35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A Speech Enhancement Method for Cochlear Implant Listeners

Meng Yuan, *Member, IEEE*, Yang Sun, Haihong Feng, and Tan Lee, *Member, IEEE*

Abstract—**This paper discusses a single-channel speech enhancement method for cochlear implant listeners. It is assumed that the Fourier Transform coefficients of speech and background noise have different statistical distributions. A statistical-model-based method is adopted to update the signal-to-noise ratio and estimate the background noise so that the musical noise and speech distortion induced by traditional spectral subtraction method can be effectively reduced. This enhancement method was evaluated on seven postlingually deaf Chinese cochlear implant listeners in comparison with other two speech enhancement methods. Test materials were Mandarin sentences corrupted by three different types of background noise. Experimental results showed that the proposed speech enhancement method could benefit the speech intelligibility of Chinese cochlear implant listeners. The results suggest that different noise types may affect the performance of different speech enhancement algorithms.**

I. Introduction

Cochlear implant is a medical device which can electrically stimulate auditory nerve to restore partial hearing for severe to profoundly deaf persons. In the past three decades, tremendous improvement of cochlear implant (CI) technologies has been achieved enabling many CI users to enjoy high levels of speech understanding in quiet environments; however, for most CI listeners, listening under noisy conditions remains challenging [1]. Many studies reported that CI listeners are considerably more vulnerable to noise than normal-hearing (NH) listeners [2][3]. Even a small amount of noise may cause them uncomfortable and lose the target sound entirely, whereas it may not be a problem for NH people. This difficulty appears to be related to a variety of abnormalities in the perception of sound. One main abnormality is the limited spectral resolution that CI could provide. Typically, there are maximally 16-22 spectral channels (or electrodes) in a state-of-the-art CI system, which is much less than the number of frequency bands used in a normal-hearing person. The consequence of limited spectral resolution means that (i) the spectral features of speech sounds are less well resolved; (ii) background noise has a greater masking effect, since broader frequency bands generally can pass more background noise. A straightforward approach to solve this problem is to provide more frequency bands (or electrodes) in a CI device. However, due to technological constrains, little progress has been reported on the increment of spectral channels in CIs to date. Therefore, speech enhancement will play an important role for the hearing impaired people who are wearing CIs.

Several authors have described attempts to improve speech intelligibility for the hearing impaired [2], [4]-[8]. Some of these algorithms were based on the assumption that two or more microphones were available. Studies showed that an adaptive beam-forming algorithm based on multiple-microphone could substantially benefit the speech intelligibility of CI listeners when the speech and noise signals were from different directions [4]. However, due to the constrained dimension, it is not practicable to implement a second microphone for unilateral CI recipients. Therefore, single-microphone noise-reduction algorithms are more appealing and more feasible for implementation. As is known, several single-microphone noise-reduction algorithms have been proposed for cochlear implants [5][6]. Some of these algorithms were implemented on old CI processors, which were based on feature extraction strategies (F0/F1/F2 and MPEAK strategies) by enhancing the spectral features. However, the latest speech processors are based on vocoder-type strategies. For such strategies, e.g. Continuous Interleaved Sampling (CIS) and Advanced Combination Encoder (ACE), no features are extracted, as the signal is bandpass-filtered into *n* bands (8 to 22), and the temporal envelopes of the signal are extracted from each band. Several single-microphone pre-processing noise-reduction algorithms were also proposed. These algorithms are based either on spectral subtraction [8], or on statistical-model-based methods [7][9], or on subspace method [10]. Although the above noise-reduction algorithms have provided little benefit for normal hearing listeners [9], small but significant improvements have been reported for CI listeners [6]-[9]. However, a substantial performance gap still remains between the speech recognition in noisy conditions and in quiet. For a multi-channel spectral subtraction, which is often used in the current CI system, the noise levels in different frequency bands are estimated and the Signal-to-Noise Ratio (SNR) in each band of the noisy speech is determined. The speech signal is estimated by subtracting the estimated noise spectral magnitude from the noisy speech spectral magnitude. A gain function is used to determine a level of attenuation to be applied to the signal to optimally remove the noise. The phase of the noisy speech spectrum is preserved based on the best estimate of the clean speech in a least mean square sense. The main disadvantage of this method is the "musical noise" artifact, which is mainly due to the inaccuracy of the spectrum estimation.

In the present study, a modified spectral subtraction method is proposed with the use of statistical-model-based SNR update and noise update. Subjective evaluation of the

*Research supported by National Natural Science Foundation of China (11104316), Shanghai Natural Science Foundation (11ZR1446000) and Chairman Foundation of Institute of Acoustics, CAS (Y154221701).

Meng Yuan, Yang Sun, and Haihong Feng are with the Medical Center for Hearing and Speech, Shanghai Acoustics Laboratory, Chinese Academy of Sciences, Shanghai 200032, China (Corresponding author: Meng Yuan, phone: 86-021-64174235; e-mail: ym@mal.ioa.ac.cn).

Tan Lee is with the Department of Electronic Engineering, Chinese University of Hong Kong, Shatin, N. T., Hong Kong (e-mail: tanlee@ee.cuhk.edu.hk).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

speech intelligibility on seven CI listeners was carried out. A set of Chinese sentences were utilized as the test materials. Statistical analysis will be performed to evaluate the performance of the proposed speech enhancement method compared with other two commonly used noise-reduction algorithms.

II. Statistical-Model-Based Speech Enhancement Method

The proposed speech enhancement method is based on the traditional spectral subtraction algorithm [11]. The SNR was updated by the use of a statistical-model-based method [12]. This method assumes that the short-time Fourier Transform parameters of speech obey Rayleigh distribution while short-time Fourier Transform parameters of noise obey Gaussian distribution. Based on the minimum mean square error rule (MMSE), the SNR was updated as follows:

$$R_{prio}(m,k) = (1-\theta)P\left(R_{post}(m,k)-1\right) + \theta\frac{\left|G(m-1,k)\right|^2}{\left|\overline{D}(m-1,k)\right|^2} \quad (1)$$

and

$$R_{post}(m,k) = \frac{\left|X(m,k)\right|^2}{\left|\overline{D}(m-1,k)\right|^2} \quad (2)$$

where m represents the frame index, k represents the frequency point index after Fourier Transform in each frame, $R_{post}(m,k)$ represents the posterior SNR, $R_{prio}(m,k)$ represents the prior SNR, $\overline{D}(m-1,k)$ is the discrete-time Fourier Transform parameter of the estimated noise in the previous frame, $G(m-1,k)$ is the enhanced speech signal of the previous frame, and $X(m,k)$ represents the magnitude energy of the Fourier Transform parameter of the m*th* frame. The posterior SNR $R_{post}(m,k)$ is the correction factor of the prior SNR $R_{prio}(m,k)$. In (1), $P[x]=0$ when $x\le 0$; otherwise, $P[x]=x$. And θ is artificially set to 0.98.

For noise update, we adopted the Maximal Likelihood Estimation rule (MLE):

$$\overline{\Lambda}_m = \frac{1}{N}\sum_{k=0}^{N-1}\Lambda(m,k) \quad (3)$$

and

$$\Lambda(m,k) = \frac{1}{1+R_{prio}(m,k)} \bullet \exp\left(\frac{R_{prio}(m,k)R_{post}(m,k)}{1+R_{prio}(m,k)}\right) \quad (4)$$

In (3), $\overline{\Lambda}_m$ is the averaged possibility of noise occurrence in the frame m. In (4), $\Lambda(m,k)$ represents the noise occurrence possibility in the kth frequency point of frame m. N represents the total number of points in the current frame. If $\overline{\Lambda}_m$ is larger than a threshold η, it is assumed that the current frame only contains clean speech signal and the noise will be updated as the same as the previous frame; otherwise, the noise will be updated as:

$$\left|\overline{D}(m,k)\right|^2 = \gamma\left|\overline{D}(m-1,k)\right|^2 + (1-\gamma)\left|X(m,k)\right|^2 \quad (5)$$

Fig. 1 shows the diagram of the proposed speech enhancement method in each frame of noisy speech signal.

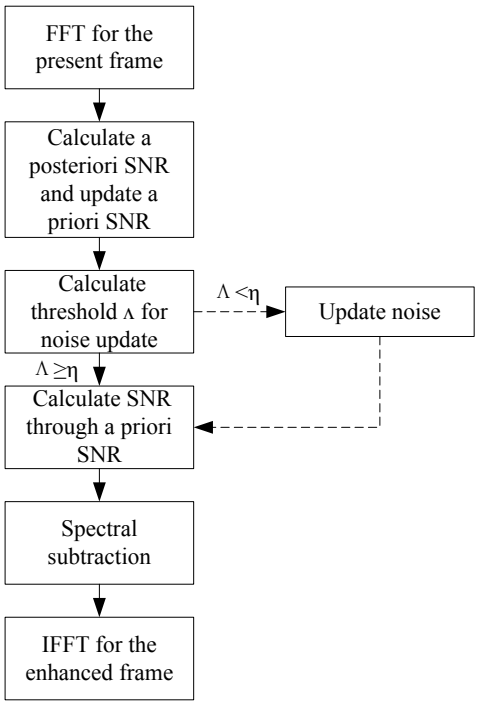

Figure 1. Block diagram of the proposed speech enhancement method in each frame of noisy speech signal.

Finally, the enhanced speech signal $G(m,k)$ is derived as

$$\text{if } \left|\overline{X}(m,k)\right|^2 \ge \beta(m,k)\delta(m,k)\left|\overline{D}(m,k)\right|^2,$$
$$\left|G(m,k)\right|^2 = \left|\overline{X}(m,k)\right|^2 - \beta(m,k)\delta(m,k)\left|\overline{D}(m,k)\right|^2,$$
$$\text{else, } Floor \bullet \left|\overline{X}(m,k)\right|^2. \quad (6)$$

, where *Floor* is set to a small positive value to avoid over-subtraction (typically set to 0.002); $\beta(m,k)$ is the original spectral subtraction factor. Compared to the traditional spectral subtraction method [11], an additional spectral subtraction factor $\delta(m,k)$ was used to reflect the different distribution of noise in each frequency band [13].

III. Materials And Methods

A. Subjects

Seven postlingually deafened native Mandarin-speaking listeners participated in this experiment (S1-S7). All subjects used the same type of cochlear implant devices (see Table 1

for more information about on the participants). The implant contains 24 inner-cochlea electrodes and 2 outer-cochlea reference electrodes. All subjects had at least 2-years' experience since their CI devices were opened. Their speech perception evaluation was generally good (overall recognition accuracy > 70% in quiet condition).

B. Speech Materials

The speech material was adopted from [14] which includes 20 lists of Chinese sentences. Each list contains 10 sentences. The energy of all the sentences has been RMS-equalized and additively corrupted by three different types of background noise (speech-spectrum-shaped noise, car noise, and babble noise) at different SNRs.

C. Speech Processing

There are three speech enhancement methods evaluated in this study, which included Multi-band Spectral Subtraction (MBSS), Minimum Mean-Square Error Log-Spectral Amplitude Estimator (Log-MMSE) [15], and the proposed Statistical-Model-based Spectral Subtraction (SMSS). All of the noisy speech signals were processed with the three speech enhancement methods at different SNRs. The original unprocessed noisy speech signals were also included as control in this study.

TABLE I. SUBJECT INFORMATION

Subject ID	Etiology	Implant Ear	Deaf Duration(yrs)	Date of Implantation
S1	Parotitis	L	9	2010.4
S2	Sudden	L	2	2010.3
S3	Noise	L	6	2009.12
S4	Unknown	L	Unknown	Unknown
S5	Sudden	R	8	2010.3
S6	Ototoxic	R	11	2010.3
S7	Unknown	L	18	2010.3

D. Psychophysical Procedure

All of the CI subjects used their own speech processors in this study. They were seated in a sound-proof room and listened to acoustic stimuli presented from a high-quality loudspeaker at 65 dBA SPL. For each subject, the speech stimuli from 4 speech processing conditions (3 speech enhancement methods + 1 original noisy speech), 3 noise conditions (speech-spectrum-shaped noise, car noise, and babble noise) formed 12 test sessions. The presentation order of these sessions was randomized. In each session, one of the 20 lists was randomly selected and tested. The initial SNR was set to 20 dB. After presenting a sentence through the loudspeaker, the subject was required to repeat what he heard. If the subject could repeat the sentence over 50% correct, this sentence was regarded as intelligible and the SNR will be decreased at a certain value. Otherwise, the SNR will be increased. The step size for the first 5 sentences was set to 4 dB while the step size was set to 2 dB for the rest 15 sentences. The final result was decided as the average SNR of the last 16 sentences.

IV. RESULTS

The results of sentence recognition of the 7 subjects were analyzed using a two-way repeated-measures analysis (ANOVA) with the 2 factors of noise type (speech spectrum shaped noise, babble noise, and car noise) and speech enhancement method (MBSS, MMSE, SMSS, and the unprocessed noisy speech).

The results revealed that noise type had no significant main effect of Chinese sentence recognition (p = 0.7), while the speech enhancement method significantly affected Chinese sentence recognition [F(3, 18) = 6.16, p = . 0045]. There is no significant effect on the interaction of noise type and speech enhancement method.

Fig. 2 shows the average sentence recognition performance over all 7 CI subjects under different test conditions. Mean SNRs for the three unprocessed noisy speech conditions were 17.34 dB, 18.53 dB, and 17.70 dB, respectively. Generally speaking, all of the three speech enhancement methods could enhance the speech intelligibility under different noisy environments. *Post-hoc* LSD tests were carried out to examine the effects of different speech enhancement methods. For babble noise, the post-hoc analysis revealed that the MBSS method could significantly enhance the speech intelligibility of the noisy counterpart and so as for MMSE method (p < .05). For car noise conditions, the MMSE method showed significantly better speech intelligibility than the unprocessed noisy condition and the MBSS method (p < .05).

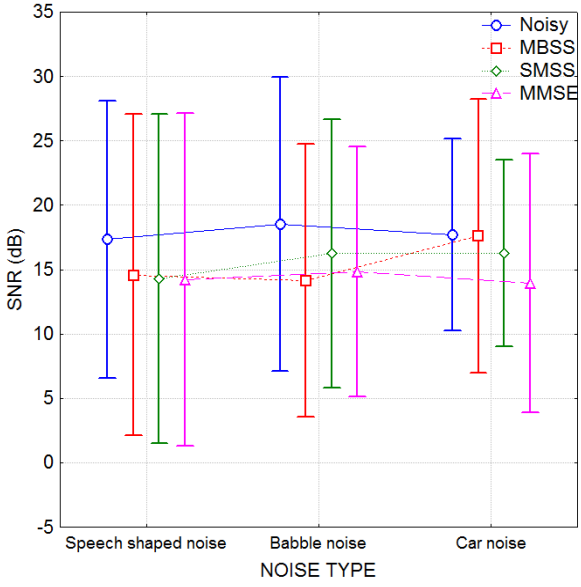

Figure 2. Chinese sentence recognition result with different noise types and speech enhancement methods.

V. Discussion

A. Comparison of speech enhancement methods towards speech intelligibility

In this study, we presented three speech enhancement methods and evaluated their performance on the 7 CI subjects. It was shown that all of the speech enhancement methods could benefit the speech intelligibility in different noisy environments. For speech spectrum shaped noise, the three methods showed similar performance. For babble noise, MBSS and MMSE showed better performance than SMSS, while SMSS demonstrated a better speech intelligibility of 2.2 dB than the unprocessed noisy condition. For car noise, MMSE showed the best performance, while MBSS didn't show enhancement on speech intelligibility in comparison to the noisy speech. In all the three noisy environments, SMSS always show an improvement on speech intelligibility.

B. Performance variability across CI subjects

From Fig. 2 we can find that the performance variability across CI subjects was very large. One reason may be due to the limited number of subjects that we have recruited in this experiment. Another more possible reason may be due to the large difference of speech intelligibility among these CI recipients. Fig. 3 shows the individual performance of these 7 CI subjects. It is shown that the performance among these subjects was quite different. For S1 and S4, their speech intelligibility performance in noise was quite poor compared to other 5 subjects. This individual difference caused the large variability in Fig. 2. It was also shown that the speech enhancement methods could improve the speech intelligibility in most cases, except for S4 whose performance was the worst. It seems that the speech enhancement methods could not benefit those persons that have severely poor speech intelligibility in noise.

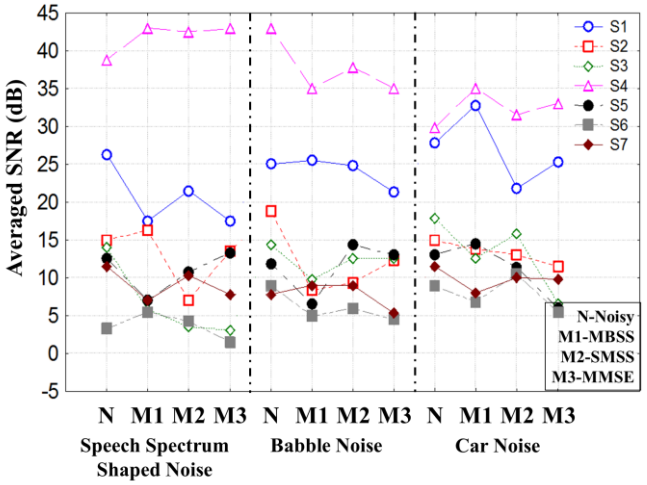

Figure 3. Individual speech intelligibility results of the 7 CI subjects.

VI. Conclusion

From this study, we can draw the following conclusions:

(1) Speech enhancement methods could benefit the speech intelligibility of CI recipients;

(2) The proposed statistical-model-based spectral subtraction algorithm is a robust speech enhancement method and shows a relatively good performance to speech intelligibility;

(3) Large individual difference was observed among CI subjects, suggesting that more should be concerned when dealing with CI subjects. Further efforts should be taken to fit the subjects with a suitable speech enhancement treatment.

Acknowledgment

We thank all the cochlear implant recipients who participated in this study. Also thanks are given to our audiologist, Miss Jin Sun, for her help during the test.

References

[1] P. Loizou, "Speech processing in vocoder-centric cochlear implants," in *Cochlear and Brainstem Implants*, vol. 64, A. Moller, Ed. Adv. Otorhinolaryngol., Basel, Karger, 2006, pp. 109–143.

[2] B. L. Fetterman, and E. H. Domico, "Speech recognition in background noise of cochlear implant patients," *Otolaryngol. Head Neck Surg.*, vol. 126, pp. 257–263, 2002.

[3] L. M. Friesen, R. V. Shannon, D. Baskent, and X. Wang, "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.*, vol. 110, pp. 1150–1163, 2001.

[4] R. van Hoesel and G. Clark, "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Am.*, vol. 97, pp. 2498–2503, 1995.

[5] M. Weiss, "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," *J. Rehab. Res. Dev.*, vol. 30, no. 1, pp. 117-128, 1993.

[6] I. Hochberg, A. Boorthroyd, M. Weiss, and, S. Hellman, "Effects of noise and noise suppression on speech perception by cochlear implant users," *Ear Hear.*, vol. 13, no. 4, pp. 263-271, 1992.

[7] P. W. Dawson, S. J. Mauger, and A. A. Hersbach "Clinical evaluation of signal-to-noise ratio based noise reduction in nucleus cochlear implant recipients," *Ear Hear.*, vol. 32, pp. 382–390, 2011.

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

[9] Y. Hu and P. Loizou, "Subjective evaluation and comparison of speech enhancement algorithms," *Speech Commun.*, vol. 49, pp. 588–601, 2007.

[10] Y. Hu, and P. Loizou, "A generalized subspace approach for enhancing speech corrupted with colored noise," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 4, 334–341, 2003.

[11] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-27, no. 2, pp. 113-120, 1979.

[12] Y. Ephraim and D. Malah, "Speech enhancement using optimal non-linear spectral amplitude estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Boston, 1983, pp. 1118-1121.

[13] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Processing*, Orlando, FL, 2002.

[14] X. Xi and F. Ji, "Mandarin speech test: monosyllabic word test CD," in *PLA Health Audiovisual Press*, 2009.

[15] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah Noise Suppressor," *IEEE Trans. Speech Audio Processing*, vol. 2, no. 2, pp. 345-349, 1994.