INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Understanding Periodically Interrupted Mandarin Speech

Jing Liu [1], *Rosanna H.N. Tong* [2], *Fei Chen* [3]

[1] Department of Biology, Southern University of Science and Technology, Shenzhen, China
[2] Division of Speech and Hearing Sciences, The University of Hong Kong, Hong Kong
[3] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

fchen@sustc.edu.cn

Abstract

This study investigated the effects of two parameters (i.e., interruption rate, and duty cycle of interruption) on the perception of periodically interrupted Mandarin speech. Normal-hearing listeners were instructed to identify consonant/vowel/tone/word from isolated Mandarin words and recognize Mandarin sentences when they were temporally interrupted by square wave. Results showed that consistent with earlier findings obtained with English speech, interruption with a large rate or duty cycle favored the perception of periodically interrupted Mandarin speech. In addition, for isolated Mandarin word, the perception of vowel or tone was less affected by periodical interruption than that of consonant, and under periodical interruption the perception of consonant could largely account for the recognition of Mandarin word. For Mandarin sentence, the tonal characteristics and the simpler syllable structure in Mandarin might facilitate spectral-temporal integration of the target words, which contributed to a sentence intelligibility advantage of Mandarin over English under interrupted conditions.

Index Terms: Speech intelligibility, periodical interruption.

1. Introduction

Adverse listening conditions, such as environments with significant background noise or competing speech, might pose challenges to our daily speech communication. Despite the discontinuities of spectro-temporal information in the original signal, speech perception is robust for young normal-hearing (NH) listeners in most listening environments [e.g., 1-3]. Previous findings from various interrupted speech studies proposed "glimpsing" or "multiple looks" models of temporal integration, to investigate how listeners fused audible spectro-temporal fragments of the original signal for meaning restoration over a wide range of interruption conditions [e.g., 1, 3-7]. A "glimpse" or "look" is defined as "an arbitrary time-frequency region which contains a reasonably undistorted view of the target signal" [4]. Many research findings about the effects of interruption parameters (e.g., speech proportion, interruption rate, and duty cycle) on "glimpses" were consistent, and indicated that any change in these parameters would exhibit certain effects on speech understanding.

Speech proportion has been found to be the prime predictor of the speech recognition performance in isolated word and sentence contexts, as the effects of other interruption parameters [e.g., interruption rate, and duty cycle (the proportion of a given cycle during which speech is presented)] on speech understanding varied depending on the speech proportion presented [1, 3, 6, 8-9]. The higher the proportion of speech that listeners could "glimpse", the greater the improvement in speech intelligibility [e.g., 3]. Yet, when the speech proportion was held constant, speech recognition score tended to be worse at slower interruption rates less than 4 Hz, and got better as the rate increased above 4 Hz until reaching a certain rate at which the score started to roll-off [1, 5, 10-11]. This variance in speech recognition score depended on how many phonemes could be "glimpsed" by listeners and how much interference with temporal envelope [1, 10]. Extraneous interruptions at low (i.e., less than 4 Hz) and very fast modulation rates might disrupt the temporal envelope cues and eliminate the entire phonemes of the target word, causing a reduction in speech intelligibility [3, 12-14].

Regarding duty cycle, previous findings revealed that larger speech duty cycle would improve speech recognition performance when the speech proportion was ≥0.5, and this performance tended to deteriorate when the speech proportion was low (i.e., <0.5) [3, 6, 9]. This variance in speech recognition performance depended on the number of phonemes sampled. Larger speech duty cycle for this proportion (i.e., <0.5) limited the available "glimpses" of vowels or consonants. In contrast, smaller speech duty cycle for this proportion (i.e., <0.5) could give rise to multiple samples of different phonemes, thus facilitating temporal integration process for interrupted speech and improving listeners' speech intelligibility.

Some previous findings also suggested that the location of the "glimpses" and the language examined would make a difference to speech intelligibility. In isolated English word context, Owren and Cardillo suggested that listeners could discriminate English lexical meaning better with consonant-only isolated words using a same/different judgement task [15]. However, regarding Mandarin, it was found that there was a greater contribution of vowels than consonants to the intelligibility of isolated word context [16]. Regarding English sentence context, Cole *et al*. showed a 2:1 intelligibility advantage of vowels over consonants [17]. As for Mandarin, the vowel-only sentences showed a much larger intelligibility advantage over the consonant-only sentences [18]. The greater contribution of vowels over consonants to Mandarin speech intelligibility might be attributed to the phonetic differences between the two languages. The tonal characteristics and the simpler syllable structure (i.e., without consonant cluster) of Mandarin help to restrict lexical competition in both isolated

http://dx.doi.org/10.21437/Interspeech.2016-176

word and sentence contexts, thus giving an intelligibility advantage of interrupted words and sentences over English [18-19].

Most of previous studies concerning the effects of interruption parameters on speech recognition were English-based studies for the Western NH or hearing-impaired population. Although the perceptual contributions of vowels and consonants to Mandarin word and sentence intelligibility were assessed using segmental interruption [e.g., 16, 18], there is very limited research evidence regarding the effect of periodical interruption on Mandarin speech understanding. Therefore, the purpose of this study is to examine the effect of periodical interruption on the recognition of Mandarin speech in both isolated word and sentence contexts and with the use of square-wave-based periodical interruption, in order to justify how NH listeners make use of the available "glimpses" to restore meaning. Although this type of stimulus interruption may be different from the noisy listening conditions that listeners encounter in their daily life, it can introduce longer transients in speech waveform and thus enable a greater control over the available "glimpses" by ensuring that speech cues are limited to the preserved speech intervals [14].

2. Experiment 1: Isolated Mandarin Words

The purpose of this experiment was to investigate the effect of periodical interruption on consonant/vowel/tone identification and isolated word recognition in Mandarin.

2.1 Subjects and materials

Sixteen NH native-Mandarin participants attended this experiment. The test materials used in this experiment were taken from a database of 1128 isolated Mandarin (monosyllabic) words, covering almost all daily-used words in Mandarin Chinese. All the words were spoken in isolation by a female native-Mandarin talker at a normal speaking rate and with broadcaster's voice quality. The fundamental frequency (F0) of recorded words ranged from 130 to 330 Hz [16].

2.2 Signal processing

A square wave was first generated with values taking 0 and 1, and its initial phase wave was randomly assigned. The duty cycle of the square wave is the percent of the period in which the wave value is 1. This experiment chose two duty-cycle values of 50% and 33%, and three rates of square wave, i.e., 4, 8 and 16 Hz. Finally the periodical square wave (with the same length of the target speech signal) was multiplied with the target speech signal to generate the periodically interrupted stimulus, where the rate of the square wave was the interruption rate applied to the target speech signal.

2.3 Procedure

The experiment was conducted in a sound-proof booth. Participants listened to the stimuli through a circumaural headphone and played at a comfortable listening level. Each participant listened to six [= 3 interruption rates (i.e., 4, 8 and 16 Hz) × 2 duty-cycle values (i.e., 50% and 33%)] conditions which were presented as blocks in a random order to minimize the effects of practice across participants. For each condition, 40 words were randomly taken from the word database for signal processing. Participants were allowed to listen to each stimulus at most three times. They were asked to choose the corresponding HanYu PinYin (a standard phonetic writing system in Mainland Chinese) from a MATLAB interface showing the 21 consonants, 35 vowels, and 4 tones in Mandarin [18]. Before the experiment, all participants listened to 40 processed words (interrupted with 4 Hz and 50% duty cycle) as practice to familiarize them with the periodically interrupted stimuli and experimental procedure, and feedback was given in the practice session. The responses collected in the experiment were scored based on four aspects, i.e., consonant, vowel, tone and word. The score for each condition was computed as the ratio between the number of the correctly recognized consonants/vowels/tones/words and the total number (i.e., 40) of stimuli tested in each condition. Note that a word was correctly recognized only when all its consonant, vowel and tone were correctly identified. A 5-minute break was given in every 30 minutes. The whole experiment took about one hour per participant.

2.4 Results and discussion

Figure 1 shows the mean consonant, vowel, tone and word recognition scores at all interrupted conditions. Statistical significance was determined by using the percent correct score as the dependent variable, and interruption rate (i.e., 4, 8 and 16) and duty cycle (i.e., 50% and 33%) as the two within-subjects factors.

For consonant identification result in Fig. 1 (a), two-way analysis of variance (ANOVA) with repeated measures indicated a significant effect of interruption rate (F[2,30]=7.91, p<0.01), duty cycle (F[1,15]=70.82, p<0.001), and a non-significant interaction between interruption rate and duty cycle (F[2,30]=0.58, p>0.05). At the same interruption rate, post hoc pairwise comparisons showed significant (ps<0.05) difference between paired scores at the two duty cycles.

For vowel identification result in Fig. 1 (b), two-way ANOVA with repeated measures indicated a significant effect of interruption rate (F[2,30]=1.85, p>0.05), duty cycle (F[1,15]=16.99, p<0.01), and a non-significant interaction between interruption rate and duty cycle (F[2,30]=0.63, p>0.05). At interruption rate 4 Hz, post hoc pairwise comparisons showed significant (p<0.05) difference between paired scores at the two duty cycles; however, no significant (p>0.05) difference was found between paired scores at the two duty cycles at interruption rates 8 and 16 Hz.

For tone identification result in Fig. 1 (c), two-way ANOVA with repeated measures indicated a significant effect of interruption rate (F[2,30]=3.93, p<0.05), a non-significant effect of duty cycle (F[1,15]=2.44, p>0.05), and a non-significant interaction between interruption rate and duty cycle (F[2,30]=1.42, p>0.05). At interruption rate 4 Hz, post hoc pairwise comparisons showed significant (p<0.05) difference between paired scores at the two duty cycles. However, no significant (p>0.05) difference was found between paired scores at the two duty cycles at interruption rates 8 and 16 Hz.

For word recognition result in Fig. 1 (d), two-way ANOVA with repeated measures indicated a significant effect of interruption rate (F[2,30]=6.40, p<0.01), duty cycle (F[1,15]=69.00, p<0.001), and a non-significant interaction between interruption rate and duty cycle (F[2,30]=0.11, p>0.05). At the same interruption rate, post hoc pairwise comparisons showed significant (ps<0.001) difference between paired scores at the two duty cycles.

The present results show that consonant identification is more susceptible to interruption influence than vowel or lexical tone identification. Furthermore, lexical tone identification is the least affected by periodical interruption with both duty-cycle values tested. This might be due to the relatively easy task of choosing one out of four candidates for lexical tone identification. In addition, when reducing the value of duty cycle, periodical interruption leads to a more

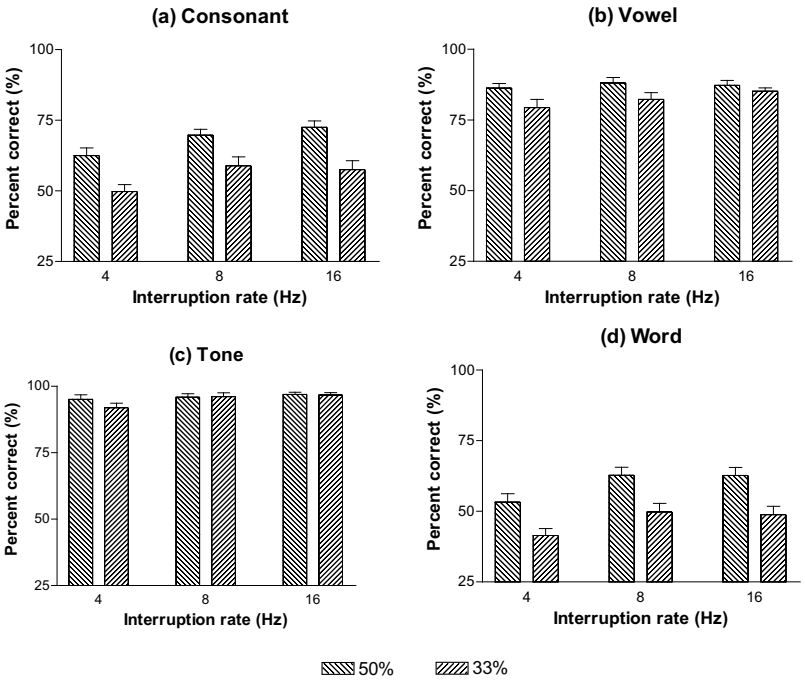

Figure 1. *Mean scores of all interrupted conditions in a) consonant identification, b) vowel identification, c) tone identification, and d) word recognition. The error bars denote ±1 standard error of the mean.*

severe negative influence to consonant and vowel identifications. Analysis was carried out to assess the correlation between Mandarin word recognition and consonant/vowel/tone identification in this study, and the Pearson's correlation coefficients are 0.99, 0.91, and 0.67, respectively. This suggests that the recognition scores of periodically interrupted Mandarin words are more correlated with those of periodically interrupted consonants. This is not surprising, as consonant identification is more easily affected by interruption interference than vowel or tone identification. This might be attributed to the shorter duration and less number of consonants than vowels in Mandarin Chinese [18].

3. Experiment 2: Mandarin Sentences

The purpose of this experiment was to investigate the effect of periodical interruption on Mandarin sentence recognition.

3.1 Subjects and materials

Twenty new NH native-Mandarin participants attended this experiment. The stimuli employed consisted of recordings from the Mandarin Hearing in Noise Test (MHINT) [20]. There are a total of 240 MHINT sentences, and each sentence consists of 10 keywords. An adult male native-Mandarin speaker with F0 of 75-180 Hz spoke the sentences. Steady-state speech-spectrum shaped noise (SSN) and two-talker babble (2T) were used to corrupt test sentences at 5 dB signal-to-noise ratio (SNR), which was chosen by a pilot study to avoid the ceiling effect in sentence recognition.

3.2 Signal processing

The same signal processing in Experiment 1 was used in Experiment 2. The interruption rate was selected as 1, 2, 4 and 8 Hz, and the duty-cycle value was chosen to be 50% and 33%.

3.3 Procedure

The listening experiments were conducted in a sound-proof booth. Stimuli were played to the participants through a circumaural headphone binaurally at a comfortable listening level. Sixty processed sentences (interrupted with 1 or 8 Hz and 33% duty cycle) were played to the participants as a practice session (i.e., with feedback) before the experiment to familiarize the participants with the experiment procedure. Participants attended 24 [= 3 noisy condition (i.e., clean or CLN (i.e., clean), SSN at 5 dB, and 2T at 5 dB) × 4 interruption rates (i.e., 1, 2, 4 and 8 Hz) × 2 duty-cycle values (i.e., 50% and 33%)] testing conditions. The order of the 24 testing conditions was randomized across participants. Participants were allowed to listen to each sentence 3 times at most, and then were instructed to verbally repeat all the words that they could recognize.

3.4. Results and discussion

Figure 2 shows the mean sentence recognition scores at all interrupted conditions. Statistical significance was determined by using the percent correct score as the dependent variable, and interruption rate (i.e., 1, 2, 4, and 8 Hz), duty cycle (i.e., 50% and 33%) and noise type (i.e., clean, SSN and 2T) as the three within-subjects factors. Three-way ANOVA with repeated measures indicated a significant effect of interruption rate (F[3,57]=170.77, p<0.001), duty cycle (F[1,19]=2272.10, p<0.001), and noise type (F[2,38]=789.01, p<0.001), a significant interaction between interruption rate and duty cycle (F[3,57]=16.30, p<0.001), between interruption rate and noise type (F[6,114]=63.94, p<0.001), and between duty cycle and noise type (F[2,38]=4.88, p<0.05), and a significant interaction among interruption rate, duty cycle and noise type (F[6,114]=13.51, p<0.001). Post hoc pairwise comparisons showed significant (ps<0.05) difference between paired scores

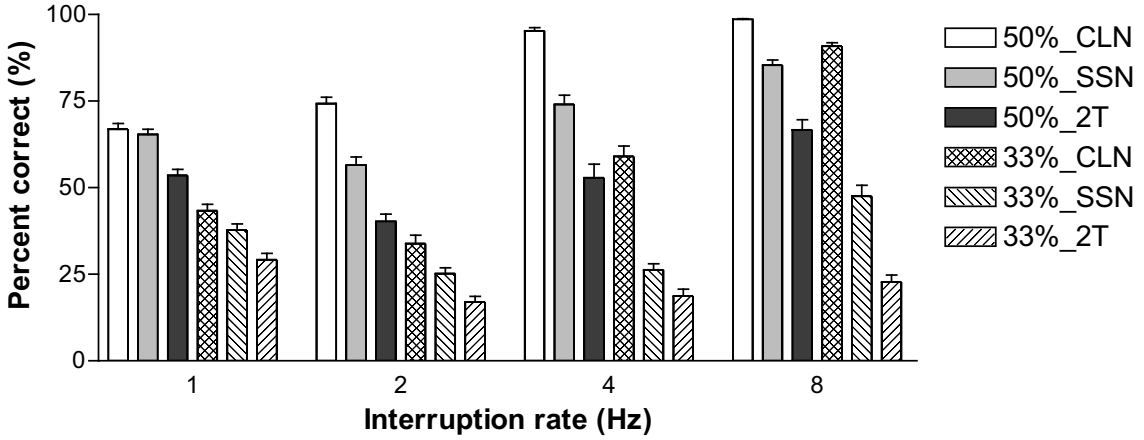

Figure 2. *Mean sentence recognition scores at all interrupted conditions. The error bars denote ±1 standard error of the mean.*

at the same interruption rate and the same duty cycle, except the difference between clean and SSN at 1 Hz interruption rate and 50% duty cycle (p>0.05) and between clean and SSN at 1 Hz interruption rate and 33% duty cycle (p>0.05).

First, findings of the current study show that the duty cycle of speech primarily predicts the speech recognition performance in Mandarin sentence context. The recognition performance of the Mandarin sentences improves significantly as the duty cycle of speech increases from 33% to 50%. This finding is consistent with previous English-based studies [e.g., 1, 3, 6, 8-9], suggesting a significant positive correlation between the duty cycle of speech and speech recognition performance.

Second, this work also suggests that there is a significant effect of interruption rate on Mandarin sentence perception. In general, the recognition of Mandarin sentences tends to improve as the rate of periodical interruption increases. More specifically, regardless of whether the duty cycle of speech was 50% or 33%, sentence recognition score increased as the interruption rate increased, except for the increment from interruption rate 1 Hz to 2 Hz, at which the lowest sentence recognition score is found. This finding is generally consistent with those reported in previous studies [e.g., 1, 5, 10]. For example, they all demonstrated the worst intelligibility occurred at about 2 Hz, and the intelligibility score improved as the rate of interruption increased. At 2 Hz interruption rate, the interference on the perception of phonemes and temporal envelope cues caused a significant drop in speech perception ability [1, 3, 5-6, 10]. As the interruption rate continued to increase, less extrapolation was needed in order to "glimpse" the words in each sentence, leading to an improvement in sentence recognition performance [1-2, 5-6, 10].

Earlier studies also found that modulation rates between 4 and 16 Hz were important for speech recognition and speech became relatively intelligible when the interruption rate increased above 4 Hz [12-13]. However, in the present study, when the duty cycle is 50%, Mandarin sentences are quite intelligible (i.e., ⩾70%) at 4 Hz interruption rate for both CLN and SSN conditions. These differences between the present and previous findings could be attributed to the use of different stimuli. Due to the difference between Mandarin and English in terms of the intrinsic nature of language, the tonal characteristics and the simpler syllable structure of Mandarin may help to restrict lexical competition and facilitate the integration of the spectro-temporal fragments of the original signal in the sentence context. The current findings also show that for 2T condition, unlike previous findings, the interruption rate 8 Hz is not fast enough for sentence intelligibility, indicating that the type of noise may have a detrimental effect on the recognition of periodically interrupted Mandarin sentences.

Finally, this study shows that the type of noise may significantly affect Mandarin sentence intelligibility under periodically interrupted conditions. This could be seen in the testing conditions having the same duty cycle of speech and interruption rate. When the duty cycle and interruption rate are held constant, the sentence recognition is the best for clean speech, except at 1 Hz interruption rate, at which similar performance is found for CLN and SSN conditions. The sentence recognition performance is the worst under interrupted conditions masked with 2-talker babble noise. This finding generally supports previous results [e.g., 1].

4. Conclusions

The present work examined the effects of interruption rate and duty cycle on the recognition of periodically interrupted Mandarin isolated words and sentences. Experimental results with NH listeners showed that consistent with earlier findings in English speech, interruption with a large rate or duty cycle favored the perception of periodically interrupted Mandarin speech. For recognizing isolated Mandarin words in the context of periodical interruption, the perception of vowels or tones was less affected by periodical interruption than that of consonants, and the perception of consonants could largely account for the variance in recognizing of isolated Mandarin words. For understanding periodically interrupted Mandarin sentences, the tonal characteristics and the simpler syllable structure in Mandarin might facilitate spectral-temporal integration of the target words, which contributed to a sentence intelligibility advantage of Mandarin over English under interrupted conditions.

5. Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 61571213). Part of this study was the basis for the Master's dissertation of the second author (R.H.N.T.).

6. References

[1] Miller, G. A. and Licklider, J. C. R., "The intelligibility of interrupted speech," J. Acoust. Soc. Am., 22(2): 167–173, 1950.

[2] Shafiro, V., Sheft, S., and Risley, R., "Perception of interrupted speech: Effects of dual-rate gating on the intelligibility of words and sentences," J. Acoust. Soc. Am., 130(4): 2076–2087, 2011.

[3] Wang, X. and Humes, L. E., "Factors influencing recognition of interrupted speech," J. Acoust. Soc. Am., 128(4): 2100–2111, 2010.

[4] Cooke, M., "Glimpsing speech," J. Phonetics, 31(3): 579–584, 2003.

[5] Jin, S. H. and Nelson, P. B., "Interrupted speech perception: The effects of hearing sensitivity and frequency resolution," J. Acoust. Soc. Am., 128(2): 881–889, 2010.

[6] Kidd, G., R. and Humes, L. E., "Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences," J. Acoust. Soc. Am., 131(2): 1434–1448, 2012.

[7] Moore, B. C. J., "Temporal integration and context effects in hearing," J. Phonetics, 31(3): 563–574, 2003.

[8] Fogerty, D., "Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration," J. Speech Lang. Hear. Res., 56(5): 1402–1408, 2013.

[9] Li, N. and Loizou, P., C., "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am., 122(2): 1165–1172, 2007.

[10] Bhargava, P. and Başkent, D., "Effects of low-pass filtering on intelligibility of periodically interrupted speech," J. Acoust. Soc. Am., 131: EL87–EL92, 2012.

[11] Wong, K., Li, W., Xu, R., and Zhang, Z., "Introduction to Chinese natural language processing," San Rafael: Morgan and Claypool Publishers, 2010.

[12] Drullman, R., Festen, J. M., and Plomp, R., "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am., 95(5): 2670–2680, 1994.

[13] Drullman, R., Festen, J. M., and Plomp, R., "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am., 95(2): 1053–1064, 1994.

[14] Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C., "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," J. Acoust. Soc. Am., 122(3): 1336–1339, 2007.

[15] Owren, M. J. and Cardillo, G. C., "The relative roles of vowels and consonants in discriminating talker identity versus word meaning," J. Acoust. Soc. Am., 119(3): 1727–1739, 2006.

[16] Chen, F., Wong, M. L. Y., Zhu, S., and Wong, L. L. N., "Relative contributions of vowels and consonants in recognizing isolated Mandarin words," J. Phonetics, 52: 26–34, 2015.

[17] Cole, R.A., Yan, Y., Mak, B., Fanty, M., and Bailey, T., "The contribution of consonants versus vowels to word recognition in fluent speech," Proceedings of the ICASSP'96, pp. 853-856, 1996.

[18] Chen, F., Wong, L. L. N., and Wong, E. Y. W., "Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility," J. Acoust. Soc. Am., 134(2): EL178–EL184, 2013.

[19] Fogerty, D. and Humes, L. E., "Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors," J. Acoust. Soc. Am., 128(5): 3114–3125, 2010.

[20] Wong, L., Soli, S., Liu, S., Han, N., and Huang, M., "Development of the Mandarin hearing in noise test (MHINT)," Ear Hear., 28: 70S–74S, 2007.