LINGUISTIC RELEASE FROM MASKING IN SIMULATED ELECTRIC HEARING

Huizi Lung [1], Shangdi Liao [2], Fei Chen [2]

[1] Unit of Human Communication, Development, and Information Sciences, Faculty of Education, The University of Hong Kong, Hong Kong, China
[2] Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China

fchen@sustech.edu.cn

ABSTRACT

Many works revealed linguistic release from masking (LRM) in speech perception from normal-hearing listeners. The present work investigated the LRM in understanding vocoded Mandarin speech simulating cochlear implants (CIs) listening. Mandarin sentences were corrupted by 2- and 6-talker babble maskers spoken in languages with varying degrees of linguistic similarity to Mandarin, namely Mandarin, Cantonese, and English. Target and masker sentences were mixed and processed by an 8-channel noise vocoder to simulate CI listening, and presented to normal-hearing listeners to recognize. Mandarin recognition scores were significantly higher in English babble masker than in Cantonese and Mandarin babble maskers, and significantly higher in Cantonese than Mandarin talker babble masker, demonstrating LRM effect in simulated CI listening. Listeners benefited from a target-masker linguistic mismatch even though spectral resolution was low, and when competing talkers also spoke Mandarin, more talkers might be less disruptive to Mandarin speech recognition in CI listening.

Keywords: Mandarin speech perception, cochlear implants, linguistic release of masking (LRM).

1. INTRODUCTION

Speech perception in noise is a challenging task for people with normal hearing, and even more so for people with hearing loss. Normal-hearing (NH) listeners experience a release from masking (MR) when the masker is fluctuating, compared to when the masker is a steady state noise (SSN) [e.g., 1-2]. The target speech can be "glimpsed" during temporal dips in a fluctuating masker when the overall intensity of the masker is low or during spectral dips when target and masker differ in their frequency composition [2].

Linguistic dissimilarity between the target and masker speech has been shown to improve speech understanding. Such improvement is known as linguistic release from masking (LRM). The target-masker linguistic similarity hypothesis states that the more dissimilar the target and the masker language, the easier it is for the listener to segregate the target from the masker, especially at challenging signal-to-noise ratios (SNRs) [3]. LRM has been found in English [e.g., 4] and Mandarin speech perception [e.g., 5] among NH listeners. Both adults and children can benefit from a linguistic mismatch between target and masker [e.g., 6].

LRM studies also showed that masker meaningfulness plays a role in speech masking. Van Engen and Bradlow found that at SNR of 0 dB or lower, English sentence recognition by monolingual English speakers was worse in English 2-talker babble masker than in Mandarin 2-talker babble masker [7]. Bilingual listeners who can understand both the target and masker languages also showed LRM. Calandruccio and Zhou found that English-Greek bilinguals whose dominant language was English showed improved English speech recognition in Greek maskers compared to in English maskers [8]. In [5], Mandarin-English bilinguals showed better Mandarin sentence recognition when English rather than Mandarin maskers were used. These results suggested that a more familiar masker language interfered more with target speech perception.

For listeners with profound-to-severe hearing impairment, cochlear implants (CIs) provide an efficient way for them to restore their hearing [9]. CIs extract useful acoustic information (i.e., primarily multi-channel temporal envelope waveforms, or discarding spectral details) from the original speech input, and electrically stimulate the auditory nerves to deliver the acoustic information and subsequently evoke sound perception in brain.

Vocoder processing has been long used to simulate CI listening in NH subjects [10]. Real CI users may vary significantly due to factors such as etiology of hearing loss and implant device differences. Vocoder processing is useful in measuring the effect of a processing parameter

change in CI as the amount of spectral and temporal cues, such as the number of frequency channels, can be manipulated independent of patient-related factors.

In CI users, speech perception is worse in modulated noise than SSN or not significantly different in the two kinds of noise, showing negative MR or no MR in a fluctuating masker [e.g., 11-12]. Liu et al. found that children who used CIs also showed less MR than NH adults and children [13]. Fu and Nagoki proposed that the reduced or absent MR benefit may be due to reduced spectral resolution and spectral smearing in CI listening [11], as they found that NH subjects listening to vocoded speech performed similarly to CI users when spectral resolution was low (i.e., 4 channels). CI signal processing is limited by the number of electrodes implanted and the number of spectral channels. Ihlefeld et al. also found that NH listeners demonstrated decreased MR when presented with vocoded stimuli [14].

While there is ample evidence that hearing-impaired listeners show reduced MR benefit than NH listeners, few studies investigated LRM in hearing-impaired listeners. Viswanathan et al. reported LRM benefit in simulated CI listening [15]. NH listeners listened to vocoded English sentences in competing English or Dutch masker sentences. English speech recognition performance was better when the masker was Dutch.

The purpose of this study was to investigate whether Mandarin-speaking CI users may experience LRM by using a vocoder processing to simulate CI listening. The target language in this study was Mandarin (which is a tonal language), and three masker languages, Mandarin, Cantonese, and English, which vary in linguistic similarity to Mandarin, were used.

2. METHODS

2.1. Subjects and materials

Sixteen native Mandarin Chinese listeners (10 males and 6 females, age range 21 to 27 years) were recruited from the Southern University of Science and Technology, and paid to participate in the study. Participants were students at the university and were bilingual (Mandarin and English), and had hearing within normal range. The experimental procedure involving human subjects was approved by the Institution's Ethical Review Board of Southern University of Science and Technology.

The speech material comprised sentences extracted from the Mandarin Hearing in Noise Test (MHINT) database [16]. The MHINT corpus includes 24 lists, each with 10 sentences and 10 keywords per sentence. All sentences were spoken by a male native Mandarin Chinese speaker with a fundamental frequency (F0) of 75–180 Hz and recorded at a sampling rate of 16 kHz. The target MHINT sentences were corrupted by competing talker babble masker (2-talker or 6-talker babble, in Mandarin, Cantonese, or English) at SNR levels of 0 or –5 dB.

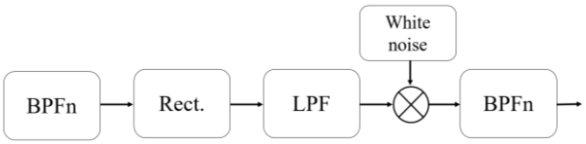

Figure 1: Schematic of Implementing the Noise Vocoder at the n^{th} Channel. Rect.: Waveform rectification.

Six native-Mandarin speakers (3 male and 3 female) and 6 native-Cantonese speakers (3 male and 3 female) recorded sentences in Mandarin and Cantonese respectively for the Mandarin and Cantonese maskers. The speakers were each instructed to read 8 sentences extracted from the newspaper in a natural style. Recording was done in a sound-proof booth with a digitization rate of 16 kHz. All sentences were equated for root-mean-square (RMS) level. All 8 sentences spoken by each speaker were combined to form a one-talker masker signal, resulting in 6 different masker signals for each language (Mandarin and Cantonese). English masker speech was taken from the TIMIT database [17]. Eight sentences were taken from each of 6 (3 male and 3 female) speakers. Sentences from each speaker were combined to form 6 one-talker masker signals. Each of the 18 (=6 speakers × 3 languages) one-talker masker signal ranged from 15 to 25 seconds in duration.

2.2. Signal processing

Two male voices from each language were used to produce the 2-talker babble signals to match the speaker gender of the target materials (MHINT sentences spoken by a male speaker). A random segment of masker speech with duration equal to that of the target MHINT sentence was selected from each of the 2 male voice masker signals. Then, the two segments were equalized for RMS level and summed up. The 2-talker signal was then mixed with the target MHINT sentence at SNR levels of 0 and –5 dB. The corrupted MHINT sentence was then adjusted to have the same RMS value as the original sentence. For the 6-talker babble maskers, segments randomly selected from each of 6 masker sentences (in the same language) were used to corrupt the

MHINT sentence at SNR levels of 0 and –5 dB. The processing was the same as that used in the 2-talker babble (i.e., RMS equalization, summation, and mixture with the MHINT sentence).

Vocoder processing was used to simulate CI listening [e.g., 10]. MHINT sentences, after being corrupted with the maskers, were processed by a noise vocoder. Figure 1 shows the schematic of implementing the noise vocoder in one channel. Speech signals were first processed through a pre-emphasis high-pass filter with 1200 Hz cut off frequency. Then, signals were band-pass filtered (BPF) into 8 frequency channels between 80 and 6000 Hz. The corner frequencies for the 8 channels were 80, 221, 425, 724, 1158, 1790, 2710, 4050, and 6000 Hz. The temporal envelope was extracted by half-wave rectification followed by low-pass filtering (LPF) with 250 Hz cut off frequency. The extracted envelope from each frequency band was then used to modulate a white noise signal, and BPF again into 8 frequency bands. The BPF signals at each frequency channel were then summed up to generate the noise-vocoded speech stimulus with its amplitude readjusted to match the RMS value of the original signal.

2.3. Procedure

Participants took part in a training session before the speech recognition experiment. The training session and the experiment were conducted in a soundproof room. Speech materials were played to listeners through circumaural headphones at a comfortable listening level indicated by the listeners. Participants were instructed that they would listen to Mandarin sentences that had been processed. Each sentence could be played three times at maximum.

In the training session, listeners listened to vocoded MHINT sentences while reading the transcript for the sentences to become familiarized with the processed speech materials. In the experiment, participants listened to vocoded MHINT sentences without reading. They were instructed to repeat the words they heard, and to guess if they were not sure. The experimenter scored participants' response by the number of words correctly identified in the sentence. The intelligibility score for each condition was computed as the percentage of words in the target MHINT sentences correctly identified in each list of MHINT sentences.

In the experiment, all subjects participated in a total of 12 conditions [=3 masker languages × 2 types of talker babbles × 2 SNR levels]. The babble maskers in different languages with varying linguistic contents are denoted as "Eng 2", "Eng 6", "Can 2", "Can 6", "Man 2" and "Man 6". "Eng",

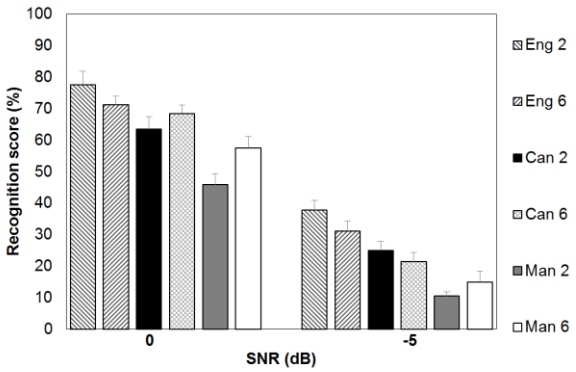

Figure 2: Mean Sentence Recognition Scores for All Conditions. Error bars denote 1 standard error of the mean.

"Can" and "Man" refer to English, Cantonese, and Mandarin respectively, and "2" and "6" indicated the number of talkers in each masker babble. A different list of MHINT sentences was used in each condition and none of the target sentences was repeated across conditions. The order of the test conditions was randomized across subjects to minimize potential learning effects. Subjects were given a 5-minute break every 30 minutes during testing. The whole experiment took about 2 to 3 hours to complete for each subject.

3. RESULTS

Figure 2 shows the mean Mandarin sentence recognition scores for all conditions. Statistical significance was determined using percent correct score as the dependent variable, and masker language (English, Cantonese, or Mandarin), type of talker babble (2-talker or 6-talker), and SNR level (0 dB or –5 dB) as within-subject factors.

Three-way repeated measures analysis of variance (ANOVA) indicated a significant effect of masker language ($F_{2,30}$ = 70.890, p < 0.001, η^2_P = 0.825), a non-significant effect of type of talker babble ($F_{1,15}$ = 0.498, p = 0.491, η^2_P = 0.032), and a significant effect of SNR level ($F_{1,15}$ = 984.877, p < 0.001, η^2_P = 0.985). Analysis also revealed a significant interaction between masker language and type of talker babble ($F_{2,30}$ = 7.745, p = 0.002, η^2_P = 0.341). The interaction between masker language and SNR level ($F_{2,30}$ = 0.742, p = 0.485, η^2_P = 0.047) and between type of talker babble and SNR level ($F_{1,15}$ = 3.164, p = 0.096, η^2_P = 0.174) were not significant. The interaction among masker language, type of talker babble, and SNR level ($F_{2,30}$ = 0.660, p = 0.524, η^2_P = 0.042) was not significant.

Post-hoc analysis with Bonferroni correction showed that with respect to the effect of masker language, Mandarin sentence recognition score was

significantly better (higher) in English maskers than in Cantonese maskers ($p < 0.001$) or Mandarin maskers ($p < 0.001$), and significantly better in Cantonese maskers than in Mandarin maskers ($p < 0.001$). Comparisons across conditions showed that this was true when the type of talker babble and the SNR level were the same, except in the condition pairs "Eng 6" vs. "Can 6" at 0 dB SNR, and "Can 6" vs. "Man 6" at –5 dB SNR. With respect to the effect of SNR level, recognition score was significantly better at 0 dB SNR than at –5 dB SNR ($p < 0.001$). Comparisons across conditions indicated that this was true in all condition pairs when the masker language and type of talker babble were the same.

4. DISCUSSION AND CONCLUSION

The present study investigated whether CI users experience LRM by using vocoder processing to simulate CI listening in NH listeners. The results indicate that simulated CI listening benefited from a linguistic mismatch between target and masker languages and showed LRM in Mandarin speech recognition.

Mandarin sentence recognition scores were significantly better in conditions with English talker babble masker compared to Cantonese or Mandarin babble masker, and significantly better in Cantonese compared to Mandarin babble masker, demonstrating LRM effect in simulated CI listening. The most effective masker was Mandarin, and English was the least effective masker. These findings are similar to results seen in NH listeners, which indicated that when the masker language was different from the target language, listeners showed MR benefit, compared to when the masker and target speech were spoken in the same language [e.g., 4-5, 18]. Furthermore, findings were consistent with results from Viswanathan et al. [15] which also demonstrated LRM benefit in simulated CI listening.

Despite a lack of access to F0 cue and reduced spectral resolution in the 8-channel noise-vocoded stimuli, listeners were able to benefit from a linguistic mismatch between target and masker speech. This suggests that linguistic properties which are not represented through F0 or spectral cues may contribute to the LRM benefit in a target-masker mismatch condition.

Mandarin speech recognition may be more vulnerable to interference from competing talkers also speaking Mandarin in CI listening because they may not be able to resolve concurrent tonal information in competing streams of speech. Luo and colleagues investigated concurrent vowel and tone recognition in NH listeners, simulated CI listeners and real CI listeners in a series of experiments [e.g., 19-20]. NH listeners achieved nearly perfect vowel and tone identification across conditions. But when the same group of subjects listened to 8-channel or 4-channel noise-vocoded speech simulating CI listening, tone recognition performance in single syllable or concurrent syllables became significantly poorer [19]. Real CI users also performed the recognition tasks, and their tone recognition performance in the single-syllable condition was significantly poorer than 4- and 8-channel simulation results [20], indicating that real CI users were even less able to make use of F0 and pitch information for tone recognition. CI users' performance of concurrent tone recognition was also poorer than performance of NH subjects and simulation performance.

In [5] and [21], NH listeners showed better Mandarin speech perception with fewer number of talkers in competing speech. Their findings were also consistent with findings from studies of English speech-on-speech masking in NH English speakers [e.g., 7]. A masker with fewer talkers may provide more opportunities for glimpsing the target speech in NH listeners. However, Chen et al. found that adult Mandarin-speaking CI users performed worse in a single-talker masker than in a 2-talker or 4-talker masker [21]. Results from the present study with regards to the effect of number of competing talkers were somewhat consistent with results from [21], but in contrast to studies of English speech masking [22]. In the present study, when the masker language was Mandarin, subjects performed worse with fewer number of competing talkers in the Mandarin babble masker. This effect was the same at both SNR levels used.

In conclusion, consistent with early findings from NH listeners, the present study simulating CI speech perception showed the effect of linguistic release from masking. The recognition of vocoder-processed Mandarin sentences in speech masker was notably influenced by the linguistic content of masker, with an increased influence from English, Cantonese to Mandarin. As CI speech perception primarily relies on limited amount of temporal envelope information, the findings in this work provide evidence on the perceptual impact of temporal envelope in linguistic release of masking.

5. ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 61971212). Part of this study was the basis for the Master's dissertation of the first author (H.Z.L.).

6. REFERENCES

[1] Festen, J. M., Plomp, R. 1990. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *The Journal of the Acoustical Society of America*, 88, 1725-1736.

[2] Peters, R. W., Moore, B. C. J., Baer, T. 1998. Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, 103, 577-587.

[3] Brouwer, S., Van Engen, K. J., Calandruccio, L., Bradlow, A. R. 2012. Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, 131, 1449-1464.

[4] Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., Bradlow, A. R. 2013. Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *American Journal of Audiology*, 22, 157-164.

[5] Chen, F., Li, J., Wong, L. L. N., Yan, Y. 2013. Effect of linguistic masker on the intelligibility of Mandarin sentences. *Proc. of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, 2098-2102.

[6] Calandruccio, L., Leibold, L. J., Buss, E. 2016. Linguistic masking release in school-age children and adults. *American Journal of Audiology*, 25, 34-40.

[7] Van Engen, K. J., Bradlow, A. R. 2007. Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121, 519-526.

[8] Calandruccio, L., Zhou, H. 2014. Increase in speech recognition due to linguistic mismatch between target and masker speech: Monolingual and simultaneous bilingual performance. *Journal of Speech, Language, and Hearing Research*, 57, 1089-1097.

[9] Chen, F., Ni, W., Li, W., Li, H. 2019. Cochlear implantation and rehabilitation. In: Li, H., Chai, R. (eds), *Hearing Loss: Mechanisms, Prevention and Cure.* Springer, Singapore, 129-144.

[10] Chen, F., Zheng, D. C., Tsao, Y. 2017. Effects of noise suppression and envelope dynamic range compression to the intelligibility of vocoded sentences for a tonal language. *The Journal of the Acoustical Society of America*, 142, 1157–1166.

[11] Fu, Q. J., Nogaki, G. 2005. Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *Journal of the Association for Research in Otolaryngology*, 6, 19-27.

[12] Zirn, S., Polterauer, D., Keller, S., Hemmert, W. 2016. The effect of fluctuating maskers on speech understanding of high-performing cochlear implant users. *International Journal of Audiology*, 55, 295-304.

[13] Liu, J. S., Liu, Y. W., Yu, Y. F., Galvin, J. J., Fu, Q. J., Tao, D. D. 2021. Segregation of competing speech in adults and children with normal hearing and in children with cochlear implants. *The Journal of the Acoustical Society of America*, 150, 339-352.

[14] Ihlefeld, A., Deeks, J. M., Axon, P. R., Carlyon, R. P. 2010. Simulations of cochlear-implant speech perception in modulated and unmodulated noise. *The Journal of the Acoustical Society of America*, 128, 870-880.

[15] Viswanathan, N., Kokkinakis, K., Williams, B. T. 2018. Listeners experience linguistic masking release in noise-vocoded speech-in-speech recognition. *Journal of Speech, Language, and Hearing Research*, 61, 428-435.

[16] Wong, L. L. N., Soli, S. D., Liu, S., Han, N., Huang, M. W. 2007. Development of the Mandarin Hearing in Noise Test (MHINT). *Ear and Hearing*, 28, 70S–74S.

[17] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., Zue, V. 1993. TIMIT acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium, 1993.*

[18] Calandruccio, L., Dhar, S., Bradlow, A. R. 2010. Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, 128, 860-869.

[19] Luo, X., Fu, Q. J. 2009. Concurrent-vowel and tone recognition in acoustic and simulated electric hearing. *The Journal of the Acoustical Society of America*, 125, 3223–3233.

[20] Luo, X., Fu, Q. J., Wu, H. P., Hsu, C. J. 2009. Concurrent vowel and tone recognition by Mandarin-speaking cochlear implant users. *Hearing Research*, *256*, 75–84.

[21] Chen, B., Shi, Y., Zhang, L., Sun, Z., Li, Y., Gopen, Q., Fu, Q. J. 2020. Masking effects in the perception of multiple simultaneous talkers in normal-hearing and cochlear implant listeners. *Trends in Hearing*, 24, 1-12.

[22] Cullington, H. E., Zeng, F. G. 2008. Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects. *The Journal of the Acoustical Society of America*, 123, 450-461.