Use of a sigmoidal-shaped function for noise attenuation in cochlear implants

Yi Hu, Philipos C. Loizou,[a] Ning Li, and Kalyan Kasturi
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688
yihuyxy@utdallas.edu, loizou@utdallas.edu, lining@student.utdallas.edu, kalyan@utdallas.edu

Abstract: A new noise reduction algorithm is proposed for cochlear implants that applies attenuation to the noisy envelopes inversely proportional to the estimated signal-to-noise ratio (SNR) in each channel. The performance of the proposed noise reduction algorithm is evaluated with nine Clarion CII cochlear implant patients using IEEE sentences embedded in multi-talker babble and speech-shaped noise at 0–10 dB SNR. Results indicate that the sigmoidal-shaped weighting function produces significant improvements to speech recognition compared to the subjects' daily strategy. Much of the success of the proposed noise reduction algorithm is attributed to the improved temporal envelope contrast.

PACS numbers: 43.66.Ts, 43.66.Sr [QJF]
Date Received: May 16, 2007 **Date Accepted:** July 10, 2007

1. Introduction

Although many cochlear implant (CI) users are enjoying high levels of speech understanding in quiet environments, noisy listening conditions remain challenging for most. A number of preprocessing noise-reduction algorithms have been proposed for cochlear implants over the years (Yang and Fu, 2005; Loizou *et al.*, 2005; Loizou, 2006; Van Hoesel and Clark, 1995; Wouters and Vanden Berghe, 2001). The preprocessing approach to noise reduction, however, has three main drawbacks: (1) preprocessing algorithms sometimes introduce unwanted distortion in the signal, (2) some algorithms (e.g., subspace algorithms) are computationally complex (and consequently power hungry) and do not integrate well with existing CI strategies, and (3) it is not easy to optimize the operation of a particular algorithm to individual users.

Ideally, noise reduction algorithms should be easy to implement and be integrated into existing coding strategies. In this paper, we propose a simple noise reduction algorithm that can be easily integrated in existing strategies used in commercially available devices. The proposed algorithm fits into the general category of algorithms that perform noise suppression by spectral modification (e.g., spectral subtraction, Wiener filtering—see review in Loizou, 2007). The enhanced envelopes are obtained by applying a weight (taking values in the range of 0 to 1) to the noisy envelopes of each channel. The weights are chosen to be inversely proportional to the estimated SNR of each channel. Envelope amplitudes in channels with high SNR are multiplied by a weight close to one (i.e., left unaltered), while envelope amplitudes in channels with low SNR are multiplied by a weight close to zero (i.e., heavily attenuated). The underlying assumption is that channels with low SNR are heavily masked by noise and therefore contribute little, if any, information about the speech signal. As such, these low-SNR channels are heavily attenuated (or annihilated) leaving only the high-SNR channels, which likely contribute more useful information to the listener.

[a]Author to whom correspondence should be addressed: Department of Electrical Engineering, University of Texas at Dallas, P.O. Box 830688, EC 33, Richardson, TX 75083-0688.

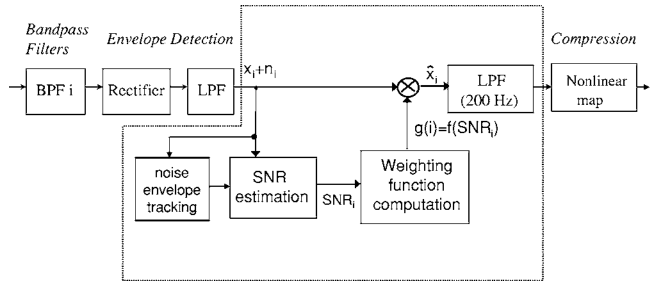

Fig. 1. Block diagram of the proposed noise reduction for the *i*th channel.

2. Experimental design

2.1 Subjects

A total of nine postlingually deafened Clarion CII implant users participated in this experiment. All subjects had at least 3 years of experience with their implant device. Most subjects visited our lab two times. During the first visit, all nine subjects were tested on music perception tasks and other psychophysical tasks unrelated to the current investigation. Due to the limited time available, the subjects were also tested on a single noise condition (5 dB SNR babble). Subjects were subsequently invited again to our lab, but due to various reasons (e.g., health, schedule conflicts), only five of the nine subjects were able to visit our lab to continue the testing for the other noise conditions.

2.2 Noise-suppression algorithm

Figure 1 shows the block diagram of the proposed noise reduction algorithm. The noisy speech signal is bandpass filtered into 16 channels and the envelopes are detected in each channel after full-wave rectification and low-pass filtering (200 Hz, sixth-order Butterworth). The noisy envelopes in each channel are multiplied by channel-specific weighting functions taking values in the range of zero to one depending on the estimated SNR of that channel. The envelopes attenuated by the channel weighting functions are smoothed with a low-pass filter (200 Hz) and log-compressed to the subject's electrical dynamic range. The low-pass filter is used to ensure that the enhanced envelopes are smoothed and are free of any abrupt amplitude changes that may be introduced by the application of the time-varying weighting function.

There are two major components in the proposed algorithm: SNR estimation and computation of the weighting function, which in turn depends on the estimated SNR. These components are discussed next.

2.2.1 Weighting function

We considered using a weighting function that applies heavy attenuation in channels with low SNR and little or no attenuation in channels with high SNR. With that in mind, we chose to use the following sigmoidal-shaped function:

$$g(i,l) = e^{-\beta/\xi(i,l)}, \tag{1}$$

where $\beta=2$, $g(i,l)$ is the weighting function $(0<g(i,l)\leq 1)$, and $\xi(i,l)$ denotes the *estimated* instantaneous SNR in the *i*th channel at stimulation cycle *l*. This weighting function plateaus at one for SNR$>$20 dB and floors to 0 for SNR<-5 dB. The above function was chosen as it has a sigmoidal shape similar to the human listener's psychometric function of intelligibility versus SNR. Other functions with similar shape could alternatively be used. Following the weighting function computation in Eq. (1), the enhanced temporal envelope is obtained by

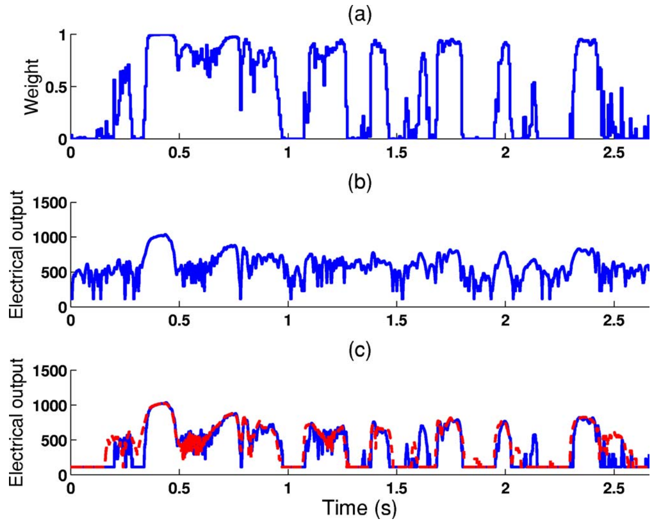

Fig. 2. (Color online) Example plots of attenuation (top panel) applied over time to the noisy envelope (middle panel) to obtain the enhanced envelope (bottom panel) for an IEEE sentence embedded in 5 dB SNR babble. The clean envelope is superimposed (dashed lines) in the bottom panel for comparison.

$$\hat{x}(i,l) = g(i,l) \cdot y(i,l), \tag{2}$$

where $y(i,l)$ is the noisy envelope of the ith channel at cycle l and $\hat{x}(i,l)$ contains the enhanced signal envelope. The enhanced envelope $\hat{x}(i,l)$ is smoothed by the same low-pass filter (200 Hz cutoff frequency) used in the envelope detection, and finally log compressed to produce the electrical amplitudes for stimulation (see Fig. 1). The low-pass filter is used to ensure that the enhanced envelopes are smoothed and are free of any abrupt amplitude changes that may be introduced by the application of the time-varying weighting function.

Figure 2 shows an example plot of the noisy speech envelope in comparison with the enhanced speech envelope for an IEEE sentence ("Every word and phrase he speaks is true.") in 5 dB SNR babble. Only the envelopes for channel 1 (center frequency=398 Hz) are plotted. The top panel (Fig. 2) shows the attenuation values applied to the noisy envelope. For the most part, the attenuation values (i.e., weights) were near one for envelope peaks and close to zero for envelope valleys. The envelope peaks were thus left unaltered while the valleys were attenuated. The resulting enhanced envelope (bottom panel in Fig. 2) had improved temporal envelope contrast. It should be noted that the processed stimuli were not intermittent despite the envelope dips seen in Fig. 2. When a weight of zero is applied to the noisy envelopes, the resulting envelope amplitude is set to the corresponding (electrical) threshold value and not to zero. The dips seen in the enhanced envelopes (Fig. 2) of channel 1 are due to the silence/closures already present in the original clean stimulus (also shown in bottom panel of Fig. 2) or to high-frequency consonants (e.g., /s/).

2.2.2 SNR estimation

The computation of the weighting function in Eq. (1) depends on the estimation of the instantaneous SNR $\xi(i,l)$ of channel i at cycle l, which is obtained using a variant approach reported by Ephraim and Malah (1984):

$$\xi(i,l) = \begin{cases} \alpha \dfrac{\hat{x}^2(i,l-1)}{\hat{n}^2(i,l-1)} + (1-\alpha)\max(\gamma(i,l)-1,0), & i \le 10 (f \le 1.8 \text{ kHz}), \\ \max(\gamma(i,l)-1,\varepsilon), & i > 10 (f > 1.8 \text{ kHz}), \end{cases} \tag{3}$$

where $\hat{x}(i,l-1)$ is the enhanced signal envelope obtained in the last stimulation cycle, ε is a small constant (10^{-6}) needed to avoid possible division by zero in Eq. (1), $\hat{n}(i,l)$ is the estimated envelope amplitude of the noise obtained using a noise-tracking algorithm (Cohen and Berdugo, 2002), α is a smoothing parameter $(0<\alpha<1)$, and $\gamma(i,l)=y^2(i,l)/\hat{n}^2(i,l)$. The reason for using different methods for estimating the SNR for high $(i>10)$ and low $(i\leq 10)$ frequency channels is to allow for faster tracking of sudden changes to the instantaneous SNR in the high frequency channels and relatively slower changes to the instantaneous SNR in the low frequency channels. Unlike Ephraim and Malah (1984) who used a high value for α $(\alpha=0.98)$, we found that a smaller smoothing constant was necessary for faster tracking of the instantaneous SNR, which is computed on a sample-by-sample basis rather than every 20-ms frame. In our study, the smoothing parameter α was set to $\alpha=0.4$ and $\alpha=0.6$ based on pilot experimental data. A noise-estimation algorithm (Cohen and Berdugo, 2002) is used to continuously track and update the noise envelope amplitude $\hat{n}(i,l)$ even during speech activity.

2.3 Procedure

The listening task involved sentence recognition in noise. IEEE sentences (IEEE, 1969) corrupted in multi-talker babble (ten female and ten male talkers) and speech-shaped noise were used in the test. Subjects were tested in four different SNR levels: 5 and 10 dB SNR in babble and 0 and 5 dB SNR in speech-shaped noise. Lower SNR levels were chosen for the speech-shaped noise conditions to avoid ceiling effects, as most subjects performed very well at the 10 dB SNR level. The IEEE sentences were recorded in our lab in a double-walled sound-attenuating booth and are available from Loizou (2007). The babble recording was taken from the AUDITEC CD (St. Louis, MO). Two sentence lists (ten sentences per list) were used for each condition. The sentences were processed offline in MATLAB by the proposed algorithm and presented to the subjects using the Clarion CII research platform at a comfortable level. For comparative purposes, subjects were also presented with unprocessed noisy sentences using the experimental processor. Sentences were presented to the listeners in blocks, with 20 sentences per block per condition. Different sets of sentences were used in each condition. Subjects were instructed to write down the words they heard. The presentation order of the processed and control (unprocessed sentences) conditions was randomized among subjects.

3. Results

The sentences were scored in terms of percent of words identified correctly (all words were scored). Figure 3 shows the individual scores for all subjects for the multi-talker babble (5–10 dB SNR) conditions and Fig. 4 shows the individual scores for a subset of the subjects tested in the speech-shaped noise (0–5 dB SNR) conditions. The scores for the subjects (S1, S5, S6, S8, S9) who visited our lab two times were averaged across the two visits (5 dB SNR, Fig. 3, top panel).

ANOVA (with repeated measures) showed a highly significant $(F[2,16]=14.4, p<0.0005)$ effect of the noise reduction algorithm on speech intelligibility for the 5-dB SNR babble condition. Scores obtained with the proposed noise reduction algorithm were significantly $(p<0.005)$ higher than the scores obtained with the subject's daily processor for both values of α. As shown in Fig. 3, all subjects benefited to some degree with the noise-reduction algorithm. Subjects S7 and S8 in particular received large benefits as their scores nearly doubled. With the exception of subjects S3, S5, S6, and S9, subjects performed equally well with $\alpha=0.4$ and $\alpha=0.6$ in Eq. (3). ANOVA (with repeated measures) performed on the 10-dB SNR babble data also showed a highly significant $(F[2,8]=24.3, p<0.0005)$ effect of the noise reduction algorithm on speech intelligibility. The mean improvement in performance with the proposed noise reduction algorithm was substantially larger than that obtained in the 5-dB SNR condition. Mean scores improved from 47% correct to 71% correct, with a small variability among subjects. The large benefit in intelligibility was consistent for all five subjects.

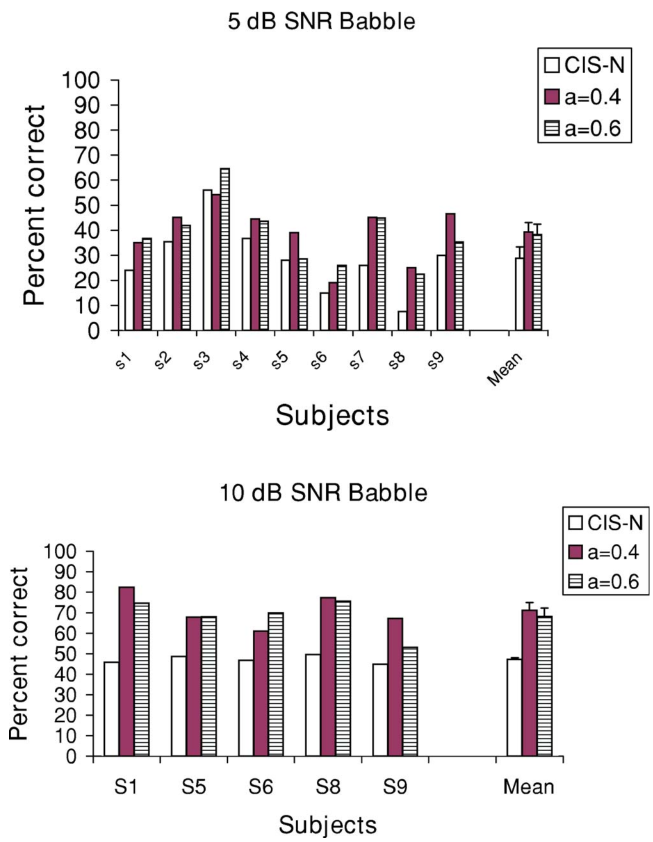

Fig. 3. (Color online) Subjects' performance on identification of words in sentences embedded in 5 and 10 dB SNR multi-talker babble and processed by the proposed algorithm with $\alpha=0.4$ and $\alpha=0.6$. Subjects' baseline performance (CIS-N) on unprocessed noisy stimuli is indicated by white bars. Error bars indicate standard errors of the mean.

ANOVA (with repeated measures) performed on the 0-dB SNR speech-shaped noise data showed a highly significant ($F[2,8]=9.5, p=0.008$) effect of the noise reduction algorithm on speech intelligibility. Particularly large improvements in performance were noted for subjects S1, S8, and S9. ANOVA (with repeated measures) performed on the 5-dB SNR speech-shaped noise data did not show a significant effect ($F[2,8]=0.9$, $p=0.445$). Performance obtained with the proposed noise reduction algorithm was as good as that obtained with the subject's daily processor.

4. Discussion

The above analysis clearly indicates that the proposed sigmoidal-shaped function provided significant benefits to CI users in nearly all conditions. We believe that much of the success of the proposed noise reduction algorithm can be attributed to the improved temporal envelope contrast. As shown in the example in Fig. 2, the sigmoid function preserves the envelope peaks and deepens the envelope valleys, thereby increasing the effective envelope dynamic range within each channel. Note, for instance, in Fig. 2, the change in the noisy envelope at $t=2.3$ s corresponding to the phoneme /s/ in the latter portion of the sentence. The depicted noisy envelope of channel 1 (center frequency=382 Hz) is near 50% of the electrical dynamic range when, in fact, it should have been near threshold level, since much of the energy of /s/ is concentrated in the high-frequency channels. After applying the sigmoidal-shaped function, the noisy envelope was

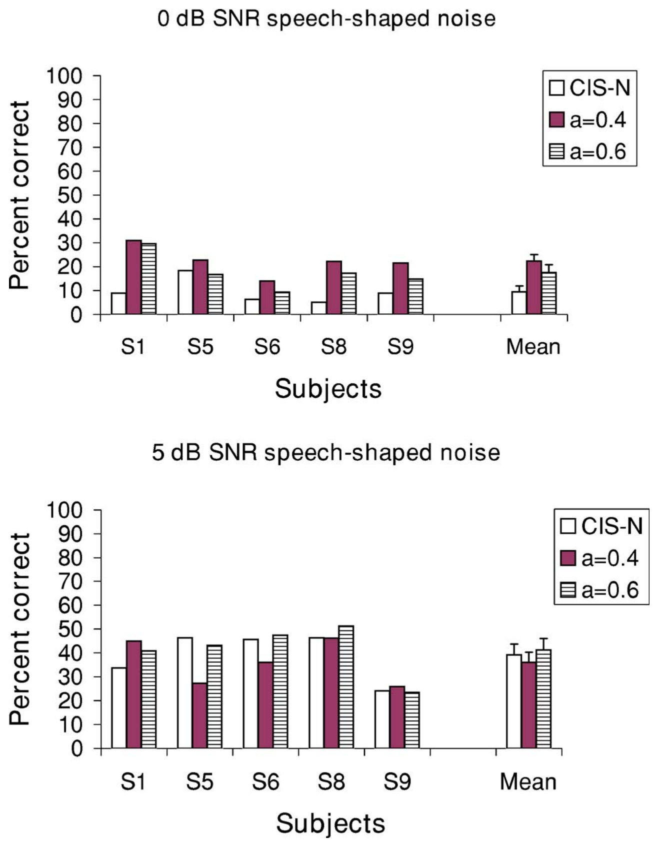

Fig. 4. (Color online) Subjects' performance on identification of words in sentences embedded in 0 and 5 dB SNR speech-shaped noise and processed by the proposed algorithm with $\alpha=0.4$ and $\alpha=0.6$. Subjects' baseline performance (CIS-N) on unprocessed noisy stimuli is indicated by white bars. Error bars indicate standard errors of the mean.

reduced to near threshold level. In brief, the proposed weighting function can make the envelope peaks and envelope valleys more discernible and perhaps more accessible to the CI users (compare the middle and bottom panels in Fig. 2).

The attenuation applied to the noisy envelopes is computed inversely proportional to the estimated SNR of each channel and is applied to both the speech and noise envelope signals. Clearly the attenuation itself cannot eliminate the noise as it is applied to both speech and noise signals. Yet, the applied attenuation, estimated according to the sigmoidal-shaped function, improves speech intelligibility for two main reasons. First, it is applied selectively and inversely proportional to the estimated SNR of each channel. More attenuation is applied to low-SNR channels where the speech signal is heavily masked by the noise, and is probably unintelligible. These channels provide unreliable and perhaps distracting information to the CI users and should therefore be attenuated. In contrast, no attenuation is applied to the channels in which the SNR is sufficiently high, thus enabling the listeners to ignore the masker. The high-SNR channels carry perhaps the most reliable information about the underlying signal and are therefore left unaltered. From Fig. 2 we observe that the SNR estimation does not need to be very accurate in terms of computing the exact weight to be applied to the noisy envelopes. It suffices if the SNR estimation algorithm performs sufficiently well in terms of discriminating high from low SNR envelopes, since the assigned weight will be either near the value of one (for SNR >20 dB) or near the value of zero (for SNR<-5 dB), respectively. Secondly, the temporal envelope contrast is improved (see example in Fig. 2). As a result, the envelope peaks and valleys are more discernible to the CI users who have a limited electrical dynamic range. This in turn makes the consonant/vowel boundaries clearer and more accessible to the CI users.

Overall, our algorithm compares favorably against other single-microphone methods proposed for cochlear implants (e.g., Yang and Fu, 2005; Loizou *et al.* 2005). A larger improvement (10–25 percentage points) in performance was obtained with our proposed method in multi-talker babble compared to the improvement (7 percentage points) obtained by the preprocessing method reported in Yang and Fu (2005). Other advantages of the proposed method include the lack of algorithmic delay associated with preprocessing techniques, low computational complexity, and ease of integration in existing CI strategies.

Acknowledgments

Research was supported by Grant No. R01 DC007527 from the National Institute of Deafness and other Communication Disorders, NIH.

References and links

Cohen, I., and Berdugo, B. (**2002**). "Noise estimation by minima controlled recursive averaging for robust speech enhancement," IEEE Signal Process. Lett. **9**, 12–15.
Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**(6), 1109–1121.
IEEE. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17(3)**, 225–246.
Loizou, P. (**2006**). "Speech processing in vocoder-centric cochlear implants," Adv. Oto-Rhino-Laryngol. **64**, 109–143.
Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL).
Loizou, P., Lobo, A., and Hu, Y. (**2005**). "Subspace algorithms for noise reduction in cochlear implants," J. Acoust. Soc. Am. **118**(5), 2791–2793.
Van Hoesel, R., and Clark, G. (**1995**). "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," J. Acoust. Soc. Am. **97**(4), 2498–2503.
Wouters, J., and Vanden Berghe, J. (**2001**). "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," Ear Hear. **22**(5), 420–430.
Yang, L., and Fu, Q. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**(3), 1001–1004.