BACKGROUND NOISE CLASSIFICATION USING RANDOM FOREST TREE CLASSIFIER FOR COCHLEAR IMPLANT APPLICATIONS

Fatemeh Saki and Nasser Kehtarnavaz

Department of Electrical Engineering, University of Texas at Dallas, USA

ABSTRACT

This paper presents improvements made to the previously developed noise classification path of the environment-adaptive cochlear implant speech processing pipeline. These improvements consist of the utilization of subband noise features together with a random forest tree classifier. Three commonly encountered noise environments of babble, street, and machinery are considered. The results using actual noise signals indicate that this classification method provides 10% improvement in the overall classification rate compared to the previously developed classification while maintaining the real-time implementation aspect of the entire speech processing pipeline.

Index Terms— Background noise classification, random forest tree classifier, subband noise features, cochlear implants.

1. INTRODUCTION

Since the introduction of cochlear implants (CIs) that has brought hearing sensation to profoundly deaf people, many advances have been made to improve their capabilities. It is well known that the hearing sensation of patients wearing cochlear implants degrades considerably in noisy background environments. In [1-5], we developed a speech processing pipeline that performs automatic classification of different background noises for the purpose of tuning the speech enhancement component of CIs according to the classified noise type.

Many studies have been reported on noise classification consisting of the two major components of feature extraction and classification. Table I provides a representative listing of recent studies where different features and classifiers have been used to achieve background noise classification. The classification rates for the listed studies varied from 80% to 95% using different datasets. One issue that has not been specifically addressed in these studies is the computational complexity or the real-time implementation aspect for actual deployment on a CI processing platform. In [1], it was shown that the use of mel-frequency cepstral coefficients (MFCC) features together with a Gaussian mixture model (GMM) classifier provided a balance between noise classification rate and real-time implementation on the PDA platform approved by the US Food and Drug Administration (FDA) for cochlear implant studies [3].

This work involves improving the results reported in [1] in two ways: First, in place of a GMM classifier, a tree classifier is used in order to improve the overall classification rate. This is the first time tree classifier is utilized for the purpose of achieving background noise classification. Second, alternative features to MFCC are considered in order to improve the overall classification rate. In this study, the noise classes have been limited to three widely encountered noise environments of babble noise (e.g., restaurant, mall), street noise, and machinery noise. Although there are other noise types or classes, by limiting the noise classes to the above three major noise environments, the computational complexity is kept low making the real-time deployment feasible.

2. PREVIOUSLY DEVELOPED ENVIRONMENT-ADAPTIVE COCHLEAR IMPLANT PIPELINE

The previously developed environment-adaptive speech processing pipeline for cochlear implants described in [1-5] is briefly mentioned here to provide an overview of the components involved. The pipeline consists of two parallel paths, see Fig. 1: a speech processing path and a noise classification path, both running in real-time. The speech processing path includes a parameterized noise suppression component whose parameters are set according to the noise class identified by the noise classification path. This path includes a voice activity detector (VAD) to determine whether signal frames are pure noise or speech+noise. Then, MFCC features are extracted from these durations and fed into a GMM classifier to determine the noise class or noise type.

In sections 3 and 4, the modification of the noise classification path is presented. The performance results of the modifications are then reported in section 5.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

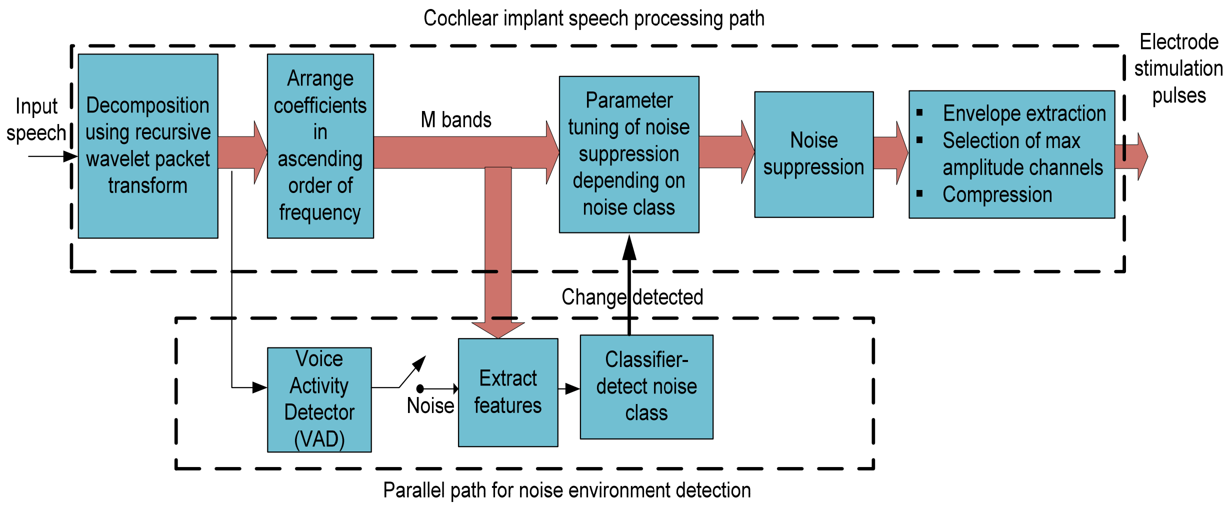

Fig.1. Cochlear implant speech processing pipeline implemented in real-time [1]

3. MODIFIED NOISE CLASSIFICATION

The use of tree classifier has been growing for real-time applications due to their recall computational efficiency. There are different training methods for tree classifiers. It has been shown that ensemble training methods such as boosting and bagging are effective training methods for tree classifiers. In particular, in [18], the method of random forest (RF) was shown to provide higher or more accuracy than the other ensemble methods. In this method, ensemble of trees are grown independently using randomly selected subsets of the training data. For training, the entropy of the root node (starting point of training) starts high since all the training samples from all the classes are included at this level. Then, the tree is grown in a way that the amount of entropy is decreased at each level, finally reaching the leaves having the lowest entropy.

Let $H(Q)$ denote the entropy at node Q,

$$H(Q) = -\sum_{c} P(\omega_c)\log_2(P(\omega_c)) \quad (1)$$

where $P(\omega_c)$ represents the portion of samples from class ω_c at node Q. It is desired for this value to be 0, that is all the samples reaching this node corresponding to the same class; otherwise this value would be high when all the

TABLE I. PREVIOUS WORKS ON NOISE CLASSIFICATION

References	Year	Features	Classifier
Khunarsal et al. [6]	2013	spectrogram, LPC and MP	NN (neural network) classifier
Chu et al. [7]	2012	MFCC and matching pursuit (MP)	Deep belief network classifier
Li et al. [8]	2010	MFCC, rhythm pattern (RP) and matching pursuit (MP)	SVM
Lozano et al. [9]	2010	MFCC, zero crossing rate, centroid and roll-off point with multi-resolution window size	GMM
Chu et al. [10]	2009	matching pursuit (MP) and MFCC	GMM
Byeong et al. [11]	2009	traditional features (TFs), change detection features (CDFs), and acoustic texture features (ATFs)	SVM
Ntalampiras et al. [12]	2008	MFCC and MPEG-7 features	Hidden Markov Model (HMM)
Kraetzer et al. [13]	2007	63 statistical features computed by AAST	Bayes classifier
Eronen et al. [14]	2006	zero-crossing rate (ZCR), MFCCs, delta-MFCC, band energy, spectral roll-off, linear prediction coefficients (LPCs) and linear prediction cepstral coefficient	k-NN (k nearest neighbor) and one-state HMM
Wang et al. [15]	2006	spectral centroid, spectral spread, and spectral flatness	Hybrid SVM and k-NN classifier
Malkin et al. [16]	2005	64 dimenstional MFCC and spectral centriod	Auto-encoder NN and GMM
Toyoda et al. [17]	2004	instantaneous spectrum at power peak and the power pattern in the time domain	NN

classes appear equally. Let us consider being at node Q and the samples are to be split between the left and right side of node Q. The information gain after the split can be expressed by

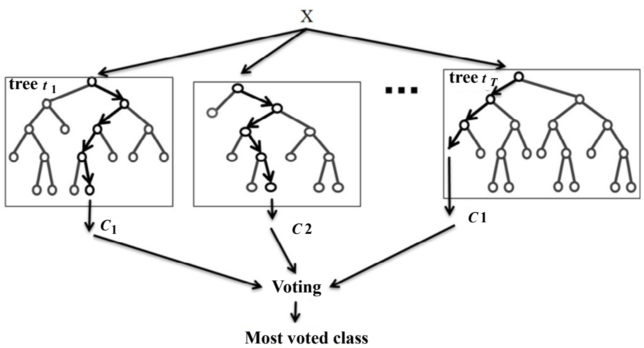

Fig.2. Recall process of Random Forest

$$I_Q = H(S_Q) - \sum_{d \in \{Right, Left\}} \frac{|S_Q^d|}{|S_Q|} H(S_Q^d) \tag{2}$$

where S_Q is the subset of training samples reaching node Q, and *Right, Left* indicate the left and right side. Among all possible splitting values, the value which provides the maximum information gain is selected. In other words, the split point which leads to a higher entropy reduction is used for growing the tree.

The classification decision is then made based on the most voted class over all the trees. Each decision tree of RF is grown on a bootstrap training sample using a learning algorithm such as CARTS [19]. During the recall, a test input X gets pushed through all the trees (starting at the root) until it reaches the leaves. Fig. 2 shows the recall or test process of RF.

As the first contribution of this work, the RF tree classifier is used in place of the GMM classifier previously used in [1] in order to improve the classification outcome. As the second contribution of this work, alterative noise features are considered in place of the MFCC features previously used in [1] in order to improve the classification outcome. These alternative noise features are discussed next.

4. SUBBAND NOISE FEATURES

The alternative noise features considered include band periodicity (BP) and band entropy (BE). Band periodicity was used in [20] to distinguish between music and background noise based on the periodicity characteristics in each subband of a signal. This feature is utilized here to capture the periodicity aspect of the machinery noise signal whose characteristics remain mostly constant or stationary over time. Fig. 3 illustrates the difference in the probability densities of this feature in the first subband for the three noise classes considered.

The periodicity of each subband can be represented by the maximum local peak of the normalized correlation function. The normalized correlation function between two adjacent frames is calculated as follows:

$$C_{b,n}(k) = \frac{\sum_{m=0}^{M-1} f(m-k) f(m)}{\sqrt{\sum_{m=0}^{M-1} f^2(m-k) \sum_{m=0}^{M-1} f^2(m)}} \tag{3}$$

where $C_{b,n}(k)$ denotes the normalized correlation function between two frames with b denoting band index and n frame index, *f(.)* is the subband signal associated with the two consecutive frames, and M indicates the frame length. Let the maximum local peak of the correlation of two adjacent frames be $C_{b,n}(k_p)$. Then, the band periodicity of noise signal frames at each subband is calculated as follows:

$$BP_b = \frac{1}{N} \sum_{n=1}^{N} C_{b,n}(k_p) \tag{4}$$

where N indicates the total number of frames.

Band entropy is a feature that provides a measure of entropy at each subband of noise signal frames, that is

$$BE_b = \frac{1}{N} \sum_{n=1}^{N} H(n) \tag{5}$$

where $H(n)$ denotes the entropy of nth frame. This feature is meant to capture the non-stationary characteristics of the babble and street noise types.

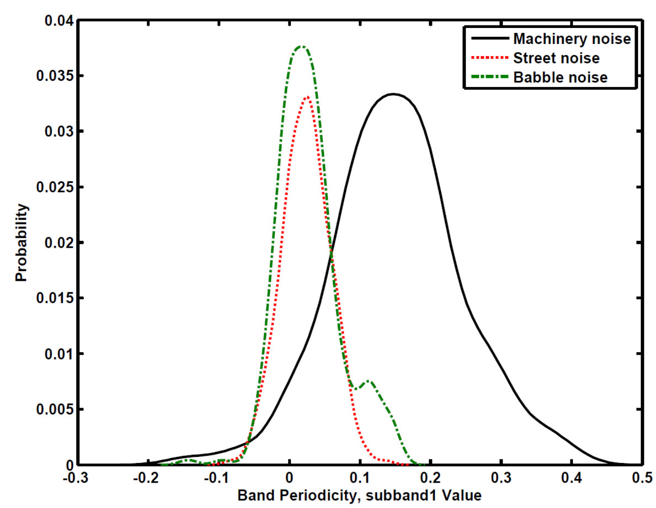

Fig. 3. Probabilty density curves of band periodicity for babble, machinery and street noise classes

TABLE II. COMPARISON OF GMM AND TREE CLASSIFIER

	Babble		Street		Machinery	
	GMM	*Tree*	*GMM*	*Tree*	*GMM*	*Tree*
Babble	**90.1%**	**94.6%**	7.2%	1.7%	2.7%	3.7%
Street	5.5%	2.6%	**91.6%**	**97.2%**	2.9%	0.2%
Machinery	10.2%	1.3%	4.8%	0.2%	**85 %**	**98.5%**

5. CLASSIFICATION RESULTS

To examine the effectiveness of the modifications made, noise data corresponding to the three noise environments of babble, street, and machinery were collected at a sampling frequency of 44,100Hz. Randomly selected 80% of the dataset was used for training and the remaining 20% of the dataset was used for testing. This selection was repeated 100 times and the classification outcomes were averaged.

For extracting MFCC features, the signals were windowed into 11-ms frames as done in [1] via a Hamming window with 6-ms overlap. 50 CART trees were used for the RF tree classifier. Table II provides the classification outcome when using the GMM and the tree classifier while keeping the features the same as the ones in [1], i.e. 13 MFCC features. These results show that the tree classifier provided a higher classification rate than the GMM classifier. In addition, it was found that the computation time associated with the RF tree classification was approximately 30% lower than that of the GMM classifier. In other words, the entire speech processing pipeline could still be run in real-time.

TABLE III. CONFUSION MATRIX USING SUBBAND NOISE FEATURES

	Babble	Street	Machinery
Babble	**98.3%**	1.7%	0.0%
Street	1.5%	**98.5%**	0.0%
Machinery	0.1%	0.0%	**99.9%**

When using the subband features, as done in [20], the noise signals were segmented into 1s window frames across 8 subbands. These segments were then divided into forty 25-ms non-overlapping frames. It was found that the band periodicity of the first 6 subbands and the band entropy of the first 4 subbands provided the highest discriminatory power than the other bands. As a result, these 10 subband features were used for the classification. Table III shows the classification confusion matrix via the tree classifier when using the 10 subband features in place of the original MFCC features. As can be seen from this table, the overall classification rate was improved by 10% over the previous classification rate as a result of the two improvements made in this work.

As stated earlier, to keep the complexity low, the noise classes were limited to the three widely encountered noise environments of babble noise, street noise, and machinery noise. Another experiment was carried out to examine the performance of our developed classification approach in the presence of other noise types. Table IV provides the classification results for this experimentation. As can be seen from this table, other noise types were placed into the closest noise class with similar noise feature characteristics.

6. CONCLUSION

In this paper, two modifications were made to the previously developed noise classification path of the environment-adaptive speech processing pipeline of cochlear implants. The first modification involved the utilization of a random forest tree classifier in place of a GMM classifier. The second modification involved the utilization of subband features to capture periodicity and entropy of noise signals. These modifications led to 10% increase in the overall classification rate while at the same time generating a lower computational burden, thus maintaining the real-time implementation aspect on the FDA approved PDA research platform for cochlear implant studies. It is planned to carry out a study on patients by turning on and off the classification path developed in this work.

7. ACKNOWLEDGEMENT

This work was supported by a contract from Cochlear Limited to the University of Texas at Dallas.

TABLE IV. TREATMENT OF OTHER NOISE ENVIRONMENTS

Noise Environment	Mapped Noise Environment		
	Babble	Street	Machinery
Quiet office with PC fan running	0%	0%	100%
Bus on road	0%	3%	97%
Party	100%	0%	0%
Hood in kitchen	0%	0%	100%
Market	90%	0%	10%
Church	96%	4%	0%
Airport	70%	10%	20%

8. REFERENCES

[1] V. Gopalakrishna, N. Kehtarnavaz, T. Mirzahasanloo, and P. Loizou, “Real-time automatic tuning of noise suppression algorithms for cochlear implant applications,” *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 1691-1700, 2012.

[2] T. Mirzahasanloo and N. Kehtarnavaz, “Real-time dual-microphone noise classification for environment-adaptive pipelines of cochlear implants,” *Proceedings of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, July 2013.

[3] T. Mirzahasanloo, V. Gopalakrishna, N. Kehtarnavaz, and P. Loizou, “Adding real-time noise suppression capability to the cochlear implant PDA research platform,” *Proceedings of 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, San Diego, CA, Aug 2012.

[4] T. Mirzahasanloo, N. Kehtarnavaz, and I. Panahi, “Adding quiet and music detection capabilities to FDA-approved cochlear implant research platform,” *Proceedings of 8th International Symposium on Image and Signal Processing and Analysis*, Trieste, Italy, Sept. 2013.

[5] V. Gopalakrishna, N. Kehtarnavaz, P. Loizou, and I. Panahi, “Real-time automatic switching between noise suppression algorithms for deployment in cochlear implants,” *Proceedings of 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Buenos Aires, Argentina, Sept. 2010.

[6] P. Khunarsal, C. Lursinsap, and T. Raicharoen, “Very short time environmental sound classification based on spectrogram pattern matching,” *Journal of Information Sciences*, vol. 243, pp. 57-74, 2013.

[7] S. Chu, S. Narayanan, C. Kuo, “Composite-dbn for recognition of environmental contexts,” *Proceedings of the Signal Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Hollywood, CA, pp. 1 –4, 2012.

[8] Y. Li, Y. Li, “Eco-environmental sound classification based on matching pursuit and support vector machine,” *Proceedings of the 2nd International Conference on Information Engineering and Computer Science (ICIECS),* Wuhan, China, pp. 1–4, 2010.

[9] H. Lozano, I. Hernaez, A. Picon, J. Camarena, and E. Navas, “Audio classification techniques in home environments for elderly/dependant people,” *Proceedings of the 12th International Conference on Computers Helping People with Special Needs*: *Part I, ICCHP’10*, Springer-Verlag, pp. 320–323, 2010.

[10] S. Chu, S. Narayanan, C. Kuo, “Environmental sound recognition with time-frequency audio features,” *IEEE Transactions on Audio, Speech and Language Processing*, vol.17 pp. 1142–1158, 2009.

[11] H. Byeong-jun, H. Eenjun, “Environmental sound classification based on feature collaboration,” *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, NY, pp. 542–545, 2009.

[12] S. Ntalampiras, I. Potamitis, N. Fakotakis, “Automatic recognition of urban environmental sounds events,” *Proceedings of International Association for Pattern Recognition Workshop on Cognitive Information Processing*, pp. 110–113, 2008.

[13] C. Kraetzer, A. Oermann, J. Dittmann, and A. Lang, “Digital Paudio forensics: a first practical evaluation on microphone and environment classification,” *Proceedings of the 9th Workshop on Multimedia and Security*, Dallas, TX, 2007.

[14] A. Eronen, V. Peltonen, J. Tuomi, A. Kalpuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi, “Audio-based context recognition,” *IEEE Transactions on Audio, Speech and Language Processing,* vol. 14, pp. 321–329, 2006.

[15] J. Wang, J. Wang, K. He, C. Hsu, “Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio lowlevel descriptor,” *Proceedings of IEEE, International Joint Conference on Neural Networks*, pp. 1731–1735, 2006.

[16] R. Malkin, and A. Waibel, “Classifying user environment for mobile applications using linear autoencoding of ambient audio,” *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Philadelphia, PA, pp. 509–512, 2005.

[17] Y. Toyoda, J. Huang, S. Ding, Y. Liu, “Environmental sound recognition by multilayered neural networks,” *Proceedings of the 4th International Conference on Computer and Information Technology, CIT ’04,* Washington, DC, pp. 123–127, 2004.

[18] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.

[19] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*, Wadsworth Int. Group, 1984.

[20] L. Lu and H. Zhang “Content analysis for audio classification and segmentation,” *IEEE Transactions on Speech and Audio Processing,* vol. 10, pp. 504-516, 2002.