IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.Volume 11, Issue 3, Ver. II (May-Jun .2016), PP 57-64
www.iosrjournals.org

Hybrid Masking Algorithm for Universal Hearing Aid System

H.Hensiba[1], Mrs. V. P. BIRLA[2]
[1] *Pg Student , Applied Electronics , C.S.I Institute Of Technology*
[2]*Assistant Professor, Department of Electronics and Communication, C.S.I Institute Of Technology*

Abstract: *Technological advances witnessed in cochlear implant (CI) devices, most CI users can now achieve reliable speech intelligibility in controlled quiet scenarios, particularly in predictable conversations. Environmental distortions, such as reverberation and additive noise, on the other hand, are known to significantly degrade speech intelligibility. This paper validates a novel approach to predict speech intelligibility for Cochlear Implant users (CIs) in reverberant environments. First, energy thresholding is proposed to reduce the variability caused by the differences in modulation spectral representations for different phonemes and speakers, as well as speech enhancement algorithm artifacts. Second, a narrower range of Fourier domain based echo Filter is employed to reduce fundamental frequency effects. Results indicate substantial improvements in intelligibility over that attained by human listeners with unprocessed stimuli. The findings from this study show that algorithm can estimate reliably the SNR and can improve speech intelligibility.*

Keywords: *Cochlear Implants(CIs),Noise Reduction, Speech Enhancement, Wiener filter(WF),SNR.*

I. Introduction

UNDERSTANDING speech in adverse listening conditions with auditory prostheses like hearing aids or CIs is a very which try to remove as much noise as possible from the mixture of the target speech and the interfering sound, with the objective of increasing SI and/or improving the speech quality of the processed signal. A typical constraint of such strategies is that distortions of the target signal should be avoided.

Usually, noise reduction algorithms operate upon a time frequency representation of the input (noisy) signal, applying a gain to each time frequency point to suppress the noise. The pattern of the gain function over all time frequency points is often called mask. Most of these time frequency domain approaches derive their gains as a function of the short-term signal-to-noise ratio (SNR) in the respective time frequency point. There is an ongoing discussion about the choice of the perfect gain function that improves SI and speech quality in NH listeners very popular choice is the so-called binary mask (BM).

The mask is motivated by the auditory masking phenomenon and preserves with its binary values time frequency points, where the target is dominant (i.e., the short- term SNR is above a threshold). The BM exploits the spar sity and disjointness of the target and interferer spectra. When a priori knowledge of the signal and noise spectra is used in the derivation of the mask, the mask is often called IBM. Under certain listening conditions, approaches based on BMs with and without a priori knowledge for the mask computation can increase SI in NH , and hearing impaired listeners .

In contrast to the hard-decision approach of the BM, state-of- the-art noise reduction algorithms derive mask .Such algorithms demonstrate improved speech quality as compared to BM processed output . A popular representative of this class of algorithms is the Wiener filter (WF), which was shown to be very promising in terms of quality improvement. When a priori knowledge was used to calculate the gain function of the WF, the approach is referred to as IWF. It was shown in that the IWF restored perfect intelligibility with a Bark-scale frequency resolution even at very low SNRs in both multi talker babble noise and interfering talker scenarios. This was in stark contrast to the performance of the IBM, which yielded intelligibility scores of around 60% at the low SNRs. In this study, the potential of the IWF and the IBM approaches in terms of SI and speech quality is investigated with regard to its application in CIs.

The tests are carried out on two groups of participants: a group of NH subjects listening to noise vocoder simulations a same model of CI processing of the processed signals, and a group of CI users. Because of the relative ease with which NH volunteers may cruited, testson NHlisteners, presented with noise vocoded versions of the processed signals comparable to CI processing, are often used in a first step to evaluate speech enhancement strategies for application in CIs. In our case, the inclusion of such tests with noise vocoded sounds (also referred to as CI simulations) allows us to investigate if the intelligibility scores obtained on NH listeners translate to the scores of CI users as well.

The aims of this study are the following: we wish to vestigate in terms of SI, which mask pattern is more beneficial for NH subjects listening to noise vocoder CI simulations and for CI users. It is interesting in particular for CI users, because noise reduction approaches based on time frequency masks can be added to the signal processing chain of existing clinical coding strategies without significant effect on other stages.

Furthermore, we study the influence of estimation errors on the SI for both groups of listeners, as it was shown in that CI users are less sensitive to speech distortions. The design of the study allows us to investigate if the SI results ob tained with NH listeners using CI simulations can be translated to that of CI users. Additionally, we want to study the potential for speech quality improvement of both mask patterns in CI users.

II. Signal Processing

In Cochlear ,Ltd.,upto N=22 envelopes are extracted in the frequency range up to 8 kHz. Therefore, such CIs usually operate with a frequency resolution that is close to the Bark- scale spectrum used in. The signal model and processing used in this study are very similar to that in purposes of completeness, we present these briefly below.

A. Signal Model

Denote the time discrete signal recorded by the microphone as y(t), where t is the sample index. The signal y(t) consists of the target signal s(t) and the additive interference v(t). This additive signal model for the recorded signal can be written as

$$y(t)=s(t)+v(t). \quad (1)$$

Due to the fact that the IBM and IWF speech enhancement approaches operate in the frequency domain, the short time frequency representation of the signal in (1) can be written, with the frame index n and the frequency index k, as

$$Y(n,k)=S(n,k)+V(n,k). \quad (2)$$

Y (n,k) is the microphone signal in the time frequency domain, S (n,k) and V (n,k) represent the target signal and the interferer, respectively.

The estimate $\hat{}$ S(n,k) of the target signal is obtained by applying the time frequency mask G(n,k) $\in$ [0,1] yielded by the IBM and/or IWF approach, to Y (n,k). Thus, the output of the speech enhancement step can be written as $\hat{}$

$$S(n,k)=G(n,k)Y(n,k). \quad (3)$$

Both the IBM and the IWF approaches derive their respective masks as a function of the short-term SNR ξ (n,k), which is defined as the ratio between the power spectral density (PSD) of the target signal ΦSS (n,k) and the PSD of the interferer ΦVV (n,k)

$$\xi(n,k)=\Phi SS(n,k)\ \Phi VV(n,k). \quad (4)$$

Usually, the PSD of the target signal and the interfering sound are computed by using the Welch method, one implementation which is a first order recursive smoothing of the respective periodograms. Since we deal with ideal estimates of the parameters of the IBM and the IWF approach, we can approximate the PSD with

$$\Phi SS(n,k)=|S(n,k)|2 \quad (5)$$
$$\Phi VV(n,k)=|V(n,k)|2. \quad (6)$$

1) Ideal Binary Mask: The IBM GIBM consists of binary weights. GIBM is equal to 1 when the SNR is above a threshold value, and 0 when the SNR is lower than this threshold. In this study, the threshold used was the global input SNR ξin. The BM GIBM can be written as

$$GIBM(n,k) = \begin{cases} 1, & \text{if } \xi(n,k) \geq \xi in \\ 0, & \text{else.} \end{cases} \quad (7)$$

Note that for a given combination of s(t) and v(t), the mask pattern is constant and independent of the SNR. This is termed as the local threshold . The binary gain function GIBM is applied to the input signal Y to obtain the enhanced output $\hat{}$ SIBM $\hat{}$

$$SIBM(n,k) = GIBM(n,k)Y(n,k). \quad (8)$$

2) Ideal Wiener Filter: The gain function GIWF of the WF approach is a continuous value between 0 and 1. It is obtained as the minimum mean-squared error estimate of the complex spectral amplitude

$$\min E\{|S(n,k)- \hat{} S(n,k)|2\} \quad (9)$$

and can be written as

$$GIWF(n,k) = \xi(n,k)\ 1+\xi(n,k) \qquad (10)$$

The corresponding estimate ˆ SIWF may then be written as ˆ

$$SIWF(n,k)\ GIWF(n,k)Y(n,k) \qquad (11)$$

3) Simulation of Estimation Errors: To investigate the influence of estimation errors in the mask patterns of WF and the BM on SI, such errors in the mask pattern were simulated. Due to the fact that over and under estimation errors influence SI differently we use the approach first described is to generate a balanced pattern of estimation errors. For the mask derivation, the spectra of the target and the noise signal were corrupted with an additional noise term which can be written as ˜

$$S(n,k) = S(n,k)+S(k) \qquad (12)$$
$$V(n,k) = V(n,k)+V(k) \qquad (13)$$

where S (k) and V (k) are complex randomly distributed variables with zero mean and power equal to the respective clean signal in the frequency band k. The corrupted spectra influence the short-term SNR estimation in and, thereby, the mask computation. This results in corrupted mask patterns GBM and GWF. When referring to results and patterns obtained in the condition with perturbed estimates, the masks are called BM and WF. The output signals for the BM and the WF mask are ˆ

$$SBM(n,k)=GBM(n,k)Y(n,k) \qquad (14)$$
$$SWF(n,k)=GWF(n,k)Y(n,k) \qquad (15)$$

The corrupted parameter estimates are only used for the mask pattern estimates. The corrupted masks are applied to the original, unperturbed mixture in (14) and (15). This manner of simulating estimation errors allows for both under and over estimation of the instantaneous PSD estimate. Additionally, such a perturbation of the underlying spectrum has the advantage that it does not preserve the silence periods of the speech and/or interference. Thus, the musical-noise phenomenon will be present in such speech/interference pauses [26]. This lends realism to the simulation.

B. General Processing Steps

The processing steps to generate the stimuli that are presented to the listener are shown in Fig. 1. The first six processing steps are the same for both groups of listeners. The processing steps that are different between NH listeners and CI users are represented by the black trace for the noise band vocoder CI simulation with NH listeners and by the dashed gray trace for the electrical stimulation with CI users.

In the first step, the target signal s(t) and the interfering signal v(t) (sampled at 16 kHz) are filtered with a pre emphasis filter that consists of the frequency response of the SP12 microphone of the Freedom speech processor of Cochlear, Ltd. The result of this pre emphasis is a boost of the higher frequencies. A square-root Hann window was used as the analysis window. The envelope extraction is done by grouping the magnitude-squared DFT coefficients into N frequency bands. This process is applied to the target and the interfering signal to calculate the power spectral density (PSD) estimates used for the gain computation in (7) and (10).

C. Noise Band Vocoder as CI Simulation

The number of channels used for the noise band vocoder CI simulation was set to N =8, because asymptotic SI performance for most CI users is reached with the current clinical speech processing strategies with eight effective channels [27]. The cut off frequencies to obtain the band pass filtered envelopes are 187.5 , 437.5 7937.5 Hz. These cutoff values for the band pass filters correspond to bandwidths of 250 , 250 , 375 , 500 , 750 , 1125 , 1750 , and 2750 Hz for the eight channels. The signal components under 187.5 Hz are not considered in the signal processing. Finally, all noise vocoded channels are added to obtain the final audio stimulus that can be presented acoustically to the NH listener

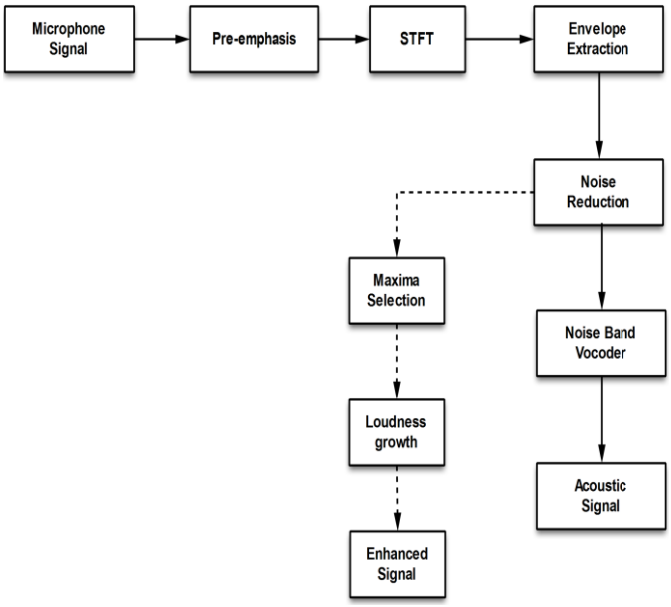
.

D. Cochlear Implants

The current clinical CI device of Cochlear, Ltd., can stimulate 22 channels. Therefore for most patients, N = 22frequency bands are processed in the envelope extraction stage. In this study,all six patients used a frequency resolution of 22 channels. The advanced combination encoder (ACE) strategy that is the default speech processing strategy in CIs of Cochlear, Ltd., does not stimulate all available channels in each time frame.

The ACE strategy consists of a maxima selection stage in each frame, where the M<N channels with envelopes of the highest amplitude in the respective frame are selected and only these channels are stimulated. In clinical practice, the number M of selected maxima varies between 7 and 12. In this study, all patients used M =8 design must contain these enable conditions in order to use and benefit from clock gating. This clock gating process can also save significant die area as well as power, since it removes large numbers o replaces them with clock gating logic. This clock gating logic is generally in the form of "Integrated clock gating" (ICG) cells. However, note that the clock gating logic will change the clock tree s

Fig. 1. Processing chain for experiments with a noise band vocoder in the top line for NH listeners (black trace) and the bottom line for CI listeners (dashed gray trace)



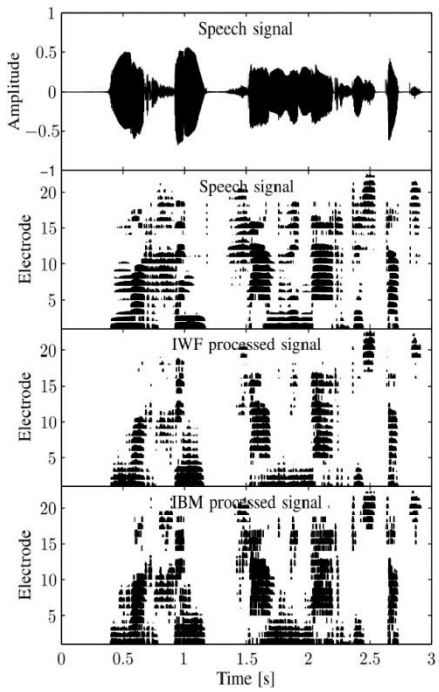


Fig: 2 Example of the speech signal

III. Methods

A. Subjects

The sentence recognition task with the noise vocoded CI simulation was conducted with six NH listeners. All NH subjects were between 16 and 22 years old (mean age 20.1 years) and had hearing thresholds below 20 dB hearing level (HL), on both ears, for the octave frequencies between 125 and8000 Hz. They were not paid for their travel expenses. The second group of listeners consisted of six CI users. The group of CI users was paid for their participation. All subjects were Dutch speaking adults. They signed an informed consent form before the tests were conducted.

B. Test Material

The Leuven Intelligibility Sentence Test (LIST) sentences were selected as the target speech material. These are Dutch/Flemish sentences spoken by a male and a female speaker. The female LIST sentence material consists of 35 lists of ten Dutch/Flemish sentences. There are 38 lists of ten Dutch/Flemish sentences available in the male LIST sentence material of which the vocabulary of 20 lists is different from the female lists. Only these non-overlapping lists were used during the experiments. Each sentence for the male and the female speaker consists of four to eight words. Keywords are marked in each sentence which results in 32 to 33 keywords per list.

Each list is balanced according to the phonetic distribution of conversational speech. The scoring was done on sentence level, where a sentence was marked as correctly recognized if all keywords of the respective sentence were repeated correctly. In the speech recognition task in multi talker babble noise, the female LIST sentences were used as the target speech material and the Auditec multitalker babble noise (from the CD Auditory Tests(Revised),Auditec, St. Louis,MO,USA) was used as the interfering sound. In the speech in speech scenario, the male LIST sentences served as the target speech material, while the female lists were the interfering speaker. The lists that did not serve before as the target speech material in the sentence recognition task in multi talker babble noise were used as the interferer.

C. Procedure

The SI improvement was evaluated in two sentence recognition tasks: a sentence recognition task when the mask patterns were derived with ideal parameter estimates, and a sentence recognition task with simulated estimation errors in the mask calculation. Both groups of listeners participated in these tasks. A quality rating was performed to as the potential for speech quality improvement by the IWF and the IBM with CI users. All listening tasks were conducted in a test-retest design of the study with a break of at least one week between the two sessions, each session lasting around 90 min. In all listening tasks, the processed signals for all algorithms were rescaled to the same presentation level of the clean target speech. This was done to prevent audibility issues of the pro- cessed signals at low SNRs. The presentation level was set to 65 dB sound pressure level. After the processing steps described previously, the signals were presented monaurally (left ear) using Sennheiser HDA200 headphones, in a sound-proof booth, for the NH participants. With CI users, direct stimulation of the channels of the CI was done with the L34 research processor.

1) Speech Intelligibility: In this first sentence recognition task, the signals were mixed at SNRs of 0 dB to −20 dB or −25 dB in steps of 5 dB for the noice vocoded speech and the CI users, respectively. The mask patterns for the IWF and IBM were constructed with ideal parameter knowledge. After the processing steps described previously, the signals were presented monaurally (left ear) using Sennheiser HDA200 headphones, in a sound-proof booth, for the NH participants.

2)Robustness to Estimation Errors: In this second task, both groups of listeners were tested at SNRs of 0 and −5 dB for the WF and the BM processed signals. The interfering signal was multi talker babble noise. The SNRs chosen in this study represent more realistic listening environments as compared to the first task.

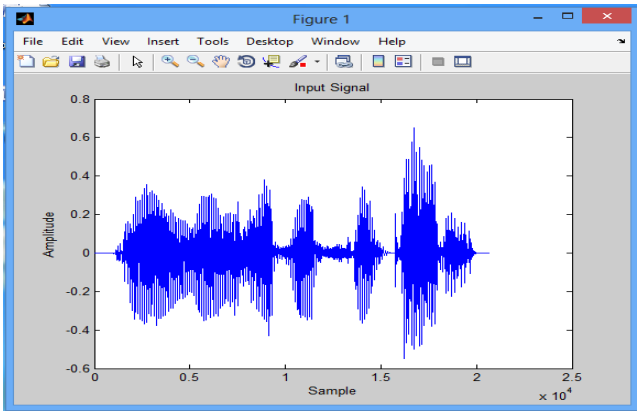
3) Preference Rating: We also investigated the speech quality of the IWF and the IBM approaches. For this, a preference rating was done only with the CI users. Preference is often used as a model of quality .Therefore it is assumed that a preference in a pair wise comparison correlates with a quality advantage of the preferred signal. For the quality improvement, the same procedure was used. This consists of a two-stage pair wise preference rating test. The pair wise comparison was administered across

1) clean speech and IWF processed output;
2) clean speech and IBM processed output;
3) IWF processed output and IBM processed output.

Clean speech, as implied here, corresponds to the clean tar- get sentence without any interfering signal present and without being subject to any noise reduction. Two noise conditions were tested (speech-in-speech, speech-in-babble) at an SNR of 0 dB . First, the signals of the respective pair wise comparison were played one after the other and the subject had to make a decision on which one was preferred in terms of quality.

IV. Results

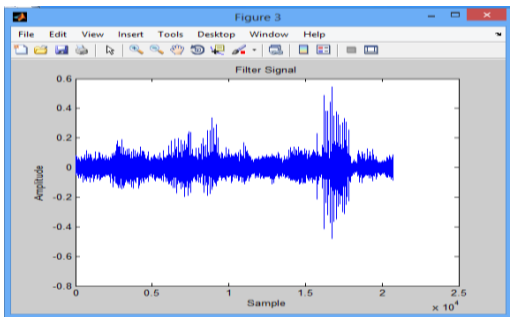
Input signal:



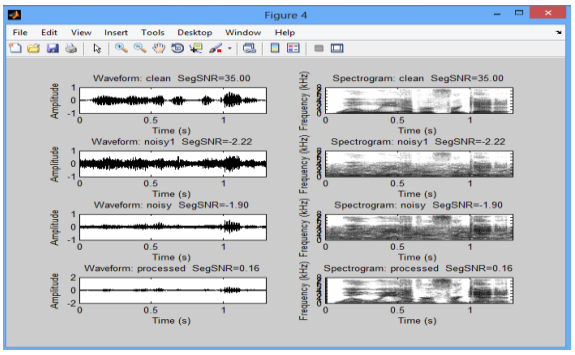
The input signal to select the normal speech in audio signal.It is using to normal speech processing at the signal module compaired to audio speech.

Filter signal:

Add the babble noise and filter to the normal speech.In non hearing person get the normal speech signal to add the sum noise and clearing to the audio.And filtering to the noise to send the clear speech in CI users.

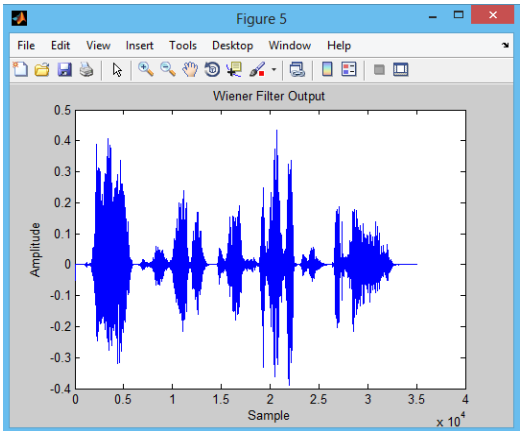


Speech output:

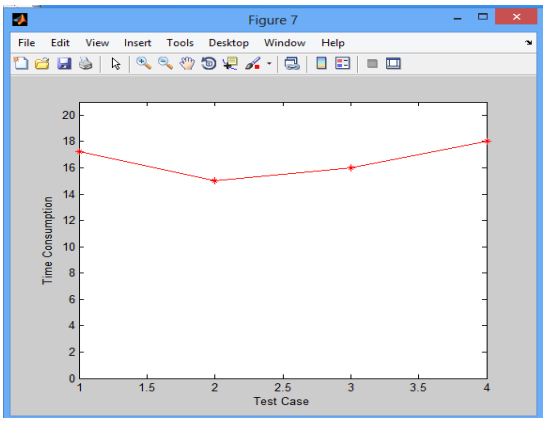


Speech output that compared to Ideal Binary Mask and Ideal Wiener Filter.IBM compared to binary weights in amplitude value.IWF compared to continuous weights in frequency value. Multi talker babble was used as the interference.

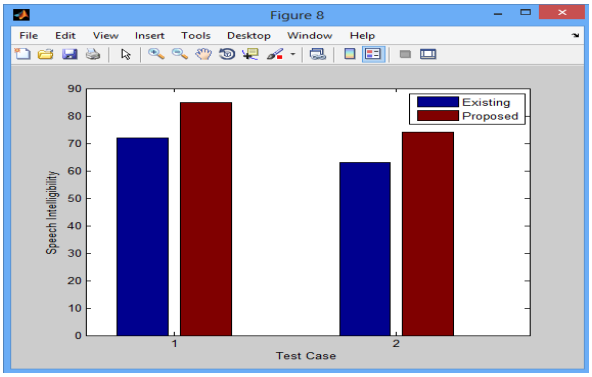
Wiener Filter Output:



Time Consumption:



Speech Intelligibility:



V. Conclusion

The investigated the potential of the IWF and the BM approaches for SI and speech quality improvement in CIs. The results of NH listeners presented with noise vocoded CI simulations are consistent and favor the soft decision approach over the BM. However, the outcomes of this study suggest that for CI users with the ACE channel selection strategy the choice between a hard and a soft decision approach is not important in terms of SI. This study also points out that the frequency resolution of the noise reduction algorithm is a less important parameter in the application of CI than it seems to be in NH listeners. The obtained results are a reference for SI evaluations of time frequency masking algorithms applied to noisy signals without a priori knowledge.

Reference

[1] D. S. Brungart, P. S. Chang, B. D. Simpson, and D. Wang, "Isolating the energetic component of speech-on-speech masking with ideal time-frequency,"J. Acoust. Soc. Amer., vol. 120, pp. 4007–4018, 2006.

[2] I. Brons, R. Houben, and W. A. Dreschler, "Perceptual effects of noise reduction by time-frequency masking of noisy speech," J. Acoust. Soc. Amer., vol. 132, no. 4, pp. 2690–2699, 2012.

[3] D. Wang, "Time-frequency masking for speech separation and its potential for hearing aid design,"Trends Amplification, vol. 12, no. 4, pp. 323–353, 2008.

[4] N. Madhu, A. Spriet, S. Jansen, R. Koning, and J. Wouters, "The potential for speech intelligibility improvement using the ideal binary mask and the ideal wiener filter in single channel noise reduction systems: Applica-tion to auditory prostheses,"IEEE Trans. Audio, Speech, Lang. Process., vol. 21, no. 1, pp. 63–72, Jan. 2013.

[5] D. Wang, "On ideal binary mask as the computational goal of auditory scene analysis," inSpeech Seperation by Humans and Machines. Norwell, MA, USA: Kluwer, 2005, pp. 181–197.

[6] D. Wang and G. J. Brown,Computational Auditory Scene Analysis: Principles, Algorithms, and Applications, 1st ed. New York, NY, USA: Wiley, 2006.

[7] M. C. Anzalone, L. Calandruccio, K. A. Doherty, and L. H. Carney, "De-termination of the potential benefit of time-frequency gain manipulation," Ear Hearing, vol. 27, no. 5, pp. 480–492, 2006.

[8] S. Cao, L. Li, and X. Wu, "Improvement of intelligibility of ideal binary-masked noisy speech by adding background noise,"J. Acoust. Soc. Amer., vol. 129, no. 4, pp. 2227–2236, 2011.