Real-Time Automatic Tuning of Noise Suppression Algorithms for Cochlear Implant Applications

Vanishree Gopalakrishna, Nasser Kehtarnavaz, *Fellow, IEEE*, Taher S. Mirzahasanloo, and Philipos C. Loizou*, *Senior Member, IEEE*

***Abstract*—The performance of cochlear implants deteriorates in noisy environments compared to quiet conditions. This paper presents an adaptive cochlear implant system, which is capable of classifying the background noise environment in real time for the purpose of adjusting or tuning its noise suppression algorithm to that environment. The tuning is done automatically with no user intervention. Five objective quality measures are used to show the superiority of this adaptive system compared to a conventional fixed noise-suppression system. Steps taken to achieve the real-time implementation of the entire system, incorporating both the cochlear implant speech processing and the background noise suppression, on a portable PDA research platform are presented along with the timing results.**

***Index Terms*—Automatic tuning of noise suppression, characterization of noisy environments, noise adaptive cochlear implants, real-time implementation of cochlear implant speech processing.**

I. Introduction

MORE than 118, 000 people around the world have received cochlear implants (CIs) [1]. Since the introduction of CIs in 1984, their performance in terms of speech intelligibility has considerably improved. However, their performance in noisy environments still remains a challenge. Speech understanding with cochlear implants is reportedly good in quiet environments but is shown to greatly degrade in noisy environments [2], [3]. Several speech enhancement algorithms, e.g., [4], [5], have been proposed in the literature to address the performance gap in noisy environments. However, no real-time strategy has been offered to automatically tune these algorithms in order to obtain improved performance across different kinds of background noise environments encountered in daily lives by CI patients.

In [6]–[10], a number of speech enhancement algorithms are discussed which provide improved performance for a number of noisy environments. In this paper, we have developed an automatic mechanism to tune or adjust the noise suppression component to different noisy environments in a computationally efficient (real-time) manner. The motivation here has been to improve performance of CIs by allowing them to automatically adapt to different noisy environments. The real-time requirement is the key aspect of our developed solution as any computationally intensive approach is not practically useable noting that the processors that are often used in CIs have limited computing and memory resources.

More specifically, a real-time CI system is developed in this study, which is capable of automatically classifying the acoustic environment with the intent of adopting noise suppression parameters that are optimized for the selected environment. The classification is done in such a way that the computation burden to the CI speech-processing pipeline is kept to a minimum. Depending on the output of the noise classification stage, the system automatically and on-the-fly, switches to those parameters which provide optimal performance for a specific noisy environment. For the speech-processing pipeline, our previously developed n-of-m strategy using the recursive wavelet decomposition method is utilized [11], [12]. It is worth mentioning that this method can be easily replaced by the classical n-of-m strategy using fast Fourier transform (FFT).

The rest of the paper is organized as follows. Section II describes the developed noise adaptive CI system. Section III covers a detailed explanation of the components, which are introduced in this paper, namely noise detector, noise feature extraction, noise classification, and noise suppression. Section IV includes a discussion on the real-time implementation of the complete CI system as shown in Fig. 1 and the steps taken to ensure its real-time operation on a PDA platform. Section V discusses the performance of the newly introduced components or blocks of the developed system. Finally, the conclusions are stated in Section VI.

II. Noise Adaptive Cochlear Implant System

The proposed CI system is capable of detecting a change in the background noise with no user intervention, and changes the noise suppression parameters to previously determined (during training) optimal parameters for that particular background noise. A block diagram of the proposed adaptive system is shown in Fig. 1. First, the input speech signal is windowed and decomposed into different frequency bands. Most commercial CIs use a filterbank or FFT to achieve this decomposition [13]. As discussed in [11], we showed the advantages of using the recursive wavelet packet transform (WPT) for the decomposition. Based on the previously developed noise-suppression algorithm in [8]–[10], noise is suppressed by appropriately applying

Manuscript received November 9, 2011; revised February 8, 2012 and March 2, 2012; accepted March 5, 2012. Date of publication April 3, 2012; date of current version May 18, 2012. This research was supported by the National Institute of Deafness and other Communication Disorders, National Institute of Health, under Grant R01 DC010494. *Asterisk indicates corresponding author*.

V. Gopalakrishna, N. Kehtarnavaz, and T. Mirzahasanloo are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: vani@utdallas.edu; kehtar@utdallas.edu; mirzahasanloo@utdallas.edu).

*P. C. Loizou is with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: loizou@utdallas.edu).

Digital Object Identifier 10.1109/TBME.2012.2191968

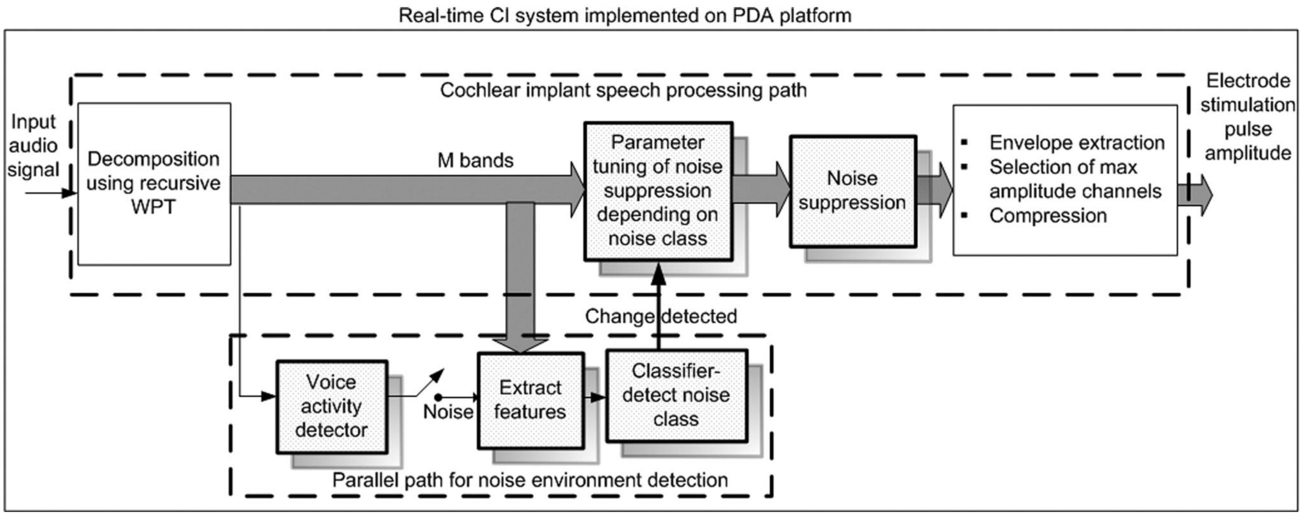


Fig. 1. Block diagram of the developed noise-adaptive cochlear implant system implemented on a PDA platform in real time; highlighted blocks indicate the new blocks that were introduced in this study.

a noise-suppressive gain function to the magnitude spectrum. From the suppressed magnitude spectrum, channel envelopes are extracted by combining the wavelet packet coefficients of the bands, which fall in the frequency range of a particular channel. Finally, the envelopes are compressed using a logarithmic compression map. Based on these compressed channel envelopes, the amplitude of stimulating pulses for CI implanted electrodes is determined.

In a parallel path to the aforementioned speech processing path, the first stage of the WPT coefficients of the windowed signal are used to detect whether a current speech segment is voiced/ unvoiced speech or noise via a noise detector. If the input windowed segment is found to be noise, signal features are extracted using the wavelet packet coefficients that are already computed from the speech processing path. The extracted feature vector is fed into a Gaussian mixture model (GMM) classifier to identify the background noise environment. When a change in the background noise is detected, the noise suppression parameters of the system switch to the optimized parameters of the detected environment.

According to the hearing aid study done in [14], hearing aid users spend about 25% of their time, on average, in quiet environments while the remaining 75% of their time is distributed among speech, speech in noise and noisy environments. The different background noise environments encountered in the daily lives of hearing-aid users depend on many demographic factors such as age, life style, living place, working place, etc. Hearing aid data logging studies have provided usage statistics in different environments. The study reported in [15] discusses commonly encountered environments in which hearing aid patients expressed that it is important for them to be able to hear clearly in those environments.

Using similar data logging studies for CIs, it would be possible to get usage statistics of CIs in different environments. However, in the absence of such studies for CIs, here we have chosen ten commonly encountered environments mentioned in [15] with the assumption that the most frequently visited environments of CI and hearing aid users are similar. The ten background noise classes considered in this study include car, office, apartment living room, street, playground, mall, restaurant, train, airplane, and place of worship. Our system is designed in such a way that additional noise classes can be easily incorporated into it. It should be pointed out that in response to a noise class, which is not present in the aforementioned noise classes, the system selects the class with the closest matching noise characteristics.

III. System Components

A. Voice Activity Detector

For extracting noise features, it is required to determine if a captured data frame contains speech plus noise or noise only. After deciding that it is a noise-only frame, noise signal features get extracted and a noise classifier gets activated. In order to determine the presence of noise-only frames, a voice activity detector (VAD) is used. There are a number of VADs that have been proposed in the literature. Some of the well-known ones include ITU recommended G.729b, signal-to-noise ratio (SNR)-based, zero-crossing-rates-based, statistical-based, and HOS-based VADs [16]–[19].

In this paper, we have considered a noise detector based on the WPT since this transform is already computed as part of our CI speech-processing pipeline in order to limit the computational burden on the overall system. This noise detector or VAD was proposed in [19], where the subband power difference is used to distinguish between speech and noise frames. Subband power is computed using the wavelet coefficients from the first level WPT coefficients of the input speech frame. Then, the subband power difference (SPD) between the lower frequency band and the higher frequency band is computed, as given in (1). Next, SPD is weighted as per the signal power, as shown in (2), and the result is compressed such that it remains in the same range for different speech segments as indicated in (3). A first-order low-pass filter is also used at the end to smooth out any fluctuations

$$\mathrm{SPD}\,(m) = \left| \sum_{n=1}^{N/2} \left(\psi_{1,m}^{0}\,(n)\right)^2 - \sum_{n=1}^{N/2} \left(\psi_{1,m}^{1}\,(n)\right)^2 \right| \tag{1}$$

$$Dw\,(m) = \mathrm{SPD}\,(m) \left[\frac{1}{2} + \frac{16}{\log(2)} \log\left(1 + 2\sum\nolimits_{n=1}^{N} y_m\,(n)^2\right) \right] \tag{2}$$

$$Dc\,(m) = \frac{1 - e^{-2Dw(m)}}{1 + e^{-2Dw(m)}} \tag{3}$$

where $y_m\,(n)$ is the input speech signal of the *m*th window with each window containing *N* samples, $\psi_{1,m}^{0}\,(n)$ and $\psi_{1,m}^{1}\,(n)$ are the wavelet coefficients corresponding to the lower and higher frequency bands, respectively, at the first level of the decomposition.

To differentiate between noise and speech, a threshold $T_v(m)$ is computed using an adaptive percentile filtering approach. Percentile filtering is applied to a sorted array of smoothed and compressed subband power difference D_c. The sorted array D_{cs} has *B* number of D_c values corresponding to past 1-s segments. The threshold is computed using the first value of D_{cs} as given in (4) which satisfies the condition shown in (5). Considering that statistics of sustained noise do not change as fast as speech, the threshold value is updated slowly using a single-pole low-pass filter as indicated in (6) with $\alpha_v = 0.975$. A speech or noise decision is made if the $D_c(m)$ value is greater than or less than the threshold value $T_v(m)$

$$\widetilde{T_v}\,(m) = D_{\mathrm{cs}}\,(b) \tag{4}$$

$$D_{cs}\,(b) - D_{\mathrm{cs}}\,(b-4) > 0.008 \quad \forall b = 4 \ldots B \tag{5}$$

$$T_v\,(m) = \alpha_v T_v\,(m-1) + (1-\alpha)\,\widetilde{T_v}\,(m). \tag{6}$$

Unvoiced segments are generally difficult to detect and they are often mistaken as noise-only frames. Unvoiced frames often occur before or after voiced frames. Hence, the frames which are detected as noise frames just after voiced frames are still treated as speech. In other words, a guard time of 200 ms after voiced segments is considered noting that most consonants do not last longer than 200 ms on average [20]. This reduces the likelihood of treating unvoiced frames as noise. It should be mentioned that this noise detector is not used to update the noise spectrum in the noise suppression component. Thus, this extra guard time does not harm the noise tracking speed and its bias over detecting speech. It is also important to note that this noise detector does not depend on any training and it can operate across various SNR levels. Fig. 2 shows the noise detector applied to a stimulus consisting of two IEEE sentences “The birch canoe slid on the smooth planks” and “Glue the sheet to the dark blue background”, recorded at 8 kHz and produced by a male speaker. There is a 1-s pause between the two sentences. The bottom two plots in Fig. 2 show the noise detector output with the guard time for the same signal without noise (i.e., in quiet) and when corrupted by car noise at 5-dB SNR.

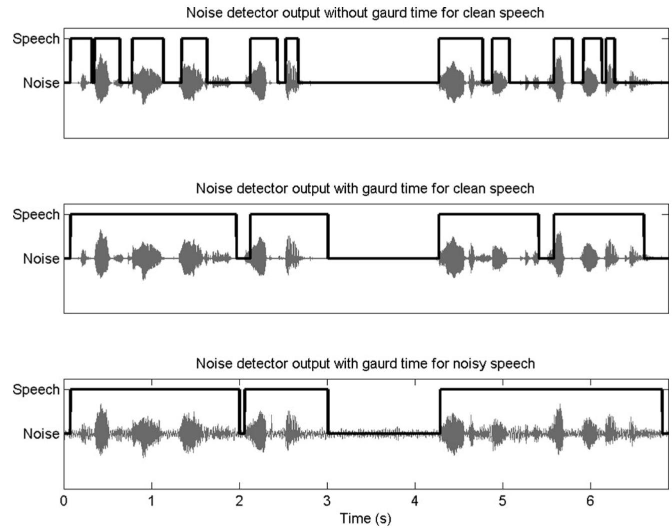


Fig. 2. (Top to bottom) Noise detector output of clean speech signal without guard time correction, noise detector output of clean speech signal with guard time correction, and noise detector output of corrupted speech signal by car noise at 5-dB SNR with guard time correction.

B. Noise Features

Various features have been utilized in the literature for noise characterization. For example, time domain features including zero-crossing rate, short-time energy, energy entropy, envelope modulation spectra in auditory critical bands have been used [22], as well as spectral domain features such as spectral roll off, spectral centroid, spectral flux, and harmonicity measure [23]. Noise features derived from LPC and wavelet transforms are also widely used [24]–[26]. In our previous work [27], we introduced Markov random field-based features operating on spectrograms [28]. For the developed system, we examined various combinations of the aforementioned time domain, spectral domain, mel-frequency cepstral coefficients (MFCC), and Markov random field-based features. Among various feature combinations examined, it was found that the MFCC + ΔMFCC features (26-dimensional feature vector) provided the best compromise between a high classification rate and a low computational complexity allowing the real-time implementation of the entire system. Other combinations either did not provide as high classification rates or were computationally intensive and did not allow a real-time throughput to be obtained.

To compute the MFCC coefficients, an overlapping triangular filter is applied to the magnitude spectrum of the WPT in order to obtain a mel-scale spectral representation. Here, 40 triangular filters are used, i.e., the 64-frequency bands magnitude spectrum is mapped to 40 bins in mel scale. The first 13 filters are spaced linearly and the remaining 27 filters are placed such that the bandwidth increases logarithmically. A discrete cosine transform is then applied to the logarithm of the magnitude spectrum in mel scale, thus generating 13 MFCCs in total. The first derivatives of MFCCs (ΔMFCC) are also computed as described in the following:

$$\Delta\mathrm{MFCC}\,(m,p) = \mathrm{MFCC}\,(m,p) - \mathrm{MFCC}\,(m-1,p) \tag{7}$$

where $\mathrm{MFCC}(m, p)$ represents the *p*th MFCC coefficient of the *m*th window.

C. Environmental Noise Classifier

Different classifiers have been used to classify speech, noise, and music, or different sound classes. The main classifiers studied consist of neural network (NN), K-nearest neighbor (KNN), support vector machine (SVM), GMM, and hidden Markov model [22]–[26], [29]. In our previous work [30], we used an SVM classifier with radial basis kernel and showed that this classifier provided high classification rates among a number of different classifiers for a two-class noise classification problem. However, the implementation of an SVM classifier is computationally expensive for the multiclass noise classification problem of interest here due to the large number of projections of features. We examined NN, KNN, Bayesian, SVM, and GMM classifiers and found that the GMM classifier with two clusters per class yielded the right balance between computational complexity for real-time implementation and classification performance.

The GMMs were trained as follows. The mean, covariance, and the prior probability of the GMM clusters are first determined for each noise class. For each noise class, k-means clustering is used to determine initial values of the aforementioned cluster parameters. These values are then fed into the expectation maximization (EM) algorithm to reach the optimum parameters. In each EM step, an expectation or the probability of training data generated from the current set of parameters is computed. The parameters are then updated for next iteration such that the expectation is increased. The training process is stopped when the log likelihood computed on training data does not increase significantly from the previous iteration. A fivefold cross validation is used to ensure that the trained model is not dependent on any specific training data set. It is worth pointing out that training is carried out offline and is not an issue for the real-time operation of the system.

D. Noise Suppression

As stated earlier, several environment-specific noise-suppression algorithms have appeared in the literature. Most of these algorithms are computationally intensive and do not meet our real-time requirement. For our system, we have deployed a combination of the noise suppression algorithms appearing in [8]–[10], which model the noise statistics using a data-driven approach. The primary idea is to apply a lower weight to those frequency bins which are masker dominated compared to target dominated such that target dominated bands get selected for the stimulation of electrodes. The challenge here is to accurately track noise so that noise power is not overestimated or underestimated. Overestimation leads to excessive removal of speech in the enhanced signal leaving the speech distorted and unintelligible, and underestimation leads to greater amount of residual noise. There are several methods for tracking the noise spectrum. In general, these methods attempt to update the noise spectrum using the corrupted speech spectrum with a greater amount of confidence when the probability of speech presence goes low. In what follows, we briefly describe our deployment of the data-driven approach for noise tracking, which was proposed in [9] and [10]. It should be noted that other tunable noise-suppression algorithms can be used in our system provided that they can be made to run in real time.

Let us consider an additive noise scenario, (8) with clean, noise and noisy received signals represented by $x_m(n), d_m(n)$ and $y_m(n)$, respectively, where *m* denotes the window number. The equivalent short-time DFT is given in (9), where *k* represents the frequency bin of FFT. *A priori* and *a posteriori* SNRs for the speech spectral estimation are given as follows:

$$y_m(n) = x_m(n) + d_m(n) \tag{8}$$

$$Y_m(k) = X_m(k) + D_m(k) \tag{9}$$

$$\xi_m(k) = \frac{\lambda_x(k)}{\lambda_d(k)}, \qquad \gamma_m(k) = \frac{Y_m^2(k)}{\lambda_d(k)} \tag{10}$$

where $\xi_m(k)$ denotes the *a priori* SNR, $\gamma_m(k)$ denotes *a posteriori* SNR at the frequency bin k, λ_d denotes the noise variance and λ_x denotes the clean speech variance. *A priori* SNR and *a posteriori* SNRs are obtained by using the "decision-directed" approach as

$$\widehat{\xi_m}(k) = \alpha_{dd}\frac{\widehat{X_{m-1}^2}}{\widehat{\lambda_d}(k)} + (1-\alpha_{dd})\max\left(\frac{\widehat{Y_m^2}(k)}{\widehat{\lambda_d}(k)} - 1, \xi_{\min}\right) \tag{11}$$

$$\widehat{Y_m}(k) = \frac{\widehat{Y_m^2}(k)}{\widehat{\lambda_d}(k)} \tag{12}$$

where α_{dd} is a smoothing parameter [9], [10], and $\xi_{\min}$ is a small number greater than 0. According to [10], the use of the nonideal *a priori* SNR estimate, which is derived using the speech spectral estimation of the previous window leads to erroneous spectral estimates. This error gets fed back into the system. To minimize this error, a modified *a priori* SNR estimate, $\widehat{\xi_{NT\,m}}$, based on the previous noisy speech spectra (rather than enhanced spectra) is considered as shown in the folowing:

$$\widehat{\xi_{NT\,m}}(k) = \alpha_{NT}\frac{Y_{m-1}^2(k)}{\widehat{\lambda_d}(k)} + (1-\alpha_{NT})\max\left(\frac{Y_m^2(k)}{\widehat{\lambda_d}(k)} - 1, \xi_{\min}\right). \tag{13}$$

where α_{NT} is a smoothing parameter set to 0.98.

The noise variance and speech spectra are then obtained according to the weighted spectra specified in (14) and (15), where the weight (gain) is a function of *a priori* and *a posteriori* SNR estimates

$$\widehat{\lambda_d}(k) = G_D\left(\widehat{\xi_{NT\,m}}(k), \widehat{\gamma_m}(k)\right) Y_m^2(k) \tag{14}$$

$$\widehat{X_m^2}(k) = G_X\left(\widehat{\xi_m}(k), \widehat{\gamma_m}(k)\right) Y_m^2(k) \tag{15}$$

where G_D is derived using the data-driven approach with the gain function determined using the minimum mean square error (MMSE) criteria, and G_x is derived using the log-MMSE

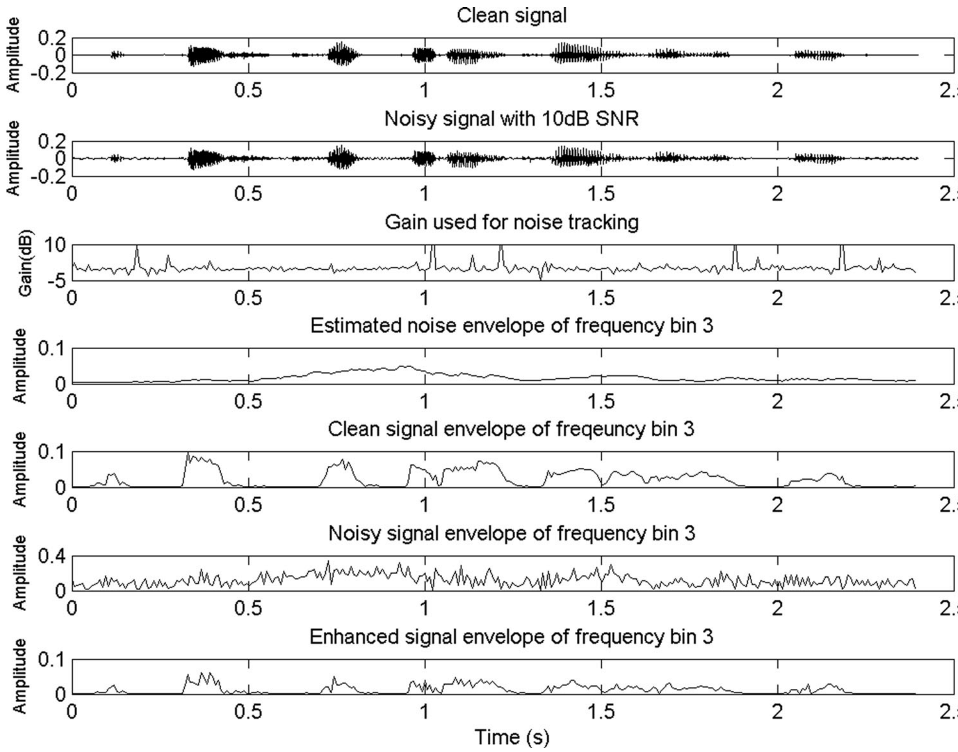


Fig. 3. (Top to bottom) Plots showing clean speech signal, noisy speech signal corrupted by car noise at 10-dB SNR, gain used for noise tracking, estimated noise envelope, clean signal envelope, noisy signal envelope, and enhanced signal envelope of frequency band 3.

estimator [31] as indicated

$$G_X\left(\widehat{\xi_m}(k), \widehat{\gamma_m}(k)\right) = \frac{\widehat{\xi_m}(k)}{\widehat{\xi_m}(k)+1} \exp\left\{\int_{vx}^{\infty} \frac{e^{-t}}{t} dt\right\}$$

$$vx = \frac{\widehat{\xi_m}(k) \cdot \widehat{\gamma_m}(k)}{\widehat{\xi_m}(k)+1}. \tag{16}$$

A gain table is derived during training for each noise class for *a priori* SNR values ranging from -20 to 40 dB and for *a posteriori* SNR values ranging from -30 to 40 dB in 1 dB steps, as proposed in [9] and [10]. The training procedure and all the parameters used match the ones reported in [9] and [10]. In other words, the G_D lookup table that is used for tuning becomes of size 61×71 for each noise class. To illustrate the working of the noise-tracking algorithm, Fig. 3 shows the clean speech, the noisy speech corrupted by car noise, the selected gain function G_D for frequency band 3, and the enhanced speech.

IV. Real-Time Implementation

The system was implemented on a PC and a PDA platform. The PDA platform had limited computational and memory resources as compared to the PC platform and has been previously used as a research platform for cochlear implants [11]. The PDA platform has been recently approved by FDA for clinical trials. The input speech, sampled at 22 050 Hz, using the PDA platform is windowed into 11.6-ms windows (256-sample windows). The analysis rate can be set more than that of the required stimulation rate by adjusting the overlap between windows; thus, the overlap between windows for computing the recursive WPT can be decided depending on the required stimulation rate. The detail and analysis coefficients from the first stage of WPT are used to compute the subband power difference measure for the VAD. The MFCC features are computed for every alternate noise-only window using the WPT coefficients at the sixth stage, which are already computed during the signal decomposition. This was done to ensure real-time implementation on the PDA platform. The MFCC feature vector, after normalization, was used as the input feature vector to the trained GMM classifier.

The decision made by the GMM classifier for 20 consecutive noise frames is used to generate a class decision. Median filtering of the decisions made by the classifier is considered due to the nonperfect behavior of the noise detector as some of the voiced sections might be labeled as noise. The number of windows for median filtering was chosen to be 20 because any further increase in the number of windows did not show much improvement in the classification performance. Reacting to transient noise by frequently switching from one noise class to another produces unpleasant distortions. Hence, a median filter with a duration of 2 s was used to eliminate such frequent switching. As a result, a switch is only made when the noise environment is sustained for more than 2 s. Clearly, this duration depends on user comfort and can be easily changed in the system for any lesser or longer duration.

The system implementation was done in C and an interactive GUI was added using LabVIEW. The PC platform used for

TABLE I
CLASSIFICATION RATES OF THE NOISE ADAPTIVE CI SYSTEM AVERAGED OVER 10 NOISE CLASSES AT DIFFERENT SNRS

SNR (dB)	Classification rate (%)
0	97.1
5	96.8
10	96.2
15	96

implementation had a processor clock rate of 3.33 GHz with 4-GB RAM, and the PDA platform had a processor clock rate of 624 MHz with 512-MB RAM.

Due to the limited computing and memory resources of the PDA platform, several code optimizations had to be done in order to achieve a real-time throughput. The rate at which the classifier was activated was reduced to every other noise frame instead of every noise frame. Since the PDA processor was a fixed-point processor, the implementation was done using fixed-point integer arithmetic. Parts of the code, where the accuracy was crucial and a large dynamic range was required, were implemented using 32-bit word length, while the other parts were implemented using 16-bit word length to save processing time. In addition, the exponential integral [used in ([16])] was implemented as a lookup table, and the lookup table was designed in such a way that the size of the table was minimized at the expense of negligible loss in accuracy. Different sections of the table were created with different resolutions to save memory and were arranged in a tree structure to speed up the lookup table search.

V. PERFORMANCE EVALUATION AND DISCUSSION

In this section, we provide the real-time timing as well as the performance results of the noise adaptive CI system described in the previous sections. The performance of both the classifier and the noise suppression blocks are reported. To assess classification accuracy, 100 audio signals were formed using sentences provided in [21] with each sentence of approximately 3-s duration. All the speech sentences were concatenated to form speech segments of 30-s duration with a 1-s pause between them. A pause was deliberately added between sentences so that the noise classification decision was made based on the noise present during speech pauses. These concatenated sentences were used to serve as the speech material. Ten noise classes with 5-min recording for each class were considered as the noise database. Both noise and speech were sampled at 8 kHz. 50% of the data were randomly selected and used for training and the remaining 50% for testing. The noise added to the speech sentences was randomly changed every 3 s. A deliberate frequent change in the background noise was only done to determine the performance of the classifier. Table I shows the correct classification rates averaged across all the classes at various SNRs. Table II shows the classification confusion matrix at SNR = 0 dB for ten classes of noise.

To study the performance of the adaptive-noise suppression approach, we compared it against two other scenarios: one without any noise suppression and the other with a fixed (non-environment specific) noise-suppression algorithm. A total of 30-s long concatenated speech sentences were added to each noise at a particular SNR. For the fixed-noise suppression, the minimum search algorithm was used to track the noise variance in place of using the lookup table that was generated via the data-driven approach. The speech quality measures of perceptual evaluation of speech quality (PESQ) and log-likelihood ratio (LLR) were considered to examine the quality of the noise suppressed output signals. In addition, the three composite measures of signal distortion (C_{sig}), background intrusiveness (C_{bak}), and overall quality (C_{ovl}), which have been shown to correlate highly with subjective speech quality [32] were computed. These composite measures [32] have been shown to be reasonably close to the subjective quality ratings made by normal hearing listeners. These measures were computed using the clean speech signal and the enhanced reconstructed signal. The comparative results are shown in Fig. 4. This figure shows the data for the 5-dB SNR condition with the standard deviation displayed as an error bar. A one-way analysis of variance (ANOVA) was conducted which showed a statistically significant ($F(2,117) > 9.8, p < 0.001$) of processing on the measures examined. Post-hoc tests were run, according to Tukey's HSD (Honesty Significant Difference) test (with Bonferroni correction), to assess differences between the values of the measures obtained in the various conditions. The notation "$*$"on the adaptive noise suppression bars represent the confidence with which the null hypothesis was rejected when comparing the means of the adaptive and the fixed-noise suppression. Similar improvements were observed for other SNR conditions. As can be seen from this figure, the adaptive-noise suppression approach provided significantly better performance according to the aforementioned measures as compared to the no-noise suppression and fixed-noise suppression systems. For the playground environment, for instance, the PESQ improved from 2.3 with the fixed-noise suppression system to 2.6 with the adaptive system. It should be noted that all these objective measures were computed using the acoustic waveforms generated by the adaptive noise-suppression approach discussed in Section III-D. For visualization purposes, Fig. 5 shows an electrodogram, derived using the 8-of-22 stimulation strategy for the speech segment "asa" spoken by a female talker. More specifically, this figure shows the electrodogram of a clean speech, a noisy speech with street noise added at 5-dB SNR, and enhanced electrodogram with the adaptive and fixed-noise suppression algorithms. The enhanced electrodogram was obtained by passing the noisy speech signal through the CI system illustrated in Fig. 1. The noisy and clean speech electrodograms were obtained using the CI system without noise suppression. As can be seen from this figure, the adaptive system is more effective in suppressing noise than the fixed-suppression system. This is evident, for instance, in electrodes 8–12 at segments $t = 0.2$–0.4 s and $t = 0.55$–0.8 s. It is worth mentioning that although following a misclassification a different gain function than the one corresponding to the correct noise class might be selected, we found that this did not degrade performance. That is, the enhanced speech was found to still have higher quality (as assessed by the aforementioned objective measures) than that of noisy speech obtained without noise suppression. It should be noted that based on the earlier evaluation of the proposed

TABLE II
CLASSIFICATION CONFUSION MATRIX OF THE NOISE ADAPTIVE CI SYSTEM AT SNR = 0 dB

	Apartment	**Car**	**Flight**	**Mall**	**Office**	**Place of worship**	**Playground**	**Restaurant**	**Street**	**Train**
Apartment	**99.7**	0.02	0.03	0.08	0.02	0.02	0.02	0.06	0.01	0
Car	0.14	**96.24**	0.19	0.04	0.31	0.45	0.38	0.23	1.05	0.96
Flight	0.01	0.16	**97.87**	0.01	0.49	0.65	0.19	0.03	0.08	0.48
Mall	0.22	0.01	0	**98.34**	0.05	0.28	0.86	0.20	0.01	0
Office	0.94	0.36	0.18	0.02	**95.04**	0.39	0.63	0.06	1.07	1.27
Place of worship	0.13	0.34	0.24	0.04	0.14	**98.03**	0.13	0.19	0.46	0.37
Playground	0.19	0.16	0.47	0.35	0.46	0.02	**97.17**	0.45	0.32	0.38
Restaurant	0.31	0.73	0.05	0.71	0.22	0.83	0.78	**94.96**	1.39	0
Street	0.27	0.49	0.72	0	0.55	0.55	0.57	0.24	**96.47**	0.12
Train	0	1.28	0.27	0	0	0	0.39	0.09	0.34	**97.60**

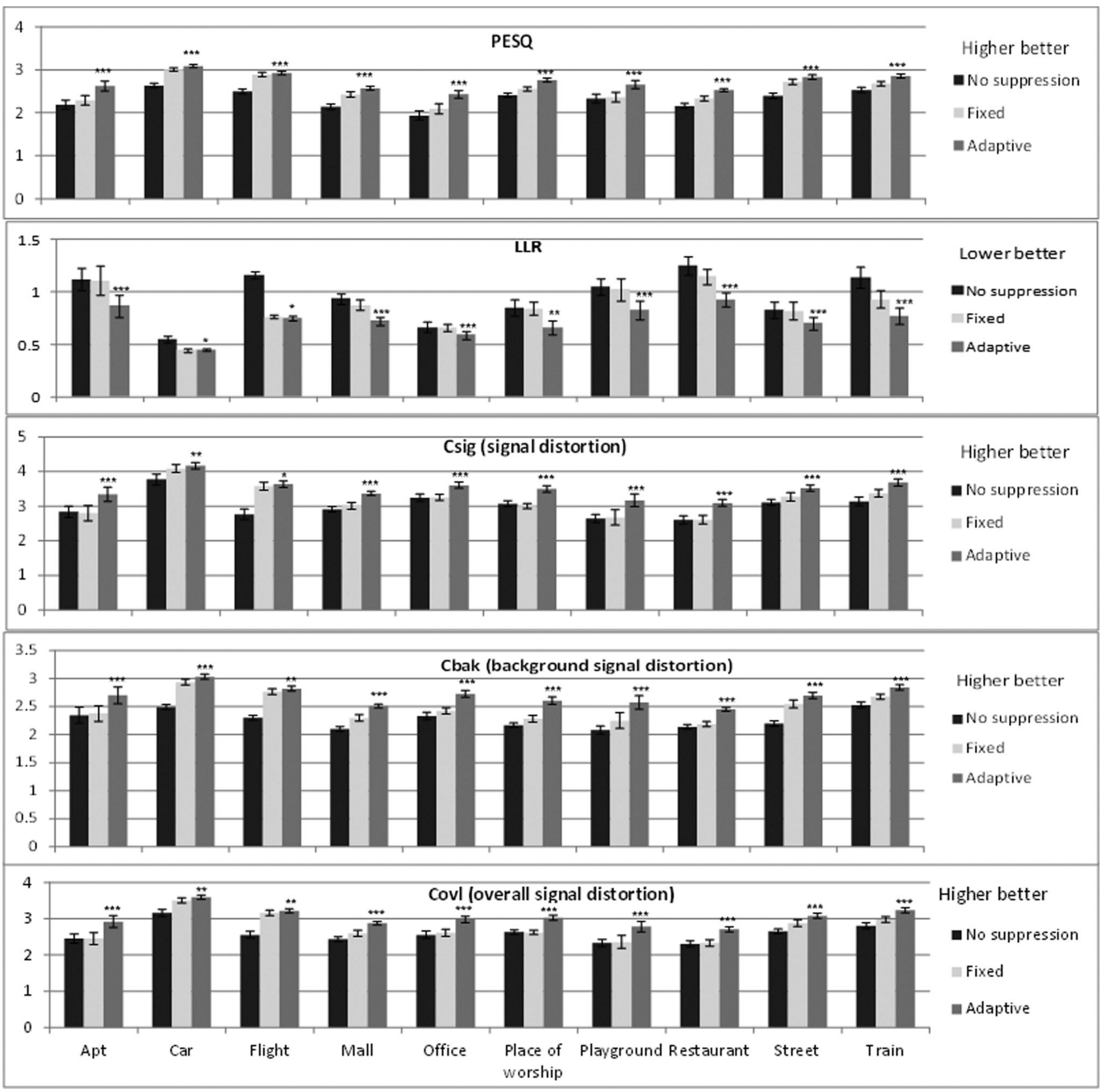


Fig. 4. Bar charts showing the performance of the adaptive noise suppression, fixed-noise suppression and no-noise suppression algorithms in terms of the objective measures PESQ, LLR, $C_{\rm sig}$, $C_{\rm bak}$, and $C_{\rm ovl}$. Error bars represent the standard deviation. The asterisk over the adaptive noise suppression bar indicates the confidence with which the means of adaptive and fixed-noise suppression are significantly different with $*$, $**$ and $***$ indicating $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

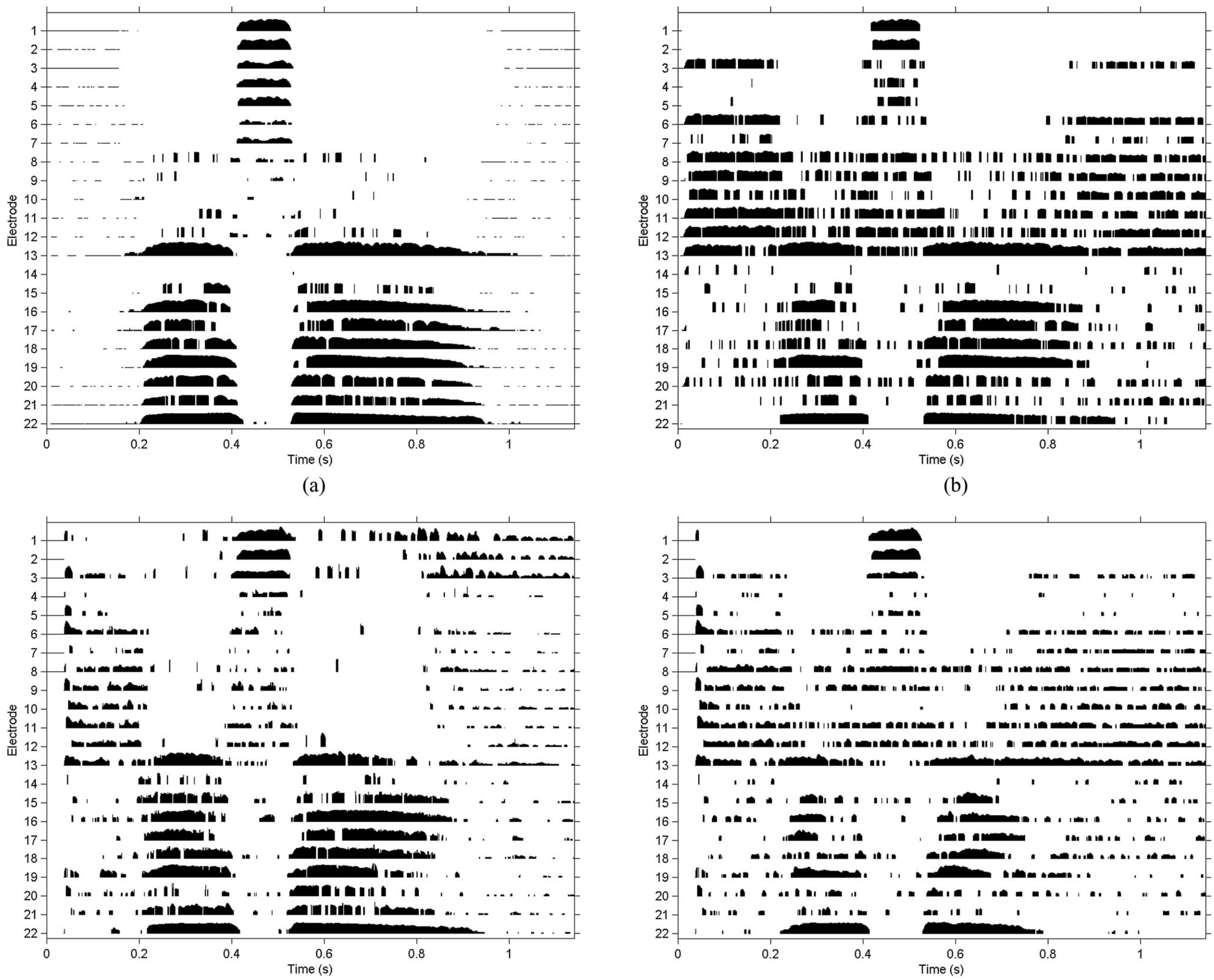


Fig. 5. Electrodograms of the utterance 'asa': (a) clean signal, (b) noisy signal with street noise at 5-dB SNR, (c) after adaptive noise suppression, and (d) after fixed-noise suppression.

TABLE III
REAL-TIME TIMING PROFILE OF THE CI SYSTEM COMPONENTS FOR 256-SAMPLE FRAMES (=11.6 MS AT 22 KHZ SAMPLING FREQUENCY)

Processing time in ms on	Proposed CI system	Recursive WPT	Voice activity detector	Feature extraction, classifier	Noise suppression	Envelope computation
PDA	8.41	1.24	0.91	2.03	2.40	1.83
PC	0.70	0.12	0.03	0.14	0.36	0.05

adaptive noise-suppression system, we cannot infer that there will be any concomitant improvements in speech intelligibility. Further clinical testing of the proposed system is needed to answer this question.

Table III shows the real-time profiling of the complete system components on both the PC and PDA platforms. The Table lists the times required for the specified components in the system to process 11.6-ms frames (256 samples). As expected, the PDA platform took a much longer processing time than the PC platform to process 11.6-ms frames due to its limited processing power. However, it still achieved a real-time throughput by processing 11.6-ms frames in about 8.5 ms.

VI. CONCLUSION

A real-time noise classification and tuning system along with the n-of-m speech processing strategy using the WPT has been implemented for cochlear implant applications. The system is capable of automatically detecting noise environment changes and selecting the optimized parameters of a noise suppression algorithm in response to such changes. The feature vector and the classifier deployed in the system to automatically identify the background noise environment are carefully selected so that the computation burden is kept low to achieve a real-time throughput. The results reported indicate improvement in speech enhancement when using this adaptive real-time cochlear implant system. In our future work, we plan to carry out a clinical testing of the enhanced cochlear implant system introduced in this paper.

References

[1] National Institute on Deafness and Other Communication Disorders. (2009, Aug.). "Cochlear Implants," National Institutes of Health, publication no. 09-4798. [Online]. Available: http://www.nidcd.nih.gov/health/hearing/coch.asp

[2] J. Remus and L. Collins, "The effects of noise on speech recognition in cochlear implant subjects: Predictions and analysis using acoustic models," *EURASIP J. Appl. Speech Process.: Spec. Issue DSP Hear. Aids Cochlear Implants*, vol. 18, pp. 2979–2990, 2005.

[3] B. Fetterman and E. Domico, "Speech recognition in background noise of cochlear implant patients," *Otolaryngol. Head Neck Surg.*, vol. 126, no. 3, pp. 257–263, 2002.

[4] Y. Hu, P. Loizou, N. Li, and K. Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," *J. Acoust. Soc. Amer.*, vol. 128, no. 4, pp. 128–134, 2007.

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

[6] Y. Hu and P. Loizou, "Environment specific noise suppression for improved speech intelligibility by cochlear implant users," *J. Acoust. Soc. Amer.*, vol. 127, no. 6, pp. 3689–3695, 2010.

[7] G. Kim and P. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 8, pp. 2080–2090, Nov. 2010.

[8] T. Fingscheidt, S. Suhadi, and S. Stan, "Environment-optimized speech Enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 825–834, May 2008.

[9] J. Erkelens, J. Jensen, and R. Heusdens, "A data-driven approach to optimizing spectral speech enhancement methods for various error criteria," in *Proc. Speech Commun., Spec. Iss. Speech Enhancement*, vol. 49, no. 7–8, pp. 530–541, 2007.

[10] J. Erkelens and R. Heusdens, "Tracking of non-stationary noise based on data-driven recursive noise power estimation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 6, pp. 1112–1123, Aug. 2008.

[11] V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, "A recursive wavelet-based strategy for real-time cochlear implant speech processing on PDA platforms," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2053–2063, Aug. 2010.

[12] V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, "Real-time implementation of wavelet-based advanced combination encoder on PDA platforms for cochlear implant studies," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 2010, pp. 1670–1673.

[13] P. Loizou, "Speech processing in vocoder-centric cochlear implants," in *Cochlear Brainstem Implants* (Adv. Otorhinolaryngol Series). Basel, Switzerland: Karger, vol. 64, pp. 109–143, 2006.

[14] S. Eddie, "Hearing aid usage in different listening environments," M.S. thesis (Audiology), Univ. Canterbury, Christchurch, New Zealand, 2007.

[15] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *Hear. J.*, vol. 63, no. 1, pp. 19–32, 2010.

[16] "A Silence Compression Scheme for G.729 Optimized for Terminals Conforming to Recommendation V.70," ITU-T Rec. G.729-Annex B, 1996.

[17] J. Ramirez, J. Segura, C. Benitez, L. Garcia, and A. Rubio, "Statistical voice activity detection using a multiple observation likelihood ratio test," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 689–692, Oct. 2005.

[18] E. Nemer, R. Goubran, and S. Mahmoud, "Robust voice activity detection using higher-order statistics in the LPC residual domain," in *IEEE Trans. Speech Audio Process.,* vol. 9, no. 3, pp. 217–231, Mar. 2001.

[19] M. Stadtschnitzer, T. Pham, and T. Chien, "Reliable voice activity detection algorithms under adverse environments," in *Proc. IEEE 2nd Int. Conf. Commun. Electron.*, 2008, pp. 218–223.

[20] S. Jovicic and Z. Saric, "Acoustic analysis of consonants in whispered speech," *J. Voice*, vol. 22, no. 3, pp. 263–274, 2008.

[21] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC Press, 2007.

[22] J. Kates, "Classification of background noise for hearing-aid applications," *J. Acoust. Soc. Amer.*, vol. 97, pp. 461–470, 1995.

[23] E. Alexandre, L. Cuadra, L. Alvarez, M. Zurera, and F. Ferreras, "Automatic sound classification for improving speech intelligibility in hearing aids using a layered structure," in *Lecture Notes in Computer Science*. vol. 4224, New York: Springer-Verlag, 2006.

[24] M. Buchler, S. Allergo, S. Launer, and N. Dillier, "Sound classification in hearing aids inspired by auditory scene analysis," *EURASIP J. Appl. Signal Process.*, vol. 2005, pp. 2991–3002, 2005.

[25] J. Xiang, M. McKinney, K. Fitz, and T. Zhang, "Evaluation of sound classification algorithms for hearing aid applications," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process.*, 2010, pp. 185–188.

[26] L. Ma, B. Milner, and D. Smith, "Acoustic environment classification," *ACM Trans. Speech Lang. Proc.*, vol. 3, pp. 1–22, 2006.

[27] V. Gopalakrishna, S. Yousefi, N. Kehtarnavaz, and P. Loizou, "Markov random field-based features for background noise characterization in hearing devices," presented at the 14th Appl. Stochastic Models Data Anal. Conf., Rome, Italy, 2011.

[28] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. PAMI-9, no. 1, pp. 39–55, Jan. 1987.

[29] C. Lin, S. Chen, K. Truong, and Y. Chang, "Audio classification and categorization based on wavelets and support vector machine," *IEEE Trans. Speech Audio Proc.*, vol. 13, no. 5, pp. 644–651, Sep. 2005.

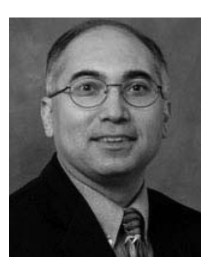
[30] V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, "Real-time automatic switching between noise suppression algorithms for deployment in cochlear implants," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, 2010, pp. 863–866.

[31] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, Apr. 1985.

[32] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

Vanishree Gopalakrishna received the B.E. degree in electronics and communication from Visvesvaraya Technological University, Karnataka, India, in 2005, the M.S. degree in electrical engineering from the University of Texas at Dallas, Richardson, TX, in 2008, where she is currently working toward Ph.D. degree in the Department of Electrical Engineering.

Her research interests include pattern recognition and real-time speech processing for cochlear implants on PDA platforms.



Nasser Kehtarnavaz (S'82–M'86–SM'92–F'12) received the Ph.D. degree from Rice University, Houstan, TX, in 1987.

He is a Professor of electrical engineering and the Director of the Signal and Image Processing Lab, University of Texas at Dallas, Richardson, TX. His research interests include digital signal and image processing, real-time signal and image processing, and pattern recognition. He has authored or co-authored eight books and more than 220 papers related to these areas.

Dr. Kehtarnavaz is a Fellow of the International Society for Optical Engineering. He is currently the Chair of the Dallas Chapter of the IEEE Signal Processing Society, and Coeditor-in-Chief of *Journal of Real-Time Image Processing*. From 2009 to 2010, he served as a Distinguished Lecturer of the IEEE Consumer Electronics Society.

Taher S. Mirzahasanloo is currently pursuing the Ph.D. degree in electrical engineering at the University of Texas at Dallas, Richardson, TX.

His current research interests include real-time implementation of speech processing algorithms for bilateral cochlear implants.

Philipos C. Loizou (S'90–M'91–SM'04) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Arizona State University, Tempe, AZ, in 1989, 1991, and 1995, respectively.

From 1995 to 1996, he was a Postdoctoral Fellow in the Department of Speech and Hearing Science, Arizona State University, working on research related to cochlear implants. He was an Assistant Professor at the University of Arkansas, Little Rock, from 1996 to 1999. He is now a Professor and holder of the Cecil and Ida Green Chair in the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX. He is the author of the textbook *Speech Enhancement: Theory and Practice* (Boca Raton, FL: CRC Press, 2007) and coauthor of the textbooks: *An Interactive Approach to Signals and Systems Laboratory* (Austin, TX: National Instruments, 2008) and *Advances in Modern Blind Signal Separation Algorithms: Theory and Applications* (San Rafael, CA: Morgan & Claypool, 2010). His research interests include areas of signal processing, speech processing, and cochlear implants.

Dr. Loizou is a Fellow of the Acoustical Society of America. He is currently an Associate Editor of the *International Journal of Audiology*. He was an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING (1999–2002), IEEE SIGNAL PROCESSING LETTERS (2006–2009), IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING (2009–2011), and a member of the Speech Technical Committee (2008–2010) of the IEEE Signal Processing Society.