ROBUST AND EFFICIENT ENVIRONMENT DETECTION FOR ADAPTIVE SPEECH ENHANCEMENT IN COCHLEAR IMPLANTS

Oldooz Hazrati⋆*, Seyed Omid Sadjadi, and John H.L. Hansen*

Center for Robust Speech Systems (CRSS),
The University of Texas at Dallas, Richardson, TX 75080-3021, USA
{hazrati, sadjadi, john.hansen}@utdallas.edu

ABSTRACT

Cochlear implant (CI) recipients require alternative signal processing for speech enhancement, since the quantities needed for intelligibility and quality improvement differ significantly when direct stimulation of the basilar membrane is employed for CIs. Here, a robust feature vector is proposed for environment classification in CI devices. The feature vector is directly computed from the output of the advanced combination encoder (ACE), which is a sound coding strategy commonly used in CIs. Performance of the proposed feature vector is evaluated in the context of environment classification tasks under anechoic quiet, noisy, reverberant, and noisy reverberant conditions. Speech material taken from the IEEE corpus are used to simulate different environmental acoustic conditions with: 1) three measured room impulse responses (RIR) with distinct reverberation times (T_{60}) for generating reverberant environments, and 2) car, train, white Gaussian, multi-talker babble, and speech-shaped noise (SSN) samples for creating noisy conditions at 4 different signal-to-noise ratio (SNR) levels. We investigate 3 different classifiers for environment detection, namely Gaussian mixture models (GMM), support vector machines (SVM), and neural networks (NN). Experimental results illustrate the effectiveness of the proposed features for environment classification.

***Index Terms*—** Advanced combination encoder, cochlear implants, environment detection, noise, reverberation

1. INTRODUCTION

Although cochlear implant (CI) recipients are able to identify speech in anechoic quiet environments to a large extent, their speech identification performance drops significantly in the presence of noise and/or reverberation [1, 2]. Here, speech enhancement strategies that can alleviate the impact of noise and/or reverberation are of great interest. Several single- and multi-microphone techniques have been proposed towards suppressing adverse effects of noise and reverberation on speech which have resulted in substantial speech intelligibility gains for CI users [1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15].

Despite the effectiveness of speech enhancement strategies for improving quality and/or intelligibility of noisy, reverberant, and noisy reverberant speech for CI users, mitigating the negative effects of each of the maskers requires environment-specific treatment that is dependent on the nature of distortion (e.g., convolutive vs. additive, narrowband vs. wideband, or linear vs. non-linear interferences). Therefore, environment detection becomes an essential element to adaptively provide proper speech enhancement strategies for different listening environments.

For general speech processing, the domain of "Environmental Sniffing" has merged as a means of extracting knowledge concerning noise or the environment for the subsequent speech processing solutions [16]. In addition, knowing the type of noise/distortion can be employed to adapt the necessary noise frame update rate needed to achieve a required level performance [17].

Although a few noise and/or reverberation suppression techniques in CIs have been proposed and evaluated [1, 3, 4, 5, 7, 8, 9], the literature on noisy and/or reverberant environment detection is sparse, where most studies have focused on either room impulse response (RIR) or T_{60} estimation for a given environment (e.g. see [18, 19]). Only in a recent study [20], channel-specific models were used to detect reverberation in CI stimuli. However, the goal in [20] was to classify reverberant, noisy, and anechoic quiet environments based on two features extracted at the output of 22 frequency bands before the maxima selection in advanced combination encoder (ACE) [21]. The strategy proposed in [20] resulted in a reverberation signal detection score of 100% for T_{60} = 1.2 s, and 86% for a shorter reverberation time of T_{60} = 0.5 s. The technique was only evaluated using *simulated* RIRs for reverberation and two noise types (speech-shaped and white Gaussian noise). No attempt was made to include actual or simulated noisy reverberant environments, where both noise and reverberation exist, in classification experiments of all four types, neither did they consider other *real* noise types.

In this study, following the work presented in [20], three features are proposed for the environment classification task. These features, which are extracted from the simplified output of the maxima-selection stage in the ACE, are based on the average inter-stimuli intervals (ISI), stimulation length (SL), and stimulation energy (SE) of each frequency channel in the CI device. Gaussian mixture model (GMM), support vector machine (SVM), and neural network (NN) classifiers are trained based on these features computed in different types of acoustic environments (anechoic quiet, noisy, reverberant, and noisy reverberant). To evaluate the effectiveness of the proposed features for environment detection, speech material extracted from the IEEE corpus are used to simulate various acoustic conditions along with three RIRs with different reverberation times (T_{60} = 0.3, 0.6, and 0.8 s) [22], five noise types, namely white Gaussian noise (WGN), speech-shaped noise (SSN), multi-talker babble, car, and train, at four signal-to-noise ratio levels (SNR = -5, 0, 5, 10 dB).

2. ENVIRONMENT DETECTION

2.1. Mathematical environment model

In this study, four general environments are considered for classification:

a) Anechoic quiet, where neither additive noise, nor reverberation is present and ACE strategy performs well and there is no need

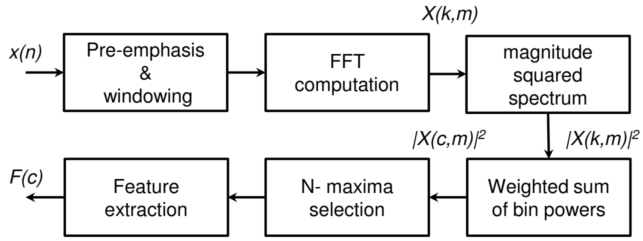

Fig. 1. Block diagram of the proposed feature extraction strategy. Here, k and c denote frequency channel indices ($k = 1, \ldots, 128$ and $c = 1, \ldots, 22$), and m denotes time-frame index.

for speech enhancement.

b) Reverberant, where the received speech signal is a delayed sum of the direct sound and its reflections in the acoustic environment. Reverberant speech can be modeled as the convolution of the RIR with the clean speech signal,

$$r(n) = \sum_{l=0}^{L-1} s(n-l)h(l) = s(n) * h(n), \tag{1}$$

where $r(n)$ and $s(n)$ are the reverberant and anechoic signals, respectively, and $h(n)$ is the RIR with the length L.

c) Noisy, where the anechoic speech signal is masked with additive (stationary or non-stationary) noise,

$$y(n) = s(n) + d(n), \tag{2}$$

where $d(n)$ and $y(n)$ are additive noise and noisy signals, respectively.

d) Noisy reverberant, in which noise (stationary or non-stationary) is added to the reverberant speech signal,

$$x(n) = (s(n) * h(n)) + d(n), \tag{3}$$

where $x(n)$ is the noisy reverberant signal. It is a common practice in engineering literature to add noise to the reverberant signal when studying their combined effects [6].

2.2. Feature extraction

A block diagram illustrating different stages of the ACE strategy is shown in Fig.1. In the ACE routine, the signal is first pre-emphasized and short-time windowed. Next, a 128-point fast Fourier transform (FFT) of each short-time frame is computed and is followed by a magnitude squaring stage. The 128-point spectrum of the signal is weighted and summed across 22 frequency bands to generate a 22 channel signal spectrum. From the 22 channel outputs, N (which is typically 8,10, or 12) maxima are selected as the stimuli and presented to the CI user. Here, three different features are extracted from the output of the N-maxima to be used for environment classification. The first feature is the stimuli energy, averaged over all frames at each frequency bin:

$$SE = \sum_{m=1}^{M} se(m)/M, \tag{4}$$

where m and M are the frame index and total number of frames, respectively, and se denotes the short-time energy of each frame.

The second feature is the average inter-stimulation interval (ISI), which is the average time during which a frequency channel is not selected by the ACE (duration of a channel being "off" between two "on"s) which is defined as,

$$ISI = \sum_{j=1}^{J} isi(j)/J, \tag{5}$$

where j and J denote the index of inter-stimulation interval and total number of short-time ISIs in a frequency bin. The short-time ISI is denoted by isi.

The third feature is the average stimulation length (SL), or the average time duration that a channel is selected by the ACE (duration of a channel being "on" between two "off"s) which is defined as,

$$SL = \sum_{k=1}^{K} sl(k)/K, \tag{6}$$

where k and K are stimulation-length index and total number of continuous stimulations in a frequency channel, respectively, and short-term stimulation length is denoted as sl.

The final feature vector for environment classification is formed by concatenating the three features described above. Note that all three features are computed at the output of ACE which is already implemented in CI devices. Hence, no additional computational load is introduced for extracting features, making them attractive for real-time processing in CI platforms. Log-normal distributions are fit to the histograms of the features and are shown in Fig. 2. The normalized distributions illustrate the discrimination power of the features for environment classification in: anechoic quiet, reverberant, noisy, and noisy reverberant conditions. All sentences from the IEEE database [23] are used for generating the histograms. As can be seen from the histogram for the averaged SE feature (left), the averaged energy concentration of the maxima channels selected by the ACE increases from anechoic quiet (solid red) condition to the noisy reverberant condition (dot-dashed brown). This is expected as reverberation smears the speech energy over time, hence the energy distribution for reverberant speech has a higher variance. In the additive noise condition, as noise energy adds to the speech energy masking of weaker consonants, the energy variances become smaller and average energy increases. In the noisy reverberant condition, the energy distribution has a structure between reverberant alone and noise alone distributions, with the energy concentration at a larger value due to the combined effects of additive noise and convolutive reverberation. A Similar trend is observed for averaged ISI and SE features, where a wider distribution is seen for reverberant-alone compared to other conditions. As noise and reverberation fill in the gaps of the speech spectrum in a complementary fashion, the averaged ISI is concentrated around a smaller value compared to the reverberant condition. However, the similarity between noise-alone and noisy reverberant ISI and SL distributions is due to averaging over different frequency bins.

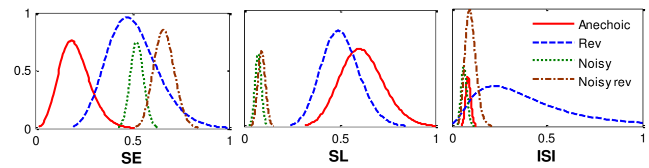

Fig. 2. Normalized distributions of histograms of averaged SE, ISI, and SL features for IEEE database (720 sentences) in anechoic quiet, reverberant (T_{60} = 0.6 s), noisy (SNR= -5 dB, babble noise), and noisy reverberant (T_{60} = 0.6 s and SNR = -5 dB, babble noise)

2.3. Classifier training

2.3.1. Support vector machines (SVM)

A SVM is a discriminative binary classifier that divides a d-dimensional real space into two half spaces with the largest margin by constructing an optimal separating hyperplane (OSH) [24]. Binary classification is the task of classifying the members of a given observation sequence into two groups on the basis of whether they have the same property or not. A separating hyperplane (also called discriminant function), divides the dataset such that all points with the same class label are on the same side of the hyperplane.

A linear decision function works well when the decision boundary between the two classes is linear. However, the training set is not always linearly separable. To achieve better generalization performance, the input data can be first mapped into a high-dimensional feature space, where the decision boundary is linear, using kernel functions such as a polynomial or Gaussian. The OSH is then constructed in the feature space. We employ a one-versus-rest strategy to perform multi-class classification with SVMs.

2.3.2. Gaussian mixture models (GMM)

A GMM is a parametric probability density function described as a weighted sum of Gaussian component densities. It is usually efficient in representing a large class of sample distributions, and is represented by three hyper-parameters: i) mean vectors, ii) covariance matrices, and iii) mixture weights of all its component densities. These hyper-parameters are estimated from training data using an iterative Expectation-Maximization (EM) algorithm [25]. We train a GMM per environment class and use a maximum likelihood criterion for classification.

2.3.3. Neural networks (NN)

NN classifiers are artificial network with processing units (neurons) that operate in parallel [26]. NNs are developed to mimic the function of neuro biological networks. NNs learn complex mappings between inputs and output and are specifically useful when the underlying statistics of the classification task are not properly understood.

In a multi-layer network, the back-propagation algorithm learns the neuron weights using gradient descend in order to minimize the squared error between the estimated output and the target outputs of the network. We train a three-layer feed-forward structure using sigmoid activation functions for the input and hidden layer, and a linear activation function for the output layer.

3. EXPERIMENTS

Performance of the proposed feature vector, which is formed by concatenating the three features described in Section 2.2, is evaluated in the context of environment detection tasks using GMM, SVM, and NN classifiers. In this study, we consider four types of environments for classification: clean, noisy, reverberant, and noisy reverberant, and report confusion matrices and classification accuracies as performance metrics. In addition, performance of the proposed feature vector is also assessed in noisy environment classification.

For classification experiments, training and test speech material are obtained from the IEEE database [23]. The IEEE speech corpus contains 72 lists of 10 phonetically balanced sentences where each sentence has 7-12 words produced by a male speaker [27]. For each classification task, half the corpus (360 sentences) is used for classifier training and the remaining half (360 sentences) is used for tests. For the environment classification task (classification of 4 types of environments), 360 training sentences are divided into four groups of 90 sentences per environment. The other half of the database (360 sentences) are also divided into four 90-sentence groups for subsequent tests. For the noise classification task, sixty sentences are used to train the classifiers for each noise type (SSN, babble, car, train, WGN). The other 360 IEEE sentences are used for evaluations (divided equally between noise types).

Multi-talker babble, SSN, train, car, and WGN are added to anechoic clean speech signals at -5, 0, 5, and 10 SNR levels to generate the noisy stimuli. Multi-talker babble, train, and car noise samples are extracted from real noise recordings (using a laptop and a MC391 microphone) sampled at 44 kHz and down-sampled to 16 kHz for this study. Train and car noise samples are recorded from outside of a train and a car cruising at a speed of 60 mph, respectively [27].

The reverberant stimuli are generated by convolving the clean signals with real RIRs recorded in a 10.06 m $\times$ 6.65 m $\times$ 3.4 m (length $\times$ width $\times$ height) room [22]. The reverberation time of the room is varied from 0.3, 0.6, and 0.8 s by adding absorptive panels to the walls and floor carpeting. The direct-to-reverberant ratios (DRR) of the RIRs are 1.5, -1.8, and -3.0 corresponding to T_{60} = 0.3, 0.6, and 0.8 s, respectively. The distance between the single-source signal and the microphone is 5.5 m, which is beyond the critical distance. Noisy reverberant stimuli are generated by adding noise to the reverberant signal for all combinations of 5 noise types, four SNR levels, and three reverberation times. For the noisy reverberant stimuli, the reverberant signal served as the target signal in the SNR computation.

For implementation of the ACE strategy, general CI parameters such as 800 pulses per second (pps), 22 frequency bins, and 8 maxima (out of 22) channels are used.

The SVM classifier is trained with a Gaussian RBF kernel. A 5-fold cross validation is used with the training data to determine the SVM parameters. As noted earlier, we employ a one-versus-rest strategy to perform multi-class classification with SVMs. Two, four, eight, and sixteen-mixture GMMs are trained and tested for the classification task. However, we only report the results with 4-mixture GMM classifiers because they performed the best in our experiments. For the NN classifier, a feed-forward multi-layer perception structure with one hidden layer (with 7, 7, and 1 nodes in the layers) with sigmoids, and linear activation functions is used (1000 iterations). The weights are initiated with small random values (between -0.5 and 0.5).

4. RESULTS

The environment classification results obtained with GMM, SVM, and NN classifiers are shown as confusion matrices in Table 1. Each row represents a specific environment type, where the diagonal elements present the correct classification score for each condition. As seen from the table, clean, reverberant, noisy, and noisy reverberant environments are accurately classified 97.79%, 97.14%, and 95.13% of time using SVM, GMM, and NN classifiers, respectively. It is clear that the proposed feature vector can yield high classification performance with the four types of clean, reverberant, noisy, and noisy reverberant environments. This is due to the robustness of ISI, SL, and SE of the N-maxima channels selected from the ACE speech coding strategy implemented in many CI devices. For all three classifiers, the best scores are obtained for the noisy reverberant environment (99.84%, 99.31%, and 98.73% for SVM, GMM,

Table 1. Environment classification confusion matrices for SVM, GMM, and NN classifiers.

Clean	**96.11**	1.19	2.70	0.00
Rev	1.30	**97.86**	0.64	0.20
Noisy	2.17	0.29	**97.33**	0.21
Noisy Rev	0.00	0.16	0.00	**99.84**

(a) Results on SVM classifiers: overall avg. = 97.79%

Clean	**94.70**	1.66	3.64	0.00
Rev	1.67	**97.37**	0.85	0.12
Noisy	2.43	0.41	**97.16**	0.00
Noisy Rev	0.03	0.27	0.40	**99.31**

(b) Results on GMM classifiers: overall avg. = 97.14%

Clean	**92.92**	4.17	2.48	0.43
Rev	2.14	**95.65**	1.65	0.56
Noisy	2.09	2.90	**93.23**	1.78
Noisy Rev	0.09	0.30	0.88	**98.73**

(c) Results on NN classifiers: overall avg. = 95.13%

and NN classifiers, respectively).

The environment classification results obtained using the NN classifier are slightly inferior to those obtained with SVM and GMM classifiers because of two main reasons. Training NNs requires large amounts of training data. However, here only 90 sentences with an average of 3 seconds duration are used for NN training. Moreover, due to computational complexity as well as limited amount of training data, the number of hidden layers and nodes are limited which consequently restricts the effectiveness of NN solution.

Although the main goal of the proposed feature vector is environment detection (clean, noisy, reverberant, and noisy reverberant), in another experiment the robustness of the proposed feature vector is evaluated in noisy environment classification. In this case, classifiers are trained only using noisy data with different noise types (as well as clean data). The results of noisy environment classification are presented in Table 2. As results suggest in the table, clean, babble, SSN, car, WGN, and train noisy environments are accurately classified 95.08%, 93.40%, and 82.76% of the time using SVM, GMM, and NN classifiers, respectively. The results indicate that ISI, SL, and SE features can not only discriminate among different types of distortions introduced to the signal (no distortion, convolutive, additive, convolutive+additive), but also can provide useful information on noise types (stationary and non-stationary) quite accurately.

As seen from Table 2, the best noise classification performance is obtained for WGN and car noise, and the worst performance is obtained for train noise. Moreover, compared to the four-environment (clean, reverberant, noisy, and noisy reverberant) detection task, less training and test data is used for this experiment (60 sentences per each noise type, approximately 180 s), which results in lower classification scores, that is especially evident for the results obtained with the NN classifier.

5. CONCLUSION

This study has proposed a feature vector based on the ACE (used in cochlear implants (CI)) for environment detection and classification in clean, noisy, reverberant, and noisy reverberant conditions. The feature vector was formed by concatenating the inter-stimulus interval (ISI), stimulation length (SL), and stimulation energy (SE) features obtained from the output of the N-maxima selection stage of the ACE speech coding strategy in CI devices. The performance of the proposed features in capturing environment-specific characteristics was assessed using GMM, SVM, and NN classifiers trained and tested on sentences from the IEEE database in anechoic quiet, moderate to relatively large reverberation times (T_{60} = 0.3, 0.6, and 0.8s), different noisy conditions (train, SSN, WGN, multi-talker babble, and car) at four signal-to-noise ratio (SNR) levels (-5, 0, 5, and 10 dB), and sixty noisy reverberant environments (all combinations of T_{60}s, noise types and SNR levels). All four environments were classified with an accuracy as high as 97% using SVM and GMM classifiers, and 95% using NN classifiers. The high accuracy in environment classification was due to the efficacy of the proposed feature vector in capturing characteristics of different environments and consequently training robust classifiers. Environment classification is of great importance in the field of speech enhancement, especially for CIs where there exist several speech enhancement algorithms that function well in specific scenarios (e.g., under additive noise), but degrade the quality/intelligibility of speech in other environments (e.g., reverberant or noisy reverberant). Incorporating each speech enhancement strategy in the environment for which it is designed for will improve overall performance as well as avoid unnecessary battery usage and excessive signal processing in the CI devices.

6. ACKNOWLEDGMENT

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

Table 2. 6-way noise classification confusion matrices for SVM, GMM, and NN classifiers.

Clean	**94.32**	0.50	3.69	0.73	0.00	0.76
BABB	3.33	**92.02**	2.58	0.00	0.00	2.07
SSN	1.57	2.67	**95.18**	0.00	0.00	0.57
CAR	1.15	0.00	0.00	**98.85**	0.00	0.00
WGN	0.00	0.00	0.00	0.00	**99.82**	0.18
TRAIN	4.13	3.19	1.69	0	0.73	**90.26**

(a) Results on SVM classifiers: overall avg. = 95.08%

Clean	**91.68**	0.72	3.75	1.29	0.00	2.55
BABB	3.24	**87.89**	3.09	0.00	0.00	5.78
SSN	1.76	2.95	**93.67**	0.00	0.00	1.62
CAR	1.01	0.00	0.00	**98.72**	0.00	0.28
WGN	0.00	0.00	0.00	0.00	**98.28**	1.72
TRAIN	3.33	2.20	1.45	0.00	2.88	**90.14**

(b) Results on GMM classifiers: overall avg. = 93.40%

Clean	**77.12**	15.72	4.47	1.40	0.53	0.75
BABB	4.90	**79.42**	11.29	1.51	1.00	1.88
SSN	1.29	14.35	**80.18**	2.73	0.65	0.80
CAR	1.52	0.66	2.44	**92.25**	2.67	0.46
WGN	0.32	0.17	0.36	1.44	**94.90**	2.82
TRAIN	2.41	3.82	4.73	4.20	12.15	**72.69**

(c) Results on NN classifiers: overall avg. = 82.76%

7. REFERENCES

[1] K. Kokkinakis, O. Hazrati, and P. C. Loizou, "A channel-selection criterion for suppressing reverberation in cochlear implants," *J. Acoust. Soc. Am.*, vol. 129, pp. 3221–3232, May 2011.

[2] O. Hazrati and P. C. Loizou, "The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners," *Int. J. Audiol.*, vol. 51, pp. 437–443, June 2012.

[3] Y. Hu, P. C. Loizou, N. Li, and K. Kasturi, "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," *J. Acoust. Soc. Am.*, vol. 122, pp. EL128–EL134, September 2007.

[4] Y. Hu and P. C. Loizou, "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users," *J. Acoust. Soc. Am.*, vol. 127, pp. 3689–3695, June 2010.

[5] ——, "A new sound coding strategy for suppressing noise in cochlear implants," *J. Acoust. Soc. Am.*, vol. 124, pp. 498–509, July 2008.

[6] P. A. Naylor and N. D. M. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[7] O. Hazrati and P. C. Loizou, "Reverberation suppression in cochlear implants using a blind channel-selection strategy," *J. Acoust. Soc. Am.*, vol. 133, pp. 4188–4196, June 2013.

[8] O. Hazrati, S. O. Sadjadi, P. C. Loizou, and J. H. L. Hansen, "Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy," *J. Acoust. Soc. Am.*, vol. 134, November 2013.

[9] O. Hazrati, J. Lee, and P. C. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *J. Acoust. Soc. Am.*, vol. 133, pp. 1607–1614, March 2013.

[10] P. Jafari, H. Y. Kang, X. Wang, Q. Fu, and H. Jiang, "Phase-sensitive speech enhancement for cochlear implant processing," in *Proc. IEEE ICASSP*. May, 2011, pp. 5104–5107.

[11] P. Dawson, S. Mauger, and A. Hersbach, "Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus cochlear implant recipients," *Ear & Hear.*, vol. 32, pp. 382–390, May/June 2011.

[12] A. Hersbach, S. Mauger, D. Grayden, J. Fallon, and H. McDermott, "Algorithms to improve listening in noise for cochlear implant users," in *Proc. IEEE ICASSP*. May, 2013, pp. 428–432.

[13] E. Healy, S. Yoho, Y. Wang, and D. Wang, "An algorithm to improve speech recognition in noise for hearing-impaired listeners," *J. Acoust. Soc. Am.*, vol. 134, pp. 3029–3038, October 2013.

[14] K. Nie, G. Stickney, and F.-G. Zeng, "Encoding frequency modulation to improve cochlear implant performance in noise," *IEEE Trans. Biomed. Eng.*, vol. 52, pp. 64–73, January 2005.

[15] A. Bhattacharya, A. Vandali, and F.-G. Zeng, "Combined spectral and temporal enhancement to improve cochlear-implant speech perception," *J. Acoust. Soc. Am.*, vol. 130, pp. 2951–2960, November 2011.

[16] M. Akbacak and J. H. L. Hansen, "Environmental sniffing: Noise knowledge estimation for robust speech systems," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, pp. 465–477, February 2007.

[17] N. Krishnamurthy and J. H. L. Hansen, "Environmental dependent noise tracking for speech enhancement," *Int. J. Speech Tech.*, vol. 16, pp. 303–312, September 2013.

[18] J. Wen, E. Habets, and P. Naylor, "Blind estimation of reverberation time based on the distribution of signal decay rates," in *Proc. IEEE ICASSP*. IEEE, 2008, pp. 329–332.

[19] Y. Lin and D. Lee, "Bayesian regularization and nonnegative deconvolution for room impulse response estimation," *IEEE Trans. Signal Process.*, vol. 54, pp. 839–847, March 2006.

[20] J. Desmond, L. Collins, and C. Throckmorton, "Using channel-specific statistical models to detect reverberation in cochlear implant stimuli," *J. Acoust. Soc. Am.*, vol. 134, pp. 1112–1120, August 2013.

[21] A. Vandali, L. Whitford, K. Plant, and G. Clark, "Speech perception as a function of electrical stimulation rate: Using the nucleus 24 cochlear implant system," *Ear Hear.*, vol. 21, pp. 608–624, December 2000.

[22] A. Neuman, M. Wroblewski, J. Hajicek, and A. Rubinstein, "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," *Ear Hear.*, vol. 31, pp. 336–344, June 2010.

[23] IEEE, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 225–246, 1969.

[24] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[25] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. Speech Audio Process.*, vol. 3, pp. 72–83, January 1995.

[26] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Prentice-Hall, 1999.

[27] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. CRC Press, 2013.