A Multi-Band Environment-Adaptive Approach to Noise Suppression for Cochlear Implants

Fatemeh Saki, *IEEE Student Member*, Taher Mirzahasanloo, and Nasser Kehtarnavaz, *IEEE Fellow*

***Abstract*— This paper presents an improved environment-adaptive noise suppression solution for the cochlear implants speech processing pipeline. This improvement is achieved by using a multi-band data-driven approach in place of a previously developed single-band data-driven approach. Seven commonly encountered noisy environments of street, car, restaurant, mall, bus, pub and train are considered to quantify the improvement. The results obtained indicate about 10% improvement in speech quality measures.**

I. INTRODUCTION

Cochlear Implants (CIs) are surgically implanted devices that enable hearing sensation in profoundly deaf people. It is known that speech understanding by CI patients drops significantly in noisy environments. The literature includes many studies, e.g. [1, 2], where noise suppression is achieved by treating all noise types as noise with no distinction in the characteristics of the noise in a particular environment.

In the previous works conducted by our research team [3-6], a more effective noise suppression in terms of speech quality was developed by automatically adapting to different noise types. In addition, the real-time implementation of our environment-adaptive speech enhancement was provided as part of the CI speech processing pipeline on the FDA-approved PDA (Personal Digital Assistant) research platform. In these works, the adaptive-environment aspect was achieved by utilizing a number of gain tables for different noise environments based on the data-driven approach in [7]. In other words, for each noisy environment, a gain table discretized over a range of priori and posteriori SNRs was obtained. This table was built without distinguishing among different frequency bands.

Noting that the spectrum of real-world noise signals varies depending on different frequency bands, this paper provides a multi-band environment-adaptive speech enhancement approach. In this approach, a number of gain tables were trained for different frequency bands. It is shown that this multi-band approach generates improved results over the previously developed single-band approach.

The rest of the paper is organized as follows: Section II provides an overview of the previously developed environment-adaptive speech processing pipeline of CIs. The new multi-band approach is then presented in section III followed by the experimental results in section IV. Finally, the conclusion is stated in section V.

F. Saki, 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.

II. OVERVIEW OF PREVIOUSLY DEVELOPED ENVIRONMENT-ADAPTIVE NOISE SUPPRESSION PIPELINE

Fig. 1 shows a block diagram of the environment-adaptive pipeline for cochlear implants that was previously developed in [3]. This environment-adaptive CI speech processing pipeline is briefly mentioned here to set the stage for the understanding of the multi-band approach. The pipeline consists of two parallel paths running in real-time: speech processing path, and noise detection/classification path. The noise detection/classification path uses a Voice Activity Detector (VAD) to determine if a current signal frame is speech+noise or pure noise. If it is found to be pure noise, mel-frequency cepstrum (MFCC) or sub-band features are extracted and fed into a trained Gaussian Mixture Model (GMM) or Random Forest (RF) classifier to determine the noise type [8]. The speech processing path includes a parameterized noise suppression component whose parameters get automatically used based on the noise class determined by the classification path.

A. Data-driven Noise Suppression

To achieve speech enhancement by noise suppression, a gain function is used to map the magnitude spectrum of the input noisy speech signal to an estimate of the associated clean spectrum according to

$$\hat{A}_k(n) = \tilde{G}(\xi_k(n), \gamma_k(n)) R_k(n) \tag{1}$$

$$\xi_k(n) = \frac{\lambda_x(k,n)}{\lambda_d(k,n)} \tag{2}$$

$$\gamma_k(n) = \frac{R_k^2(k,n)}{\lambda_d(k,n)} \tag{3}$$

where $\hat{A}_k$ and R_k are the estimated clean spectral and noisy amplitudes in the frequency bin k for the time frame n, respectively, $\tilde{G}$ denotes the optimized gain function, and ξ_k & γ_k represent the priori and posteriori SNRs, respectively. To compute these SNRs, estimations of the clean spectral variance $\lambda_x(k)$ and noise spectral variance $\lambda_d(k)$ are needed. The so called decision-directed estimator involves the use of the following rule to update the priori SNR for each frame n [9]:

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

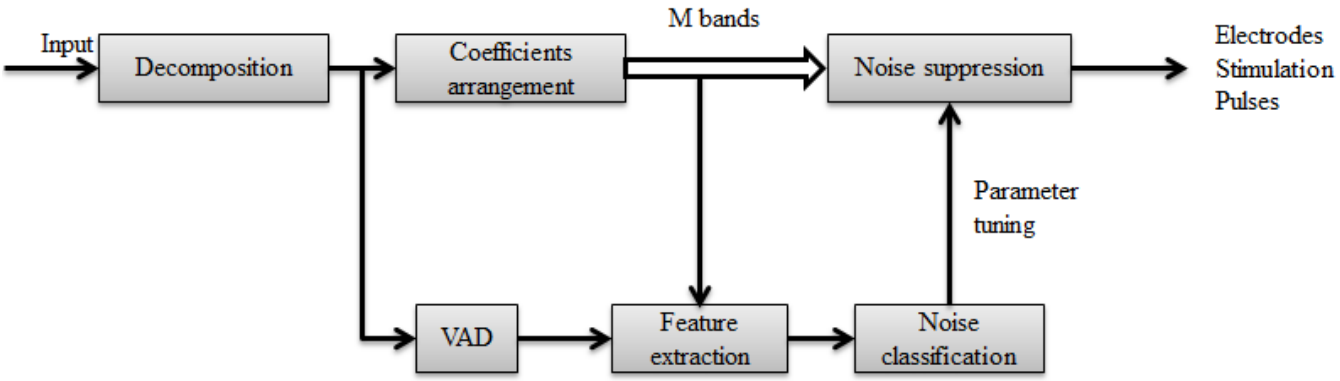

Figure 1. Cochlear implant speech processing pipeline implemented in real-time [3]

$$\hat{\xi}_k(n) = \max\left[\alpha \frac{\hat{A}_k^2(n-1)}{\lambda_d(k,n)} + (1-\alpha)[\gamma_k(n)-1], \xi_{\min}\right] \quad (4)$$

where α is a weight close to one and $\xi_{\min}$ is a lower bound on the estimated value of $\hat{\xi}_k(n)$. In this paper, the estimator and noise tracking discussed in [7] are utilized.

According to the estimated priori and posteriori SNRs, the spectral amplitude of the enhanced (clean) signal is estimated from the noisy signal based on an assumed probability density function and the optimization of an objective function. The objective function can involve MMSE, log MMSE, maximum a posteriori (MAP) estimation methods [10] or involve more recent data-driven methods [7]. In the data-driven methods, no estimation of the spectral variance is required. A brief explanation of the data-driven approach is provided next.

Let X and $\hat{X}$ be the clean and enhanced signals. In the data-driven approach, the aim is to find the function $\tilde{G}(\xi_k, \gamma_k)$ so that by applying it to the noisy signal, the estimated clean signal gets close to the clean signal. In other words, the average distortion $D(X, \hat{X})$ between clean and enhanced signals for (ξ_k, γ_k) pairs is minimized. This distortion can be any of the following: Weighted-Euclidean (WE), Log-Euclidean (LE), Weighted-Cosh (WC) or simple mean-square error (MSE) [7, 11]. Mathematically, the following equations describe the data-driven approach:

$$\tilde{G} = \{\tilde{G}_{ij}, \forall_i = 1,\ldots I, \forall_j = 1,\ldots J\} \quad (5)$$

$$\tilde{G}_{ij} = \arg\min_{G_{ij}} D(X, \hat{X}) \quad (6)$$

where $\tilde{G}$ is a look-up table discretized over a grid of priori and posteriori SNRs. A parameter cell contains the closest values of $\hat{\xi}$ and γ to a grid point with values $\tilde{G}_{ij}$ stored in matrix $\tilde{G}$. Thus, for a total of I and J priori and posteriori SNRs, respectively, the gain table consists of an $I \times J$ matrix containing the noise suppression parameters.

III. Multi-band Data-driven Noise Suppression

In the data-driven method discussed in [7], for each frame and each frequency bin, there is an $(\hat{\xi}_k, \gamma_k)$ pair that falls into one of the parameter cells of the gain table. As a result, an $(\hat{\xi}_k, \gamma_k)$ pair from different frequency bins and different frames may fall into the same parameter cell during the training of $\tilde{G}(\hat{\xi}_k, \gamma_k)$ involving a clean amplitude A_k and a noisy amplitude R_k.

In the multi-band data-driven approach introduced here, the signal is divided into M different non-overlapping frequency bands. Then, M different gain tables corresponding to M frequency bands are trained. The frequency band decomposition can be done in Fourier domain or by using a filter bank. In each frame for a frequency band, the priori and posteriori SNRs $(\hat{\xi}_{bk}, \gamma_{bk})$, with b denoting the band index, are computed. Therefore, the parameterized suppression values for each frequency band get trained separately. It is worth mentioning that the size of the gain tables is kept the same considering that each gain table covers the same prior and posterior SNR ranges. Hence, there would be M gain tables for each environment, that is:

$$\tilde{G}_b = \{\tilde{G}_{bij}, \forall_i = 1,\ldots I, \forall_j = 1,\ldots J\} \quad b = 1:M \quad (7)$$

As mentioned earlier, in the single-band noise suppression, $(\hat{\xi}_k, \gamma_k)$ pairs from different frequency bands and different frames might fall in the same cell of the gain table. This means that the corresponding suppression value $\tilde{G}(\hat{\xi}, \gamma)$ for an input frame is only a function of the estimated priori and posteriori SNRs, and thus the frames from different frequencies are treated the same. This causes some distortion in the signal. By separating the gain tables based on the frequency bands, any such distortion can be avoided. Here it is worth pointing out that the data driven suppression is performed independently in each band. As reported in [3], the suppression processing time takes only 2.4 ms out of a total processing time of 8.41 ms on the PDA platform for 11.6 ms frames. Hence, the two-band

suppression processing time is still expected to run in real-time on the PDA platform.

IV. Experimental Results and Discussion

The introduced multi-band noise suppression was evaluated by using seven commonly encountered noise types of street, car, restaurant, mall, bus, pub and train. Noise samples were collected using the same BTE (Behind-The-Ear) microphone worn by Nucleus ESPrit cochlear implant users at a sampling frequency of 8000 Hz. For training, the first 50 IEEE sentences provided in [12] (approximately 2-3s long) were used to serve as clean speech files. For each noisy environment, 50% of the noise files were added to each speech signal at several SNRs from -12.5 to 27.5 dB in steps of 5 dB to generate the training dataset. The signals were windowed into 25-ms frames via a Hamming window with 50% overlap across two non-overlapping low and high frequency bands. In the experiments reported in this paper, α and $\xi_{\min}$ were set to 0.98 and -19 dB, respectively, the prior SNR was discretized from -19 dB to 40 dB and the posterior SNR from -30dB to 40 dB in steps of 1 dB with a grid size of 60×71. It was found that the use of two bands maintained the real-time throughput of the pipeline.

The speech quality measures of Perceptual Evaluation of Speech Quality (PESQ) and Log-Likelihood Ratio (LLR) [10] were computed to provide a quantification of the improvement in the noise suppressed output signals. Fig. 2 shows the comparison of the PESQ and the LLR measures for the multi-band and single-band approaches for 0 dB SNR. The non-suppressed noisy signals are shown to serve as the baseline. The results reflect the averages on the second half of the 50 IEEE sentences which had not been used in the training dataset. This figure illustrates that the multi-band approach provides an improvement of nearly 10% in speech quality measures averaged across the noisy environments considered compared to the single-band approach. An Analysis of Variance (ANOVA) was conducted to show the statistical significance of the improvement ($p < 0.001$). In our noise dataset, the files for train and car noises had approximately uniform spectrum over all the frequency bands. That is why the improvement did not generate statistically significant improvement over the single-band approach for these two noise types while for the other noise types the improvement was found to be statistically significant.

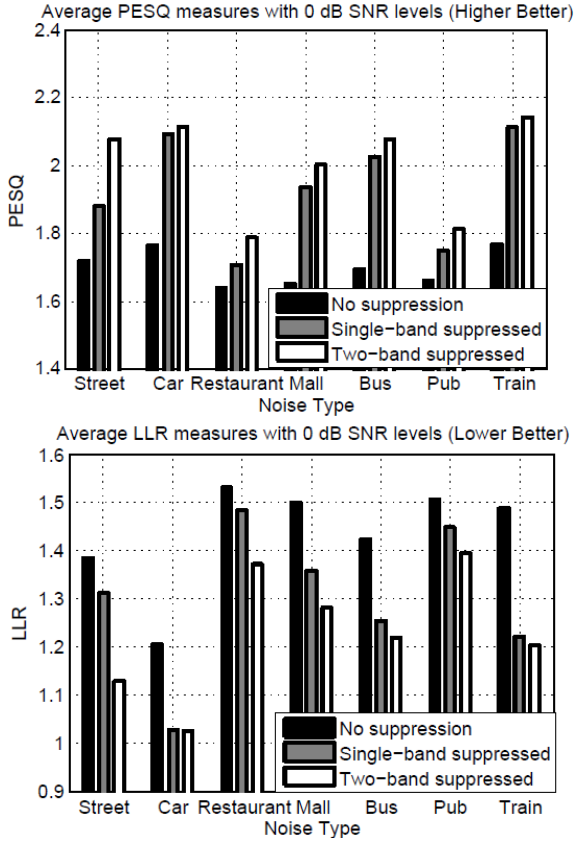

Figure 2. Bar charts showing the performance of the signle-band data-driven adaptive noise suppression, two-band data-driven adaptive noise suppresion and no-noise suppression in terms of the speech quality measures of Perceptual Evaluation of Speech Quality (PESQ) and Log-Likelihood Ratio(LLR)

Another experiment was carried out to examine the performance of the multi-band approach in the presence of other noise types which had not been considered in the original set of environments. Fig. 3 shows a comparison of the PESQ and the LLR measures exhibiting the outcome for the multi-band data-driven approach versus the noisy non-processed signals for three noise environments of airport, airplane and market. These three environments in the classification path were placed into the closest class, namely street, bus and restaurant, respectively. Consequently, the suppression parameters of these detected classes were used

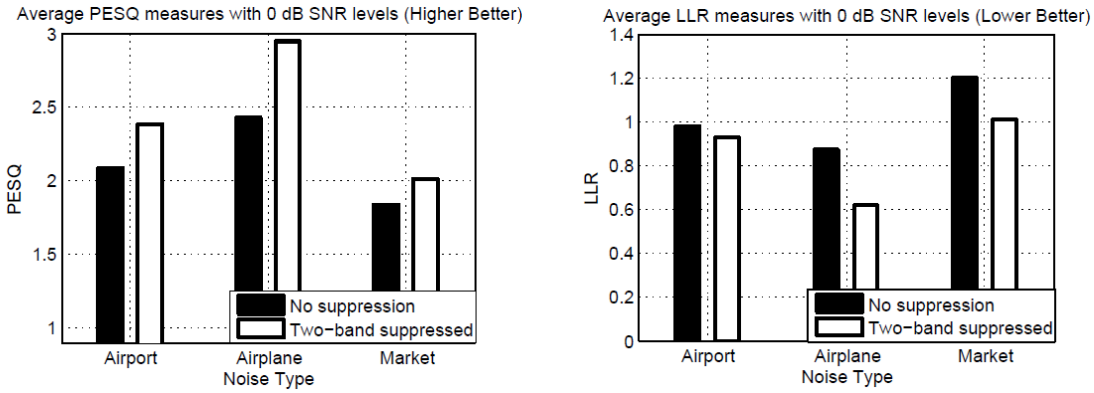

Figure 3. Comparison of PESQ and LLR quality measures when encountering unkonwn noise

for the noise suppression. A visual comparison can be made in Fig. 4 where the spectrogram of the clean, noisy, single-band data-driven suppressed and multi-band data-driven suppressed signals at 0 dB SNR are shown. This figure shows the background noise was suppressed by the developed multi-band method more than the single-band method, thus retrieving the speech signal more accurately.

V. Conclusion

A modification to the previously developed noise suppression path of the environment-adaptive speech processing pipeline of cochlear implants was introduced in this paper to improve speech enhancement via noise suppression. This modification involved the use of multiple frequency bands instead of a single-band to achieve data-driven environment-adaptive noise suppression. The experimental results showed 10% improvement in speech quality measures for seven noisy environments considered while at the same time maintaining the real-time throughput of the entire speech processing pipeline.

VI. References

[1] 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.

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

[3] 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.

[4] T. Mirzahasanloo, N. Kehtarnavaz, V. Gopalakrishna, P. Loizou, "Environment-adaptive speech enhancement for bilateral cochlear implants using a single processor," *Speech Commun.* vol. 55, pp. 523-534, 2013.

[5] 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.

[6] V. Gopalakrishna, N. Kehtarnavaz, P. Loizou, and I. Panahi, "Real-time automatic switching between noise suppression algorithms for deployment in cochlear implants," *Proc. of IEEE Int. Conf. on Eng. Med. Biol.*,Buenos Aires, 2010.

[7] 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.

[8] F. Saki, N. Kehtarnavaz, "Background noise classification using random forest tree classifier," to appear in *Proc. of IEEE ICASSP*, May 2014.

[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] P. Loizou, *Speech Enhancement: Theory and Practice.* Boca Rotan, FL: CRC, Taylor and Francis, 2007.

[11] T. Mirzahasanloo, N. Kehtarnavaz, "A generalized data-driven speech enhancement framework for bilateral cochlear implants," *Proc. of IEEE ICASSP,* Vancouver, May 2013.

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

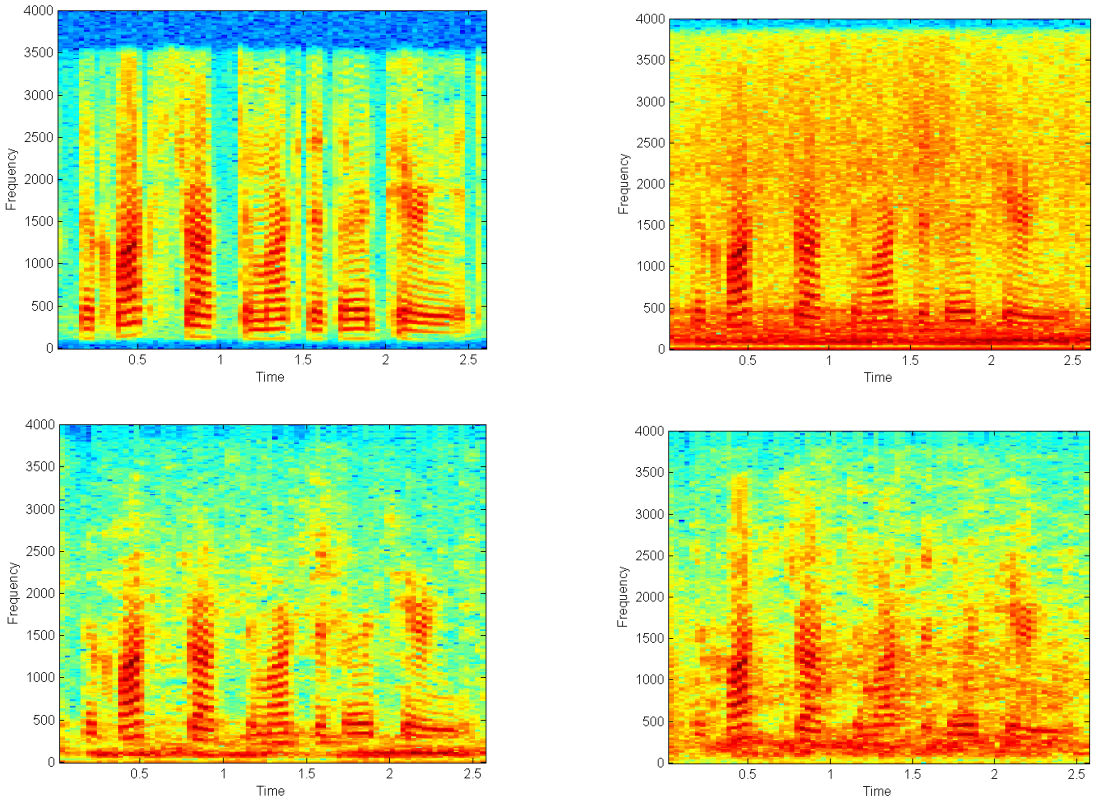

Figure 4. Sepectograms of the clean speech (top left) and noisy signals (top right) (SNR = 0 dB). Bottom left figure shows enhanced signals by the introduced two-band noise suppression approachand the bottom right one shows the single-band noise suppression approach; IEEE sentence: "*The clock struck to mark the third period*"