32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Real-Time Automatic Switching between Noise Suppression Algorithms for Deployment in Cochlear Implants

Vanishree Gopalakrishna, Nasser Kehtarnavaz, Philipos C. Loizou, Issa Panahi

Abstract—**Cochlear implant patients often complain about their difficulty in understanding speech in noisy environments. Currently a fixed noise suppression algorithm is used in cochlear implants regardless of the characteristics of the speech or noise environment. Access to an intelligent mechanism to determine the noise environment on-the-fly in order to automatically switch between different noise suppression algorithms in real-time can enhance patients experience with cochlear implants. In this paper, we report the first prototype system implementing such a real-time switching mechanism for automatic selection between two noise suppression algorithms designed for two commonly encountered noisy environments. The results obtained indicate the feasibility of this on-the-fly switching for actual deployment in cochlear implants.**

I. INTRODUCTION

COCHLEAR implants are prosthetic devices which provide a sensation of hearing to profoundly deaf people. The number of cochlear implant patients has risen to approximately 188,000 around the world as of mid 2009 [1]. Though cochlear implants provide acceptable performance when used in controlled background environments, patients express difficulty understanding speech in environments which have different noise characteristics such as restaurant babble, car noise during driving, etc. Several studies have shown that the performance of cochlear implants degrades by a large amount in the presence of noise even at 5 or 10dB SNR [2-4] indicating that cochlear implant patients are more susceptible to background noise compared to normal hearing people.

Several noise suppression algorithms have been proposed in the literature to reduce noise for single and dual microphone systems, e.g. [5-7]. Many existing cochlear implants have a single microphone as it is cosmetically more appealing, though, in general, dual microphone systems provide better results [5], [7]. Noise suppression algorithms have provided improved speech intelligibility performance in cochlear implants for a specific noise environment. A single noise suppression algorithm cannot cope with different background noise environments faced in real life situations. It is thus expected that an adaptation mechanism to different environments will enhance the performance of cochlear implants. Some noise suppression algorithms, e.g. spectral subtraction [6], perform better in stationary noise environments compared to non-stationary noise environments, while other algorithms, e.g. [8], perform better in speech shaped noise when compared to babble. Hence, naturally speech intelligibility can be improved if the noise suppression algorithm and its parameters can be chosen automatically in real-time depending on the background environment encountered by cochlear implant patients.

In this paper, we report the first prototype system built to sense a sustained change in the noise environment. This system is capable of switching between two noise suppression algorithms in real-time in an automatic manner depending on the characteristics of the two noise sources. The additional computational burden required to recognize the noise environment is minimal as the noise features used are the ones already computed as part of the cochlear implant signal processing strategy we introduced in [9],[10]. This prototype system is designed and tested for two noisy environments, namely car and restaurant babble.

The paper is organized as follows. Section II describes the components of the noise selection module. Section III describes the noise feature set and the classifier used. Section IV describes the two noise suppression algorithms which are used as part of our on-the-fly switching solution. Finally, the results and conclusions are stated in sections V and VI, respectively.

II. AUTO NOISE ENVIRONMENT SELECTION

The signal processing implemented in speech processor in a cochlear implant tries to mimic the hearing function of a normal ear. It decomposes speech signals into different frequency bands and generates stimulation pulses depending on the energy in the band corresponding to an implanted electrode. Most of the existing commercial cochlear implants use the bandpass filterbank method or the FFT method to decompose speech signals into a number of frequency bands. In our previous work, we introduced a signal processing strategy based on the wavelet packet transform to decompose speech signals in a similar way as done by the FFT method [9],[10]. In this strategy, input speech signals are passed through a 6-stage wavelet packet decomposition yielding a 64-band output. Wavelet packets of the bands which fall in the frequency range of a particular channel are summed together and passed through a rectifier and then lowpass filtered to extract the channel envelope. This envelope is further compressed using a logarithmic compression map to modulate pulse amplitudes for stimulating implanted electrodes. Daubechies or symmlet mother wavelets are used to compute the coefficients for the wavelet highpass and lowpass filters. Wavelet packets are recursively updated for each window or frame to reduce the computational burden. Figure 1 illustrates the block diagram

V. Gopalakrishna, N. Kehtarnavaz, P. Loizou, I. Panahi are with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: kehtar@utdallas.edu).

of the signal processing strategy in cochlear implants using our wavelet packet method.

The block diagram of the developed automatic switching prototype system is shown in Figure 2. Input speech signals are windowed and decomposed into M bands using the wavelet packet transform. A voice activity detector is used to determine whether the input frame contains only noise or speech with noise. If the frame is identified as noise, then its M-band wavelet packets are utilized to derive the noise feature set. Based on the derived features, the noise type (e.g., babble, car, etc.) is classified. A majority voting decision is adopted in order to increase robustness by examining several frames and selecting the noise class with the majority vote. This is carried out to avoid unnecessary switching between the noise classes and to make sure that the noise is a sustained type of noise. In our implementation, 64 bands are considered by using a 6-stage wavelet packet decomposition.

III. NOISE CLASSIFICATION

A. Voice activity detector (VAD)

Classification of noise is performed on the frames which include only noise and no speech. To determine the presence or absence of speech, a simple SNR based VAD is used in our implementation [11]. Of course, it is possible to use more sophisticated VAD algorithms. It is worth pointing out that the increase in the computational burden due to this VAD is negligible. The choice of VAD plays an important role noting that misclassification of noise is caused primarily by the presence of speech. That is why we have used majority voting to minimize this major source of noise misclassification.

B. Noise feature set

There exist many papers in the literature on various features that can be used to separate different sound environments. These include time-domain derived features such as zero crossing rate, energy based features, frequency domain features such as harmonics, spectral separation, spectral centroid [12-14]. Several other features which are commonly used include those derived from LPCC, MFCC, FFT, DCT and DWT [15]. As we discussed in [9],[10], since an improved outcome is resulted by using the wavelet packet transform for decomposition of speech signals, our feature set in the developed system are derived based on the wavelet packet transform already computed as part of our cochlear implant signal processing strategy. These features include the mean and variance of the wavelet detail coefficients. It is important to note that the use of wavelet-based features here adds little computational burden as wavelet coefficients are already computed for the generation of stimulation pulses. This feature set is normalized so that higher amplitude features do not bias the decision made by the classifier. Figure 3 shows the 6 wavelet detail coefficients whose mean and variance are used as the noise feature set in our system.

C. SVM classifier

Support Vector Machines (SVMs) have been successfully used for noise classification, e.g. [16]. Essentially, SVMs provide a higher dimensional space representation of the extracted feature set such that sample points belonging to two different classes can be more effectively separated via a hyperplane. The hyperplane is chosen such that the margin between the hyperplane and the sample points is maximized for the classes. In our implementation, a radial basis function SVM is used. Of course, it is possible to use other types of classifiers. The radial basis function is applied to the 12-dimensional input feature vectors (mean and variance of 6 detail coefficients within each frame) to find a separating hyperplane with the largest margin during the training phase.

IV. NOISE SUPPRESSION

A. Log Minimum Mean Square Error (Log MMSE)

The Log MMSE [17] [18] algorithm is commonly used to suppress noise when the noise is uncorrelated additive Gaussian noise. Thus, this suppression is used here to serve as one of the noise suppression algorithms for car noise. Let x denote clean speech, n noise and $y = x + n$ noisy speech. This method tries to obtain an estimate of x such that the mean square error of its log-magnitude spectra is minimized.

B. Sigmoidal-shaped weighting of envelopes

This second noise suppression algorithm is discussed in [19] and is shown to provide better recognition of words by cochlear-implant patients in babble and speech shaped noise conditions. Thus, this suppression is used here to serve as the other algorithm for babble noise. In this algorithm, a weighting function is computed for each channel envelope. The weighting function is made inversely proportional to the instantaneous SNR for a channel envelope as envelopes with low SNR provide less reliable information and envelopes with high SNR provide reliable information, and as such are not altered much. The weights are easily incorporated into our system.

V. REAL-TIME RESULTS AND DISCUSSION

A total of 80 noisy sentences (~3 sec in duration/sentence) embedded in 10dB SNR and sampled at 8 kHz sampling rate were considered to test the performance of the above discussed prototype system [11]. 50% of the sentences were used for training and the remaining 50% for testing with no overlap with the training set. The decision to switch to the other noise class was made only after obtaining 20 successive noise frames and using the class with the majority vote. This helped the robustness because of the presence of spurious noise and incorrect VAD decisions. This majority voting scheme ensured that unnecessary switching was not made between the two noise suppression algorithms. In other words, this ensured that the switching was made only when the noise was a sustained type of noise. Fewer numbers of frames caused the system to become sensitive to spurious noise generating higher misclassifications, and higher numbers of frames led to switching delay without any improvement in classification.

Table 1 provides the confusion matrix exhibiting an overall classification rate of 95% for the system. It is worth pointing out that even if a misclassification occurs, the noise would still be reduced by the other noise suppression algorithm though not as well as it would have been if the correct classification was made.

A real-time implementation of the entire prototype system shown in Figure 2 was achieved using C programming on a general purpose PC processor running at 3 GHz clock rate. Table 2 lists the times taken to process 256-sample frames equivalent to 11.6ms frames at a sampling frequency of 22 kHz. Three conditions are listed in this table: (a) wavelet decomposition without any noise suppression, (b) wavelet decomposition with a fixed noise suppression algorithm, and (c) wavelet decomposition with the switching module which includes the classifier. As can be seen from this table, the switching case took less than 11ms, thus achieving a real-time throughput.

Figure 4 shows a sample analysis channel output obtained by the system for sentences corrupted with car and babble noise at 10dB SNR. As depicted in Figure 4, changes in the environment were detected within 1s, which is quite acceptable in practice.

Table 1: Noise classification confusion matrix

True\Classified	Car	Babble
Car	97.5%	2.5%
Babble	6.9%	93.1%

VI. Conclusion

An automatic switching system between two noise suppression algorithms for deployment in cochlear implants has been achieved in real-time. It is shown that it is feasible to characterize the noise source on-the-fly using wavelet features in order to switch between the corresponding noise suppression algorithms with very little delay. Our ongoing work involves inclusion of other noise suppression algorithms corresponding to other types of noisy environments and using classifiers suitable for multiclass classification.

Acknowledgments

This work has been funded by the NIH contract N01-DC-6-0002 from NIDCD. The authors would like to thank Naga Devineni and Nishank Pathak for their discussions of the noise suppression algorithms.

References

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

[2] Q. Fu, and G. Nogaki, "Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing," *Journal of Association for Research in Otolaryngology*, vol. 6, pp. 19-27, 2005.

[3] B. Fetterman, and E. Domico, "Speech recognition in background noise of cochlear implant patients," *Otolaryngology-Head and Neck Surgery*, vol. 126, issue 3, pp. 257-263, 2002.

[4] G. Stickney, F. Zeng, R. Litovsky, and P. Assmann, "Cochlear implant speech recognition with speech maskers," *Journal of Acoustical Society of America*, vol. 116(2), pp. 1081-91, 2004.

[5] F. Toledo, P. Loizou, and A. Lobo, "Subspace and envelope subtraction algorithms for noise reduction in cochlear implants," *Proc. of IEEE EMBS Conf.*, vol. 3, pp. 2002-2005, 2003.

[6] J. Li, Q. Fu, H. Jianq, and M. Akaqi, "Psychoacoustically-motivated adaptive beta-order generalized spectral subtraction for cochlear implant patients," *Proc. of IEEE ICASSP*, pp. 4665-4668, 2009.

[7] M. Kompis, M. Bertram, P. Senn, J. Muller, M. Pelizzone, and R. Hausler, "A two-microphone noise reduction system for cochlear implant users with nearby microphones- part II: Performance evaluation," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, Article ID 451273, 2008.

[8] L. Yang, and Q. Fu, "Spectral subtraction-based speech enhancement for cochlear implant patients in background noise (L)," *Journal of Acoustic Society of America*, vol. 117(3), pp. 1001-1004, 2005.

[9] V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, "Real-Time implementation of wavelet-based advanced combination encoder on PDA platforms for cochlear implant studies," *Proc. of IEEE ICASSP*, Dallas, 2010.

[10] V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, "A Recursive Wavelet-based Strategy for Real-Time Cochlear Implant Speech Processing on PDA Platforms," to appear in *IEEE Trans. on Biomedical Engineering*, 2010 (available online).

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

[12] J. Kates, "Classification of background noise for hearing-aid applications," *Journal of Acoustic Society of America,* vol. 97(1), pp. 461-470, 1995.

[13] M. Buchler, S. Allergo, S. Launer, and N. Dillier, "Sound classification in hearing aids inspired by auditory scene analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 2991-3002, 2005.

[14] 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," *Lecture Notes in Computer Science,* Springer Verlag, vol. 4224, 2006.

[15] E. Didiot, I. Illina, D. Fohr, and O. Mella, "A wavelet-based parameterization for speech/music discrimination," *Computer Speech and Language*, vol. 24 (2), pp. 341-357, 2010.

[16] A. Rabaoui, M. Davy, S. Rossignol, N. Ellouze, "Using one-class SVMs and wavelets for audio surveillance," *IEEE Trans. on Information Forensics and Security*, vol. 3(4), pp. 763-775, 2008.

[17] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 33 (2), pp. 443-445, 1985.

[18] N. Pathak, I. Panahi, P. Devineni, and R. Briggs, "Real time speech enhancement for the noisy MRI environment," *Proc. of IEEE EMBS*, pp. 6950-6953, 2009.

[19] Y. Hu, P. Loizou, N. Li, and K. Kasturi, "Use of a Sigmoidal-shaped function for noise attenuation in cochlear implants," *Journal of Acoustical Society of America*, vol. 122(4), pp. 128-134, 2007.

Table 2: Timing required to process 256-sample frames (≈11.6 ms at 22kHz sampling frequency)

Wavelet decomposition without noise suppression		4 ms
Wavelet decomposition + Fixed noise suppression	logMMSE	9.1 ms
	Envelope weighting	6 ms
Wavelet decomposition + Adaptive noise suppression including classification	logMMSE	9.9 ms
	Envelope weighting	6.8 ms

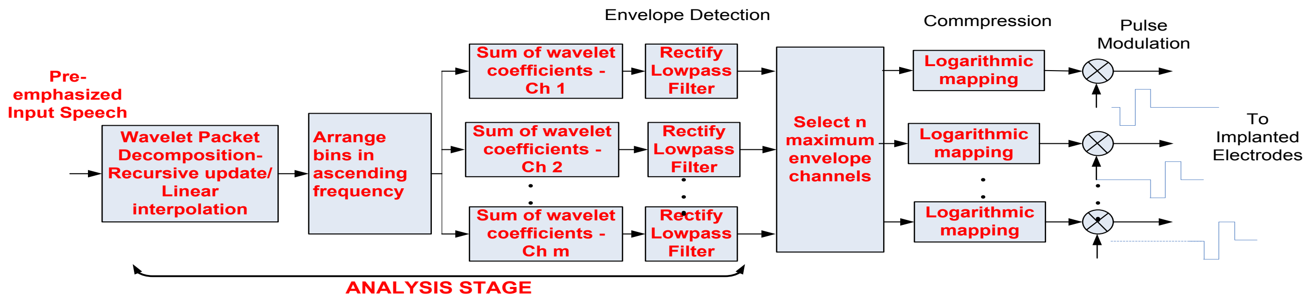

Fig. 1. Block diagram of N-of-M speech processing strategy using wavelet packet transform.

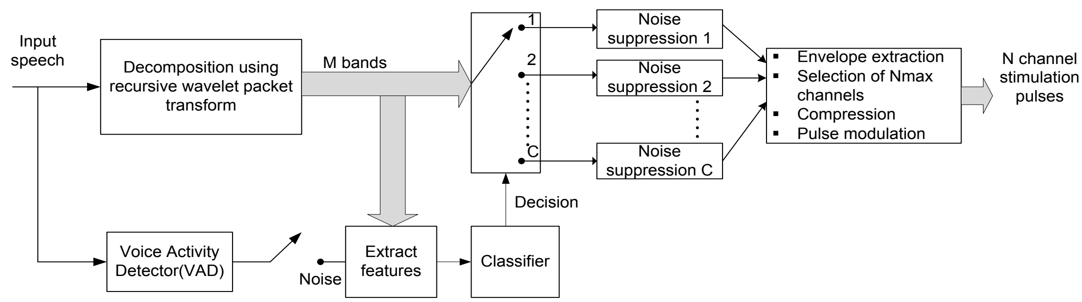

Fig. 2. Block diagram of automatic switching system based on classified noise environment.

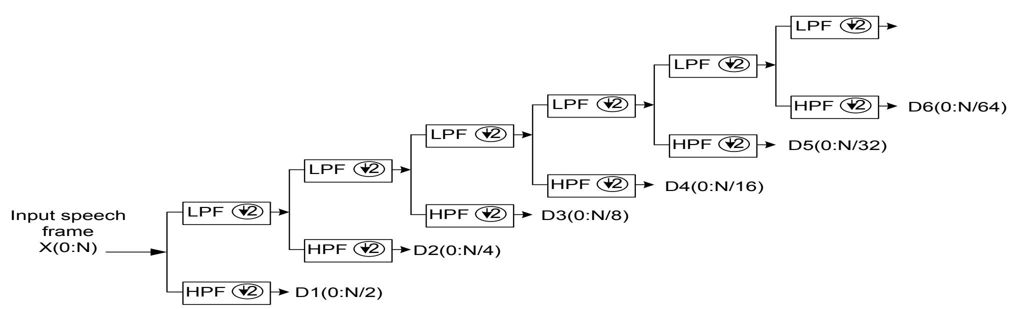

Fig. 3. Noise feature vector [D1 through D6] consisting of mean and variance of wavelet detail coefficients.

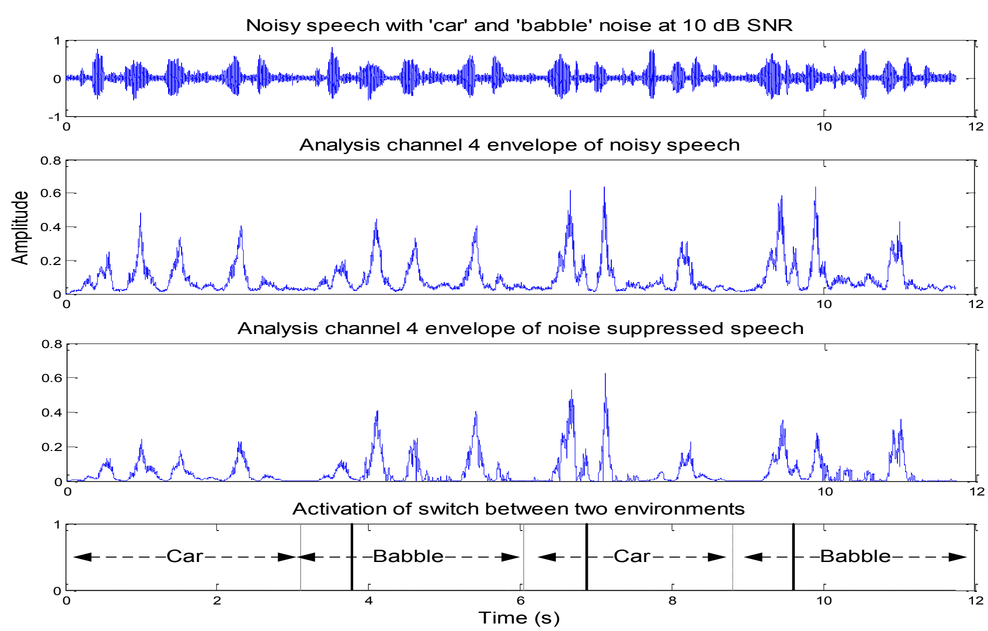

Fig. 4(a). Sample speech in the presence of car and babble noise alternating approximately every 3 secs, (b) analysis channel 4 envelope for noisy speech, (c) noise suppressed envelope, and (d) switch activation (solid line) and actual change in environment (dashed line).