



COVER PAGE

Document downloaded by @DAEL

Mon Oct 26 17:09:56 2026

For personal use

When automatic English translation is provided, only the original document is authentic.

The EAA cannot be held responsible of any translation error

Bibliographical reference

Acoustic Cues Utilized by Normal-hearing and Hearing-impaired Listeners Are Different for Mandarin Concurrent-vowels Identification, Zhen Fu, Hongying Yang, Xihong Wu and Jing Chen, *Acta Acustica* **vol. 104** (Number 5), 2018, pp. 792-795

DOI

<https://doi.org/10.3813/AAA.919225>

Acoustic Cues Utilized by Normal-hearing and Hearing-impaired Listeners Are Different for Mandarin Concurrent-vowels Identification

Zhen Fu, Hongying Yang, Xihong Wu, Jing Chen

Department of Machine Intelligence, Speech and Hearing Research Center, and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100081, China. fuzhen364@pku.edu.cn

Summary

Lexical Tones provide semantic meaning for Chinese words, and they are inherently bonded with vowels for Chinese speech. The present work examined the effect of lexical Tone contrast on concurrent-vowels identification for both normal-hearing (NH) and sensorineural-hearing-loss (SNHL) listeners. To ensure listeners utilize the primary cues of Tone perception, e.g. F0 contour, acoustic cues of mean F0 and durations were controlled for the speech stimuli. The result showed that benefit from Tone contrast existed for NH listeners, but was weak for SNHL listeners. Further analysis based on the measurement of Tone contrast suggested that F0 contour contrast and temporal envelope contrast were both highly correlated with identification performance for each of the listener groups. However, the importance of F0 contour was bigger than that of temporal envelope, especially for NH listeners. These findings would have implications for developing speech processing strategies for Mandarin-native SNHL listeners.

© 2018 The Author(s). Published by S. Hirzel Verlag · EAA. This is an open access article under the terms of the Creative Commons Attribution (CC BY 4.0) license (<https://creativecommons.org/licenses/by/4.0/>).

PACS no. 43.60.-c, 43.71.-k, 43.66.-x

1. Introduction

People usually have considerable difficulties in speech recognition in noisy environments, particularly people with hearing loss. For concurrent vowels identification, a simplified version of a noisy auditory scene, various perceptual cues are used to facilitate the segregation, e.g. F0 differences and F0 contours. It has been shown that identification could be improved by increasing the F0 difference between the two vowels in an appropriate frequency range [1]. For a given vowel pair, the frequency-modulated (FM) vowel could be perceived as more prominent than the unmodulated vowel [2]. It has been suggested that the auditory mechanism for detecting frequency modulation (FM) could alert listeners to the presence of FM sounds. These studies were based on synthesized vowels of English, but the related acoustic features for Tonal language, e.g. Mandarin Chinese, are rarely studied.

For Mandarin Chinese, the direction of F0 contour of vowels defines lexical Tones, which provide semantic meaning for Chinese words and hence are vital for speech perception [3]. There are four Tones in Mandarin, described phonetically as high level (T1), high rising (T2), rising falling (T3), and falling (T4), respectively. Naturally spoken vowels with T1 have almost flat F0 contours,

while vowels with T2-4 have varying F0 contours. In other words, the FM of F0 is an inherent characteristic for Mandarin concurrent-vowels identification, but its effect remains unclear.

Sensorineural hearing loss (SNHL), particularly cochlear hearing loss, is associated with degraded auditory functions, including FM detection and amplitude modulation detection [4, 5]. While F0 contour is the primary cue for Tone perception, listeners can also identify Tone category according to the temporal amplitude envelope of speech [6]. Therefore, how SNHL listeners utilize the acoustic cues associated with Tone contrast is the other question we addressed, which could be instructive to develop speech processing strategies of auditory prostheses for Mandarin-native users.

In summary, the present study was designed to 1) clarify the effect of Tone contrast on Mandarin concurrent vowels for both normal-hearing (NH) listeners and SNHL listeners; and 2) clarify the relative contribution of acoustic features (F0 contour vs. temporal envelope) on the effect of Tone contrast for the two listener groups.

2. Methods

2.1. Subjects

Twelve NH listeners (5 males, mean age = 23.3 years) and ten listeners with bilaterally symmetric SNHL (7 males, mean age = 43.2 years) participated in the experiment. All

Received 2 March 2018,
accepted 14 August 2018.

subjects were Mandarin-native speakers and were paid and given informed consent. The SNHL subjects were postlingually deaf with mild to moderate hearing loss. To ensure audibility of the stimuli over the frequency range important for speech intelligibility, the frequency-dependent gains prescribed by the “Cambridge” hearing-aid fitting procedure were applied to the stimuli when they were presented to SNHL subjects, based on individual audiograms [7].

2.2. Stimuli

Twenty four Mandarin single-vowel syllables comprising six vowels (IPA symbol: /a/, /ɔ/, /ɤ/, /i/, /u/, /y/) in combination with four Tones, were used. These syllable materials were originally spoken by a female Mandarin-native speaker and recorded digitally at a sampling rate of 44.1 kHz. To remove the confounding effect of duration [6, 8], duration of all syllables was fixed at 450 ms, including 15-ms raised-cosine rise and fall ramps. To avoid the effect of mean F0 difference between syllables, all original stimuli were manipulated using Praat [9] to have an equalized mean F0, while shapes of F0 contours remained unchanged. The RMS levels of all speech signals were equated before mixing. The vowel pairs were constructed by mixing any two different vowels, leading to a total of 276 (C_{24}^2) concurrent-vowels pairs.

2.3. Procedures

A custom graphical user interface (GUI) written in MATLAB was built to present stimuli and collect data. Stimuli were presented at the right ear for NH subjects, and at the better ear for SNHL subjects through a Sennheiser HD 650 headphone. The experiment was conducted in a sound-attenuation booth. The sound level was fixed at 55 dB SPL and 20 dB SL for NH subjects and SNHL subjects, respectively. Concurrent-vowel pairs were presented trial by trial in a random order, and each pair was presented once. Subjects were instructed to identify both tonal vowels (vowel plus Tone) on the GUI after one trial was presented. All subjects received a training session to familiarize them with the stimuli and procedure.

2.4. Data analysis

2.4.1. Identification score calculation

Identification scores were calculated from the percentage correct of “vowel plus Tone” identification, named as syllable identification. Each test trial received two responses, corresponding to the two individual syllables. The percentages were transformed to rationalized arcsine units for statistical purposes.

2.4.2. Tonal contrast measurement

For attributing identification scores to different acoustic features, the tonal contrast was calculated using perceptually relevant measures of F0 contour and temporal envelope. Specifically, the F0 contour of each syllable was extracted and transformed to the ERB_N scale [10]. The tem-

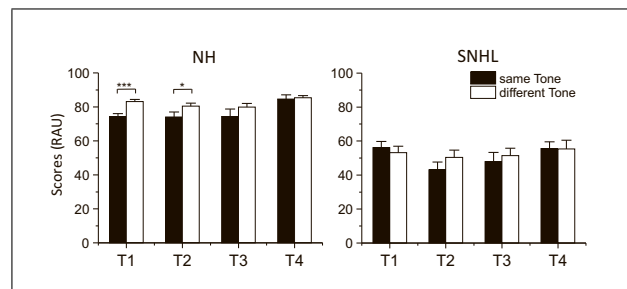


Figure 1. Averaged syllable identification scores as a function of Tone category and Tone contrast for NH (left) and SNHL (right) subjects. Error bars indicate one standard error, * $p < 0.05$ and *** $p < 0.001$.

poral envelope of each syllable was extracted and transformed to instantaneous loudness [11]. Subsequently, the contrast of each feature for all concurrent-vowels pairs was measured with Fréchet distance, as in Equation (1),

$$F(C_1, C_2) = \inf_{\alpha, \beta} \max_{t \in [0,1]} d[C_1(\alpha(t)), C_2(\beta(t))], \quad (1)$$

where C_1 and C_2 represent the feature curves for each of the syllable signals of one pair (F0 or instantaneous loudness), $C_1(\alpha(t))$ and $C_2(\alpha(t))$ are the resampling value of C_1 and C_2 at time t via the re-parameterization functions α and β , and the function d measures the distance between C_1 and C_2 . Finally the contrast values were normalized to the range [0,1]. A greater distance corresponds to a bigger contrast of the acoustic feature, and hence suggests a stronger cue for segregating the two vowels for a given concurrent-vowel pair.

3. Results and discussion

3.1. Effect of Tone contrast

Identification scores were organized as a function of Tone category and Tone contrast, as shown in Figure 1. For conditions of Tone contrast, “same Tone” (ST) and “different Tone” (DT) represent the absence and presence of Tone contrast, respectively. For NH subjects, identification scores for the DT condition is always greater than ST condition. A two-way ANOVA (4 Tone types $\times$ 2 contrast conditions) revealed that the main effects of Tone category ($F_{3,33}=9.698$, $p < 0.001$) and Tone contrast ($F_{1,11}=14.794$, $p=0.003$) were both significant. This result was consistent with previous studies that the manipulation of FM on one vowel could improve the identification performance [2]. For SNHL subjects, the main effect was significant for Tone category ($F_{3,27}=9.698$, $p=0.022$), but not for contrast ($F_{1,9}=1.798$, $p=0.213$), and their interaction was not significant ($F_{3,27}=2.416$, $p=0.088$). This results indicates a degradation of auditory function for SNHL subjects to utilize the cue of Tone contrast, and it was probably associated with their decreased FM discrimination ability [4]. It should also be noticed that for both groups, T4 scores were equally high for ST and DT conditions, suggesting that some acoustic features other than F0 contour probably played a role for the identification.

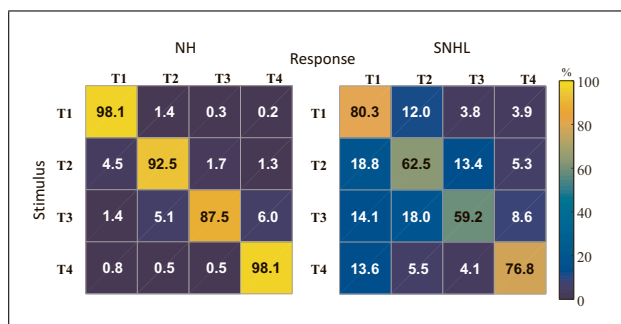


Figure 2. (Colour online) Averaged confusion matrices of Tone recognition for NH (left) and SNHL (right) subjects.

3.2. Confusion matrix of Tone identification

Details of the degradation of Tone identification for SNHL subjects are shown in the confusion matrix (Figure 2). For each matrix, the row and column represent the stimulus and the response Tones, respectively. The numerical value in each cell is the proportion of the indicated response-Tone to the indicated stimulus-Tone. Confusion of Tones (off-diagonal cells) occurred more frequently for SNHL subjects than for NH subjects. This result indicates that the SNHL subjects were less sensitive to Tonal cues.

3.3. Contributions of acoustic features

The identification scores are shown in Figure 3 as a function of F0 contour contrast (top panel) and loudness contour contrast (bottom panel), respectively. A general pattern of the data distribution is that the identification improves with increasing contrast. But for the further analysis, three Tone combinations should be carefully treated, which are T4-T4, T2-T3 for both groups and T1-T1 for the SNHL group. Identification scores for T4-T4 were high for both groups, relative to the small contrast in F0 contours. However, the contrast in loudness was relatively high. Hence, this probably indicated that the effect of temporal envelope was dominant for T4-T4 condition. The identification scores for T2-T3 were low for both groups, relative to the moderate contrast in both F0 contour and loudness contour. This result is consistent with a previous report that the perceptual cue for discriminating T2 and T3 was the timing of the turning point of T3, and they were easily confused because of the similarity of F0 contours after the turning point [12]. Hence the perceptual contrast of T2-T3 was a bit different from the contrast measurement used in this study. Scores of T1-T1 were extremely high for SNHL subjects, relative to the small contrasts. This suggested that other auditory mechanisms than Tone contrast, e.g. the preference for flat streams (associated with top-down processing), probably dominated their performance in this condition. When Tone combinations above (open symbols in Figure 3) were excluded from the linear regression analysis, scores increase linearly with the contrast for both acoustic features, and the correlation was significant for each of the groups (Pearson's r and p -values are shown in Figure 3).

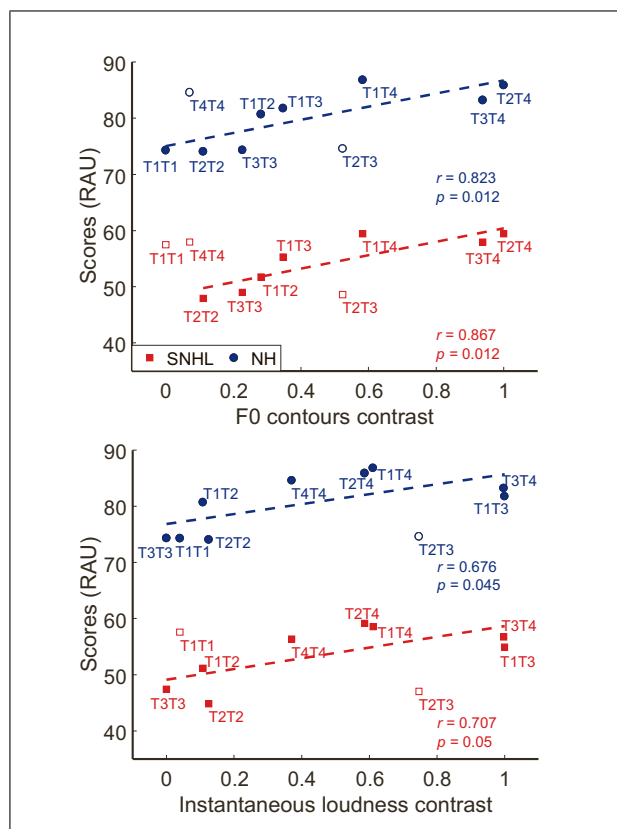


Figure 3. (Colour online) Averaged identification score as a function of F0 contours (top) and instantaneous loudness (bottom) contrast for all Tone-pairs for NH (blue circles) and SNHL subjects (red squares). Open symbols are data points excluded from the regression analysis.

A 3-dimension linear regression was conducted to explore the relative contribution of F0 contours and temporal envelopes, as in Equation (2),

$$S = W_{F0} \times C_{F0} + W_{ENV} \times C_{ENV} + b, \quad (2)$$

where S represents identification scores, W_{F0} and W_{ENV} represent the weight of F0 contours and temporal envelopes, respectively. C_{F0} and C_{ENV} represent the contrast value of F0 contour and temporal envelope, respectively. The parameter b represents the intercept value.

For the Tone combinations where two features were co-working (T1-T1, T4-T4 and T2-T3) were excluded from regression analysis as in Figure 3), the relative contribution was calculated as the ratio of W_{F0}/W_{ENV} , and the ratio was 2.4 ($p=0.04$) and 2 ($p=0.04$) for NH and SNHL subjects, respectively. These two greater-than-1 ratio values indicated that the F0 contour contrast played a major role in most Tone combinations for both groups. It was consistent with the general knowledge that F0 contour is the primary cue for perception of Mandarin Tone [6]. The decrease of ratio value from 2.4 for NH group to 2 for SNHL group indicated the importance of F0 contour weakened for SNHL listeners. It was consistent with the previous study by Wang et al. In their work, the relative contribution of temporal envelope and fine structure was investigated

with hearing impaired listeners by measuring Tone identification [13]. This trend is in agreement with the degraded ability to utilize F0 contour cues [4] and preserved ability to utilize temporal envelope cues for SNHL subjects [14].

Contemporary speech processing strategies for cochlear implants (CIs) which emphasize temporal more than spectral cues, are less favorable for tonal language speakers due to the diminution in F0 cues [8]. This was confirmed by the present result that F0 contour contrast played a more important role than temporal envelope in Mandarin concurrent-vowels identification. Although diminished spectral cues could be compensated by temporal cues based on their trade-off relation [8], encoding of varying F0 is still inadequate, especially when fluctuating maskers are presented. So processing strategies conveying more robust F0 information of signals are crucial for Mandarin speech perception. The qualitative relative importance of the two features described in the present study suggests that the representation strength should be mediated according to limits of listener's spectral and temporal resolvability.

4. Conclusion

The present work examined the effect of Tone contrast on Mandarin concurrent-vowels identification for NH and SNHL listeners, and the relative contribution of F0 contour and temporal envelope to the effect of Tone contrast was also analyzed and discussed. The main findings included: (1) The identification of vowel pairs with different Tone was significantly better than vowels pairs with same Tone for NH listeners, but the effect was not significant for SNHL listeners. (2) When the Tone contrast was measured with F0 contour and temporal envelope, identification performance was highly correlated with Tone contrast for both listener groups, consistent with the premise that Tone contrast cue worked for the vowel pairs. (3) The importance of F0 contours was greater than that of temporal envelopes for both groups. (4) SNHL listeners relied more on temporal envelopes than NH listeners.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61473008, 61771023, and 11590773).

References

- [1] P. F. Assmann, Q. Summerfield: Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J. Acoust. Soc. Am.* **88** (1990) 680–97.
- [2] J. F. Culling, Q. Summerfield: The role of frequency modulation in the perceptual segregation of concurrent vowels. *J. Acoust. Soc. Am.* **98** (1995) 837.
- [3] J. Chen, H. Yang, X. Wu, B. C. J. Moore: The effect of F0 contour on the intelligibility of speech in the presence of interfering sounds for Mandarin Chinese. *J. Acoust. Soc. Am.* **143** (2018) 864–877.
- [4] B. C. J. Moore, E. Skrodzka: Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. *J. Acoust. Soc. Am.* **111** (2002) 327–335.
- [5] S. P. Bacon, N. F. Viemeister: Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology* **24** (1985) 117–134.
- [6] D. H. Whalen, Y. Xu: Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* **49** (1992) 25–47.
- [7] B. C. J. Moore, B. R. Glasberg: Use of a Loudness Model for Hearing-Aid Fitting. I. Linear Hearing Aids. *Br. J. Audiol.* **32** (1998) 317–335.
- [8] L. Xu, Y. Tsai, B. E. Pflugst: Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J. Acoust. Soc. Am.* **112** (2002) 247–58.
- [9] P. Boersma, D. J. M. Weenink: Praat : doing phonetics by computer [Computer program]. <http://www.praat.org/> (2013).
- [10] B. R. Glasberg, B. C. Moore: Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* **47** (1990) 103–138.
- [11] B. R. Glasberg, B. C. J. Moore: A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* **50** (2002) 331–342.
- [12] X. S. Shen, M. Lin: A Perceptual Study of Mandarin Tones 2 and 3. *Lang. Speech* **34** (1991) 145–156.
- [13] S. Wang, X. Li, R. Mannell: Relative Contributions of Temporal Envelope and Fine Structure Cues to Lexical Tone Recognition in Hearing-Impaired Listeners. *J. Assoc. Res. Otolaryngol.* **12** (2011) 783–794.
- [14] J. Schlittenlacher, B. C. Moore: Discrimination of amplitude-modulation depth by subjects with normal and impaired hearing. *J. Acoust. Soc. Am.* **140** (2016) 3487.