Environment-specific noise suppression for improved speech intelligibility by cochlear implant users

Yi Hu[a)] and Philipos C. Loizou
Department of Electrical Engineering, University of Texas–Dallas, Richardson, Texas 75080

(Received 13 October 2009; revised 14 February 2010; accepted 18 February 2010)

Attempts to develop noise-suppression algorithms that can significantly improve speech intelligibility in noise by cochlear implant (CI) users have met with limited success. This is partly because algorithms were sought that would work equally well in all listening situations. Accomplishing this has been quite challenging given the variability in the temporal/spectral characteristics of real-world maskers. A different approach is taken in the present study focused on the development of environment-specific noise suppression algorithms. The proposed algorithm selects a subset of the envelope amplitudes for stimulation based on the signal-to-noise ratio (SNR) of each channel. Binary classifiers, trained using data collected from a particular noisy environment, are first used to classify the mixture envelopes of each channel as either target-dominated (SNR ≥ 0 dB) or masker-dominated (SNR < 0 dB). Only target-dominated channels are subsequently selected for stimulation. Results with CI listeners indicated substantial improvements (by nearly 44 percentage points at 5 dB SNR) in intelligibility with the proposed algorithm when tested with sentences embedded in three real-world maskers. The present study demonstrated that the environment-specific approach to noise reduction has the potential to restore speech intelligibility in noise to a level near to that attained in quiet.

© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3365256]

PACS number(s): 43.66.Ts, 43.71.Ky [JCM] Pages: 3689–3695

I. INTRODUCTION

Cochlear implant (CI) users face a number of challenging listening situations in their daily lives, some of which include listening to speech in various types of background noise (e.g., restaurant), while others include listening to speech corrupted by different degrees of reverberation (e.g., classrooms). The temporal/spectral characteristics of the various types of background noise vary widely and can be for instance modulated (e.g., train noise), can have narrowband spectra (e.g., siren noise, car noise) or relatively wideband (e.g., multi-talker babble) spectra. Given the inherent spectral variability of background noise present in realistic listening scenarios, the goal of effectively suppressing background noise in all listening conditions with a single suppression algorithm seems too ambitious. Yet, much research effort was devoted in the last two decades in developing such algorithms.

A number of noise reduction algorithms for unilateral CI users have been proposed (Hochberg *et al.*, 1992; Weiss, 1993; Yang and Fu, 2005; Loizou *et al.*, 2005; Kasturi and Loizou, 2007; Hu *et al.*, 2007). Yang and Fu (2005) tested subjects wearing the Clarion, Nucleus-22 and Med-EL devices using a spectral-subtractive noise reduction algorithm as a pre-processing step, and obtained significant improvement for recognition of speech embedded in speech-shaped noise. The improvement in multi-talker babble was modest and non-significant, and that was attributed partly to the fact that it was extremely challenging to track and estimate the masker spectrum needed in spectral-subtractive algorithms. Hu *et al.* (2007) evaluated an SNR-weighting based noise suppression algorithm that unlike other pre-processing algorithms (Hochberg *et al.*, 1992; Weiss, 1993; Yang and Fu, 2005), directly operated on the vocoded temporal envelopes. The noisy envelopes in each spectral channel were multiplied by channel-specific weighting factors that depended on the estimated SNR of that channel. A total of nine Clarion CII implant users were tested, and the results showed significant improvement in speech recognition in babble noise. The above noise suppression methods were promising with some yielding small, but significant, improvements in intelligibility. There still remains, however, a substantial performance gap between CI users' speech recognition in noisy listening conditions and in quiet.

A different approach is taken in this study to improve speech intelligibility in noise by CI users. Rather than focusing on the development of a single, universal, coding strategy that could be applied to all listening situations, we focus on the development of an environment-specific noise suppression algorithm. Such an approach can be implemented and utilized in commercially available implant speech processors (or hearing aids) in two different ways. One possibility is for the audiologist to program the speech processors with multiple MAPs, one for each listening situation that a CI user might encounter. The CI user can then switch to a different program each designed for different listening environments. A second possibility is to include a sound classification algorithm at the front-end of the CI processing, which will automatically identify the listening environment.

[a)] Author to whom correspondence should be addressed. Present address: Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201. Electronic mail: huy@uwm.edu

A number of such sound classification algorithms have already been developed for hearing aids applications (Nordqvist and Leijon, 2004). Following the identification of the listening environment, the appropriate noise reduction algorithm can be initiated automatically. Some hearing aids manufacturers (e.g., Phonak) and cochlear implant manufacturers (e.g., Cochlear Corporation's Nucleus 5 system) have recently adopted such an environment-specific approach for noise reduction, but no studies have yet been reported about the efficacy of the adopted algorithms.

The proposed noise suppression coding strategy builds upon our previous work on the study of optimum channel selection criterion (Hu and Loizou, 2008), that can potentially be used in lieu of the traditional maximum selection criterion. The ACE strategy adopted by Cochlear, Ltd, uses the maximum selection criterion whereby out of a total of up to 22 envelope amplitudes, only electrodes corresponding to the 8–12 largest amplitudes are selected for stimulation. This has been found to work well in quiet, however, in noise this criterion could be problematic: first, the selected amplitudes could include information from the masker-dominated channels; second, the maximum criterion may be influenced by the spectral distribution (e.g., spectral tilt) of the target and/or masker. The study by Hu and Loizou (2008) proposed the use of SNR as the selection criterion. Based on this criterion, target-dominated envelopes ($\mathrm{SNR} \geq 0$ dB) are retained, while masker-dominated envelopes ($\mathrm{SNR} < 0$ dB) are discarded. The results by Hu and Loizou (2008) demonstrated that the SNR channel selection criterion has the potential to *restore* the speech intelligibility in noise for CI listeners to the level attained in quiet, and for this reason, it is denoted as optimal ACE (opACE) in the present study.

Although the opACE strategy is a very promising strategy, its implementation poses a considerable challenge in real-world applications, as the SNR of each spectral channel needs to be estimated from the mixture envelopes, which is a formidable task. Conventional noise estimation algorithms have been found to perform poorly in terms of estimating the SNR (Hu and Loizou, 2008). This was not surprising, since most conventional noise estimation algorithms are not optimized for a particular listening situation, and thus do not take into account the differences in temporal/spectral characteristics of real-world maskers. By taking advantage of the distinctive temporal/spectral characteristics of different real-world maskers, which can be learned using machine learning techniques (Duda *et al.*, 2001), an algorithm can be designed to select channels based on the estimated SNRs of each channel. Such an algorithm can be optimized for a specific listening environment, and is thus expected to yield substantial improvements in intelligibility. The present study evaluates the performance by CI users of a noise suppression algorithm, which has been optimized for three different real-world environments, namely multi-talker babble, train and exhibition hall.

II. PROPOSED NOISE-REDUCTION ALGORITHM

The proposed algorithm consists of two steps: a training stage, which can be executed off-line, and an enhancement stage. The training stage uses the temporal envelopes of the clean speech signals (typically taken from a large corpus) along with the envelopes of the masker signals, to compute the true SNRs of each channel. Using the true channel SNRs, the binary status of the channels is determined as being either speech dominated (i.e., $\mathrm{SNR} \geq 0$ dB) or being masker-dominated ($\mathrm{SNR} < 0$ dB). In our present study, we found out that better classification performance can be obtained if an SNR threshold of -10 dB, rather than 0 dB, is used. Then, features extracted from the noisy mixture temporal envelopes and the corresponding binary classification of each channel are used to train a binary classifier for each channel. Gaussian mixture models (GMMs) were used in the present study as classifiers, as they were found to perform well in normal-hearing studies (Kim *et al.*, 2009). Features similar to amplitude modulation spectra (AMS; Kollmeier and Koch, 1994; Tchorz and Kollmeier, 2003) were used to train the binary classifiers. In the enhancement stage, a Bayesian classifier is used to classify each channel into two classes: target-dominated and masker-dominated channels. A channel is selected for stimulation only if it is classified as target-dominated. Figure 1 shows the block diagram of the enhancement stage of the proposed noise reduction algorithm. Note that the binary classifiers are designed and trained separately for each individual masker of interest. In doing so, we are able to achieve high classification accuracy. A different classifier is thus used for different listening environments.

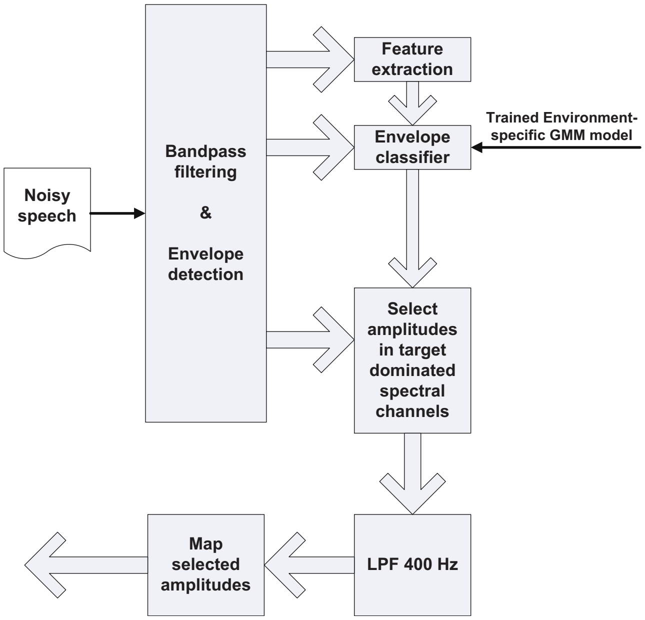

FIG. 1. (Color online) Block diagram of the proposed coding strategy.

A. Feature extraction

Figure 2 shows the block diagram of the feature extraction module. The noisy speech signal is first bandpass filtered into a number of bands corresponding to the active electrodes in the implant devices (e.g., if a CI user is using 14 electrodes, features are extracted in these 14 channels). The envelopes in each channel are computed via full-wave recti-

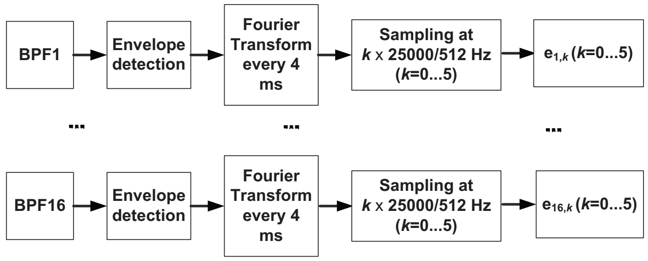
BPF1
Envelope detection
Fourier Transform every 4 ms
Sampling at k x 25000/512 Hz (k=0...5)
$e_{1,k}$ (k=0...5)
...
...
...
BPF16
Envelope detection
Fourier Transform every 4 ms
Sampling at k x 25000/512 Hz (k=0...5)
$e_{16,k}$ (k=0...5)

FIG. 2. Block diagram of the feature extraction module.

fication and lowpass filtering with a cutoff frequency of 400 Hz. A 512-point fast Fourier transform (FFT) of the past 20-ms envelope segment is then computed every 4 ms (i.e., with an 80% frame overlap). The magnitude spectrum is sampled at $k \times 25000/512$ Hz, $k=0\cdots 5$, to provide six modulation amplitudes for each channel, spanning the frequency range of 0 to 250 Hz. We denote these amplitudes as $\mathbf{e}(\tau,n)$, where τ indicates the segment index and n indicates the channel index. In addition to this AMS-like feature vector $\mathbf{e}(\tau,n)$, we also include delta features to capture variations across time and frequency (channel). The final feature vector is represented by:

$$\mathbf{E}(\tau,n)=[\mathbf{e}(\tau,n),\Delta\mathbf{e}_T(\tau,n),\Delta\mathbf{e}_N(\tau,n)], \qquad (1)$$

where

$$\Delta\mathbf{e}_T(1,n)=\mathbf{e}(2,n)-\mathbf{e}(1,n),$$

$$\Delta\mathbf{e}_T(\tau,n)=\mathbf{e}(\tau,n)-\mathbf{e}(\tau-1,n), \qquad \tau=2,\cdots,T,$$

$$\Delta\mathbf{e}_N(\tau,1)=\mathbf{e}(\tau,2)-\mathbf{e}(\tau,1),$$

$$\Delta\mathbf{e}_N(\tau,n)=\mathbf{e}(\tau,n)-\mathbf{e}(\tau,n-1), \qquad n=2,\cdots,N,$$

where $\Delta\mathbf{e}_T(\tau,n)$ and $\Delta\mathbf{e}_N(\tau,n)$ are the delta feature vectors computed across time and channel respectively, T is the total number of segments in an utterance, and N is the number of active electrodes in a CI user. The total dimension of the feature vector $\mathbf{E}(\tau,n)$ was 6×3 for each channel.

B. Training stage and enhancement stage

The probability distribution of the feature vectors in each class was represented with a GMM. As in Kim *et al.* (2009), the two classes (target-dominated envelopes λ_1 and masker-dominated envelopes λ_0) were further divided into four smaller classes λ_1^0, λ_1^1, λ_0^0, λ_0^1 to improve convergence speed and performance in GMM training. We used 256-mixture Gaussian models for modeling the feature vector distribution in each class. The *a priori* probability for each sub-class was calculated by dividing the number of feature vectors belonging to the corresponding class by the total number of feature vectors. The expectation-maximization algorithm was used to train the parameters (e.g., means, covariances, mixture weights) of the GMM binary classifier. A total of 32 IEEE lists (320 sentences) were used to train the GMMs. A different classifier was trained for each of the three maskers tested.

In the enhancement stage, using a Bayesian classifier (Duda *et al.*, 2001), the trained GMM models classify each channel envelope segment as either target or masker-dominated based on the feature vectors extracted from the mixture envelopes, and the binary status is deemed as stationary in each 4-ms envelope segment. More specifically, the envelope segments are classified as λ_1 or λ_0 by comparing two *a posteriori* probabilities, $P(\lambda_1|\mathbf{E}(\tau,n))$ and $P(\lambda_0|\mathbf{E}(\tau,n))$; $P(\lambda_1|\mathbf{E}(\tau,n))$ denotes the probability that the envelope segment belongs to class λ_1 when the feature vector $\mathbf{E}(\tau,n)$ is observed, and $P(\lambda_0|\mathbf{E}(\tau,n))$ denotes the probability that the envelope segment belongs to class λ_0 when the feature vector $\mathbf{E}(\tau,n)$ is observed. This comparison yields an estimate of the binary mask $g(\tau,n)$ as:

$$g(\tau,n)=\begin{cases}1, & P(\lambda_1|\mathbf{E}(\tau,n))\geq P(\lambda_0|\mathbf{E}(\tau,n))\\ 0, & P(\lambda_1|\mathbf{E}(\tau,n))< P(\lambda_0|\mathbf{E}(\tau,n))\end{cases},$$

where $P(\lambda_1|\mathbf{E}(\tau,n))$ is calculated using Bayes' rule (Duda *et al.*, 2001):

$$P(\lambda_1|\mathbf{E}(\tau,n))=\frac{P(\lambda_1,\mathbf{E}(\tau,n))}{P(\mathbf{E}(\tau,n))}$$
$$=\frac{P(\lambda_1^0)P(\mathbf{E}(\tau,n)|\lambda_1^0)+P(\lambda_1^1)P(\mathbf{E}(\tau,n)|\lambda_1^1)}{P(\mathbf{E}(\tau,n))},$$

and $P(\lambda_0|\mathbf{E}(\tau,n))$ is computed similarly. The channels classified as target-dominated are retained and stimulated. No stimulation is provided to the channels classified as masker-dominated.

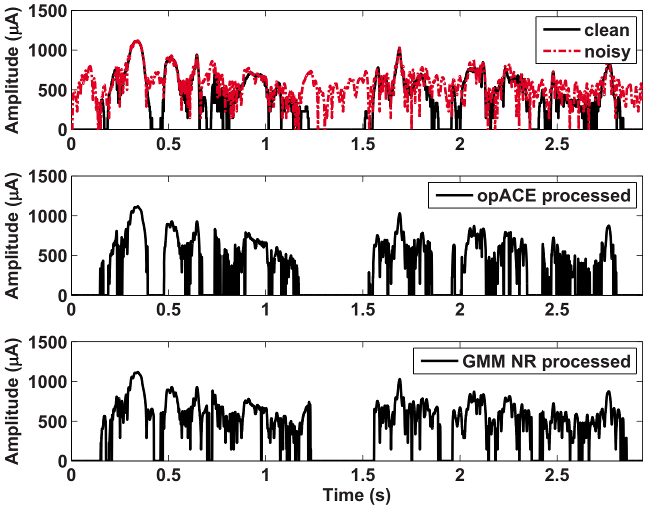

FIG. 3. (Color online) An example plot of the temporal envelopes of channel 1 (center frequency=383 Hz). The masker is 5 dB babble. The top panel shows the temporal envelopes of the clean and noisy speech. The middle panel shows the opACE-processed envelope, and the bottom panel shows the GMM-processed envelope.

Figure 3 shows example plots of the temporal envelopes of channel 1 (center frequency=383 Hz) obtained by the proposed GMM-based noise reduction algorithm and the opACE strategy in babble noise at a SNR at 5 dB SNR. As can be seen, the GMM envelopes are close to those obtained by opACE (Hu and Loizou, 2008). Figure 4 shows an example plot of the electrical stimulation pattern for a sentence processed by the GMM-based noise reduction algorithm. As can be seen, at many instances no electrode is selected for stimulation.

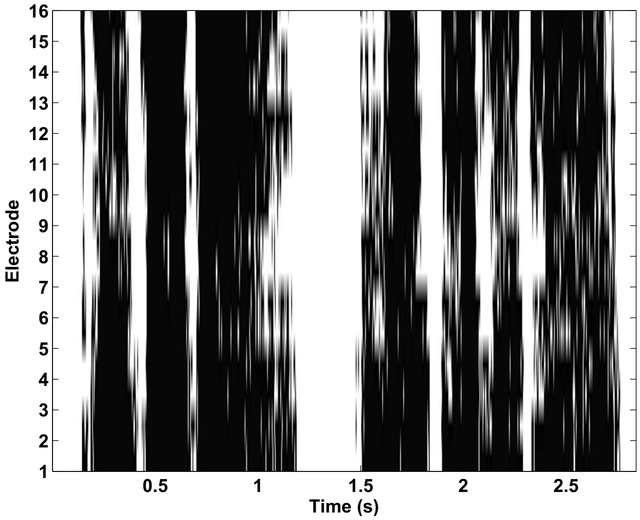

FIG. 4. An example plot of the electrical stimulation pattern obtained using the GMM-based noise reduction algorithm. Black pixels indicate active electrodes (stimulated) and white pixels indicate in-active electrodes (not stimulated).

III. LISTENING EXPERIMENTS

A. Subjects and stimuli

Seven post-lingually deafened Clarion CII implant users participated in this study. All subjects had at least five years of experience with their implant devices, and were paid an hourly wage for their participation. Table I shows the biographical data for all subjects.

The target speech materials consisted of sentences from the IEEE database (IEEE, 1969) and were obtained from Loizou (2007). The IEEE corpus contains 72 lists comprising of ten phonetically balanced sentences each. Sentences were produced by a male speaker and recorded in a double-walled sound-attenuation booth at a sampling rate of 25 kHz.

Three types of maskers were used: multi-talker babble (ten female talkers and ten male talkers), train noise and exhibition hall noise. The babble recording was taken from the AUDITEC CD (St. Louis, MO); the train noise recording and the exhibition hall noise recording were taken from the Aurora database (Hirsch and Pearce, 2000).

A total of 32 lists (320 sentences) were used to train the GMMs. These sentences were degraded by the three types of maskers at 0, 5, and 10 dB SNR. The remaining sentences in the IEEE database were used to test the CI subjects. The masker segments were randomly cut from the noise recordings and mixed with the target sentences. This was done to evaluate the robustness of the proposed GMM-based noise reduction algorithm in terms of testing sentences corrupted using different segments of the masker signal.

B. Procedure

The listening task involved sentence recognition in three types of real-world maskers. Subjects were tested at six different noise conditions: 5 and 10 dB SNR in babble noise, 5 and 10 dB SNR in train noise, and 5 and 10 dB SNR in exhibition hall noise. The SNR is defined as:

$$\mathrm{SNR} = 10 \times \log_{10} \frac{\sum_{k=1}^{k=K} s^2(k)}{\sum_{k=1}^{k=K} n^2(k)}$$

where s and n are speech and masker signals, respectively; and K is the number of samples in each speech sentence. Two sentence lists were used for each condition. The sentences were processed off-line in MATLAB (The MathWorks, Natick, Massachusetts) by the opACE algorithm and the proposed GMM-based noise reduction algorithm, and presented directly to the subjects using the Clarion CII research platform at a comfortable level. More specifically, after MATLAB processing, an encoded set of stimulus parameters for the electrical signals were stored in binary files with one binary file corresponding to one speech sentence; during testing, the binary files were downloaded to the Clarion CII research platform and presented to the subjects using customized software. The opACE condition was added as it can provide the upper bound in performance that can be achieved. For comparative purposes, subjects were also presented with unprocessed (corrupted) sentences using the experimental processor. More specifically, the corrupted sentences were processed via our own CIS implementation that utilized the same filters, same stimulation parameters (e.g., pulse width, stimulation rate, etc) and same compression functions used in the subjects' daily strategy. The opACE algorithm and the proposed GMM-based noise reduction algorithm also used the same filters, same stimulation parameters and same compression functions used in the subjects' daily strategy. In total, subjects participated in 18 conditions [= 2 SNR levels (5 and 10 dB) × 3 processing conditions (unprocessed noisy speech, opACE, and GMM-based noise reduction algorithm) × 3 types of maskers]. Subjects were also presented with sentences in quiet during the practice session. Sentences

TABLE I. Biographical data of the CI subjects tested.

Subject	Gender	Duration of deafness prior to implantation (yr)	CI use (yr)	Number of active electrodes	Etiology
S1	Male	1	5	15	Hydrops/Menier's syndrome
S2	Female	2	5	16	Unknown
S3	Female	2	5	15	Medication
S4	Female	1	>5	14	Unknown
S5	Female	<1	5	16	Medication
S6	Male	1	5	16	Fever
S7	Female	>10	6	16	Unknown

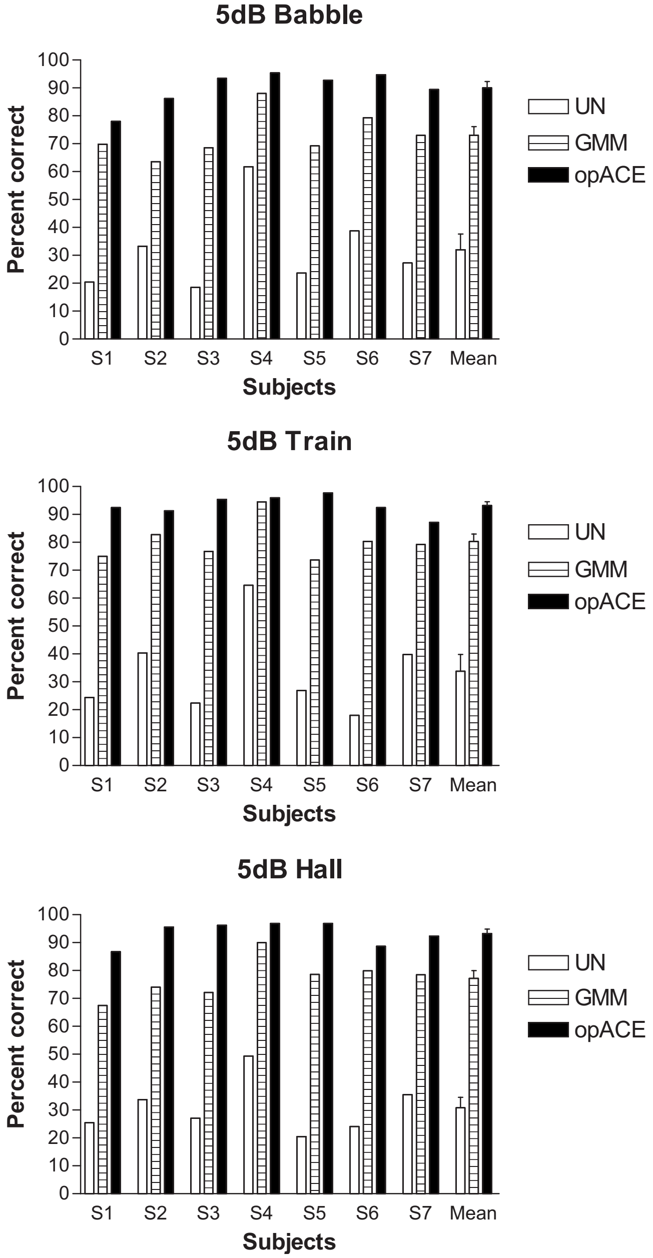

FIG. 5. Mean percent correct scores for babble noise, train noise and hall noise at 5 dB SNR. The error bars denote ±1 standard error of the mean. UN indicates the baseline condition with unprocessed (corrupted) sentences, GMM indicates the proposed GMM-based noise reduction algorithm, and opACE indicates the strategy proposed in Hu and Loizou (2008).

were presented to the listeners in blocks, with 20 sentences/block per condition.

Different sets of sentences were used in each condition. The presentation order of the processed and control (unprocessed sentences in noise) conditions was randomized for each subject. Subjects were allowed to take breaks at their leisure, and they were instructed to write down the words they heard. No feedback was given during testing.

C. Results

The mean percent correct scores for all conditions are shown in Figs. 5 and 6. Performance was measured in terms of percent of words identified correctly (all words were scored). To examine the effect of SNR level (5 and 10 dB) and processing conditions (UN, GMM, and opACE), we subjected the scores to statistical analysis using the percent correct score as the dependent variable, and the SNR levels and processing conditions as the two within-subjects factors. For babble noise, analysis of variance (ANOVA) with repeated measures indicated significant effects of both SNR level $[F(1,6)=14.16, p<0.0005]$ and processing condition $[F(2,12)=60.96, p<0.0005]$. There was significant interaction between SNR level and processing conditions $[F(2,12)=23.44, p<0.0005]$. For train noise, ANOVA with repeated measures indicated significant effects of both SNR

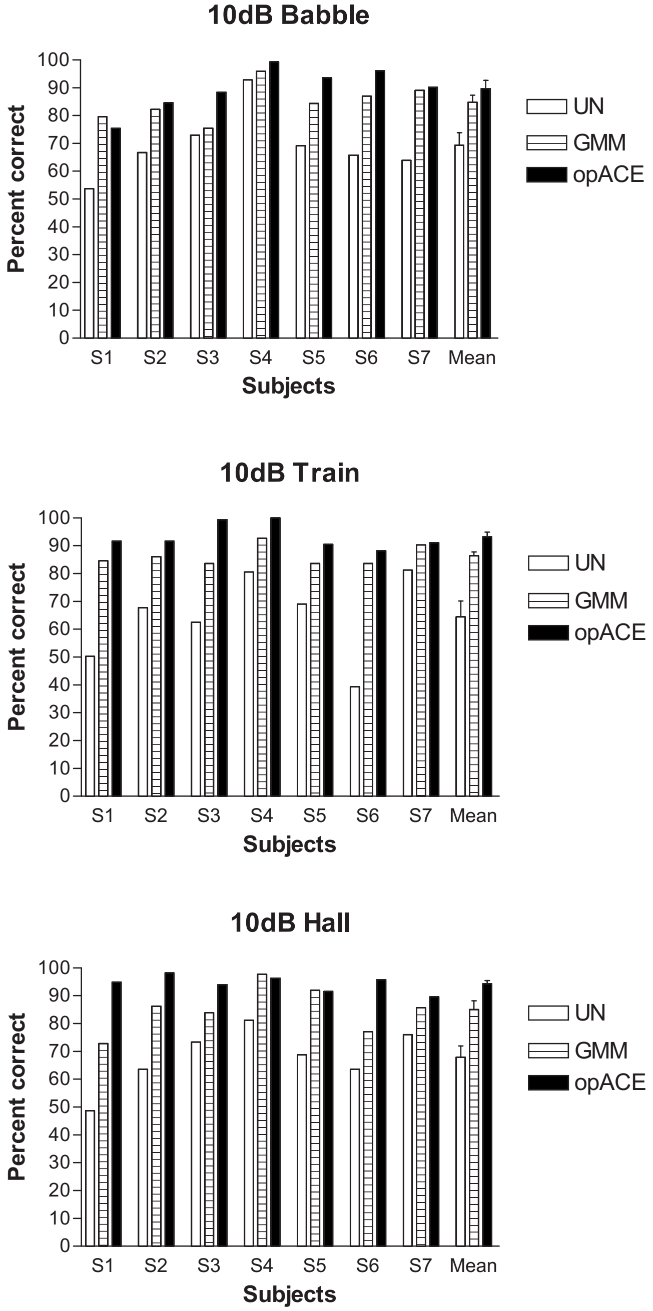

FIG. 6. Mean percent correct scores for babble noise, train noise and hall noise at 10 dB SNR. The error bars denote ±1 standard error of the mean. UN indicates the baseline condition with unprocessed (corrupted) sentences, GMM indicates the proposed GMM-based noise reduction algorithm, and opACE indicates the strategy proposed in Hu and Loizou (2008).

level $[F(1,6)=36.49,p=0.001]$ and processing conditions $(F(2,12)=115.88,p<0.0005)$. There was significant interaction between SNR levels and processing conditions $[F(2,12)=20.31,p<0.0005]$. For hall noise, ANOVA with repeated measures indicated significant effects of both SNR levels $[F(1,6)=9.44,p<0.0005]$ and processing conditions $[F(2,12)=42.78,p<0.0005]$. There was significant interaction between SNR level and processing conditions $[F(2,12)=30.49,p<0.0005]$. The above interactions were introduced by the fact that the improvement in intelligibility with the proposed algorithm was larger at 5 dB SNR than at 10 dB SNR, as shown by the *post hoc* tests below. The improvement at 10 dB SNR might have been limited by ceiling effects.

Post hoc tests (Schéffe, corrected for multiple comparisons) were run to assess the statistical significance between conditions. For all noise conditions, performance with the unprocessed sentences were significantly lower than both the GMM-based noise reduction algorithm and the opACE strategy. The performance of some subjects (e.g., S4) with sentences processed via the GMM-based algorithm at 5 dB SNR was above 90%, nearing their performance in quiet. For 5 and 10 dB train noise, 10 dB babble noise, and 10 dB hall noise, there were no significant differences between the GMM-based noise reduction algorithm and the opACE strategy; for the other noise conditions, the opACE strategy was significantly better than the GMM-based algorithm. The highest performance was obtained with the opACE strategy. This was not surprising since the opACE strategy assumes access to the true SNRs of each channel.

To quantify the accuracy of the binary Bayesian classifier, we calculated the average hit (HIT) and false alarm (FA) rates across all channels for the six noise conditions using 120 sentences processed via the GMM-based noise reduction algorithm. The HIT and FA rates were computed by comparing the estimated SNRs (using the GMM models) against the true SNRs of each channel. A false alarm error is introduced when masker-dominated envelopes (i.e., envelopes with $\mathrm{SNR}<0$ dB) are wrongly classified as target-dominated envelopes. Table II shows the HIT, FA and HIT-FA rates for the six noise conditions obtained by the proposed GMM-based noise reduction algorithm. Compared with the conventional noise reduction algorithms (Hu and Loizou, 2008), the GMM-based noise reduction algorithm produced much higher HIT rates and much lower FA rates, thus much higher HIT-FA rates. The difference metric HIT-FA is also reported because it bears resemblance to the sensitivity index, d', used in psychoacoustics [this metric was found by Kim *et al.* (2009) to correlate highly with intelligibility scores obtained with normal-hearing listeners]. As demonstrated in Li and Loizou (2008), low FA rates are required to achieve high levels of speech intelligibility, and this most likely explains the high performance of the proposed GMM-based noise reduction algorithm.

IV. GENERAL DISCUSSION AND CONCLUSIONS

Large improvements in intelligibility were observed with the proposed GMM-based noise reduction algorithm (Figs. 5 and 6). In 5 dB babble noise, for instance, mean subject scores improved from 32% to 73%; in 5 dB train noise, mean subject scores improved from 34% to 80%; and in 5 dB hall noise, mean subject scores improved from 31% to 77%. Performance approached that obtained with opACE, and the improvement in performance was consistent for all three types of maskers tested.

TABLE II. HIT and FA rates (expressed in percent) for the six noise conditions.

	Babble		Train		Hall	
	5 dB	10 dB	5 dB	10 dB	5 dB	10 dB
HIT	89.29	87.95	88.81	87.08	86.89	83.70
FA	14.19	13.83	13.18	12.18	13.03	12.46
HIT-FA	75.10	74.12	75.63	74.91	73.86	71.24

Two factors most likely contributed to the high performance of the proposed GMM-based noise reduction algorithm: first, the AMS-like features are neurophysiologically motivated (Kollmeier and Koch, 1994) and most likely capture reliably the difference between speech dominated and masker-dominated envelopes; second, GMM-based Bayesian classifiers are highly suitable for this binary mask application. Other classifiers, such as neural networks (Tchorz and Kollmeier, 2003), could alternatively be used. Our attempt, however, to use neural networks to estimate the binary masks, did not yield much improvement (Hu and Loizou, 2009), especially when randomly cut masker segments were mixed with the test sentences.

In the present study, a total of 320 sentences were used to train the GMM SNR classifiers for each masker. Alternatively, GMM classifiers can be trained incrementally. Starting with an initial GMM model (trained with a small number of sentences), the GMM parameters can be continuously updated (Huo and Lee, 1997) as more training sentences are added, and this model adaptation technique can be quite effective and more appropriate for real-world deployment of the proposed technique. Ongoing work in our laboratory is focused on further development of such model adaptation techniques.

A different GMM SNR classifier was used for each masker in the present study. Alternatively, a GMM classifier can be trained using data from multiple maskers. In other words, a GMM classifier can be built using more generalized noise models. Data from normal-hearing listeners indicated that a GMM classifier trained using data from 3 different maskers (babble, factory, speech-shaped noise) performed nearly as well as the GMM classifier trained using data from a single masker (Kim *et al.*, 2009). In realistic scenarios where the user knows *a priori* the types of background noise he or she will be encountering daily, the multiple-masker based GMM classifier could be a viable option.

There are several potential issues that warrant further investigation of the GMM-based noise reduction approach: first, the present study used speech materials produced from a male speaker, and it is not clear whether the gender of the speaker would have any impact on performance. Results from Kim *et al.* (2009), however, indicated that the speaker gender had no significant impact on performance; this is understandable as the extracted features do not carry much information about the identity of the speaker; second, the present study used an FFT-based feature extraction process and the envelope segment (20 ms duration) was not long enough to capture modulations below 20 Hz, which are important for speech intelligibility (Drullman *et al.*, 1994a, 1994b). A potential solution to this issue is to use a wavelet-based feature extraction procedure that is based on the use of different window lengths for different frequency components (Mallat, 2008); third, realistic deployment of the proposed GMM-based noise reduction method warrants further investigation. In the training stage of the proposed approach, substantial computational resources are needed to train the parameters of the GMM binary classifier, hence the training must be done in an off-line fashion; after the training, in the enhancement stage, the computational load is moderate and can be easily handled by modern cochlear implant devices; however, substantial memory space is required to store the parameters of the trained GMMs. From the above discussion, it can be seen that realistic deployment of the proposed GMM-based noise reduction algorithm needs inexpensive storage space and computational resources. As the memory cost is dropping rapidly, the requirement for storage space can be met in the near future; however from the perspective of end users, computational resources still present a formidable problem. Some solutions can be derived from the area of automatic speech recognition systems (e.g., call centers operated by computers). Commonly used GMMs, for instance, can be incorporated into the processors by the cochlear implant device manufacturers. For GMM training tasks initiated by end users, a viable solution is to use an internet-based cloud computing platform, which will become available in the near future.

In summary, an environment-optimized approach to noise reduction was proposed in the present study for cochlear implant users, and the proposed approach aligns well with existing methods used in hearing aids (e.g., Phonak's Savia), where sound classification methods are used to first identify different listening situations, and then adjust accordingly hearing-aid processing parameters (Zakis *et al.*, 2007). The data collected in the present study demonstrated that the proposed environment-optimized noise suppression algorithm has the potential to restore speech intelligibility in noise to a level near to that attained in quiet by CI listeners.

ACKNOWLEDGMENTS

This research was supported by NIH/NIDCD Grant Nos. R03-DC008887 (Y.H.) and R01-DC07527 (P.C.L.). The authors thank Dr. John Middlebrooks and the anonymous reviewer for their very helpful comments.

Drullman, R., Festen, J., and Plomp, R. (**1994a**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Drullman, R., Festen, J., and Plomp, R. (**1994b**). "Effect of temporal smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Duda, R., Hart, R., and Stork, D. (**2001**). *Pattern Classification* (Wiley, New York).

Hirsch, H., and Pearce, D. (**2000**). "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in Proceedings of the International Speech Communication Association Workshop ASR2000, Paris, France.

Hochberg, I., Boorthroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 263–271.

Hu, Y., and Loizou, P. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**, 498–509.

Hu, Y., and Loizou, P. (**2009**). "Environment-optimized noise suppression for cochlear implants," in Proceedings of the 33rd ARO MidWinter Meeting, The Association for Research in Otolaryngology, Baltimore, MD.

Hu, Y., Loizou, P., Li, N., and Kasturi, K. (**2007**). "Use of a sigmoidal-shaped function for noise attenuation in cochlear implant," J. Acoust. Soc. Am. **122**, EL128–EL134.

Huo, Q., and Lee, C.-H. (**1997**). "On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate," IEEE Trans. Speech Audio Process. **5**, 161–172.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

Kasturi, K., and Loizou, P. (**2007**). "Use of s-shaped input-output functions for noise suppression in cochlear implants," Ear Hear. **28**, 402–411.

Kim, G., Lu, Y., Hu, Y., and Loizou, P. (**2009**). "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust. Soc. Am. **126**, 1486–1494.

Kollmeier, B., and Koch, R. (**1994**). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," J. Acoust. Soc. Am. **95**, 1593–1602.

Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 2287–2294.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).

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

Mallat, S. (**2008**). *A Wavelet Tour of Signal Processing: The Sparse Way* (Academic, Burlington, MA).

Nordqvist, P., and Leijon, A. (**2004**). "An efficient robust sound classification algorithm for hearing aids," J. Acoust. Soc. Am. **115**, 3033–3041.

Tchorz, J., and Kollmeier, B. (**2003**). "SNR estimation based on amplitude modulation analysis with applications to noise suppression," IEEE Trans. Speech Audio Process. **11**, 184–192.

Weiss, M. (**1993**). "Effects of noise and noise reduction processing on the operation of the nucleus-22 cochlear implant processor," J. Rehabil. Res. Dev. **30**, 117–128.

Yang, L., and Fu, Q. (**2005**). "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise," J. Acoust. Soc. Am. **117**, 1001–1003.

Zakis, J. A., Dillon, H., and McDermott, H. (**2007**). "The design and evaluation of a hearing aid with trainable amplification parameters," Ear Hear. **28**, 812–830.