 2 (rising), Tone 3 (low and dipping), and Tone 4 (falling) (Howie 1976). These pitch patterns result in different meanings despite being phonetically identical. For example, the syllable, /fa/ in Tone 3 means ‘law (法)’, and ‘hair (发)’ in Tone 4 (Y. Chen et al. 2014).

The importance of lexical tone information

The role of the lexical tone in Chinese speech perception has recently become a research topic of interest. Patel (2010) investigated the intelligibility of monotone (i.e., flattening F0) Mandarin sentences and suggested that flattening the F0 contour of Mandarin sentences did not significantly affect the intelligibility of sentences in quiet. However, speech understanding was significantly affected when babble noise was introduced at the 0-dB signal-to-noise ratio (S/N) (i.e., 60% intelligibility with flat-F0 sentences versus 80% intelligibility with natural sentences). Thus, F0 information seems to be redundant for Mandarin sentence perception in quiet. This may be partially attributable to the influence of top-down processing on perceptual organization. That is, listeners with normal hearing (NH) may employ top-down processing of contextual cues to compensate for the absence of F0 information. This assumption is supported by Wang et al. (2013) who studied the role of context on the perception of sentences with flat-F0 speech. They found that intelligibility in both quiet and noise was significantly poorer when contextual cues were removed. They also reported that flattening F0 contours did not reduce sentence intelligibility in quiet but made a significant difference in noise.

However, these studies examined the role of F0 on Mandarin sentence perception, rather than the contribution of tone information to sentence perception, as only primary cues (i.e., F0 height and contours) were removed while secondary cues (i.e., temporal and spectral envelopes) critical for the recognition of tones remained unchanged in these studies. Thus, listeners were able to use secondary cues for tone perception, in addition to primary intrinsic cues (Xu and Zhou, 2011). For example, Kuo et al. (2008) studied the contributions of F0, amplitude envelope, and tone duration for the perception of Mandarin tonal contrasts. They observed that the amplitude envelope partly contributed to tone recognition while tone duration only had a marginal effect in the absence of explicit F0 since differences in duration among the 4 Mandarin tones are miniscule in connected speech as shown in Yang et al. (2017). The unique contribution of an amplitude envelope may be due to the correlation between F0 and amplitude contours. Similarly, Liang (1963) found that Mandarin Chinese tone perception was fairly good (i.e., 60–70%) when F0 information was absent, while the spectral envelope was preserved. Kong & Zheng (2006) speculated that participants were able to use formant frequencies represented in the spectral envelope to match the F0 information.

Feng et al. (2012) used sine-wave replicas of Mandarin sentences to remove the contribution of primary and secondary cues from tone perception. They found that the intelligibility of Mandarin sentences remained intact in quiet even though tone perception was only slightly above chance. However, the contribution of lexical tone information to sentence perception in noise is still unclear. This is because many acoustic features such as harmonic structures important for vowel and consonant perception are lost in sine-wave speech, making it difficult to separate the contribution of phonemic and lexical tone information in sentence perception.

To solve this problem, F. Chen et al. (2014) manipulated lexical tones in 3 conditions (normal, flat, and random tone contours) using a text-to-speech (TTS) synthesis engine. There were 3 tone conditions in the study. In the normal tone (NT) condition, no alterations were made to lexical tones within sentences; the sentences were presented with natural tone contours. In the flat tone (FT) condition, all lexical tones were changed into Tone 1; thus, tone contour information was absent. For example, the sentence, " 他穿了一件灰格子上衣/ He wears a gray plaid jacket /da1 tʂhuan1 lɤ1 ji2 tɕiæn4 xuei1 kɤ2 tsi1 ʂɑŋ4 ji1/" was presented as "/da1 tʂhuan1 lɤ1 ji1 tɕiæn1 xuei1 kɤ1 tsi1 ʂɑŋ1 ji1/." In the random tone (RT) condition, each word of test sentences was randomly assigned 1 of 4 Mandarin lexical tones; thus, tone contour information was distorted. For example, the same sentence, " 他穿了一件灰格子上衣/He wears a gray plaid jacket /da1 tʂhuan1 lɤ1 ji2 tɕiæn4 xuei1 kɤ2 tsi1 ʂɑŋ4 ji1/" was presented as "/da2 tʂhuan3 lɤ1 ji4 tɕiæn1 xuei3 kɤ2 tsi1 ʂɑŋ3 ji1/. This approach could decouple the tone contour and the harmonic structure, thus preserving the harmonic structures to some degree (F. Chen et al. 2014). Results showed that for NH listeners, lexical tones were relatively redundant cues for sentence perception in quiet, but they were critical for perceiving Mandarin sentences with a 0-dB signal-to-noise-ratio (S/N).

Collectively, results from the studies described above suggest that absent or distorted lexical tone information has deleterious effects on the intelligibility of Mandarin sentences in noise but not in quiet, despite the various methods used to manipulate tone information. However, these studies were all conducted on Mandarin speakers with NH. The contribution of lexical tones to the intelligibility of Mandarin sentences among hearing aid (HA) users is unclear. Therefore, the aim of this study was to remove lexical tone information and examine the effects on the intelligibility of Mandarin sentences in HA users, using the method employed by F. Chen et al. (2014).

For NH listeners, previous studies have demonstrated negligible effects of tone contour on speech intelligibility in quiet. Nevertheless, we hypothesized that for HA users, (1) the lexical tone contour would be critical for perceiving Mandarin sentences even in quiet, thus a reduction in performance was expected in the FT and RT conditions; (2) the reduction in speech intelligibility would be worse when tone contour information was distorted (i.e., when tone contours were randomized) compared to when tone contour information was absent (i.e., sentences with flat tones); and (3) those with better cognitive function would perform better, particularly in more demanding situations (i.e., in noise and the RT condition).

Several reasons underpin these hypotheses. First, listeners with sensorineural hearing loss often exhibit compromised spectral and temporal perception acuity due to deteriorated hair cell function, leading to poor speech reception. Furthermore, hearing aids do not fully compensate for psychoacoustic function (Ewert et al. 2013). Second, in the case of hearing loss, speech signals are no longer fully audible, even with amplification. Third, although compression improves the perception of low-level signals (Kates, 2010), resultant spectral distortion may affect the perception of speech and tone contour information (Dillon 2012).

As mentioned above, the ability to process degraded information depends on the ability to use contextual cues to fill in the gaps and to integrate new information with earlier information (Akeroyd 2008). When acoustic cues are degraded, ambiguous, or masked by noise, cognitive resources must be allocated to recover the information (Wong et al. 2014; Xu et al. 2013). Although sentences presented in the FT condition exhibit a lack of tone contour information, we expected them to be less distracting, thus resulting in better intelligibility than those distorted with random tones (i.e., RT condition). Absent tone information may require listeners to disregard missing tone information and may not compel listeners to expend as much effort as that required to recover the appropriate meaning from distorted tones. Thus, distorted tones were expected to cause poorer performance than flattened tones.

Emerging evidence suggests that general cognitive ability, working memory, and attention significantly correlated with speech perception in adults with mild hearing loss (Akeroyd 2008; Heinrich et al. 2015, 2016; L. L. Wong et al. 2014). Therefore, we hypothesized that those with better cognitive function (i.e., general cognitive ability, working memory and attention) would perform better in perceiving sentences with mismatched lexical tone information. Given that factors such as lower education level, older age, and poorer hearing thresholds were significantly related to declined cognitive function and sentence perception ability (Koerner & Zhang, 2018; Wallhagen et al., 2008; Raz, 2000), these demographic factors were controlled for when examining the relationship between cognitive ability and sentence perception.

METHODS

Participants

A convenient sample of 32 participants (5 female and 27 male) were recruited from the Shengkang Hearing Center (Beijing). They had been living in Beijing for more than 10 years and were native speakers of standard Mandarin. Participants exhibited moderate to severe sensorineural hearing loss with interaural differences in audiometric thresholds less than 10 dB across the octave frequencies from 250 to 8,000 Hz (see Figure 1). Participants were aged between

28 and 81 years old (mean = 63, standard deviation [SD] = 15) and were all binaurally fitted. Mean HA use was 8 years (SD = 7, range = 1–32). On average, they had 9.5 (SD = 1.8, range = 6–16) years of education. Ethical approval was obtained from the Education University of Hong Kong and the University of Hong Kong. Written informed consent was obtained prior to the study. A transportation allowance of 200 RMB (approximately USD 25) was provided to each participant.

Materials

Sentence perception materials

The speech materials consisted of sentences from the Mandarin Hearing in Noise Test (MHINT) (L. L. Wong et al. 2007). The MHINT contains 24 lists of 10 sentences; each sentence is composed of 10 words. High inter-list reliability has been demonstrated, suggesting that the lists are equivalent and consistent results may be obtained using any list (L. L. Wong et al. 2007).

Each MHINT sentence was manipulated to yield 3 conditions using the NeoSpeech TTS software program (NeoSpeech, 2010) (i.e., NT, RT, FT). In the NT condition, the lexical tone of each word remained unchanged, and sentence quality was similar to natural speech produced at a normal conversational style. In the RT condition, each word within a sentence was randomly reassigned a tone by the TTS program (T1 to T4). Specifically, characters for each MHINT sentence were first presented as a string of Pinyin (Chinese phonetic symbols). Four lexical tones were presented using the Digits 1, 2, 3, and 4. By changing Digits, the lexical tone of each word was manipulated. Similarly, in the FT condition, lexical tones of each word were changed into FT (i.e., Tone 1) by typing Digit 1 for all words into the TTS program. While tone contours are being flattened, there is a chance that some sentences with quite a few words originally with Tone 1 would be less mismatched than other sentences. Also, there is a chance that although tone contours are being randomly assigned to the words, some sentences may accidentally end up with more correct tone contours. To solve this problem, 6 sets of test materials were created with each consisting of 12 sentence lists. These sentence lists were randomly assigned into 6 test conditions in each set. As a result, each list was assigned into different test conditions among these sets and each participant was administered one of these sets of tests, thus reducing the effects of "less mismatched" sentences. Furthermore, inspection of these sentences with mismatched tone contours suggest that there was not a significant difference in the mean number of words with mismatched tones in a sentence in the FT condition (mean = 7.13, SD = 1.62) or in the NT condition (mean = 7.23, SD = 1.73); $t(78) = -0.27$, $p > .05$).

All synthesized stimuli were produced at a sampling rate of 16,000 Hz at a normal conversional speaking rate using a female voice with a mean F0 of 240 Hz. Sentences were then normalized to the same root-mean-square (RMS) power. A steady-state-spectrum-shaped noise was used to mask the synthesized MHINT sentences at 8 dB S/N. To generate this noise, a finite impulse response filter was designed based on the average spectrum of all MHINT sentences, and white noise was filtered to have the same long-term average spectrum as the MHINT sentences. A S/N of 8 dB was chosen based on information provided by L. L. Wong et al. (2018) who reported that the mean speech reception threshold obtained using the MHINT was 8 dB S/N in HA users speaking Mandarin as a native language. The demographic characteristics of their participants were similar to those in the current study. Testing at +8 dB S/N would prevent direct comparison with normative data obtained at 0 dB S/N, based on F. Chen et al. (2014). However, based on our pilot study, 3 out of 4 participants scored at the floor (i.e., less than 10%) when tested at a 0-dB S/N. Thus, we decided to conduct testing in noise at an S/N of +8 dB.

Cognitive tests

The Montreal Cognitive Assessment (MoCA; Nasreddine, et al., 2005), a cognitive screening test, and 2 subtests from the CogState Battery (CSB, http://www.Cogstate.com), a sensitive, computer based assessment battery (Maruff et al. 2013), were used to evaluate cognitive function.

The MoCA assesses several cognitive domains: short-term memory, visuospatial abilities, executive function, attention, concentration, working memory, language, and orientation to time and place. It can be completed within 10 minutes and is designed to assess general cognitive function (Zheng et al. 2012). The Mandarin version was retrieved from the official website (http://www.mocatest.org/).

The One-Back test (OBK) and the Identification test (IDN), selected from the CSB, were used to assess working memory and attention, respectively. In the IDN test, a playing card was presented face-down in the center of the screen. Participants were required to press the “yes” button as fast as possible when a red card was flipped over and press the “no” button when a black card was flipped. In the OBK test, a playing card was presented in the center of a screen. Participants were instructed to decide whether the current card was the same as the previous card. These tests have been adapted into Chinese with reliability and validity (Zhong et al. 2013).

Procedures

A new pair of Phonak Bolero V90-P HAs was used during testing to avoid variations introduced by different HAs. Participants wore HAs for the experimental conditions only during this study. Participants' own custom ear-molds were used during testing.

Participants' hearing was first checked using pure-tone audiometry following procedures described in ANSI/ASA S3.21-2004 (R 2009). Insert gain based on the average real ear coupler difference (RECD) values was prescribed using the "Adaptive Phonak Digital Tonal" algorithm, the default formula for fitting Phonak hearing aids in Mainland China since 2015. Compensation for the individual vent-out effect was applied to the initial fitting based on results from the real ear test using the "real ear and feedback measurement" tool in Phonak Target fitting software (version 4.1). Further target matching was not conducted. Single channel noise reduction function (i.e., NoiseBlock) was set at position 8, corresponding to a moderate noise suppression setting. This was because L. L. N. Wong et al. (2018) reported position 8, compared to noise reduction being turned off, was better at improving speech perception and was preferred by most HA users for listening effort, listening comfort, and speech clarity in noise. Other adaptive parameters and functions (e.g., directional microphone, frequency lowering, and tinnitus noise) were turned off. Detailed information about the "Adaptive Phonak Digital Tonal" algorithm, the NoiseBlock 8 and HA fitting procedures can be found in Wong et al. (2018).

For individualized fine-tuning or adjustment, the "Northwind and Sun" passage (Holube et al. 2010) was presented at 65 dB (A) in quiet for users to adjust the gain of the HAs to ensure listening comfort and loudness balance between ears. A short, lively orchestral piece featuring a carillon and wind instruments was then played at 70 dB (A) to ensure good music quality. Testing was conducted in a soundfield in a double-walled, sound-treated room which met ANSI/ASA S3.1-1999 (R2013) standards for maximum permissible ambient noise levels for uncovered ears.

The MoCA, OBK, IDN, and sentence perception tests were conducted according to manual instructions in a random order. For sentence perception, the synthesized sentences and noise were presented via a loudspeaker located 1 m directly in front of the participant (0-degree azimuth). Before the test, each participant attended a training session and listened to 3 lists of sentences (i.e., 1 list each in the NT, FT, and RT conditions) to become familiar with the sentences and procedures. A printed copy of these 3 lists of sentences was provided to each participant. Participants were allowed to read the transcription of sentences during the familiarization stage. In the testing session, there were 6 conditions: 2 S/N levels (quiet and 8 dB S/N) × 3 tone conditions (FT, NT, and RT). Each condition consisted of 2 lists of 10 sentences. The order of lists was randomized across participants. Participants were instructed to repeat the sentences as well as they could and make their best guesses regarding the stated words. No feedback was given. Intelligibility was scored as the number of words correctly repeated. Only exact matches in pronunciation were accepted. The whole testing session for each participant took approximately 2 hours with a 15-minute break after the first hour. Participants were given more breaks upon request.

For the MoCA, OBK, and IDN, instructions were provided with HAs at optimal settings. Participants' understanding of test instructions was checked before test administration. To ensure that participants would not have difficulty seeing test stimuli, they were instructed to wear their prescription glasses during testing. None of the participants reported any problems completing these tasks due to issues with vision. Participants were requested to complete a test trial before the actual testing for the OBK and IDN assessments to minimize bias caused by the inability to hear or comprehend test instructions or see test stimuli.

RESULTS

Figure 2 shows the mean recognition scores in HA users from the present study compared to those obtained for NH listeners from F. Chen et al. (2014). Given the presence of floor and ceiling effects, percent correct scores were first converted to rational arcsine units (RAU) (Studebaker 1985). A one-way repeated measures analysis of variance (ANOVA) was then used to examine the effects of tone-processing conditions on percent correct scores in both quiet and noise.

In quiet, a Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(2) = 8.7$, $p < .05$; therefore, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .8$). Sentence perception was significantly affected by tone-processing conditions (F(1.7, 51.8) = 192.3, $p < .01$). Bonferroni post-hoc tests revealed that sentence recognition scores in the NT condition (mean = 80.0, SD = 13.4) were significantly higher than those in the RT condition (mean = 32.1, SD = 11.1, $p < .01$) and the FT condition (mean = 42.4, SD = 19.5, $p < .01$). Sentence recognition scores in the FT condition (mean = 42.4, SD = 19.5) were significantly higher than those in the RT condition (mean = 32.1, SD = 11.1, $p < .01$).

In noise, a Mauchly's test indicated that the assumption of sphericity was not violated; therefore, degrees of freedom were not corrected. Sentence perception was significantly affected by tone-processing conditions, (F(2, 62) = 234.8, $p < .01$). Bonferroni post-hoc tests revealed that sentence recognition scores in the NT condition (mean = 70.0, SD = 16.0) were significantly higher than those in the RT condition (mean = 23.3, SD = 12.7, $p < .01$) and the FT condition (mean = 27.0, SD = 16.0, $p < .01$). However, sentence recognition scores in the FT and RT conditions were not significantly different.

Mean percent correct scores were marginally reduced (by approximately 5 percentage points) in quiet among participants with NH when mismatched lexical tone information was introduced (see Figure 2). Intelligibility declined substantially among HA users (a reduction of 36 percentage points when comparing NT to FT conditions and 47 percentage points from NT to RT conditions). The decline in sentence perception performance from NT to RT or FT conditions among HA users (approximately 42 percentage points) was significantly greater than that in listeners with NH (approximately 5 percentage points) (1-sample t-test, $t(31) = 24.1$, $p < .001$). In noise, the mean decrease in sentence recognition scores from NT to RT or FT conditions was approximately 45 percentage points in HA users and approximately 25 percentage points in listeners with NH. As the noise level was set at 8 dB S/N in the current study while it was set at 0 dB S/N in F. Chen et al. (2014), direct comparisons in intelligibility in noise between the 2 samples could not be performed.

The effects of cognitive function

To evaluate the impact of cognitive function on speech perception when tone contour information was mismatched, the difference in scores obtained in NT and FT or RT conditions was calculated and related to cognitive function. Thus, D(NT-FT), D(NT-RT), D(NTN-FTN), and D(NTN-RTN) represent the decline in recognition scores from the NT to FT conditions in quiet, NT to RT conditions in quiet, NT to FT conditions in noise, and NT to RT conditions in noise, respectively. Table 1 shows the results for D(NT-FT), D(NT-RT), D(NTN-FTN), and D(NTN-RTN) as well as the general cognitive ability, attention, and working memory measured using the MoCA, OBK, and IDN tests, respectively.

The following linear mixed effects model was developed in R (R Core Team, 2014) to determine the statistical significance of each fixed effect, including age at testing, duration of HA use, hearing thresholds, education level, general cognitive ability, attention, and working memory skills on the logit-transformed D(NT-FT), D(NT-RT), D(NTN-FTN), and D(NTN-RTN) values.

$$y_{ij} = \alpha + \beta_{age:test}x_i^{age} + \beta_{aid:test}x_i^{aid} + \beta_{thre:test}x_i^{thre} + \beta_{edu:test}x_i^{edu} + \beta_{gen:test}x_i^{gen} + \beta_{att:test}x_i^{att} + \beta_{mem:test}x_i^{mem} + \tau_i + \varepsilon_{ij}$$

where y_{ij} is the j-th test score ($i = 1, \dots 4$) and $x_i^{(\cdot)}$ is the value of the covariate of the i-th participant. Note that we have separate regression coefficient β values for each test and we assume the effect is linear for the continuous covariates. Meanwhile, τ_i is the random effect for the i-th participant and ε_{ij} is the random error for the observation, which follows $N(0, \sigma_\tau^2)$ and $N(0, \sigma^2)$ meaning that the correlation between the test score of the same participant is $\sigma_\tau^2/(\sigma^2, \sigma_\tau^2)$. ‘Age’, ‘aid’, ‘thre’, ‘edu’, ‘gen’, ‘att’, and ‘mem’ represent the age at testing, duration of HA use, hearing thresholds, education level, general cognitive ability, attention, and working memory skills, respectively.

The model revealed that there was a significant effect of education level on D(NTN-RTN) (t(68) = -2.17, $p < .05$), but cognitive function was not significantly associated with D(NT-FT), D(NT-RT), D(NTN-FTN), or D(NTN-RTN) values (see Table 2).

DISCUSSION

While the presence of mismatched lexical tone information might have partly affected performance in quiet, this effect was slightly more pronounced in noise. In fact, the mean decrement in sentence intelligibility reached 40 percentage points in quiet and 45 percentage points in noise in HA users in the presence of mismatched lexical tone information. This emphasizes the importance of accurately representing tone information for Mandarin speakers using HAs. The results of the current study support the hypothesis that lexical tone contours are important for Mandarin sentence perception both in quiet and noise for HA users. These findings are different from those of previous studies conducted in participants with NH (F. Chen et al. 2014; Feng et al. 2012; Patel 2010; Wang et al. 2013); they reported that mismatched lexical tone cues did not impair sentence perception in quiet among participants with NH. As mentioned earlier, NH listeners were able to use contextual cues to fill in the gaps (Chen et al. 2014a) but HA users were not able to do so efficiently due to deficient hearing ability. That is, HAs do not provide sufficient compensation (Ewert et al. 2013) even for listening in quiet. In other words, other phonemic and contextual cues are inadequate for HA users to perceive sentences with mismatched lexical tone information. Similar findings were reported by Chen et al. (2018) in NH Mandarin speakers listening to acoustic signals low pass-filtered to simulate HA processing. Thus, our first hypothesis that HA users would experience a greater reduction in sentence intelligibility in both quiet and noise compared to NH listeners was supported.

One feature of the current study was the use of TTS to decouple the tone contour and the harmonic structure, thus preserving the harmonic structures to some degree. One may argue that sentences manipulated with TTS may not sound natural; therefore, results could have been adversely affected. However, anecdotal reports from most participants of the present study suggested that the synthesized speech had normal intonation both in quiet and noise partly because the NeoSpech TTS engine generated speech with Mandarin tonal sandhi. Furthermore, F. Chen et al. (2014) reported 100% intelligibility in quiet among NH listeners when sentences were processed with TTS. Thus, poor sentence intelligibility in quiet in the FT and RT conditions was unlikely due to whether TTS-synthesized sentences sounded natural. However, it is possible that unnatural sounds make it more difficult for HA users to perceive sentences in noise compared to their NH peers as HAs provide less redundant acoustic cues for sentence perception.

The second hypothesis, that speech intelligibility with distorted tones (i.e., the RT condition) was significantly worse than that with absent tonal information (i.e., the FT condition), was supported only in quiet, not in noise. A flat F0 contour made it easier to "track" the target stream and the FT condition made the task less cognitively demanding than the RT condition (Chen et al. 2018). However, it was not significantly more difficult to perform tasks in the RT condition than the FT condition in noise. This finding for hearing-impaired individuals differed from previous findings for NH listeners, where Chen et al. (2018) measured intelligibility using 2 types of sentences from a corpus of Chinese nonsense sentences: single-tone sentences containing Tone 1 and Tone 0 (the neutral tone) only (i.e., having a flat F0 contour) and multi-tone sentences containing all tones (i.e., having a varying F0 contour). They reported that intelligibility in steady noise was significantly better with flat-F0 sentences than with multi-tone sentences among Mandarin speakers with NH. These discrepant findings between NH and hearing-impaired participants might be related to the fact that the advantage of a flat F0 contour of "tracking" the target stream could be counteracted by the introduction of noise for HA users.

Finally, the linear mixed effects model indicated that no cognitive function had a significant relationship with the decline in speech reception when the lexical tone contours were mismatched. Therefore, the third hypothesis that higher cognitive function would be related to better speech recognition with mismatched tone contours was not supported. This indicated that the loss of tone information could not be compensated for by those who have better cognitive function, emphasizing the importance of accurately representing tone information via HAs. Only education level was found to be significantly related to sentence intelligibility in the RT test condition in noise. This significant relationship may be mediated by the proficiency of standard Mandarin

Chinese. Participants with higher education levels are more likely to exhibit better linguistic skills, allowing them to use top-down processing more efficiently to compensate for the loss of lexical tone information.

Limitations

One limitation of the current study is the lack of proper controls in the demographics of participants in the HA and NH groups. In the current study, we compared results from these groups with those of NH listeners from F. Chen et al., (2014). In F. Chen et al., (2014), participants were primarily graduate students at the University of Hong Kong aged between 23 and 30 years while participants in this study were primarily older adults with an average of 9.5 years of education. This study found that participants with higher education levels performed significantly better in the RT test condition while age was not significantly associated with performance in any test conditions. If we had included a control group with matched education levels, it is possible that the differences in performance in the RT test condition between these 2 groups would not be as large as those found in the current study.

Another limitation of this study is that all HA users were fitted with the Phonak Beolero V90-P HA and were using them for the first time, although they previously wore other Phonak HAs. If given time to adjust, it is possible that their performance would improve as HA benefits usually stabilize 4 to 6 weeks after fitting (Humes et al. 2012); also, after a person adjusts, we would expect better performance compared to what we observed in this study. However, it is doubtful that improvements will result in such sufficient adaptation that mismatched tone contours no longer matter for sentence recognition.

In addition, signals were processed with a dual compression algorithm, leading to more linear processing for speech-like signals compared to fast acting compression. As a result, spectral characteristics were better preserved and the gain prescribed using the “Adaptive Phonak Digital Tonal” fitting algorithm was 2–5 dB higher than National Acoustic Laboratories (NAL)-NL2 (tonal) range between 300 Hz and 2000 Hz with a comparable compression ratio; therefore, it is likely that low frequency cues important for tonal language are better preserved. Thus, care should be exercised when using these findings to analyze different HAs.

Conclusion

This is the first study to examine the contribution of lexical tone information on Mandarin sentence intelligibility in HA users. We found that speech intelligibility was significantly reduced (by approximately 40 percentage points) in the presence of distorted lexical tone information in both quiet and noise. Although lexical tone information provides redundant cues for Mandarin sentence perception in quiet for NH listeners, HA users were less able to use of contextual cues to understand speech with mismatched tones, suggesting that tone information should be preserved as much as possible to facilitate speech perception. In addition, distorted tones caused greater difficulty in speech perception than when tone contour information was missing in quiet but not in noise. Furthermore, the lack of a significant relationship between cognitive function and sentence perception with mismatched lexical tone information indicated that HA users with better cognitive function cannot compensate for the loss of lexical tone information in decoding sentences. These findings highlight the need for the accurate representation of lexical tone information in HAs for Mandarin speakers.

Acknowledgements

Y.C., L.L.N.W, J.Q., V.K., and S.C.V. designed experiments. Y.C. analyzed data and wrote the paper; L.L.N.W. provided critical revision. F.C. created the test materials. All authors discussed the results and implications and commented on the manuscript at all stages.

The work described in this article was partially supported by the Research Seed Fund 2018/2019 of the Department of Special Education and Counselling at the Education University of Hong Kong and by Phonak China.

Corresponding author
Yuan Chen

Department of Special Education and Counselling, The Education University of Education, Hong Kong SAR, China.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol, 47*, S53-S71.

Chen, F., Wong, L. L., Hu, Y. (2014). Effects of lexical tone contour on Mandarin sentence intelligibility. *J Speech Lang Hear Res, 57*, 338-345.

Chen, J., Yang, H., Wu, X., et al. (2018). The effect of F0 contour on the intelligibility of speech in the presence of interfering sounds for Mandarin Chinese. *J Acoust Soc Am, 143*, 864.

Chen, Y., Wong, L. L., Chen, F., et al. (2014). Tone and sentence perception in young Mandarin-speaking children with cochlear implants. *Int J Pediatr Otorhinolaryngol, 78*, 1923-1930.

Ewert, S. D., Kortlang, S., Hohmann, V. (2013). A Model-based hearing aid: Psychoacoustics, models and algorithms. *Proc Meet Acoust, 19*, 050187.

Feng, Y. M., Xu, L., Zhou, N., et al. (2012). Sine-wave speech recognition in a tonal language. *J Acoust Soc Am, 131*, EL133-EL138.

Heinrich, A., Henshaw, H., Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front Psychol, 6*, 782.

Heinrich, A., Henshaw, H., Ferguson, M. A. (2016). Only Behavioral But Not Self-Report Measures of Speech Perception Correlate with Cognitive Abilities. *Fronti Psychology, 7*, 576.

Holube, I., Fredelake, S., Vlaming, M., et al. (2010). Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol, 49*, 891-903.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones / John Marshall Howie*. Cambridge [Eng.] ; New York: Cambridge University Press.

Humes, L. E., Krull, V. (2012). Hearing aids for adults. In L. L. Wong, L. Hickson (Eds.), *Evidence-Based Practice in Audiology: Evaluating Interventions for Children and Adults with Hearing Impairment* (pp. 67). San Diego, CA: Plural Publishing.

Koerner, T. K., & Zhang, Y. (2018). Differential effects of hearing impairment and age on electrophysiological and behavioral measures of speech in noise. *Hear Res, 370*, 130-142. doi:10.1016/j.heares.2018.10.009

Kong, Y. Y., Zeng, F. G. (2006). Temporal and spectral cues in Mandarin tone recognition. *J Acoust Soc Am, 120*, 2830-2840.

Kuo, Y. C., Rosen, S., Faulkner, A. (2008). Acoustic cues to tonal contrasts in Mandarin: implications for cochlear implants. *J Acoust Soc Am, 123*, 2815-2824.

Liang, Z. A. (1963). Tonal discrimination of Mandarin Chinese. *Acta Physiologica Sinica, 26*, 85-91.

R Core Team (2018). R: *A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Maruff, P., Lim, Y. Y., Darby, D., et al. (2013). Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC Psychol, 1*, 30.

Patel, A. D., Xu, Y., and Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences. In *In Proceedings of Speech Prosody 2010-Fifth International Conference*. Chicago.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res, 28*, 455-462.

Team, R. C. (2014). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wang, J., Shu, H., Zhang, L., et al. (2013). The roles of fundamental frequency contours and sentence context in Mandarin Chinese speech intelligibility. *J Acoust Soc Am, 134*, EL91-EL97.

Wong, L. L., Soli, S. D., Liu, S., et al. (2007). Development of the Mandarin hearing in noise test (MHINT). *Ear Hear, 28*, 70S-74S.

Wong, L. L., Yu, J. K. Y., Chan, S. S., et al. (2014). Screening of cognitive function and hearing impairment in older adults: A preliminary study. *Biomed Res Int, 2014*, 867852.

Wong, L. L. N., Chen, Y., Wang, Q., et al. (2018). Efficacy of a Hearing Aid Noise Reduction Function. *Trends Hear, 22*, 2331216518782839.

Xu, L., Zhou, N. (2011). Tonal languages and cochlear implants. In *Auditory Prostheses* (pp. 341-364). New York, NY: Springer.

Yang, J., Zhang, Y., Li, A., et al. (2017). On the Duration of Mandarin Tones. In *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech 2017)* (pp. 1407-1411). Stockholm, Sweden.

Zheng, L., Teng, E. L., Varma, R., et al. (2012). Chinese-language Montreal cognitive assessment for Cantonese or Mandarin speakers: age, education, and gender effects. *Int J Alzheimers Dis, 2012*, 204623.

Zhong, N., Jiang, H., Wu, J., et al. (2013). Reliability and validity of the CogState battery Chinese language version in schizophrenia. *PLoS One, 8*, e74258.

Figure Legends

Figure 1. Mean pure-tone thresholds with standard deviations (SDs) presented as error bars.

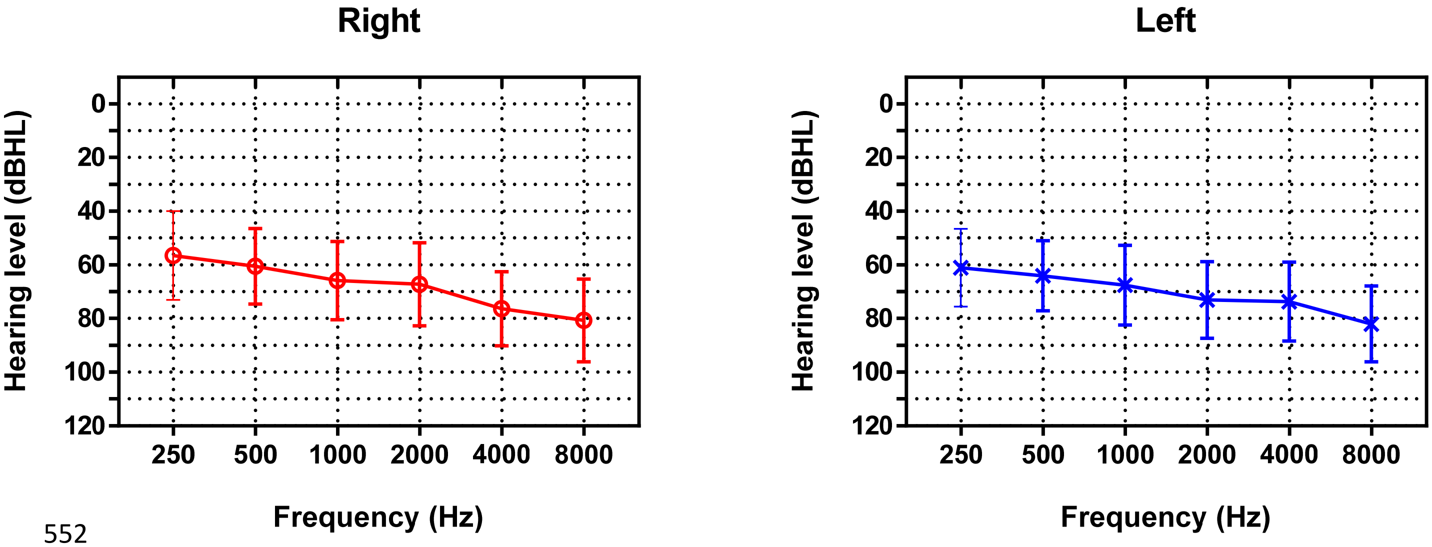

Figure 2. Sentence intelligibility of hearing aid (HA) users from the current study and normal hearing (NH) listeners from F. Chen et al. (2014). The * symbol indicates significantly worse within-individual performance in the FT and RT conditions than in the NT condition. A double arrow (⟷) suggests that HA users performed worse than NH listeners in the same tone condition. It should be noted that a direct comparison of performance in noise could not be made between HA users and NH listeners because results were obtained at different signal-to-noise ratios (S/Ns).

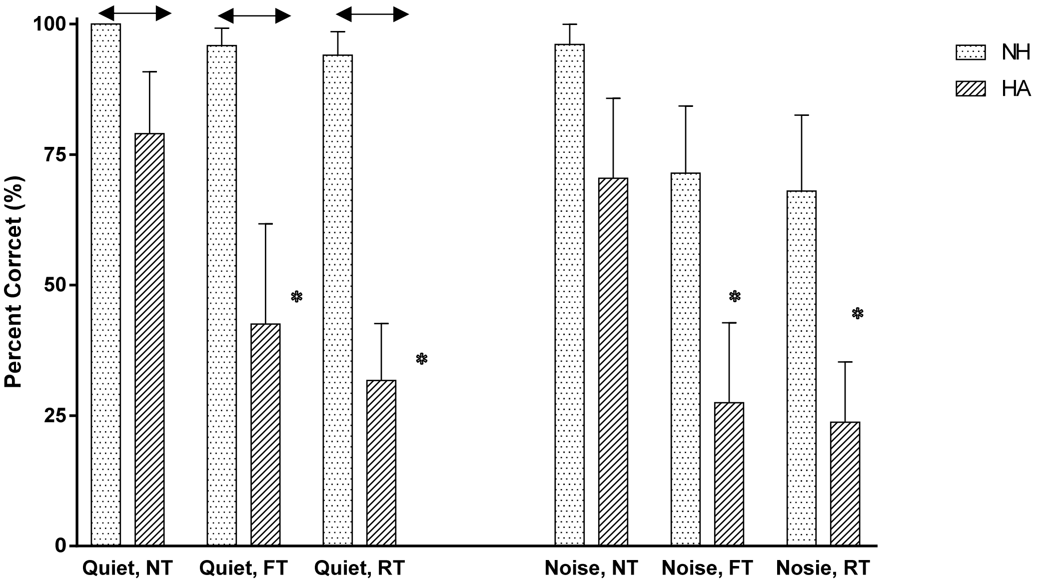

Table 1. Mean scores, standard deviations, and ranges of D(NT-FT), D(NT-RT), D(NTN-FTN), D(NTN-RTN), general cognitive ability, attention, and working memory.

	Mean	Standard deviation	Minimum	Maximum
General cognitive ability (MoCA score)	23.75	3.4	17	30
Attention (OBK score)	2.91	0.11	2.7	3.13
Working memory (IDN score)	3.07	0.16	2.6	3.32
D(NT-FT)	0.36	0.13	0.12	0.61
D(NT-RT)	0.47	0.12	0.2	0.7
D(NTN-FTN)	0.43	0.14	0.2	0.68
D(NTN-RTN)	0.47	0.13	0.2	0.7

D(NT-FT), D(NT-RT), D(NTN-FTN), D(NTN-RTN) represent decline in sentence recognition scores from NT to FT condition in quiet, NT to RT condition in quiet, NT to FT condition in noise, and NT to RT condition in noise, respectively.

Table 2. *t* statistics of the effects of age at testing, duration of hearing aid use, hearing thresholds, education level, general cognitive ability, attention, and working memory skills on the logit-transformed D(NT-FT), D(NT-RT), D(NTN-FTN), and D(NTN-RTN) values.

Effects	**D(NT-FT)**	**D(NT-RT)**	**D(NTN-FTN)**	**D(NTN-RTN)**
Age at testing	0.62	-0.38	-1.53	-0.24
Duration of hearing aid use	0.03	1.6	1.18	0.55
Hearing thresholds	0.59	0.21	-0.55	-0.25
Education level	1.90	-1.16	-2.17*	-1.28
General cognitive ability	-0.27	0.76	-0.73	1.08
Attention	0.13	0.87	1.34	1.66
Working memory	0.38	0.32	0.88	-0.62

**p<0.05*