PSYCHOACOUSTICALLY-MOTIVATED ADAPTIVE β-ORDER GENERALIZED SPECTRAL SUBTRACTION FOR COCHLEAR IMPLANT PATIENTS

Junfeng Li[1]*, Qian-Jie Fu*[2]*, Hui Jiang*[3] *and Masato Akagi*[1]

[1] School of Information Science, Japan Advanced Institute of Science and Technology, Japan
[2] Division of Communication and Auditory Neuroscience, House Ear Institute, USA
[3] Department of Computer Science and Engineering, York University, Canada
Email: junfeng@jaist.ac.jp, qfu@hei.org, hj@cse.yorku.ca, akagi@jaist.ac.jp

ABSTRACT

Many *cochlear implant* (CI) users are able to understand speech in quiet listening conditions, however, CI users' speech recognition deteriorates rapidly as the level of background noise increases. To make CI more applicable in real-life environments, noise reduction is needed in CI processor. Recently, we presented a psychoacoustically-motivated adaptive β-order *generalized spectral subtraction* (GSS) which deals with the weakness of the traditional SS algorithms [9, 10]. To apply this adaptive β-order GSS into CI processor, in this paper, we investigate the effects of noise estimation approaches and residual noise components for the proposed adaptive β-order GSS. Word-in-sentence recognition in steady white noise and speech babble noise was measured in four CI users. Experimental results showed that 1) noise estimation significantly affected performance of the proposed algorithm, 2) the algorithm with the least residual noise components was preferred by CI subjects, and 3) the proposed psychoacoustically-motivated adaptive β-order GSS outperformed the traditional SS algorithms.

Index Terms— Speech intelligibility, Band-importance function, Adaptive β-order GSS, Cochlear implant.

1. INTRODUCTION

Assistive hearing devices (e.g., hearing aids, cochlear implants) have restored hearing sensation to many hearing-impaired individuals. Many *cochlear implant* (CI) users are able to understand speech in optimal, quiet listening conditions. However, most devices provide limited benefit in the presence of background noise or competing speech [1].

To reduce the effects of background noise, many single- and multi-microphone noise reduction algorithms have been utilized to improve the quality and/or intelligibility of the noisy speech [2]. In CI devices, single-microphone techniques are more desirable and appealing than those based on multi-microphone inputs [3]. Within the single-microphone technique, *spectral subtraction* (SS) has widely been used due to its simplicity in implementation, and improved to overcome its shortcomings in different ways [4, 5, 6, 7]. Yang and Fu investigated the benefits of preprocessing the noisy signal by a standard SS algorithm and showed the significant benefits in CI users' word-in-sentence recognition [8]. In the original SS and its modifications, however, the spectral order β is usually fixed to some constants, which greatly limits the noise-reduction benefits in complex and real-world environments [9]. Moreover, the traditional SS algorithms are primarily designed to improve speech quality rather than speech intelligibility, which is a crucial demand for CI users [4, 5, 6, 7]. To overcome these drawbacks and improve speech intelligibility, we recently proposed an adaptive β-order *generalized spectral subtraction* (GSS) method that incorporated the psychoacoustic knowledge on speech intelligibility, e.g., band-importance function [10]. In the proposed adaptive β-order GSS, the spectral order β is updated frame by frame within each subband according to the input local *signal-to-noise ratios* (SNRs) in the time-frequency domain [10]. In its original implementation, the noise spectrum is assumed to be known *a priori*. However, to be beneficial in real-world environments, the noise spectrum must be estimated from the noisy speech signal.

To make the proposed adaptive β-order GSS more applicable for CI users, in this paper, we investigate the effects of different noise estimation approaches and residual noise components for the proposed adaptive β-order GSS on CI users' speech recognition performance in noise. The effect of noise estimation was examined by utilizing two noise estimation approaches: 1) soft-decision based noise estimation and 2) robust *voice activity detector* (VAD) based noise estimation. The effect of residual noise components was investigated by applying a VAD-based technique to remove all residual noise components during speech absence periods. Finally, CI users' word-in-sentence recognition performance was compared between the proposed adaptive β-order GSS algorithm and the traditional SS algorithms.

This study is supported by a Grant-in-Aid for Young Scientists (B) (No. 19700156) from the Ministry of Education, Science, Sports and Culture of Japan.

2. PSYCHOACOUSTICALLY-MOTIVATED ADAPTIVE β-ORDER GSS

2.1. β-order GSS

The β-order GSS is defined as [7]

$$\left|\hat{S}_\beta(k,\ell)\right|^\beta = a_\beta(k,\ell)\left|X(k,\ell)\right|^\beta - b_\beta(k,\ell)E\left[\left|N(k,\ell)\right|^\beta\right], \quad (1)$$

where β denotes the spectral order; $a_\beta(k,\ell)$ and $b_\beta(k,\ell)$ are two parameters; k and ℓ are the frequency bin index and the time frame index; $X(k,\ell)$, $N(k,\ell)$, and $\hat{S}_\beta(k,\ell)$ are the STFTs of the noisy signal, the noise signal and the enhanced signal by the β-order GSS. Under the complex Gaussian assumption of the spectra of the clean and noise signals, the gain function of the β-order GSS is derived as [7]

$$\hat{G}_\beta(k,\ell) = \left\{\frac{\left[\xi_\beta(k,\ell)\right]^\beta}{1+\left[\xi_\beta(k,\ell)\right]^\beta}\right\}^{\frac{1}{\beta}}\left\{1-\left(1-\left[\xi_\beta(k,\ell)\right]^{\frac{-\beta}{2}}\right)\Gamma\left(\frac{\beta}{2}+1\right)\left(\frac{1}{\gamma(k,\ell)}\right)^{\frac{\beta}{2}}\right\}^{\frac{1}{\beta}}, \quad (2)$$

where $\xi_\beta(k,\ell)$ and $\gamma(k,\ell)$ are the *a priori* SNR and the *a posteriori* SNR as defined in [11]; $\Gamma(\cdot)$ denotes the Gamma function. The estimate of $\xi_\beta(k,\ell)$ is updated in a decision-directed scheme, greatly decreasing the residual "musical" noise [11].

2.2. Psychoacoustically-motivated adaptive β-order GSS

The proposed adaptive β-order GSS is derived based on the following two observations [10]:

1. SNRs vary greatly with time due to the time-varying characteristics of speech and noise signals, and also significantly vary in different subbands because of the colorness of noise signals and the non-uniform distribution of spectral energy in speech signals. As a result, speech signal corrupted by real-world noises is characterized by different local SNRs, which results in that the appropriate value of spectral order β must be adaptively determined according to the local SNRs for different partitions in the time-frequency domain.

2. Different frequency bands contribute different amounts to speech intelligibility, which is defined by the band-importance function [12]. As a result, the band-importance function should be integrated when determining the appropriate value of spectral order β.

Then, we propose to optimize the spectral order β by minimizing the overall intelligibility-weighted distance between the spectral amplitude $\left|S(k,\ell)\right|$ of the clean signal and that of its estimate $\left|\hat{S}_\beta(k,\ell)\right|$ summed across all subbands, that is,

$$\beta^{\text{opt}} = \underset{0.1\le\beta\le3.0}{\arg\min}\left(\sum_{m=1}^{M}\sum_{k=\omega_m}^{\omega_{m+1}} I_m\left[\left|S(k,\ell)\right|-\left|\hat{S}_\beta(k,\ell)\right|\right]^2\right), \quad (3)$$

where I_m is the band-importance function in the m-th subband [12], M is the number of subbands, ω_m denotes the boundary frequency of the m-th subband, and the range of β is empirically confined to $[0.1, 3.0]$. Through a data-driven optimization processing, it is shown that the optimized value of the spectral order $\beta(m,\ell)$ should be adaptively updated frame by frame in each subband, given by [10]

$$\hat{\beta}(m,\ell) = \frac{B}{1+e^{-A\left[\rho(m,\ell)-D\right]}}, \quad (4)$$

where $\rho(m,\ell)$ is the local SNR in the m-th subband and ℓ-th frame, the parameter A controls the changing speed of the value of $\hat{\beta}(m,\ell)$, B determines the range of the value of $\hat{\beta}(m,\ell)$, and D denotes the shift along the SNR axis. According to the optimization results in [10], these parameters were optimized as $A=0.1$, $B=2.0$ and $D=7.0$. The local SNR $\rho(m,\ell)$ is calculated as

$$\rho(m,\ell) = 10\log_{10}\left(\frac{\sum_{k=\omega_m}^{\omega_{m+1}}\left|\left|X(k,\ell)\right|-\left|\hat{N}(k,\ell)\right|\right|^2}{\sum_{k=\omega_m}^{\omega_{m+1}}\left|\hat{N}(k,\ell)\right|^2}\right), \quad (5)$$

where $\hat{N}(k,\ell)$ is the noise spectrum estimate that has to be calculated from the observed noisy signal.

2.3. Noise spectrum estimation

The accuracy of noise spectrum estimation would be expected to affect the performance of the proposed adaptive β-order GSS. To investigate this effect, two noise spectrum estimation approaches were tested: 1) the soft-decision based noise estimation approach [13], and 2) the robust *voice activity detection* (VAD) based noise estimation approach [14].

2.3.1. soft-decision based noise estimation

The soft-decision based noise estimation approach is given by

$$E\left[|N(k,\ell)|^2\right] = \alpha E\left[|N(k,\ell-1)|^2\right] + (1-\alpha)E\left[|N(k,\ell)|^2\middle|X(k,\ell)\right], \quad (6)$$

where α ($0<\alpha<1$) is a forgetting factor. Under speech presence uncertainty, the second term in the right side of Eq. (6) can be estimated as

$$E\left[|N(k,\ell)|^2\middle|X(k,\ell)\right] = q(k,\ell)|X(k,\ell)|^2 + \left(1-q(k,\ell)\right)E\left[|N(k,\ell-1)|^2\right]. \quad (7)$$

where $q(k,\ell)$ is the speech absence probability defined in [11].

2.3.2. VAD based noise estimation

To determine speech activity, the following VAD decision rule was used [14]

$$\frac{1}{\Omega_m}\sum_{k=\omega_m}^{\omega_{m+1}}\left[\frac{\gamma(k,\ell)\xi(k,\ell)}{1+\xi(k,\ell)}-\log\left(1+\xi(k,\ell)\right)\right]\gtrless_{H_0}^{H_1}\delta, \quad (8)$$

where Ω_m denotes the number of frequency bin in the m-th subband, H_1 and H_0 denote the hypothesis of speech presence and absence, respectively; δ is a fixed threshold set to $\delta = 0.5$ [14]. Finally, the noise spectrum is updated during speech-absence periods.

2.4. Residual noise processing

In the proposed adaptive β-order GSS, the residual noise is expected to be more stationary due to the use of the decision-directed *a priori* SNR estimation. To test the effects of the residual noise components, a VAD-based post-processing technique was implemented to remove all the noise components during speech pauses based on the output of the adaptive β-order GSS.

3. PERFORMANCE EVALUATIONS

3.1. Subjects

Four post-lingually deafened adults using the Nucleus and Clarion cochlear implant device participated in this study. All were native speakers of American English and had at least nine years experience with the device. All implant subjects had extensive experience in speech recognition experiments.

3.2. Test materials and procedures

Word-in-sentence recognition was measured using IEEE sentence materials [15]. Sentence stimuli were digitized recordings produced by 1 male talker (recorded at House Ear Institute). Sentence materials were normalized to have the same long-term RMS (65 dB). Word-in-sentence recognition was tested in the presence of two types of noise: 1) computer-generated white Gaussian noise and 2) multi-talker speech babble noise. Noise was added to the clean speech at 5dB and 0 dB SNRs, where the long term average of each sentence was used to calculate the SNR. Speech and noise were mixed together at the target SNR and then processed by the noise-reduction algorithms.

Baseline performance was first tested for all noise types and SNRs without any noise reduction. Three experimental adaptive β-order GSS algorithms were tested: 1) soft-decision based noise estimation (SD), 2) voice activity detection based noise estimation (VAD) and 3) VAD with post-filter processing to remove all residual noise (VADPF). In addition, three traditional SS algorithms were tested with fixed β-order values: β= 0.5, 1.0, and 2.0. All stimuli were presented over a single loudspeaker at a comfortably-loud listening level (fixed speech level = 65 dBA). Subjects were tested using their clinically assigned CI speech processors; subjects were asked to set their microphone sensitivity and volume settings for comfortably-loud speech and to not change these settings or their processor program for the duration of the experiment. Subjects were tested while seated in a sound-treated booth. During testing, a list was chosen randomly (without replacement) from among 60 lists, and sentences were chosen randomly (without replacement) from among the 10 sentences within that list. Subjects responded by repeating the sentence as accurately as possible; the experimenter tabulated correctly identified words and sentences. At least two sentence lists were tested for each condition. Processing conditions, noise types and SNR levels were randomized within and across subjects.

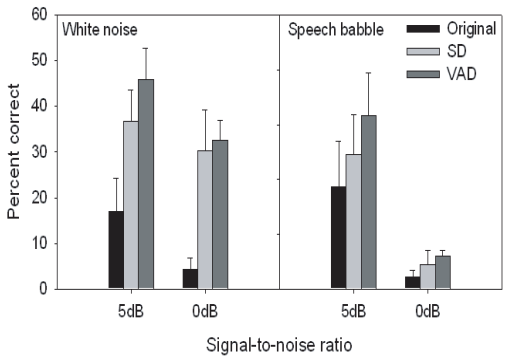

Fig. 1. Word-in-sentence recognition by four CI users for original speech-in-noise and the proposed adaptive β-order GSS algorithms with noise estimation based on either soft-decision (SD) or voice activity detector (VAD) noise estimation techniques. The error bars show one standard error of the mean.

3.3. Evaluation results and discussions

3.3.1. Effect of noise estimation techniques

Speech recognition performance was compared between original speech-in-noise and the proposed adaptive β-order GSS algorithm with different (SD and VAD) noise estimation approaches. As shown in Fig. 1, both noise estimation techniques produced markedly better speech performance. A two-way repeated measures analysis of variance (RM ANOVA) showed that both algorithms significantly improved performance in noise [F(2,18)=25.25, p=0.001].

3.3.2. Effect of residual noise components

To examine the effect of residual noise components on performance, word-in-sentence recognition was compared between original speech-in-noise, and the proposed adaptive β-order GSS with (VADPF) and without (VAD) post-filter processing. The VADPF processing removed all residual noise components, as described in section 2.4. As shown in Fig. 2, both techniques provided improved speech performance in noise. A two-way RM ANOVA showed that both algorithms significantly improved performance in noise [F(2,18)=17.18, p=0.003]. While performance was slightly better with the VAD algorithm, post-hoc Bonferroni t-tests showed no significant effect for post-processing.

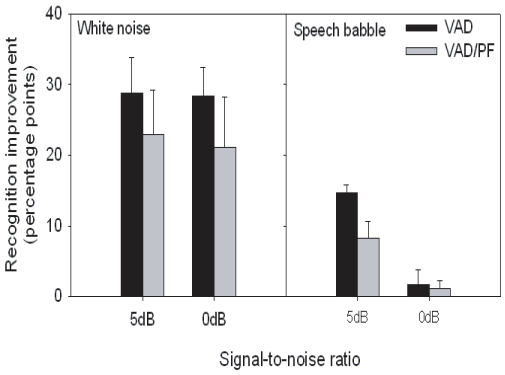

Fig. 2. Improvement in performance (relative to original speech-in-noise) for the VAD noise estimation algorithm, with (VADPF) or without (VAD) post-filtering to remove all residual noise. The error bars show one standard error of the mean.

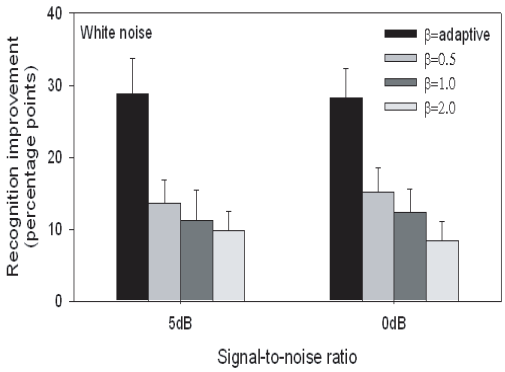

Fig. 3. Improvement in performance (relative to original speech-in-noise) for the proposed adaptive β-order GSS and three traditional SS algorithms with fixed β values. The error bars show one standard error of the mean.

3.3.3. Superiority of the adaptive β-order GSS

Performance of the proposed adaptive β-order GSS was compared to that of the traditional SS methods with fixed spectral order β values. For the traditional SS methods, β was fixed at 2.0 (power SS), 1.0 (amplitude SS) or 0.5 in Eq. (2). In all these algorithms, the VAD based noise estimation was exploited without the post processing for removing the residual noise components. As shown in Fig. 3, the adaptive β-order GSS provided a much greater improvement in speech understanding in noise. A two-way RM ANOVA showed that all the algorithms significantly improved performance in noise [$F(4,12)=22.92$, $p<0.003$]. Post-hoc Bonferroni t-tests showed that the adaptive β-order GSS provided significantly better performance than the traditional SS algorithms with fixed β values ($p<0.05$), and that there was no significant difference between the SS algorithms with the fixed β values.

4. CONCLUSION

In this paper, we proposed and tested a psychoacoustically-motivated adaptive β-order GSS. Speech recognition in noise was tested in CI users with and without the proposed noise-reduction algorithms. Results showed that the adaptive β-order outperformed the traditional SS algorithms with fixed β-order values. Experimental results also showed that noise estimation plays a crucial role in the proposed adaptive β-order GSS, and that CI users preferred the proposed adaptive β-order GSS algorithm without residual noise processing.

5. REFERENCES

[1] P.C. Loizou, "Introduction to cochlear implants," *IEEE Eng. in Medicine and Biology Magazine,* vol. 18, pp. 32-42, 1999.

[2] J. Benesty, S. Makino and J. Chen, Speech Enhancement, Springer, 2005.

[3] P.C. Loizou, A. Lobo and Y. Hu, "Subspace algorithms for noise reduction in cochlear implants," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 2791-2793, 2005.

[4] S. F. Boll, "Suppression of acoustic noise in speech using spectral subraction," *IEEE Tans. ASSP*, vol. 27, pp. 113-120, 1979.

[5] V. Schless and F. Class, "SNR-dependent flooring and noise overestimation for joint application of spectral subtraction and model combination," In *Proc. ICSLP*, pp. 721-725, 1998.

[6] S.D. Kamath and P.C. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *Proc. ICASSP*, pp. 4164-4167, 2002.

[7] B.L. Sim, *et al.* "A parametric formulation of the generalized spectral subtraction," *IEEE Trans. SAP*, vol. 6, no. 4, pp. 328-337, 1998.

[8] L. Yang and Q. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," *J. Acoust. Soc. Am.*, vol. 117, no. 3, pp. 1001-1004, 2005.

[9] J. Li, *et al.*, "Noise reduction based on adaptive β-order generalized spectral subtraction for speech enhancement," in *Proc. Interspeech2007*, pp. 802-805, 2007.

[10] J. Li, *et al.*, "Psychoacoustically-motivated adaptive β-order generalized spectral subtraction based on data-driven optimization," in *Proc. Interspeech2008*, pp. 171-173, 2008.

[11] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. ASSP*, vol. 32, no. 6, pp. 1109-1121, 1984.

[12] ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index," 1997.

[13] N.S. Kim and J.H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letter*, vol. 7, no. 5, pp. 108-110, 2000.

[14] J. Sohn and N. Kim, "Statistical model-based voice activity detection," *IEEE Signal Processing*, vol. 6, no. 1, pp. 1-3, 1999.

[15] IEEE Subcommittee, "IEEE recommended practice for speech quality measures," *IEEE Trans. Audio and Electroacoustics*, pp. 225-246, 1969.