Subspace and Envelope Subtraction Algorithms for Noise Reduction in Cochlear Implants

F. Toledo, P. Loizou and A. Lobo
Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75083, USA

Abstract—**The performance of two noise reduction algorithms is evaluated using 14 subjects fitted with the Clarion S-Series and Clarion II implant devices. The first algorithm, based on signal subspace principles, is used for pre-processing sentences embedded in +5 dB noise. The second algorithm is based on the subtraction of the noisy speech envelopes from an estimate of the noise envelopes. The noise envelopes are estimated continuously using a variation of the minimum statistics algorithm. Results showed that the subspace algorithm produced significant improvements in sentence recognition scores compared to the subjects' daily strategy. Small improvements were also obtained for a few subjects with the envelope subtraction algorithm.**

Keywords—Clarion device, cochlear implants, noise reduction •

I. Introduction

Several noise-reduction algorithms have been proposed for cochlear implant (CI) users [1]-[5]. Most of these algorithms, however, were based on the assumption that two or more microphones were available. Hoesel and Clark [1] tested an adaptive beamforming technique with four Nucleus-22 implantees using signals from two microphones – one behind each ear- to reduce noise coming from 90^o to the left of the patients. Results indicated that adaptive beamforming with two microphones can bring substantial benefits to CI users in conditions for which reverberation is moderate and only one source is predominantly interfering with speech. Hamacher *et al*. [2] evaluated the performance of two adaptive beamforming algorithms in different everyday-life noise conditions. The mean benefit obtained by the beamforming algorithms for four CI users (wearing the Nucleus device) varied between 6.1 dB improvement in SNR for meeting-room conditions to 1.1 dB for cafeteria noise conditions. Similar SNR improvement of about 10-dB was also reported recently by Wouters and Berghe [3] using a 2-channel adaptive filtering noise-reduction algorithm evaluated with four LAURA implantees.

In the above studies, it was assumed that two (or more) microphones were available, one behind each ear. Adding, however, a second microphone contralateral to the implant is ergonomically difficult without requiring the CI users to wear headphones or a neckloop [bilateral implants might provide the means, but their benefit is still being investigated]. Single-microphone noise reduction algorithms are therefore more desirable and cosmetically more appealing. While a few single-microphone noise-reduction strategies [4][5] have been proposed for cochlear implants, those strategies were implemented on old cochlear implant processors, which were based on feature extraction strategies (F0/F1/F2 and MPEAK strategies). Weiss [4] demonstrated that preprocessing the signal with a standard noise reduction algorithm could reduce the errors in formant extraction. The latest speech processors, however, are not based on feature extraction strategies but are based on vocoder-type strategies. In vocoder-type strategies, no features need to be extracted, as the signal is bandpass filtered into n bands ($8 \leq n \leq 22$), and the envelopes of the signal are extracted from each band. Hence, it is not clear whether preprocessing the signal could benefit vocoder-type strategies, such as the CIS and SPEAK strategies, commonly used today. This question is addressed in Experiment 1, where the noisy signal is preprocessed through a subspace noise reduction algorithm and presented to CI users.

Preprocessing noisy speech and presenting the "enhanced" speech to CI listeners might sometimes prove beneficial, but might not be the best approach. For one, pre-processing algorithms do not exploit or work synergistically with existing CI strategies. Secondly, we do not have much control on the effect of the pre-processing algorithms on the fine structure and/or envelope cues. In fact, in some cases those cues might be distorted. Ideally, we would like the noise reduction algorithm to be simple to implement and, most importantly, to be embedded in the existing coding strategies rather than being used as a pre-processor. To that end, we propose in Experiment 2 a signal processing algorithm which can be incorporated in current coding strategies.

II. Experiment 1: Evaluation of Subspace Algorithm

In this experiment, we investigate the potential benefits of first preprocessing the noisy signal with a noise reduction algorithm and then feeding the "enhanced" signal to the CI processor. For noise reduction, we use a custom subspace-based algorithm [6] designed to minimize speech distortion.

A. Subjects

A total of 14 Clarion implant users participated in this experiment. Nine Clarion CII patients and 5 Clarion S-Series patients were used as subjects. The majority of the CII patients were fitted with the CIS strategy, and the S-Series patients were fitted with the SAS strategy.

• This work was supported by NIDCD/NIH under Grant No. R01 DC03421. Corresponding author's email: loizou@utdallas.edu

B. Subspace algorithm

The signal subspace algorithm was originally developed by Ephraim and VanTrees [7] for white input noise and was later extended to handle colored noise (e.g., speech-shaped noise) by Hu and Loizou [6]. The underlying principle of the subspace algorithm is based on the projection of the noisy speech vector (consisting of, say, a segment of speech) onto two subspaces: the "signal" subspace and the "noise" subspace. The noise subspace contains only signal components due to the noise, and the signal subspace contains primarily the clean signal. Therefore, an estimate of the clean signal can be made by removing the components of the signal in the noise subspace and retaining only the components of the signal in the signal subspace.

Let $\mathbf{y}$ be the noisy vector, and let $\hat{\mathbf{x}} = \mathbf{Hy}$ be an estimate of the clean signal vector, where $\mathbf{H}$ is a transformation matrix. The noise reduction problem can be formulated as that of finding a transformation matrix **H,** which when applied to the noisy vector would yield the clean signal. After applying such a transformation to the noisy signal, we can express the error between the estimated signal $\hat{\mathbf{x}}$ and the true clean signal $\mathbf{x}$ as: $\boldsymbol{\varepsilon} = \hat{\mathbf{x}} - \mathbf{x} = (\mathbf{H} - \mathbf{I})\mathbf{x} + \mathbf{H}\,\mathbf{n}$, where $\mathbf{n}$ is the noise vector. Since the transformation matrix will not be perfect, it will introduce some speech distortion, which is quantified by first term of the error term, i.e. by $(\mathbf{H}-\mathbf{I})\mathbf{x}$. The second term ($\mathbf{H}\,\mathbf{n}$) quantifies the amount of noise distortion introduced by the transformation matrix. As the speech and noise distortion (as defined above) are decoupled, one can find the optimal transformation matrix $\mathbf{H}$ that would minimize the speech distortion subject to the noise distortion falling below a preset threshold. The solution to this constrained minimization problem for colored noise is given by [6]:

$$\mathbf{H}_{opt} = \mathbf{V}^{-T}\boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \mu\mathbf{I})^{-1}\mathbf{V}^{T} \tag{1}$$

where μ is a parameter (typical values for μ=5−20), $\mathbf{V}$ is an eigenvector matrix and $\boldsymbol{\Lambda}$ is a diagonal eigenvalue matrix obtained from the noisy speech vector (more details can be found in [6]). The above equation has the following interesting interpretation. The matrix $\mathbf{V}^{T}$ acts like a data-dependent transform and projects the noisy speech vector into the noise and signal subspaces. The diagonal matrix $\boldsymbol{\Lambda}(\boldsymbol{\Lambda} + \mu\mathbf{I})^{-1}$ multiplies the components of the signal in the signal subspace by a gain while zeroing out the components of the signal in the noise subspace. Finally, the matrix $\mathbf{V}^{-T}$ transforms back the projected signal (i.e., it acts like an inverse transform).

The implementation of the above signal subspace algorithm can be summarized into two steps. Step 1. For each frame of noisy speech ($\mathbf{y}$), use the above transformation given in Eq. 1 to obtain an estimate of the clean signal vector $\hat{\mathbf{x}}$, i.e., $\hat{\mathbf{x}} = \mathbf{H}_{opt}\,\mathbf{y}$. Step 2: Use the estimated signal $\hat{\mathbf{x}}$ as input to the CI processor.

The above estimator was applied to 4-ms duration frames of the noisy signal, which overlapped each other by 50%. The enhanced speech vectors were Hamming windowed and combined using the overlap and add approach. No voice activity detection algorithm was used in our approach to update the noise covariance matrix needed to compute the matrix **V**. The noise covariance matrix was estimated using speech vectors from the initial silence frames of the sentences.

C. Procedure

For testing, we used HINT sentences [8] corrupted in +5 dB S/N speech-shaped noise. Six lists (60 sentences) were processed off-line in MATLAB by the subspace noise reduction algorithm. The sentences were presented directly to the subjects via the auxiliary input jack of their CI processor at a comfortable listening level. Subjects were fitted with their daily strategy. For comparative purposes, subjects were also presented with six different lists (60 sentences) of HINT sentences corrupted in +5 dB speech-shaped noise, i.e., unprocessed sentences. The presentation order of pre-processed and un-processed sentences was randomized between subjects.

D. Results

The percent correct scores for all subjects are given in Figure 1. The sentences were scored in terms of percentage of words identified correctly (all words were scored). The mean score obtained using sentences pre-processed by the subspace algorithm was 43.8 (SEM=6.2), and the mean score obtained using unprocessed sentences was 19 (SEM=6.6). The sentence scores obtained with the subspace algorithm were significantly higher [$F(1,13)=33.1$, $p<0.0005$] than the scores obtained with the un-processed sentences. As can be seen from Fig. 1, most subjects benefited from the noise reduction algorithm. Subject's SS4 score, for instance, improved from 0% correct to 40% correct. Similarly, subjects' SS1 and SS2 scores improved from nearly 0% to 50% correct.

The above results indicate that the subspace algorithm can provide significant benefits to CI users in sentence recognition in noise. It should be noted that the above signal subspace algorithm was formulated to minimize speech distortion. We therefore believe that this algorithm is more suitable for CIs than other conventional algorithms (e.g., spectral subtraction and Wiener filtering), which might introduce spectral distortion.

III. EXPERIMENT 2: EVALUATION OF ENVELOPE SUBTRACTION ALGORITHM

In this Experiment, we investigate the performance of a noise reduction algorithm, which can be incorporated in current CI signal processing strategies. Compared to the subspace algorithm presented in Experiment 1, the proposed envelope subtraction algorithm is much easier to implement in real-time.

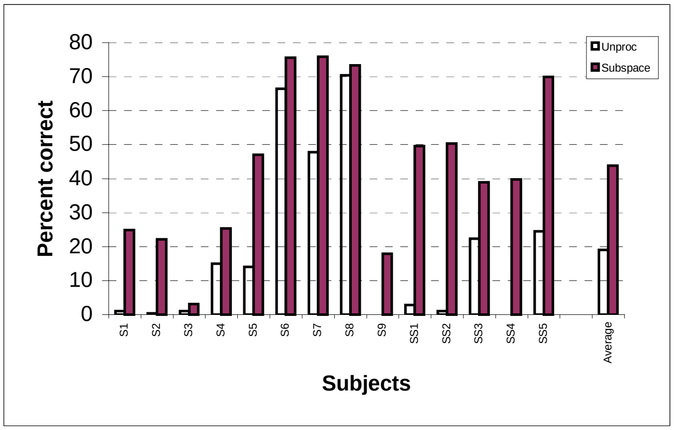

Figure 1. Subjects' performance on identification of words in sentences embedded in +5 dB S/N speech-shaped noise and preprocessed (dark bars) by the subspace algorithm or left un-processed (white bars). Subjects S1-S9 were Clarion II patients and subjects SS1-SS5 were Clarion S-Series patients.

A. Subjects

A total of four Clarion CII implant users participated in this experiment. The majority of the users were fitted with the CIS strategy.

B. Envelope subtraction algorithm

The noisy speech envelope (y_i) in the *i*th band can be approximately represented as the sum of the clean speech envelope (x_i) and the envelope due to noise (n_i), i.e., $y_i \approx x_i + n_i$. The approximation is due to the non-linearity of the full-wave rectification typically used in envelope detection. If we could somehow estimate the envelope of the noise signal (i.e., n_i), then the clean speech envelope could be simply estimated by: $x_i \approx y_i - n_i$.

The noise envelope (n_i) could conceivably be estimated (and updated) every time a speech pause is encountered. That would require, however, a reliable speech/noise detector. Although such a detector might perform well in stationary noise environment, it would perform terribly in a multi-talker babble listening situation (e.g., in a cafeteria environment). In a realistic listening situation the noise spectrum will most likely be changing constantly even during speech activity. Hence, an algorithm is needed for tracking the noise spectrum (or in our case, the noise envelope) continuously. Such an algorithm, based on minimum statistics [9] is used in this paper. This algorithm was modified to accommodate for the signal processing involved in CI strategies (note that the algorithm was originally developed and applied in the frequency domain).

The minimum statistics approach [9] is based on the observation that the power spectrum of the noisy speech signal, even during speech activity, frequently decays to the power spectrum level of the additive noise. It is therefore possible to derive a relatively accurate estimate of the noise spectrum (noise envelope, in our case) by tracking the minimum (within a finite window large enough to encompass high power speech segments) of the noisy speech signal spectrum.

The minimum tracking is done using the following algorithm [10]. Let S(*k*,*m*) denote the smoothed envelope amplitude of the *m*th channel estimated at frame *k* according to the following first-order recursive equation:

$$S(k,m) = \alpha\, S(k-1,m) + (1-\alpha) Y(k,m) \qquad (2)$$

where α (0<α<1) is a smoothing constant, and *Y(k,m)* is the noisy speech envelope amplitude of the *m*th channel. Perform pair wise comparisons between adjacent frames (present and previous) to obtain the minimum envelope amplitude value of the current frame:

$$S_{\min}(k,m) = \min(S_{\min}(k-1,m), S(k,m)) \qquad (3)$$

The local minimum is based on a window of at least L frames but no more than 2L frames. [Note that in the context of cochlear implants, a frame corresponds to one cycle of electrical stimulation, and is a function of the stimulation rate.] $S_{\min}(k,m)$ in the above equation contains the estimate of the envelope of the noise at frame *k*. Figure 2 shows an example of the noise envelope estimation for a sentence embedded in +5 dB multi-talker babble. After estimating the noise envelope in the *m*th band, we can estimate the clean envelope at frame *k* by:

$$X(k,m) = \begin{cases} Y(k,m) - \mu(k) S_{\min}(k,m) & \text{if } \; Y(k,m) > \mu(k) S_{\min}(k,m) \\ 0 & \text{if } \; Y(k,m) < \mu(k) S_{\min}(k,m) \end{cases} \qquad (4)$$

where $\mu(k)$ is an "oversubtraction" factor [11], which in our implementation varied between 1 and 2 depending on the estimate of the instantaneous *a posteriori* SNR.

The proposed envelope subtraction algorithm can be implemented in four steps: Step 1: Bandpass filter the noisy signal into *M* bands, and extract the envelopes of each band. Step 2: Smooth the noisy speech envelopes according to Eq. 2, and use Eq. 3 to update and estimate the envelope of the noise. Step 3: Estimate the clean envelope of the *m*th band using Eq. 4. Step 4: Map the estimated clean envelopes *X(k,m)* to electrical amplitudes using a log type compression.

C. Procedure

The above envelope subtraction algorithm was implemented offline in MATLAB using the following parameters: α=0.8 in Eq. 2, and L=150 corresponding to 52.1 ms. MATLAB routines were written which took as input the CI patients' MAP information (e.g., threshold levels, most-comfortable levels, pulse width) and generated patient specific amplitude files for each sentence processed. Custom software was used to "play back" the amplitude files to the implant patients using the Clarion Research Interface II platform.

For testing, we used HINT sentences [8] corrupted in +5 dB S/N multi-talker babble (taken from the AudiTec CD).

Three HINT lists (30 sentences) were processed through the envelope subtraction algorithm and another set of three lists (30 sentences) was processed through a standard implementation of the CIS algorithm. The sentences were presented directly to the subjects using the Clarion Research Interface II platform at a comfortable level.

D. Results

The individual subject's performance is shown in Figure 3. Overall, there was a substantial variability in performance between subjects, with some subjects showing an improvement in performance while others showing no improvement. Subject S6, for instance, showed a 25% improvement in performance with the envelope subtraction algorithm compared to the CIS algorithm. Subject S8, on the other hand, showed a small decrement in performance.

Overall, the proposed envelope subtraction algorithm is promising in that it may provide benefit to some subjects. Further work needs to be done, however, to find out why some subjects did not perform well with the envelope subtraction algorithm. We suspect that this might be due to inaccurate estimates of the noise envelope, which in turn, might have produced (envelope) distortion. More accurate noise envelope estimation algorithms might be required to minimize the possibility of any type of distortion. Further improvements to the noise envelope estimation are currently being investigated.

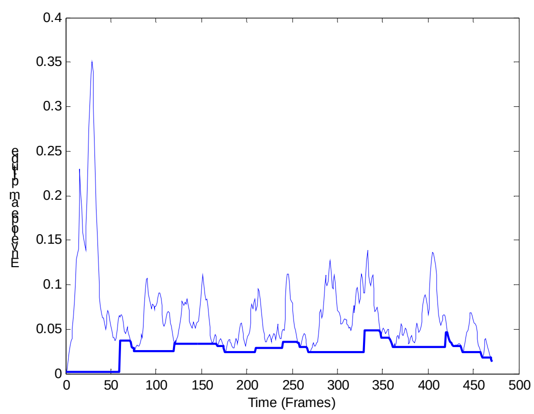

Figure 2. Example of noise envelope estimation for channel 1 (350-421 Hz). The thick line shows the estimate of the noise envelope for channel 1 and the thin line shows the smoothed noisy speech envelope estimated according to Eq. 2.

IV. Discussion and Conclusions

Two noise reduction algorithms (subspace and envelope subtraction) for cochlear implants were presented in this paper. Of the two algorithms, the subspace algorithm produced significant improvements in sentence recognition in noise for the 14 Clarion implant users tested. Small improvements in sentence recognition scores were also produced with the envelope subtraction algorithm at least for two out of the four subjects tested. The largest improvements in performance obtained by the subspace algorithm can be attributed to the fact that it was formulated to minimize speech distortion (a common artifact of conventional noise reduction algorithms). Envelope distortion might be the reason that the envelope subtraction algorithm did not perform as well as the subspace algorithm. Further work needs to be done on the envelope subtraction algorithm to obtain more accurate estimates of the noise envelope.

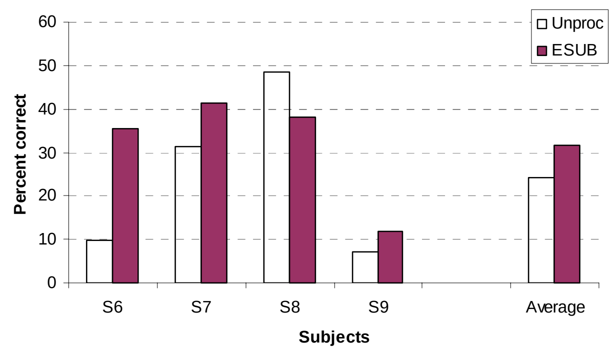

Figure 3. Subjects' performance on identification of words in sentences embedded in +5 dB S/N multi-talker babble and processed by the envelope-subtraction (ESUB) algorithm (dark bars) or left un-processed (white bars).

REFERENCES

[1] Hoesel, R. and Clark, G. "Evaluation of a portable two-microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Am.* , vol. 97, no. 4, pp. 2498-2503, 1995.

[2] Hamacher, V., Doering, W., Mauer, G., Fleischmann, H. and Hennecke, J. "Evaluation of noise reduction systems for cochlear implant users in different acoustic environments," *Am. J. Otol*., vol. 18, S46-S49, 1997.

[3] Wouters, J. and Berghe, J. "Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system," *Ear Hear*., vol. 22, no. 5, pp. 420-430, 2001.

[4] Weiss, M. "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.

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

[6] Hu, Y. and P. Loizou, "A subspace approach for enhancing speech corrupted with colored noise," *IEEE Signal Processing Letters*, vol. 9, no. 7, pp. 204-206, 2002.

[7] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions of Speech and Audio Processing*, vol. 3, pp. 251-266, July 1995.

[8] M. Nilsson, S. Soli, and J. Sullivan, " Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer*., vol 95, pp. 1085-1099, 1994.

[9] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. 7th EUSIPCO '94*, Edinburgh, U.K., Sept. 13-16, 1994, pp. 1182-1185.

[10] Cohen, I. and B. Berdugo. "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Letters*, vol. 9, no. 1, pp. 12-15, 2002.

[11] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. of the IEEE Conf ASSP*, pp. 208-211, April 1979.