35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Real-Time Dual-Microphone Noise Classification for Environment-Adaptive Pipelines of Cochlear Implants

Taher Mirzahasanloo, and Nasser Kehtarnavaz, *Fellow, IEEE*

Abstract— **This paper presents an improved noise classification in environment-adaptive speech processing pipelines of cochlear implants. This improvement is achieved by using a dual-microphone and by using a computationally efficient feature-level combination approach to achieve real-time operation. A new measure named Suppression Advantage is also defined in order to quantify the noise suppression improvement of an entire pipeline due to noise classification. The noise classification and suppression improvement results are presented for four commonly encountered noise environments.**

I. Introduction

It is shown that speech understanding of patients who have been fitted with Cochlear Implants (CIs) decreases significantly in noisy conditions [1, 2]. Some studies, e.g. [3, 4], have used speech enhancement algorithms in speech processing pipelines of CIs to address this issue. In [5-7], environment-adaptive solutions have been developed to automatically identify noise classes and optimally tune noise suppression parameters in real-time. The overall performance of these solutions depends not only on the noise suppression component, but also on the effectiveness of noise classification.

In this work, we present an improvement of the noise classification component of the environment-adaptive speech processing pipelines that have been previously developed in [5-7]. This improvement is achieved by using signals from a dual-microphone instead of a single microphone. Our solution is designed in such a way that the computational efficiency aspect of the previously developed pipelines is maintained allowing their real-time operation. In addition, to quantitatively evaluate the overall performance of an entire pipeline, a new measure is defined in this paper.

Section II provides an overview of the previously developed environment-adaptive pipelines of CIs. The dual-microphone classification improvement is then presented in section III. Section IV includes a new quality measure named Suppression Advantage followed by the results in section V. Finally, the conclusion is stated in section VI.

II. Environment-Adaptive Pipeline of CIs

As shown in Fig. 1, the basic components of our previously developed environment-adaptive noise suppression pipelines [5-7] includes two parallel paths running in real-time. The main path consists of a parameterized noise suppression component utilizing a speech decomposition method involving recursive wavelet packet transform [8]. Each parameter set is optimized using statistical data-driven methods [9–11] for a specific noise type in an offline manner and then is stored and used for noise suppression in a real-time manner.

Furthermore, the parallel path consists of an automatic noise detection and classification component that controls the optimal parameterization of the main path. It consists of a Voice Activity Detector (VAD) to identify low-energy speech frames and a noise classification module which activates when noise-only frames are identified by the VAD. This path loads the optimal noise suppression parameters into the main path based on the detected noise class. A Gaussian Mixture Model (GMM) classifier using a 26-dimensional feature vector is used to classify the noise. Also, a majority voting of classification decisions is utilized to increase reliability of the noise detection.

III. Noise Classification using Dual-Microphone

In this section, we consider the use of a dual-microphone where two input signals are captured. A comparison of two approaches when using a dual-microphone is made, leading to the selection of the more computationally and memory efficient approach.

The first approach consists of combining decisions given by two classifiers running in parallel each classifying one signal source independently, then using a decision combination module to generate a combined decision outcome. The second approach consists of fusing the feature information extracted from each signal and then using only one classifier.

Decision-level combination can be implemented by training a right and a left GMM classifier independently and combining their decisions at the majority voting step. This requires training two independent GMM classifiers and having enough memory space to store two sets of GMM parameters. Feature-level combination can be implemented by appending the feature vectors to form a single feature vector with twice the dimension. This approach would only require the use of one GMM classifier.

Table I compares the decision-level and feature-level classification approaches when using a dual-microphone in terms of memory efficiency, computational efficiency, and offline training workload. As the total number of GMM parameters for classifying a (26+26)-dimensional vector is less than that of 2 sets of GMM parameters for classifying a 26-dimensional vector, the feature-level combination requires less memory. The table also shows that the feature-level approach outperforms the decision-level approach in terms of

T. Mirzahasanloo and N. Kehtarnavaz are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (phone: 972-883-6838; fax: 972-883-2710; e-mail: kehtar@utdallas.edu).

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

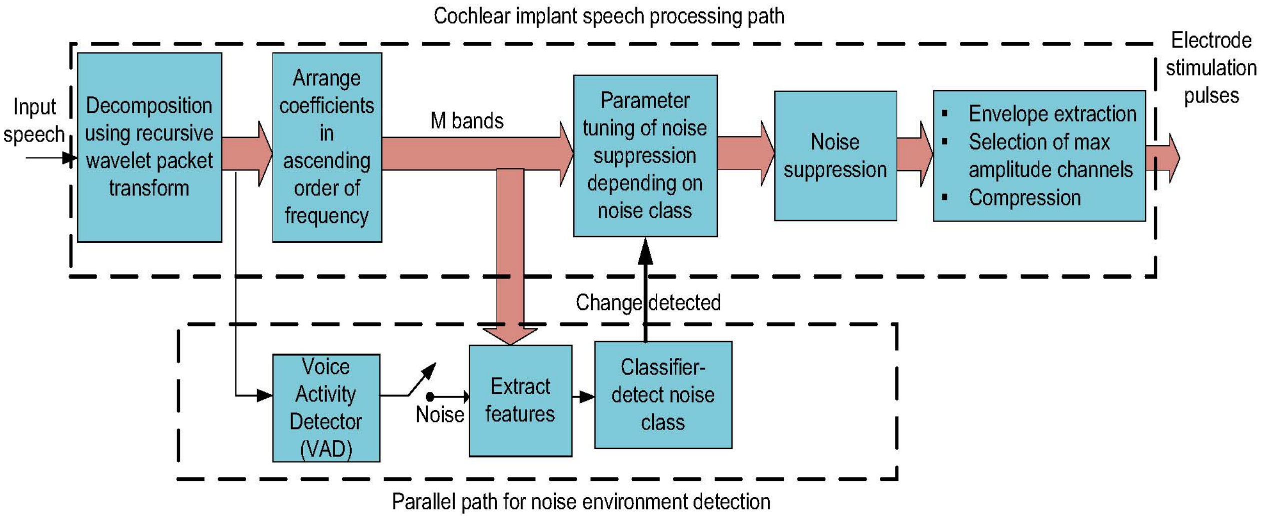
Cochlear implant speech processing path
Input speech
Decomposition using recursive wavelet packet transform
Arrange coefficients in ascending order of frequency
M bands
Parameter tuning of noise suppression depending on noise class
Noise suppression
▪ Envelope extraction
▪ Selection of max amplitude channels
▪ Compression
Electrode stimulation pulses
Change detected
Voice Activity Detector (VAD)
Noise
Extract features
Classifier-detect noise class
Parallel path for noise environment detection

Figure 1. Cochlear implant speech processing pipeline implemented in real-time [5, 7]

the computation or speed aspect. Another advantage is that the offline training is performed for only one classifier when using the feature-level approach.

Therefore, due to the memory and computational efficiency advantages of the feature-level approach, we have adopted this approach in order to improve the classification performance of the environment-adaptive pipelines of CIs in a real-time manner.

IV. Suppression Advantage Measure

A. Definitions

Let $\mathbf{P}=[P_{ij}]_{N\times N}$ be the confusion matrix associated with the above classifier, where N is the total number of environment classes and let

$$P_{ij} \triangleq P(C_i \mid C_j) \tag{1}$$

be the probability that the classifier decides class C_i while the true class is C_j with

$$\sum_{i=1}^{N} P_{ij} = 1. \tag{2}$$

Also, let $\mathbf{Q}=[Q_{ij}]_{N\times N}$ be the quality matrix associated with the noise suppression component, where

$$Q_{ij} \triangleq Q(C_i \mid C_j) \tag{3}$$

denotes the quality measure achieved when using the suppression parameters associated with class C_i while the true class is C_j.

Based on the above definitions, the expected quality for each class can be defined as follows:

$$\overline{Q}_j \triangleq \sum_{i=1}^{N} P_{ij} Q_{ij}, \quad \forall j = 1,\ldots N. \tag{4}$$

By writing $\mathbf{Q}$ and $\mathbf{P}$ as these matrices

$$\mathbf{Q}=[\mathbf{Q}_1,\ldots,\mathbf{Q}_j,\ldots,\mathbf{Q}_N], \tag{5}$$

$$\mathbf{P}=[\mathbf{P}_1,\ldots,\mathbf{P}_j,\ldots,\mathbf{P}_N], \tag{6}$$

we can write

$$\overline{Q}_j = \mathbf{P}_j^T \mathbf{Q}_j, \quad \forall j = 1,\ldots N. \tag{7}$$

The overall expected quality of the pipeline can then be stated as

$$\overline{Q} = \sum_{j=1}^{N} P_0(C_j).\overline{Q}_j. \tag{8}$$

where $P_0(C_j)$ denotes the prior probability of class C_j.

TABLE I. Comparison of Feature-Level and Decision-Level Classification Approaches

Comparisons/Approaches	**Feature-Level Combination**	**Decision-Level Combination**
Memory Efficiency	1 Set of GMM parameters for 1 input of 52-dimensional feature vector	2 Sets of GMM parameters for 2 inputs of 26-dimensional feature vectors
Computational Efficiency	1 GMM classification + 1 majority voting	2 GMM classifications + 2 majority voting + 1 decision combination
Offline Training Workload	1 GMM training	2 GMM training

B. *Fixed and Adaptive Expected Quality*

The expected values of different classes, $\overline{Q}_j$'s, depend on both the classifier and suppression components of the pipeline, thus $\overline{Q}$ evaluates the joint performance of the classifier and suppression components. Now, it is of interest to know how utilizing a noise classifier in the pipeline translates to a better suppression performance of the entire pipeline. To answer this question, we introduce a measure named Suppression Advantage (SA) here that quantifies the amount of improvement in quality measure when using an environment-adaptive suppression pipeline. This measure allows one to quantify how the overall performance improves when the classification performance improves.

Let $\overline{Q}\{\mathrm{A}\}$ be the expected quality associated with the adaptive suppression pipeline using a noise classifier with a confusion matrix of $\mathbf{P}$ as defined in (4), and $\overline{Q}\{\mathrm{F}\}$ correspond to the fixed suppression using the same fixed suppression parameter set for all noise classes, then

$$\overline{Q}\{\mathrm{A}\} = \sum_{j=1}^{N} P_0(C_j).\overline{Q}_j\{\mathrm{A}\}, \tag{9}$$

$$\overline{Q}\{\mathrm{F}\} = \sum_{j=1}^{N} P_0(C_j).\overline{Q}_j\{\mathrm{F}\}. \tag{10}$$

For adaptive suppression, from (7) and (9), we have

$$\overline{Q}_j\{\mathrm{A}\} = \mathbf{P}_j^T\mathbf{Q}_j\{\mathrm{A}\}, \quad \forall j = 1,\ldots N. \tag{11}$$

For fixed suppression, we have

$$Q_{ij}\{\mathrm{F}\} = \text{constant} = Q_j\{\mathrm{F}\}, \quad \begin{array}{l}\forall i = 1,\ldots, N,\\ \forall j = 1,\ldots, N.\end{array} \tag{12}$$

Therefore, based on (2) and from (4), we can write

$$\overline{Q}_j\{\mathrm{F}\} = Q_j\{\mathrm{F}\}, \quad \forall j = 1,\ldots N. \tag{13}$$

It can be easily seen that $\overline{Q}_j\{\mathrm{F}\}$ is independent of the confusion matrix, i.e. the expected quality is independent of the classifier performance.

C. *Suppression Advantage Measure*

To quantify the quality improvement, a base expected quality measure value is computed when there is no suppression, and then the SA measure is defined as the amount of increase in the quality measure for fixed or adaptive suppression pipelines.

Let $\overline{Q}_j\{\mathrm{N}\}$ be this base quality measure value corresponding to the noise class C_j. This value is the one given by the quality measure Q when no suppression is performed on speech signal. Then, SA of an environment-adaptive or fixed pipeline with respect to the quality measure Q can be stated as

$$SA^Q\{\mathrm{A}\} \triangleq \overline{Q}\{\mathrm{A}\} - \overline{Q}\{\mathrm{N}\}, \tag{14}$$

$$SA^Q\{\mathrm{F}\} \triangleq \overline{Q}\{\mathrm{F}\} - \overline{Q}\{\mathrm{N}\}. \tag{15}$$

Furthermore, it can be easily derived that the suppression advantage of a pipeline for each noise class is

$$SA_j^Q\{\mathrm{A}\} = \overline{Q}_j\{\mathrm{A}\} - \overline{Q}_j\{\mathrm{N}\}, \quad \forall j = 1,\ldots N, \tag{16}$$

$$\begin{aligned} SA_j^Q\{\mathrm{F}\} &= \overline{Q}_j\{\mathrm{F}\} - \overline{Q}_j\{\mathrm{N}\} \\ &= Q_j\{\mathrm{F}\} - Q_j\{\mathrm{N}\}, \quad \forall j = 1,\ldots N. \end{aligned} \tag{17}$$

V. Real-Time Implementation Results

We used noise data recorded by the BTE microphone worn by Nucleus ESPrit cochlear implant users which were sampled at a rate of 22050 Hz in four commonly encountered noise environments of Street, Car, Restaurant and Mall. These data were recorded in real noise environments using the FDA-approved PDA research platform for CI studies as described in [6, 7]. In all our classification tests, we used 50% of the data for training and 50% for testing with no overlap between the training and testing data sets. We also used the CIPIC HRTF database [12] to generate the left and right microphone signals as explained in [6]. For each microphone signal, a 26-dimensional feature vector consisting of 13 Mel-Frequency Cepstrum Coefficients (MFCC) and 13 ΔMFCC features were used as discussed in [5]. For enhancement evaluations, the collected real noise data were used to generate noisy signals of the IEEE speech sentences in [13].

Table II compares the Correct Classification Rates (CCRs) using our dual-microphone classification and the feature-level approach with that of our previously developed single-microphone classification in [5]. Using the dual-microphone classification, CCR improved by about 9.4%. Although using majority voting over a number of past classification decisions improved the classification performance considerably, time delays were introduced as a result of considering past decisions. The dual microphone approach allowed us to lower the number of past decisions leading to less time delays compared to the single microphone approach. As shown in Table II, when using 10 frames for majority voting, 7% classification improvement was achieved while getting 50% less time delay. Note that this improvement became less pronounced as more frames or a longer history of past decisions was considered for majority voting at the expense of more time delay which ultimately limited the real-time operation of the entire pipeline.

Table III provides the feature extraction and classification processing times for 11.6ms speech frames on both the FDA-approved PDA platform with a 624 MHz clock rate as well as the PC platform with a 3.0 GHz clock rate while using the majority voting over past 20 frames. As can be seen from this table, the extra computation time due to the

TABLE II. CORRECT CLASSIFICATION RATES OF DUAL-MICROPHONE CLASSIFICATION COMPARED TO SINGLE-MICROPHONE CLASSIFICATION FOR DIFFERENT NUMBER OF PAST DECISIONS OR FRAMES IN MAJORITY VOTING

Correct Classification Rate (%)	**Without majority voting**	**With majority voting over last 10 decisions**	**With majority voting over last 20 decisions**
Single-mic	74.3	81.6	91.5
Dual-mic	81.3	87.7	92.1

TABLE III. AVERAGE TIMING PROFILE OF THE ENTIRE PIPELINE FOR 11.6 MS FRAMES (IN MS)

Platform	**Total Time**	**A**	**B**	**C**	**D**	**E**
PDA (single-mic)	8.52	2.41	*1.34*	2.03	*0.91*	1.83
PDA (dual-mic)	10.39	2.41	*2.62*	2.03	*1.80*	1.83
PC	0.89	0.41	0.21	0.14	0.07	0.06

A: *FFT computation and suppression*; **B**: *Speech decomposition*;
C: *VAD decision*; **D**: *Feature extraction and classification*;
E: *Channel envelope computation*.

dual-microphone classification did not limit the real-time operation of the entire pipeline, i.e. the processing time stayed less than the frame length of 11.6ms (256 samples at 22050Hz sampling rate).

The dual-microphone approach also led to a better suppression performance of the environment-adaptive pipeline for all the noise classes as shown in Fig. 2. The same rule for all the classes was used for fixed noise suppression and the ideal system was assumed to have a perfect classification accuracy. This figure shows the SA values with respect to Perceptual Evaluation of Speech Quality (PESQ) [14], an objective quality measure proposed by ITU-T. One can see that the dual-microphone approach provided better SA over the single microphone approach when using the environment-adaptive pipeline and also when using the fixed pipeline.

VI. CONCLUSION

A real-time dual-microphone noise classification approach for environment-adaptive noise suppression in cochlear implants has been introduced in this paper. When using a dual-microphone, it was shown that the feature-level combination approach was more suitable for real-time implementation than the decision-level combination approach due to its computational and memory efficiencies. It was also shown that the classification accuracy was improved as a result of using a dual-microphone compared to using a single microphone. A new measure named Suppression Advantage was also introduced to evaluate fixed and adaptive suppression pipelines of cochlear implants and it was shown that the dual-microphone classification provided better suppression advantage.

ACKNOWLEDGEMENT

This research was supported in part by the grant no. DC010494 from the National Institutes of Health.

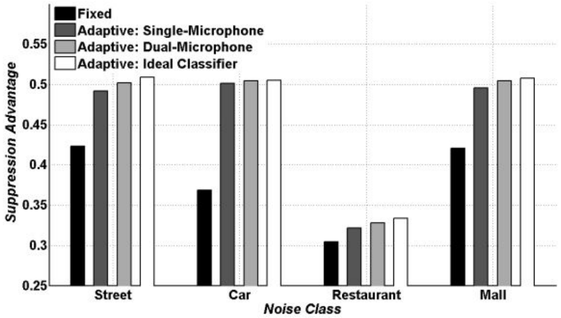

Figure 2. Suppression Advantage (SA) values with respect to the Perceptual Evaluation of Speech Quality (PESQ) [14], for a fixed noise suppression using a log-MMSE (log-Minimum Mean Squared Error) estimator [9], environment adaptive suppression using single microphone classification [5], introduced approach using dual-microphone classification, and ideal classification for environment-adaptive suppression.

REFERENCES

[1] 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. Signal Process.: Special issue on DSP in Hearing Aids and Cochlear Implants*, vol. 18, pp. 2979-2990, 2005.

[2] B. Fetterman, and E. Domico, “Speech recognition in background noise of cochlear implant patients,” *Otolaryngol. Head Neck Surg.* 126, pp. 257-263, 2002.

[3] Y. Hu, P. Loizou, N. Li, and K. Kasturi, “Use of a sigmoidal-shaped function for noise attenuation in cochlear implants,” *J. Acoust. Soc. Am.* 128, pp. 128-134, 2007.

[4] P. Loizou, A. Lobo, and Y. Hu, “Subspace algorithms for noise reduction in cochlear implants,” *J. Acoust. Soc. Am.* 118, pp. 2791-2793, 2005.

[5] V. Gopalakrishna, N. Kehtarnavaz, T. Mirzahasanloo, and P. Loizou, “Real-time automatic tuning of noise suppression algorithms for cochlear implant applications,” *IEEE Trans. Biomed. Eng.* 59, pp. 1691-1700, 2012.

[6] T. Mirzahasanloo, N. Kehtarnavaz, V. Gopalakrishna, P. Loizou, “Environment-adaptive speech enhancement for bilateral cochlear implants using a single processor,” to appear in *Speech Commun.*, 2013.

[7] T. Mirzahasanloo, V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, “Adding real-time noise suppression capability to the cochlear implant PDA research platform,” *Proc. of IEEE Int. Conf. on Eng. in Med. and Biol.*, San Diego, Aug 2012.

[8] 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, pp. 2053–2063, 2010.

[9] 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, pp. 443-445, 1985.

[10] J. Erkelens, J. Jensen, and R. Heusdens, “A data-driven approach to optimizing spectral speech enhancement methods for various error criteria,” *Speech Commun.*, vol. 49, pp. 530-541, 2007.

[11] J. Erkelens, and R. Heusdens, “Tracking of nonstationary noise based on data-driven recursive noise power estimation,” *IEEE Trans. Audio, Speech Lang. Process.* , vol. 16, pp. 1112-1123, 2008.

[12] V. Algazi, R. Duda, D. Thompson, and C. Avendano, “The CIPIC HRTF database,” *Proc. of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 99-102, 2001.

[13] IEEE Subcommittee, “IEEE recommended practice for speech quality measurements,” *IEEE Trans. Audio and Electroacoust.* AU-17, pp. 225-246, 1969.

[14] ITU, “Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” ITU, ITU-T rec. P. 862, 2000.