# A GENERALIZED DATA-DRIVEN SPEECH ENHANCEMENT FRAMEWORK FOR BILATERAL COCHLEAR IMPLANTS

*Taher S. Mirzahasanloo, and Nasser Kehtarnavaz*

Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

## ABSTRACT

This paper examines environment-adaptive noise suppression algorithms for computationally efficient or real-time implementation in bilateral cochlear implants using a single processor. A generalized framework is introduced that allows one to train suppression and head-related transfer function gain tables not only for different noise environments but also for different distortion measures. This generalization incorporates any differentiable measure with unilateral data-driven enhancement methods becoming its special cases. Specifically, the solutions for three distortion measures of Weighted-Euclidean, Log-Euclidean and Weighted-Cosh are provided. These solutions are evaluated in six commonly encountered noise environments for a wide range of directionalities.

***Index Terms***— Bilateral cochlear implants, data-driven speech enhancement, generalized data-driven speech enhancement, environment-adaptive noise suppression

## 1. INTRODUCTION

It is shown that patients fitted with Cochlear Implants (CIs) exhibit a good understanding of speech in quiet and controlled listening conditions, but in noisy environments, their speech understanding decreases significantly [1, 2]. Many studies, e.g. [3-5], have addressed this issue by incorporating noise suppression algorithms in the CI speech processing pipeline. In our previous work [6], it was shown that because different noise types have different characteristics, using an environment-adaptive noise suppression strategy is more effective when operating under realistic conditions.

In unilateral CIs, patients face difficulties locating sound sources as no directional information is perceived [7, 8]. Bilateral CIs not only provide a sense of directionality, but also improve speech understanding [9, 10], [11-13]. However, bilateral speech enhancement normally demands more processing resources, that is to say it is computationally more expensive and needs more memory as discussed in [14].

In our previous works [15, 16], we achieved the real-time implementation of an environment-adaptive speech enhancement approach for unilateral CIs as part of the speech processing pipeline on the FDA-approved PDA (Personal Digital Assistant) research platform [17]. In [15], it was shown that our adaptation approach improved the speech quality compared with a similar fixed noise suppression approach. We extended our automatic enhancement approach to bilateral CIs in [14] by taking advantage of the two signal sources. This extension was done by using only a single processor while retaining the environment-adaptability aspect of the unilateral pipeline leading to comparable speech quality scores in different noise environments. In addition, it was shown that the extension was computationally more efficient than processing the bilateral input signals independently and required almost the same amount of storage as in the unilateral enhancement. These characteristics provide a suitable solution for achieving a real-time implementation of the bilateral environment-adaptive enhancement via only a single processor.

Previous works [6, 14, 15, 18] have considered the environment-adaptability aspect by optimizing different gain tables for different noise environments, but have not studied the effects of different optimization criteria for different noise types. The problem of bilateral speech enhancement using a single processor becomes more challenging when considering non-Euclidean distortion measures. In this work, we introduce a generalized optimization framework that allows the consideration of other distortion measures [19, 20]. Specifically, we formulate our optimization for three most common distortion measures including the traditional Weighted-Euclidean (WE), Log-Euclidean (LE) and Weighted-Cosh (WC) [18-21]. The solutions provided are general purpose in the sense that they cover different problems with different weights over reference and non-reference signals and with different parameter weights for the measures. Then, the solutions for the data-driven unilateral enhancement gain optimization based on the WE, LE and WC criteria become special cases of our generalized framework. Although we have considered three most common distortion measures, the discussed framework can be applied to any differentiable measure.

## 2. CI ENVIRONMENT-ADAPTIVE NOISE SUPPRESSION PIPELINE

Our environment-adaptive CI speech processing pipeline discussed in [6, 15] includes two parallel paths: speech processing, and noise detection/classification. The speech processing path uses a recursive wavelet packet transform to decompose the input speech into different frequency bands as described in [22, 23]. After applying a gain function to the magnitude spectrum to suppress noise, it extracts channel envelopes and then generates stimulation pulses for implanted electrodes by going through the three steps of rectification, low-pass filtering and envelope compression. On the other hand, the noise detection path first uses a Voice Activity Detector (VAD) to determine if a current frame is speech or noise. If noise, a 26-dimensional MFCC feature vector is used to classify noise and to suppress it from speech according to the noise class. The data-driven method as described in [24-26] is then used to train the suppression gain function independently for each environment using a corresponding noise dataset.

### 2.1. Unilateral data-driven enhancement

In spectral domain speech enhancement algorithms, a gain function maps magnitude spectrum of the input noisy speech signal to an estimate of the associated clean spectrum. This gain is a function of prior and posterior SNRs which are found analytically by minimizing the mean squared error between noisy and clean spectral amplitudes as described in [24-26]. To account for modeling and estimation errors, in [18, 21], it was proposed to use a lookup table discretized over the course of prior and posterior SNR estimates as the gain function. The gain values of cells of a grid are obtained via a minimization procedure involving an appropriate error criterion over a training set of noisy and clean sample pairs. The data-driven nature and lower computationally complexity of this approach makes it a suitable choice for real-time implementation.

### 2.2. Extension to bilateral CI speech enhancement using a single processor

As stated above, our extension in [14] provided environment-adaptive speech processing for bilateral CIs using only a single processor. In general, there exists a delay between the two speech signals captured by the bilateral microphones. The signal that is captured first is called the reference signal. A time delay estimation based on Generalized Cross Correlation [27] determines which one is the reference signal. This estimation also provides part of the direction information required for binaural reconstruction of the non-reference spectral amplitudes from the processed reference amplitudes. The main path processes only the reference signal while the non-reference information is used to increase the reliability of the VAD and the classification components. An enhancement gain table $\mathbf{G}$ provides estimates of the reference clean spectral amplitudes and an estimation of Head-Related Transfer Function (HRTF) gain table $\mathbf{H}$ provides the reconstruction of the non-reference amplitudes based on the information provided by a direction estimation component. As described in [14], this approach eliminates the need for a second processor to provide bilateral CI stimulation pulses.

## 3. GENERALIZED DATA-DRIVEN FRAMEWORK FOR BILATERAL ENHANCEMENT

Let the suppression gain table $\mathbf{G}$ be discretized over $I$ different prior SNRs and $J$ different posterior SNRs, that is

$$\mathbf{G} = \{G_{ij}, \forall i = 1,\ldots,I, \forall j = 1,\ldots,J\} \tag{1}$$

Similarly, let the HRTF gain table $\mathbf{H}$ be discretized over $L$ different directions,

$$\mathbf{H} = \{H_l, \forall l = 1,\ldots,L\} \tag{2}$$

The total distortion is considered to be a linear combination of distortions associated with the reference and the non-reference spectral errors ( $D_r$ and $D_{nr}$ , respectively) as follows:

$$D = D_r + \beta D_{nr}, \;\; 0 \le \beta \le 1 \tag{3}$$

The parameter $\beta$ determines the relative importance of non-reference errors over reference ones. $D_r$ is the mean of distortions associated with different prior and posterior SNRs,

$$D_r \equiv \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} D_{r,ij} \tag{4}$$

where $D_{r,ij}$ is the mean distortion over data observed at the $i$ -th prior SNR and the $j$ -th posterior SNR. Non-reference errors depend not only on the suppression gain parameters but also on the HRTF gains, that is

$$D_{nr} \equiv \frac{1}{IJL}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{L} D_{nr,ijl} \tag{5}$$

where $D_{nr,ijl}$ is the mean distortion over data observed at the $l$ -th direction along with the $i$ -th prior SNR and the $j$ -th posterior SNR.

The gradient of the total distortion with respect to the suppression and HRTF gain parameters can be written as follows:

$$\frac{\partial D}{\partial G_{ij}} = \frac{1}{IJ}\{\frac{\partial D_{r,ij}}{\partial G_{ij}} + \beta \frac{1}{L}\sum_{l=1}^{L}\frac{\partial D_{nr,ijl}}{\partial G_{ij}}\} \tag{6}$$

$$\frac{\partial D}{\partial H_l} = \frac{1}{IJL}\beta\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\partial D_{nr,ijl}}{\partial H_l} \tag{7}$$

Different distortion functions can be used to compute $D_{r,ij}$ and $D_{nr,ijl}$ values. Consider $d(A,\hat{A})$ to be such a function computing the distortion between the clean spectral amplitude $A$ and the estimated enhanced counterpart $\hat{A}$. The enhanced reference signal is then given by mapping the noisy reference amplitudes via $\mathbf{G}$. Therefore,

$$D_{r,ij} \equiv \frac{1}{M_{ij}} \sum_{m=1}^{M_{ij}} d(A_{r,ij}(m), G_{ij} R_{r,ij}(m)) \quad (8)$$

where $A_{r,ij}(m)$ is the $m$-th data sample of the reference clean spectral amplitude observed at the prior and posterior SNRs corresponding to the $(i,j)$-th cell of the suppression gain table, $R_{r,ij}(m)$ is its noisy counterpart and $M_{ij}$ is the total number of data collected for this cell.

The non-reference clean amplitudes are estimated by mapping the estimated reference amplitudes and using the HRTF gain $\mathbf{H}$. Hence,

$$D_{nr,ijl} \equiv \frac{1}{M'_{ijl}} \sum_{m'=1}^{M'_{ijl}} d(A_{nr,ijl}(m'), G_{ij} H_l R_{r,ijl}(m')) \quad (9)$$

where $A_{nr,ijl}(m')$ and $R_{r,ijl}(m')$ are, respectively, the $m'$-th data sample of the non-reference clean and the reference noisy amplitudes corresponding to the $(i,j)$-th cell of the suppression gain table and the $l$-th cell of the HRTF gain table. The total number of data is assumed to be $M'_{ijl}$ for each set.

**3.1. Weighted-Euclidean distortion criterion**

The WE distortion function with weight $p$ is given by

$$d_{\mathrm{WE}}(A, \hat{A}) \equiv A^p.(A-\hat{A})^2 \quad (10)$$

Let us define the following terms to obtain a simpler representation,

$$S_{r,ij,1} \equiv \sum_{m=1}^{M_{ij}} A_{r,ij}^{p+1}(m).R_{r,ij}(m) \quad (11)$$

$$S_{r,ij,2} \equiv \sum_{m=1}^{M_{ij}} A_{r,ij}^{p}(m).R_{r,ij}^2(m) \quad (12)$$

$$S_{nr,ijl,1} \equiv \sum_{m'=1}^{M'_{ijl}} A_{nr,ijl}^{p+1}(m').R_{r,ijl}(m') \quad (13)$$

$$S_{nr,ijl,2} \equiv \sum_{m'=1}^{M'_{ijl}} A_{nr,ijl}^{p}(m').R_{r,ijl}^2(m') \quad (14)$$

From (6) and (7) and based on the definitions in (8)-(14), the WE solutions can be derived to be

$$\frac{\partial D}{\partial G_{ij}} = -2\frac{1}{IJ}\{\frac{1}{M_{ij}}(S_{r,ij,1} - G_{ij}S_{r,ij,2}) + \beta\frac{1}{L}\sum_{l=1}^{L}\frac{1}{M'_{ijl}}[H_l S_{nr,ijl,1} - H_l^2 G_{ij} S_{nr,ijl,2}]\} \quad (15)$$

$$\frac{\partial D}{\partial H_l} = -2\beta\frac{1}{IJL}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{M'_{ijl}}[G_{ij}S_{nr,ijl,1} - G_{ij}^2 H_l S_{nr,ijl,2}] \quad (16)$$

**3.2. Log-Euclidean distortion criterion**

The LE distortion is defined as

$$d_{\mathrm{LE}}(A, \hat{A}) \equiv (\log[A] - \log[\hat{A}])^2 \quad (17)$$

Similarly, for this distortion measure, let us define the following terms to obtain a simpler representation,

$$P_{r,ij} \equiv \log[\prod_{m=1}^{M_{ij}} \frac{A_{r,ij}(m)}{G_{ij}R_{r,ij}(m)}] \quad (18)$$

$$P_{nr,ijl} \equiv \log[\prod_{m'=1}^{M'_{ijl}} \frac{A_{nr,ijl}(m')}{G_{ij}H_l R_{r,ijl}(m')}] \quad (19)$$

From (6) and (7) and based on the definitions in (8)-(9) and (17)-(19), the LE solutions can be derived to be

$$\frac{\partial D}{\partial G_{ij}} = -2\frac{1}{IJ}\frac{1}{G_{ij}}\{\frac{1}{M_{ij}}P_{r,ij} + \beta\frac{1}{L}\sum_{l=1}^{L}\frac{1}{M'_{ijl}}P_{nr,ijl}\} \quad (20)$$

$$\frac{\partial D}{\partial H_l} = -2\beta\frac{1}{IJL}\frac{1}{H_l}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{M'_{ijl}}P_{nr,ijl} \quad (21)$$

**3.3. Weighted-Cosh distortion criterion**

The WC distortion with weight $p$ is given by

$$d_{\mathrm{WC}}(A, \hat{A}) \equiv A^p.(A/\hat{A} + \hat{A}/A - 1) \quad (22)$$

Similarly, by defining

$$C_{r,ij,1} \equiv \sum_{m=1}^{M_{ij}} \frac{A_{r,ij}^{p+1}(m)}{R_{r,ij}(m)} \quad (23)$$

$$C_{r,ij,2} \equiv \sum_{m=1}^{M_{ij}} A_{r,ij}^{p-1}(m) R_{r,ij}(m) \quad (24)$$

$$C_{nr,ijl,1} \equiv \sum_{m'=1}^{M'_{ijl}} \frac{A_{nr,ijl}^{p+1}(m')}{R_{r,ijl}(m')} \quad (25)$$

$$C_{nr,ijl,2} \equiv \sum_{m'=1}^{M'_{ijl}} A_{nr,ijl}^{p-1}(m') R_{r,ijl}(m') \quad (26)$$

and from (6) and (7) and based on the definitions in (8)-(9) and (22), the following WC solutions can be derived,

$$\frac{\partial D}{\partial G_{ij}} = \frac{1}{IJ}\{-\frac{1}{G_{ij}^2}\left[\frac{1}{M_{ij}}C_{r,ij,1} + \beta\frac{1}{L}\sum_{l=1}^{L}\frac{1}{M'_{ijl}}\frac{1}{H_l}C_{nr,ijl,1}\right] + \frac{1}{M_{ij}}C_{r,ij,2} + \beta\frac{1}{L}\sum_{l=1}^{L}\frac{1}{M'_{ijl}}H_l C_{nr,ijl,2}\} \quad (27)$$

$$\frac{\partial D}{\partial H_l} = \frac{1}{IJL}\{-\beta \frac{1}{H_l^2}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{1}{G_{ij}}\frac{1}{M_{ijl}^{'}}C_{nr,ijl,1} + \beta \sum_{i=1}^{I}\sum_{j=1}^{J} G_{ij}\frac{1}{M_{ijl}^{'}}C_{nr,ijl,2}\} \tag{28}$$

## 4. RESULTS AND DISCUSSION

Having extracted solutions in (15)-(16) for the WE, in (20)-(21) for the LE, and in (27)-(28) for the WC distortions, any gradient-based non-linear optimization method can be used to train the gain parameters. Here, we used the simple steepest descent together with a momentum-based learning rate adaptation (used a momentum multiplier of 0.9). Learning rates were considered to be 0.5, 1e-6 and 5e-7 for WE, LE and WC, respectively. All the common settings were chosen the same as the ones in [14] for comparison purposes. The IEEE sentences [28] were used as clean speech signals and the CIPIC HRTF dataset [29] was used to generate the HRTF-convolved reference and non-reference noise, noisy and clean speech training and testing data. The CIPIC data for 13 different azimuth angles at 0° of elevation was used for training and testing. Noise data were recorded using the BTE microphones in real environments with the PDA research platform in six commonly encountered noise environments of Street, Car, Restaurant, Mall, Bus and Train, and then were added to the clean speech signals at 5 dB SNR.

Table 1 shows the PESQ (Perceptual Evaluation of Speech Quality) scores [30] for each noise environment averaged over reference and non-reference outcomes ($\beta = 1$) and over 13 different angles. In each test case, 50 IEEE speech files (not seen during training) were used (total of 650 test samples for each environment).

Segmental SNR [31] improvements are also presented here to show how each method reduced noise levels. It can be seen that although WC provided the highest SNR improvements, it did not reach the highest quality scores except in Restaurant and Train environments. These differences were statistically significant at 99% confidence level. This implied that WC reduced noise more than the other methods but also caused removal of parts of speech, thus introducing distortions and causing speech quality loss. WE and LE did not result in significant SNR+ or PESQ score differences, but both provided higher PESQ scores in Mall and higher SNR+ in Street, Mall and Train than the direct estimation method in [14] (Dir) at 95% confidence level and in all the other environments at 99% confidence level. Using a noise environment recognition approach such as the ones in [14, 15], the best performing gain for each environment was loaded to the pipeline suppression component.

Table 1 – Segmental SNR improvements and PESQ scores for different methods of direct quasi-static gain estimation in [14] (Dir), gradient-based training based on Weighted-Euclidean (WE), Log-Euclidean (LE) and Weighted-Cosh (WC) distortion measures. Corresponding values for no suppression (N/S) are also shown for comparison.

| Noise Class | | *Segmental SNR+* | *PESQ* |
|---|---|---|---|
| *Street* | N/S | 0 | 2.13 (±0.15) |
| | Dir | **1.33** (±0.68) | 2.38 (±0.13) |
| | WE | **1.50** (±0.69) | **2.40** (±0.13) |
| | LE | **1.55** (±0.69) | **2.40** (±0.13) |
| | WC | **2.83** (±1.13) | 2.38 (±0.17) |
| *Car* | N/S | 0 | 1.99 (±0.12) |
| | Dir | 1.36 (±0.39) | 2.20 (±0.10) |
| | WE | **1.54** (±0.38) | **2.22** (±0.10) |
| | LE | **1.54** (±0.40) | **2.22** (±0.10) |
| | WC | **2.69** (±0.65) | 2.12 (±0.13) |
| *Restaurant* | N/S | 0 | 2.08 (±0.14) |
| | Dir | 0.92 (±0.41) | 2.15 (±0.12) |
| | WE | **1.12** (±0.42) | **2.18** (±0.12) |
| | LE | **1.10** (±0.42) | **2.18** (±0.12) |
| | WC | **2.84** (±0.78) | **2.23** (±0.14) |
| *Mall* | N/S | 0 | 2.07 (±0.14) |
| | Dir | **1.58** (±0.43) | **2.27** (±0.12) |
| | WE | **1.72** (±0.44) | **2.29** (±0.12) |
| | LE | **1.77** (±0.44) | **2.29** (±0.12) |
| | WC | **3.02** (±0.76) | 2.14 (±0.14) |
| *Bus* | N/S | 0 | 2.04 (±0.14) |
| | Dir | 1.66 (±0.43) | 2.34 (±0.12) |
| | WE | **1.84** (±0.42) | **2.36** (±0.11) |
| | LE | **1.84** (±0.44) | **2.36** (±0.11) |
| | WC | **4.14** (±0.72) | 2.31 (±0.17) |
| *Train* | N/S | 0 | 2.01 (±0.13) |
| | Dir | **1.79** (±0.47) | 2.31 (±0.11) |
| | WE | **1.94** (±0.45) | **2.33** (±0.11) |
| | LE | **1.97** (±0.48) | **2.32** (±0.10) |
| | WC | **4.38** (±0.60) | **2.34** (±0.13) |

## 5. RELATION TO PRIOR WORK

The unilateral environment-adaptive noise suppression pipeline developed in [15] detects the environment noise type automatically and tunes parameters of the suppression gain appropriately. The entire pipeline runs in real-time [16] on the FDA-approved PDA research platform [17]. The extension proposed in [14] for bilateral CIs provides a computationally efficient bilateral stimulation version of this pipeline. It adds minimal storage requirements to the unilateral pipeline and runs on a single processor. This work addresses a generalization of the optimization framework for single-processor bilateral enhancement that was presented in [14] by providing generalized solutions where unilateral counterparts (e.g. [15, 18, 21]) become its special cases. This work is the first attempt to provide a generalized data-driven enhancement approach for bilateral cochlear implant stimulation using a single processor.


## 6. ACKNOWLEDGEMENT

This work was partially supported by the grant no. DC010494 from the National Institutes of Health.

## 7. 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 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] P. Loizou, "Speech processing in vocoder-centric cochlear implants," *Adv. Otorhinolaryngol. 64*, pp. 109-143, 2006.

[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. Am. 128*, pp. 128-134, 2007.

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

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

[7] R. Litovsky, A. Parkinson, J. Arcaroli, R. Peters, J. Lake, P. Johnstone, and G. Yu, "Bilateral cochlear implants in adults and children," *Arch. Otolaryngol. Head Neck Surg. 130*, pp. 648-55, 2004.

[8] T. Ching, E. Van Wanrooy, and H. Dillon, "Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review," *Trends Amplification 11*, pp. 161-92, 2007.

[9] H. Kühn-Inacker, W. Shehata-Dieler, J. Müller, and J. Helms, "Bilateral cochlear implants: A way to optimize auditory perception abilities in deaf children?," *Int. J. Pediatr. Otorhinolaryngol. 68*, pp. 1257-66, 2004.

[10] R. Litovsky, P. Johnstone, and S. Godar, "Benefits of bilateral cochlear implants and/or hearing aids in children," *Int. J. Audiol. 45*, pp. S78-S91, 2006.

[11] R. Van Hoesel, and R. Tyler, "Speech perception, localization, and lateralization with bilateral cochlear implants," *J. Acoust. Soc. Am. 113*, pp. 1617-30, 2003.

[12] J. Müller, F. Schon, and J. Helms, "Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system," *Ear Hear. 23*, pp. 198-206, 2002.

[13] R. Van Hoesel, "Exploring the benefits of bilateral cochlear implants," *Audiol. Neurootol. 9*, pp. 234-46, 2004.

[14] T. Mirzahasanloo, N. Kehtarnavaz, V. Gopalakrishna, and P. Loizou, "Environment-adaptive speech enhancement for bilateral cochlear implants using a single processor," *Speech Commun*, in press, 2013. DOI: 10.1016/j.specom.2012.10.004.

[15] 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.* 6, pp. 1691-1700, 2012.

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

[17] H. Ali, A. Lobo, and P. Loizou, "On the design and evaluation of the PDA-based research platform for electric and acoustic stimulation," *Proceedings of IEEE Int. Conf. on Eng. Med. Biol.*, San Diego, Aug 2012.

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

[19] P. Loizou, "Speech enhancement based on perceptually motivated Bayesian estimators of the magnitude spectrum," *IEEE Trans. Speech Audio Process. 13*, pp. 857-869, 2005.

[20] T. Fingscheidt, S. Suhadi, and S. Stan, "Environment-optimized speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process. 16*, pp. 825–834, 2008.

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

[22] 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.*, *57*, pp. 2053–2063, 2010.

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

[24] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean square error log-spectral amplitude estimator," *IEEE Trans. Acoust.,Speech, Signal Process. 33*, pp. 443–445, 1985.

[25] Y. Ephraim, and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process. 32*, pp. 1109-1121, 1984.

[26] P. Loizou, *Speech Enhancement: Theory and Practice*, Boca Raton: Florida: CRC Press LLC, 2007.

[27] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP J. Appl. Signal Process. 26*, pp. 1-19, 2006.

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

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

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

[31] J. Hansen, and B. Pellom. "An effective quality evaluation protocol for speech enhancement algorithms," *Proceedings of Int. Conf. Spoken Language Processing 7*, pp. 2819-2822. 1998.