Assessing the Effect of Noise-reduction to the Intelligibility of Low-pass Filtered Speech

Huimin Zhuge, Lei Wang, Fei Chen *Member IEEE*, Dingchang Zheng

Abstract –Given the fact that most hearing-impaired listeners have low-frequency residual hearing, the present work assessed the effect of applying commonly-used single-channel noise-reduction (NR) algorithms to improve the intelligibility of low-pass filtered speech, which simulates the effect of understanding speech with low-frequency residual hearing of hearing-impaired patients. In addition, this study was performed with Mandarin speech, which is characterized by its significant contribution of information present in (low-frequency dominated) vowels to speech intelligibility. Mandarin sentences were corrupted by steady-state speech-shaped noise and processed by four types (i.e., subspace, statistical-modeling, spectral-subtractive, and Wiener-filtering) of single-channel NR algorithms. The processed sentences were played to normal-hearing listeners for recognition. Experimental results showed that existing single-channel NR algorithms were unable to improve the intelligibility of low-pass filtered Mandarin sentences. Wiener-filtering had the least negative influence to the intelligibility of low-pass filtered speech among the four types of single-channel NR algorithms examined.

I. INTRODUCTION

Most hearing-impaired (HI) listeners suffer from precipitous sloping severe-to-profound high frequency (HF) hearing loss, but preserve some residual hearing in low frequency (LF) regions. With a severe-to-profound high frequency loss, soft HF sounds often remain inaudible because the amount of gain needed is unreachable by conventional hearing aids [e.g., 1]. Some studies also suggested that HF amplification may not benefit or even degrade speech intelligibility of patients with HF hearing loss. It is known that high frequency sounds are often consonants important to speech intelligibility, while low frequency sounds are often vowels important to the power of speech. Missing out HF components of speech can lead to difficulties understanding speech especially in noisy background, resulting in poor speech intelligibility [2].

Huimin Zhuge is with Department of Biology, Southern University of Science and Technology, Shenzhen, China (e-mail: zhugehm@126.com).

Lei Wang is with Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China (e-mail: wangl6@mail.sustc.edu.cn).

Fei Chen is with Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China (phone: 0086-75588015878; fax: 0086-75588015878; e-mail: fchen@sustc.edu.cn).

Dingchang Zheng is with the Health and Wellbeing Academy, Faculty of Medical Science, Anglia Ruskin University, Chelmsford, UK (e-mail: dingchang.zheng@anglia.ac.uk).

It is has attracted substantial attentions in the field to make use of low-frequency residual hearing to improve speech understanding of HI listeners. For instance, a new trend, i.e., frequency compression (FC) technique, is emerging in clinical hearing aids fitting. Frequency compression delivers the inaudible information from unaidable HF regions to the aidable LF regions [1], aiming to help HI listeners to use their residual hearing in LF regions to hear HF sounds. In addition, the contribution of low-frequency residual hearing for speech recognition was demonstrated in studies of combined electric-and-acoustic hearing, whereas HI patients were implanted with electrodes to restore their HF hearing, and utilized their low-frequency residual hearing (typically 20 to 60 dB hearing level up to 750 Hz, and severe-to-profound hearing loss at 1000 Hz and above). A number of studies have shown that this combined electric-and-acoustic hearing provided HI listeners much better speech understanding performance, particularly in noise, than electric-only hearing in conventional cochlear implantation [e.g., 3].

Recently, many studies suggested that vowels carry more perceptual information than consonants [e.g., 4-6]. Fogerty and Kewley-Port in [4] and Chen et al. in [5] showed that the vowel-only sentences (i.e., preserving vowel segments and replacing consonant segments with silence or white noise) were more intelligible than the consonant-only sentences (i.e., vowels replaced by silence or white noise) in both English and Mandarin. Considering that vowels are characterized by their low frequencies, studies also assessed how low-pass filtering affected the intelligibility of vowel sentences in Mandarin and English [6]. Fogerty and Chen demonstrated significant contributions of information present during vowels to sentence intelligibility and the performance was much higher for Mandarin speech [6]. Mandarin vowels contained a higher distribution of cues in the low frequencies (e.g., below 1000 Hz), which was largely attributed to the fact that Mandarin vowels carry additional cues, most likely related to lexical tone, in the low-frequencies.

Though the importance of low-frequency residual hearing has been noted in the field, our knowledge on its contribution for speech intelligibility is still limited. The present work examined the effect of applying noise-reduction (NR) algorithms (i.e., four types of existing single-channel NR algorithms) to improve the intelligibility of low-pass filtered speech. This was motivated by two reasons. First, speech understanding in noise has been a longstanding challenge for HI listeners, including those fitted with assistive hearing devices. Second, many studies have suggested that most existing single-channel NR

algorithms could not improve the intelligibility for normal-hearing (NH) listeners [e.g., 7-8]; however, they may lead to improved speech understanding for HI listeners fitted with assistive hearing devices [e.g., 9-11]. Recognizing low-pass filtered speech simulates the speech perception of HI listeners preserving low-frequency residual hearing. It is unknown whether existing single-channel NR algorithms could improve the intelligibility of low-pass filtered speech. Therefore, the aim of the present work was to assess the effect of noise-reduction to the intelligibility of low-pass filtered speech.

II. METHODS

Subjects and materials

Ten (5 male and 5 female) NH listeners participated in this experiment. All subjects were native-speakers of Mandarin Chinese, and were paid for their participations. This study was approved by the Human Research Ethics Committee for Non-Clinical Faculties of the Southern University of Science and Technology.

The speech material consisted of sentences extracted from the Mandarin Hearing in Noise Test (MHINT) database [6]. MHINT corpus has a total of 24 lists, and each list has 10 sentences (10 keywords in each sentence). All the sentences were spoken by a male native Mandarin-Chinese speaker having a fundamental frequency (F0) of 75 to 180 Hz, which was recorded at a sampling rate fs=16 kHz.

A steady-state speech-shaped noise (SSN) was used to corrupt the MHINT sentences. To generate the SSN masker, a finite impulse response filter was designed based on the average spectrum of the MHINT sentences, and white noise was filtered and scaled to the same long-term average spectrum and level as the sentences. A noise segment of the same length as the clean intact (i.e., full-length) speech signal was randomly cut out of the SSN masker, appropriately scaled to reach the desired signal-to-noise ratio (SNR), and finally added to the speech signals at –3 and –6 dB SNR. The SNR levels were carefully chosen to avoid the ceiling/floor effect in understanding noisy/noise-reduced low-pass filtered sentences in this experiment.

Signal Processing

The noise-corrupted sentences were filtered by low-pass filters (LPFs) with cut-off frequencies of 1000 and 750 Hz in this study to simulate the low-frequency residual hearing of HI listeners. The low-pass filtering was implemented by using a linear-phase FIR filter with filter order n=10×fs/fcut, where fs was the sampling rate (i.e., 16 kHz) and fcut was the LPF cut-off frequency (i.e., 1000 and 750 Hz).

The low-pass filtered noise-corrupted sentences were processed by four representative single-channel noise-reduction algorithms, including the generalized KLT approach [12], the Log Minimum Mean Square Error (logMMSE) algorithm [13], the multiband spectral subtraction algorithm (MB) [14], and the Wiener algorithm based on a priori SNR estimation (Wiener) [15]. These four algorithms were selected because they cover the four most-used types of single-channel noise-reduction methods, i.e., subspace approach, statistical-modeling approach, spectral-subtractive approach, and Wiener-filtering approach, and

they represent the state-of-the-art noise-reduction techniques. The parameters used to implement these algorithms were the same as those published. A detailed description of the algorithms tested can be found in [7, 16].

Procedure

The experiment was performed in a sound-proof room, and stimuli were played to listeners binaurally through a circumaural head-phone at a comfortable listening level. Before actual testing session, each subject participated in a training session and was given six lists of experimental sentences (i.e., 60 sentences). The 10-min training session familiarized the subjects with the testing procedure and tested conditions. During the training session, the subjects were allowed to read feedback after they heard each sentence. In the testing session, the order of the tested conditions was random across subjects, and the subjects were asked to orally repeat all the words they heard. Each subject participated in a total of 20 conditions [=2 low-pass cut-off frequencies (i.e., 1000 and 750 Hz) × 2 SNR levels (–3 and –6 dB) × 5 signal processing conditions (i.e., Noisy, Wiener, logMMSE, MB and KLT)]. One list of ten Mandarin sentences was used per tested condition, and none of the sentences was repeated across conditions. Subjects were allowed to listen to each stimulus for three times in maximum, and required to repeat as many words as they can recognize. A 5-minute break was given to the subjects for every 30 mins to avoid listening fatigue. The intelligibility score for each condition was defined as the ratio between the number of the correctly recognized words and the total number of words contained in each list of 10 MHINT sentences.

III. RESULTS

Figure 1 shows the mean sentence recognition scores for all conditions. Statistical significance was determined by using the percent recognition score as the dependent variable, and LPF cut-off frequency, SNR level and signal processing condition as the three within-subject factors. Three-way analysis of variance (ANOVA) with repeated measures indicated a significant effect ($F[1,9]=8.59$, $p<0.05$) of LPF cut-off frequency, SNR level ($F[1,9]=8.59$, $p<0.001$), and signal processing condition ($F[4,36]=11.43$, $p<0.001$), a non-significant ($p>0.05$) interaction between LPF cut-off frequency and SNR level, between LPF cut-off frequency and signal processing condition ($p>0.05$), between SNR level and signal processing condition ($p>0.05$), and among LPF cut-off frequency, SNR level and signal processing condition ($p>0.05$).

Post-hoc pairwise comparisons showed that at all tested conditions with the same LPF cut-off frequency and SNR level in Fig. 1, the mean recognition scores of noise-reduced sentences (i.e., conditions Wiener, logMMSE, MB and KLT in Fig. 1) were not significantly larger than that of noisy sentences (i.e., condition Noisy in Fig. 1). At some conditions, using noise-reduction significantly reduced the recognition score in relative to that of noisy sentences. Note that at tested condition (750, –3 dB) in Fig. 1, the mean recognition score of Wiener-processed sentences was larger than that of noisy sentences, i.e., 42.4% vs. 28.0%; however, statistical analysis indicted that the difference was not
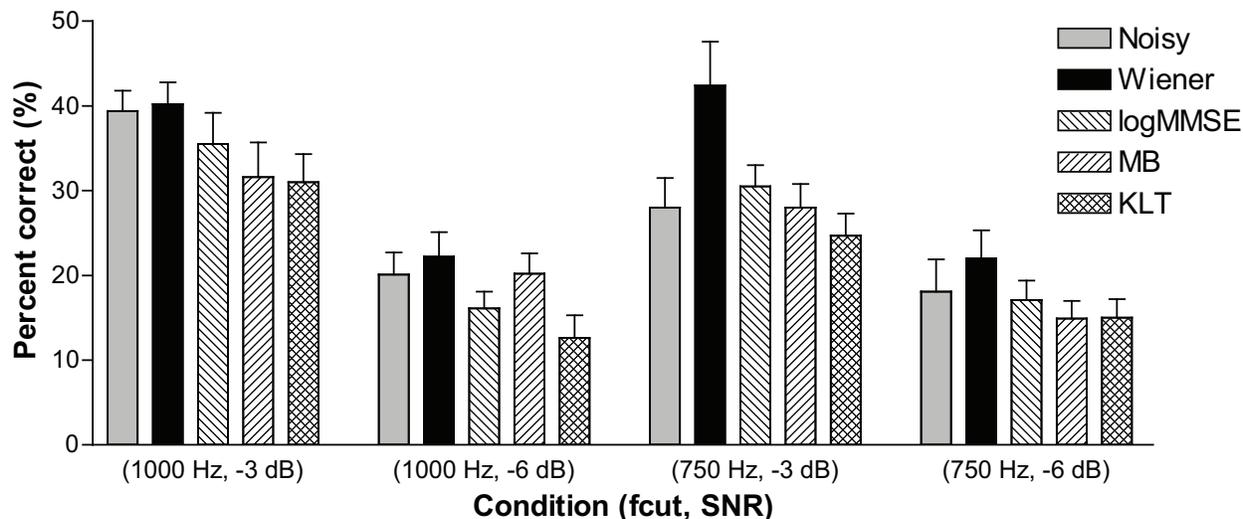
Figure 1. Mean sentence recognition scores for all conditions. The error bars denote ±1 standard error of the mean.

significant ($p>0.05$) between the two mean scores. It is further seen in Fig. 1 that at each tested condition, the score of Wiener-processed sentences was the largest among the four types of NR algorithms, suggesting that Wiener-filtering had the least negative influence to the intelligibility of low-pass filtered speech among the NR algorithms examined.

IV. DISCUSSION AND CONCLUSIONS

The present work showed that existing commonly-used single-channel noise-reduction algorithms did not improve the intelligibility of low-pass filtered speech. Earlier unpublished data showed that when the low-pass cut-off frequency was set to 1000 and 750 Hz, the intelligibility scores of clean LPF-processed were 98% and 92%, respectively. Hence, there is a large gap between the intelligibility scores of clean and noise-reduced LPF-processed speech. Taken together, it is noted that existing single-channel NR algorithms may not improve the intelligibility of wideband and low-pass filtered speech. However, it is interesting to see in Fig. 1 that Wiener filtering may slightly but not significantly improve the intelligibility of LPF-processed speech at –6 dB SNR and 750 Hz LPF cut-off frequency. The exact reason for this intelligibility benefit from Wiener filtering is unclear now. This might be attributed to the less distortion contained in noise-reduced speech due to the less aggressive noise suppression by Wiener filtering. Recent work showed that F0 contour plays a critical role for the intelligibility of LPF-processed speech. Further work may assess how different NR algorithms may cause distortion to the acoustic cues (e.g., F0 contour) contained in low-pass filtered speech.

The long-term aim of this study is to improve the speech perception for HI listeners preserving low-frequency hearing. The present work was performed with a simulation experiment, i.e., using NH listeners to listen to low-pass filtered speech to simulate the speech recognition of HI listeners. The advantage of simulation experiment is to avoid the interaction of patient-specific confounds for speech perception. Nevertheless, the speech understanding performance of hearing-impaired listeners is influenced by many factors (e.g., reduced spectral resolution and dynamic range of hearing) and their interactions. Hence, it warrants clinical studies with real hearing-impaired listeners in the future.

Note that the present work examined the effect of noise-reduction for LPF-processed Mandarin speech. Since Mandarin is a tonal language, with lexical information conveyed by its F0 contour, the spectral distribution of vowel cues to sentence intelligibility may be very different from English in which F0 does not directly convey lexical meaning. In addition, Mandarin also has a consonant-vowel syllable structure that differs significantly from the complex syllable structure of English that allows for consonant clusters [6]. Hence, it is unknown whether the same findings can be obtained when replicating this study with English sentences. In addition, this work used steady-state SSN as masker to corrupt Mandarin sentences. It is unclear how competing maskers (e.g., babble) would affect the effect of noise-reduction to the intelligibility of low-pass filtered speech.

In conclusion, the present work investigated the effect of four types of existing most-used single-channel noise-reduction algorithms to the intelligibility of low-pass filtered Mandarin sentences. Low pass filtering simulated the effect of understanding speech with low frequency residual hearing of hearing-impaired listeners. Results showed that existing single-channel NR algorithms did not improve the intelligibility of low-pass filtered speech. Among the four types of single-channel NR algorithms examined in this

study, Wiener filtering had the least negative influence to the intelligibility of low-pass filtered speech.

REFERENCES

[1] F. Kuk, D. Keenan, J. Auriemmo, and P. Korhonen, "Re-evaluating the efficacy of frequency transposition," *The ASHA Leader*, 14, 14–17, 2009.

[2] A. Simpson, "Frequency-lowering devices for managing high-frequency hearing loss: A review," *Trends in Amplification*, 13, 87–106, 2009.

[3] B. J. Gantz and C. Turner, "Combining acoustic and electric hearing," *Laryngoscope*, 113, 1726–1730, 2003.

[4] D. Fogerty and D. Kewley-Port, "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.*, 126, 847–857, 2009.

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

[6] D. Fogerty and F. Chen, "Vowel spectral contributions to English and Mandarin sentence intelligibility," *Proceedings of Annual Conference of International Speech Communication Association*, Singapore, 499–503, 2014.

[7] Y. Hu and P. C. Loizou, "A comparative intelligibility study of singlemicrophone noise reduction algorithms," *J. Acoust. Soc. Am.*, 122, 1777–1786, 2007.

[8] J. F. Li, J. P. Zhang, Y. H. Yan, Y. Hu, M. Akagi, and P. C. Loizou, "Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese and English," *J. Acoust. Soc. Am.*, 129, 3291–3301, 2011.

[9] L. Yang and Q. J. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *J. Acoust. Soc. Am.*, 117, 1001–1004, 2005.

[10] P. C. Loizou, A. Lobo, and Y. Hu, "Subspace algorithms for noise reduction in cochlear implants," *J. Acoust. Soc. Am.*, 118, 2791–2793, 2005.

[11] F. Chen, Y. Hu, and M. Yuan, "Evaluation of noise reduction methods for speech recognition by Mandarin-speaking cochlear implant listeners," *Ear Hear.*, 36, 61–71, 2015.

[12] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted with colored noise," *IEEE Trans Speech and Audio Processing*, 11, 334–341, 2003.

[13] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans on Acoust, Speech, Signal Processing*, ASSP-33, 443–445, 1985.

[14] S. Kamath and P. C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *Proceedings of International Conference on Acoust. Speech and Signal Processing*, 4164–4167, 2002.

[15] P. Scalart and J. Filho, "Speech enhancement based on a priori signal to noise estimation," *Proceedings of International Conference on Acoust. Speech and Signal Processing*, 629–632, 1996.

[16] P. C. Loizou, Speech Enhancement: Theory and Practice, 2ed, Taylor and Francis, Boca Raton, FL, 2013.