



ENHANCEMENT OF SPEECH INTELLIGIBILITY AND QUALITY IN COCHLEAR IMPLANTS USING TIME-FREQUENCY MASKING ALGORITHM

Ramya Dharshini.R¹, Senthamizh Selvi.R²

^{1,2}Dept. of Electronics and Communication Engineering,

Easwari Engineering College, Chennai, (India)

ABSTRACT

In detrimental conditions, the cochlear implants used by the hearing impaired listeners suffer from a decrease in speech intelligibility (SI). In an attempt to improve the SI, the Time-Frequency masks are used to perform noise suppression. Ideal Binary Mask (IBM) with its binary weights and Ideal Wiener Filter (IWF) with its continuous weights masks are used to enhance the speech signal. It is not known which of the masks has the highest potential for CI users in terms of SI and speech quality. The comparisons of both these filters are performed, in this study. The investigations were conducted among normal-hearing (NH) subjects listening to the noise vocoder CI simulations and also in CI users on the bases of speech quality. The ideal masks estimates the potential of SI assessed in a sentence recognition task, with an interfering talker and in multitalker babble. The estimation of the simulated errors provides the robustness of the approaches. The speech quality is assessed in CI users to a preference rate. The IWF technique outperformed the IBM technique only in NH users, while no significance was obtained in CI users. The speech intelligibility is degraded due to estimated error in Cochlear implants. The IWF has high potential for speech quality than the IBM processed signals. In this study the outcome suggest that mask pattern is not critical for CI users. The outcome suggests the new approach should consider the class of listeners considered.

Keywords: *Speech Enhancement; Ideal Binary Masking (IBM); Ideal Weiner Filter (IWF); Signal to Noise Ratio (SNR); Perceptual Evaluation of Speech Quality (PESQ).*

I. INTRODUCTION

Using the auditory prostheses like hearing aids or CIs, the understanding of speech in detrimental conditions is a very challenging task. In reverberant or noisy environment, SI decreases rapidly particularly in CI users. Therefore, the noise reduction strategies is the focused development research area, which from the target speech mixed with the interfering noise removes as much noise possible. The limitation of such approach is that the target signal should be avoided from distortions.



Generally, the operation of noise reduction algorithms is based upon a time–frequency depiction of the input (noisy) signal, by applying gain at each time–frequency point to suppress the noise. The gain function pattern over the corresponding time–frequency points is called mask. Mostly, the time–frequency based domain approaches derive their gains as the function of signal-to-noise ratio (SNR) in the respective time–frequency point. The perfect gain function obtained in NH listeners, which improves SI and speech quality is the ongoing discussion [1]–[4].

The most suited approach is the so-called binary mask (BM). The mask is derived by the auditory masking phenomenon and conserves with its binary values time–frequency points in which the target is dominant (i.e. SNR is above certain threshold). The BM exploits from the target and interferer spectra, the sparsity and disjointness. In derivation of the mask, if a priori knowledge of the signal and noise spectra is used then the mask is IBM. The approaches based on BMs by certain listening conditions with and without a priori knowledge for the mask consideration also increase SI in NH [1], [5], [6] and hearing impaired listeners [4], [5], [6].

The state-of-the-art noise reduction algorithms obtain a mask with continuous gains range within 0 and 1 in proportion to the SNR, contrast to the hard-decision approach of the BM. Such algorithms exhibit improved speech quality compared to BM processed output [7]. In this class of algorithms most popular representative is the Wiener filter (WF) which is propitious in terms quality improvement [7]. If the prior knowledge was used in computation of the gain function, then referred to as WF.

The perfect intelligibility can restored in IWF, with a Bark-scale frequency resolution even at very low SNRs conditions in both multitalker babble noise and interfering talker scenarios. This was in distinct contrast to the performance of the IBM, which provided intelligibility scores at the low SNRs of around 60%.The higher speech recognition scores more than 60% were derived for low-input SNRs, in [8], [9]. The results in both studies were based on the mask resolution and word correct scores was higher [8] or by modulating the mask pattern which discard the interfering background noise, with a stationary noise masker [9]. It also implemented, WF was more robust to estimation errors than the BM strategy. The several studies insisted that soft-decision approaches in terms of quality outperform BM [2], [10], [11]. Therefore, it was hypothesized that in auditory prostheses IWF approach should be preferred over the IBM.

For CI users in monaural and bilateral applications the multichannel speech enhancement algorithms [11]-[14], [20] have been proved to be more beneficial still the further investigation are considered in single-channel noise reduction algorithms is still relevant. Therefore in CIs application the single-channel noise reduction algorithms have been implemented, provided a lead to increase in the SI [21], [22].

In this study with regard to its application in CIs, the potential of SI and speech quality is investigated in the IBM and the IWF approaches. The performances are carried out on two classes of user: for NH listeners subjected to noise vocoder simulations of participants as a model of CI processing of the processed signals and to a group of CI user.

II. METHODOLOGY USED

In this work, Ideal Binary Mask (IBM) algorithm and Ideal Wiener Filter (IWF) algorithm implemented in the Time-Frequency mask are applied to the CI's. It performs the suppression of noise. Cochlear implant are to be implanted in the inner ear, is a prosthetic device, and can restore partial hearing to profoundly deaf people or hearing impaired (HI) listeners. Signal processing, in particular, play an important role in the development of different techniques for the electrical stimuli estimated from the speech signal. The signal model and processing in CIs of Cochlear, Ltd., extracts up to $N = 22$ envelopes, in the frequency range up to 8 kHz. Such CIs usually operate with a frequency resolution that is close to the Bark scale spectrum [2].

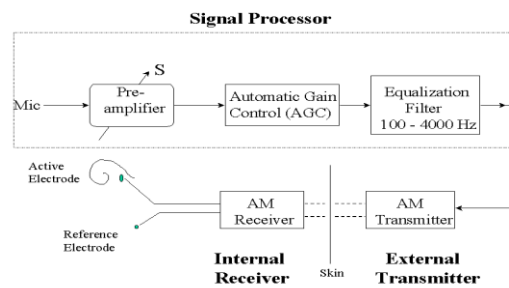


Fig. 1 Signal Processing in CI's.

A. Signal Model

A signal model is the consideration of a target speech signal embedded with the additive noise. Thus, in the discrete time domain, we obtain: $x(n) = s(n) + v(n)$, where $x(n)$ is the observed microphone signal, $v(n)$ is the additive noise, $s(n)$ is the speech signal, and n is the sample index in time. For noise suppression, signals are usually windowed in STFT representation as below:

$$X(k,l) = S(k,l) + V(k,l), \tag{1}$$

With k being the discrete frequency index and l represent the index of the time-frame. The state-of-the-art of the noise suppression algorithms compute a time-frequency mask $M(k,l)$, derived based on the power spectral density (PSD) estimation of the noise $\Psi_{VV}(k,l)$ and the speech $\Psi_{SS}(k,l)$, this mask is performed to obtain the spectrum of the (noisy) input signal. The noise-suppressed signal is then reconstructed from this masked spectrum, using standard overlap-add or overlap-save reconstruction. Note that only the amplitude of the spectrum is subjected to the masking, and further is used in the conjunction with the original (noisy) phase of $X(k,l)$ for reconstructing the noise-suppressed signal. An ideal masking can be obtained considering that the true values of Ψ_{VV} and Ψ_{SS} are known for each time-frequency point (k,l) . Two well known variants of ideal masks have been extensively studied in the field of speech recognition such as the ideal binary mask (IBM) Algorithm and the ideal Wiener Filter (IWF) Algorithm.

B. Ideal Binary Masking

The gain of IBM (GIBM) consists of binary weights. IBM is categorized at three different levels: the T-F unit level, the time frame level, and the global level. GIBM is equal to 1 when the SNR is above a threshold value and 0 when the SNR is lower than this threshold given by,

$$GIBM(n, k) = \begin{cases} 1, & \text{if } \xi \geq \xi_{in} \\ 0, & \text{else} \end{cases} \quad (2)$$

The gain is applied to the T-F representation of the mixture of target and interferer signals before recombination in a synthesis filter bank. The computation of binary mask, include separate target and interferer signals representation in T-F by using a short-time Fourier transform or a gamma tone filter bank. For each T-F unit, the power levels of the target and interferer levels are computed to determine the local signal-to noise ratio (SNR). T-F units with a local SNR above a pre-defined threshold are assigned a value of one in the mask and zero otherwise.

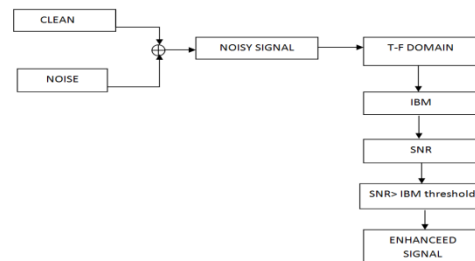


Fig. 2 Block diagram of IBM algorithm.

For the IBM, we computed the filter response magnitudes for the clean and noise signals, and then summed the energy in each band within 20 ms time frames (Hamming window with 50% overlap). For each band we assigned the mask a value of one at all 160 time samples within the time frame, if the target energy was greater than or equal to the interferer energy scaled by a threshold factor; note that we also assigned T-F units value of one if the criterion was met in either of the overlapping frames. We used a threshold of -Inf dB to simulate the unprocessed condition.

The IBM estimate produce good results in speech separation. Signal-to-noise ratio (SNR) has been widely used as a performance measure in sound separation. For sound separation, it is defined as SNR which requires estimation of target signal and the estimated target signal. It has been noted that the IBM is locally optimal in the SNR sense, as flipping in T-F unit level in the IBM always lowers the SNR in that unit. It has also been assumed that the IBM is globally optimal technique, as the output produces the highest SNR gain among all binary masks. There exist two arguments for the global optimality of the IBM implementation. It is based on the local optimality of the IBM. At each T-F unit assignment, the IBM maximally either retains target energy or removes interference energy. Therefore it minimizes the missing target energy that is discarded due to interference energy that is retained.

C. Ideal Wiener Filter

The Wiener filter algorithm is an ideal-masking technique used to estimate the desired or target random process signal by linear time-invariant (LTI) filtering, assuming known stationary signal and noise spectra, and also the additive noise. The Wiener filter generally minimizes the mean square error estimation between the random process and the desired process in speech. The Wiener filter objective is to mathematically calculate the statistical process of an unknown signal by considering the related signal as an input and then filtering that known signal, hence to produce the estimate as an output. For example, the known signal might consist of an unknown signal corrupted by additive noise signal. The Wiener filter processing involves removal of the noise from the corrupted signal, so as to provide an estimate of the underlying signal of interest. Theoretically, the IWF is a statistical approach based on minimum mean square error (MMSE) estimator article.

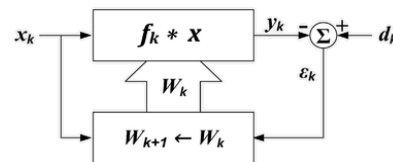


Fig. 3 Block diagram of IWF algorithm.

The Fig 3 depicts the implementation of the IWF technique process, where x is the input (noisy) signal, f_k is the impulse filter representation, W_k is the Wiener filter co-efficient, y_k is the estimated output while d_k provides the minimum MSE estimation of the process. It is a typical deterministic filter, designed for obtaining the desired frequency response. The design of the Wiener filter takes an account of different approach. One is the assumption of to have the knowledge of spectral properties in the original signal and the noise and one seeks the linear time-invariant filter whose comes close to the original signal as possible. The characterizations of Ideal Wiener filter (IWF) are:

1. Assumption: For the corresponding signal, (additive) noise are considered stationary linear stochastic processes, it provides the known autocorrelation and cross-correlation and also the known spectral characteristics.
2. Requirement: The IWF must be physically realizable/causal.
3. Performance criterion: Estimation of Minimum Mean-Square Error (MMSE) given,

$$\begin{aligned}
 E[e^2[n]] &= E[(x[n] - s[n])^2] \\
 &= E[x^2[n]] + E[s^2[n]] - 2E[x[n]s[n]] \\
 &= E\left[\left(\sum_{i=0}^N a_i w[n-i]\right)^2\right] + E[s^2[n]] - 2E\left[\sum_{i=0}^N a_i w[n-i]s[n]\right]
 \end{aligned} \tag{3}$$

Where $x(n)$ is the input speech signal being processed , $s(n)$ is the original speech signal and $w(n)$ is the weighted function of the IWF implementation technique.

The IWF gain function is defined as:

$$MIWF(k,l) = \frac{\varphi_{SS}(k,l)}{\varphi_{SS}(k,l) + \varphi_{VV}(k,l)} \quad (4)$$

Unlike the IBM function, the IWF generates a soft mask $MIWF(k,l) \in [0,1]$. The IWF has been demonstrated in CI's, to yield perfectly, the speech intelligibility (SI) improvement for different SNRs and also for different noise types, in the subjective recognition experiments. The power spectral densities $\Psi_{SS}(k,l)$ and $\Psi_{VV}(k,l)$ are estimated by recursive smoothing of the periodogram, in this studies of the IBM and the IWF quantities.

III. RESULT AND ANALYSIS

A. Objective Speech Intelligibility and Quality Prediction

The significant standardization efforts have been made by the International Telecommunications Union (ITU) for standardizing both intrusive and nonintrusive algorithms using NH listeners and CIs users. On the other hand, only a handful of algorithms that are proposed are specifically tuned to assistive listening devices. In the following sections, the choice of measures used was guided only by the applicability to the task in HA, but also by the availability of publicly available source code licensed at a reasonable cost.

The performance evaluation of this database contains IEEE sentences produced by male and 3 female speakers and was corrupted by 8 different real time noises at various levels of SNR at the input level to the Hearing Aid. Noise signals from the AURORA database is taken as input, also including the recordings from different environments such as: babble (multitalker), car, restaurant, exhibition hall, street and airport, station. The noisy signals were interpreted with the speech signals at SNRs of 0, 5, 10, and 15dB. The clean signal which is subjective to different noisy signals are given as input to the Cochlear Implants, which is then processed with the noise suppression Algorithm. This process is evaluated using Signal to Noise Ratio (SNR) and Perceptual Evaluation of Speech Quality (PESQ) metrics.

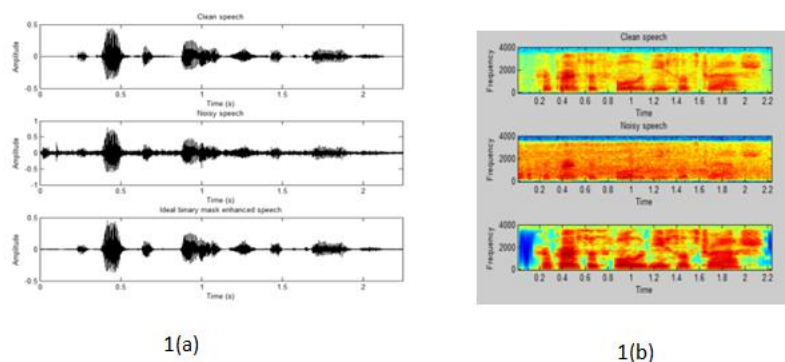


Fig. 4 Spectrogram and Waveform of corresponding clean Speech and additive noisy speech to CI producing enhanced speech signal of 0dB street noise utilizing IBM Algorithm.

Figure 4 shows Waveforms and Spectrograms for IBM approach of Street noise of 0dB. Figure 1(a) shows the Waveform of Input Clean Speech Signal, Noisy Speech Signal and the Enhanced Speech Signal

correspondingly. The Waveform of Noisy Speech Signal depicts the harmonic part of the signal is visible along with the additive residual noise, which degrades the speech signal in CI. The Waveform of the Enhanced Speech Signal correspondingly shows the enhanced speech signal, which is the output of both the IBM approach. The enhanced waveform shows the reduction in noisy part of the signal.

Figure 1(b) shows the Spectrogram of the Input Clean Speech Signal, Noisy Speech Signal and the Enhanced Speech Signal correspondingly. The Spectrogram of Noisy Speech Signal clearly displays that the harmonic part of the signal visible is corrupted by noise. The spectrogram of noisy signal corrupted with street noise of 0dB SNR. The speech signal is distinct into several frames. For this 0 dB input SNR, when the corrupted signal is run through the IBM algorithm, the output SNR is 5.0639dB. The Spectrogram of the the Enhanced Speech Signal, it depicts that the harmonic part is visible clearly. The visible harmonic parts displays that the noisy parts in the speech signal are reduced drastically.

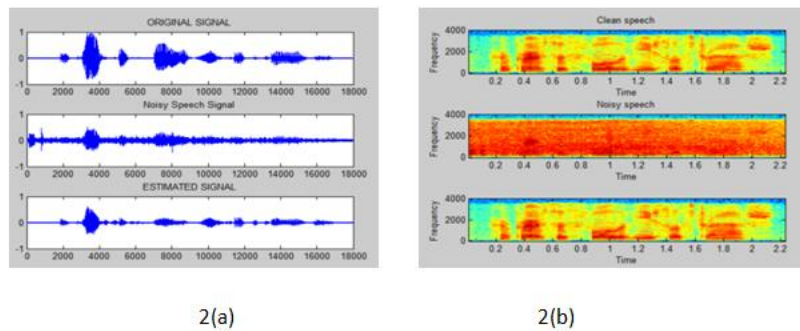


Fig. 5 Spectrogram and Waveform of corresponding clean Speech and additive noisy speech to CI producing enhanced speech signal of 0dB street noise utilizing IWF Algorithm.

Similarly Figure 5 shows Waveforms and Spectrograms for IWF approach of Street noise of 0dB. Figure 1(a) shows the Waveform of Input Clean Speech Signal, Noisy Speech Signal and the Enhanced Speech Signal correspondingly. Compared to the Enhanced Speech Signal in Figure 1 and Figure 2, IWF shows better improvisation in SNR. Compared the corresponding Enhanced Speech Signal Figure 1, the harmonic parts of the speech are clearly visible in Figure 2.

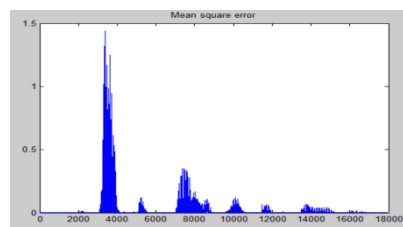


Fig. 6 Spectrogram and Waveform performance criteria of MMSE estimation in IWF algorithm.

Fig 6 shows the waveform of Minimum Mean Square Error (MMSE), which is the estimated error of the processed signal from the desired signal analysis for street noise at 0dB level. This criterion is believed to be more perceptually meaningful, between the logarithms of the spectra of the original and estimated signals.

B. Algorithmic Parameters

The intrusive and nonintrusive algorithm metric are being used in NH listeners and CIs users speech signal processing prediction. The non-intrusive intelligibility metric consists of two indices, i.e., SRMR (speech to reverberation modulation energy ratio) and ModA (modulation-spectrum area).

The intrusive metric used in speech quality measure were Perceptual Evaluation of Speech Quality (PESQ), an optimized PESQ (oPESQ), algorithm are used for reverberation degradations, while the other metric are the Kullback-Leibler Divergence (KLD) and the Frequency-Weighted Segmental Speech-to-Reverberation Ratio (FWSSRR). Among these metrics considered, the intrusive intelligibility predictors were PESQ and SNR estimator, and were fitted to NH subjects.

B. SNR Estimation

Signal to-Noise Ratio (SNR) is one of the oldest and widely used objective measures. It is simple to calculate signal mathematically, but requires both the undistorted (clean) speech and distorted (clean) speech sample-s. SNR can be calculated as follows

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^N x^2(n)}{\sum_{n=1}^N \{x(n) - \hat{x}(n)\} \Lambda^2} \quad (5)$$

where, $x(n)$ is the clean speech, $\hat{x}(n)$ the distorted speech, and N the number of samples. This classical definition of SNR is not well related to the speech quality for a wide range of distortions. Thus, several variations to the classical SNR exist which show higher correlation related with subjective quality. It was observed that classical SNR are not well correlated with speech quality although speech is not a stationary signal, SNR averages the ratio over the entire signal. Speech energy fluctuates over time and so portions, where speech energy is large, and noise which is inaudible should not be removed out as by other portions when speech energy is small, the noise can be used heard along with speech. Hence, SNR was calculated by short frames analysis, and then preceded by average. This measure is called the segmental SNR, and can be defined as

$$SNR_{seg} = 10 \log_{10} \frac{10 \sum_{n=1}^N x^2(n)}{M \sum_{n=1}^N \{x(n) - \hat{x}(n)\} \Lambda^2} \quad (6)$$

The fwSNRseg can be defined as follows,

$$fwSNR_{seg} = \frac{10 \sum_{m=0}^{M-1} \sum_{j=0}^{K-1} W(j,m) \log_{10} \frac{X(j,m)^2}{\{X(j,m) - \hat{X}(j,m)\} \Lambda^2}}{\sum_{j=0}^{K-1} W(j,m)} \quad (7)$$

where $W(j,m)$ is the weight on the j^{th} sub band in the m^{th} frame, K is the number of sub bands, $X(j,m)$ is the spectrum magnitude of the j^{th} sub band in the m^{th} frame, and $\hat{X}(j,m)$ its distorted spectrum magnitude.



TABLE I RESULTS OF IBM OUTPUT SNR WITH DIFFERENT INPUT SNR

Street Noise		
Noisy Type	Noisy SNR	Enhanced Signal SNR
0 dB	1.9534	5.0639
5 dB	4.3402	5.5183
10 dB	5.3936	5.852
15 dB	5.8306	5.8909

Hence developed system provided the satisfactory performance for different noises of 0dB, 5 dB, 10 dB, and 15dB of SNR levels. Table I shows the SNR in dB for different noises in sentence recognition task.

C. PESQ Estimation

The International Telecommunications Union (ITU-T P.862) standard, also known as Perceptual Evaluation of Speech Quality (PESQ), is a widely used objective quality measurement standard algorithm. The original PESQ algorithm was developed for 8-kHz sampling rate of narrow-band speech, while for 16 kHz sampling rate of wideband are used in this experiment. While the important emphasize was that the P.862 standard was recently superseded by ITU-T Recommendation P.863 also known as Perceptual Objective Listening Quality Assessment (POLQA), thus covering a wider scope of speech bandwidths and distortions (i.e., superwideband). POLQA, however, is not used in this study, its source code is not available at ease and its license is very costly. Perceptual Evaluation of Speech Quality (PESQ) wideband extension (50-7000 Hz, 16 kHz sample rate configuration) was applied to quantify how the perceptual quality of the distorted speech waveform varied with hearing loss severity and speech material. PESQ scores were expressed using the mean opinion score listening quality objective (MOS LQO) scale and range from 1 (worst quality) to 5 (best quality).

TABLE II RESULTS OF PESQ IN IBM WITH DIFFERENT INPUT SNR

Street Noise	
Noisy Type	PESQ
0 dB	2.8952
5 dB	3.2278
10 dB	3.5544
15 dB	3.7808

Hence developed system provided satisfactory performance of PESQ for different noise levels. Table II shows the PESQ corresponding to different noise levels.

IV. CONCLUSION

The improvement of SI potential was assessed by a sentence recognition task applicable to the CI users, with the ideal masks estimates in the multitalker street and also with an interfering talker. This study investigated the potential of the IWF and the IBM approaches for speech quality (SQ) and speech intelligibility (SI) improvising in CIs. The experimental result shows that IWF techniques outperforms the IBM technique in the case of NH listeners but this method is not so good for CI users.



It is clear that without a priori knowledge of the target and noise components, it is highly critical to rescale the original speech signal level. In conclusion, this study suggests that the future work of speech enhancement algorithms in Cochlear Implants (CI's) generally should be optimized for the respective target group of listeners. The evaluation parameter such as SNR and PESQ showed better results even at very low noise levels compared to the existing system in CI. In future this technique can be implemented in hearing aid for better performance.

REFERENCES

- [1] G. Kim, Y. Hu, and P. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," *J. Acoust. Soc. Amer.*, vol. 126, no. 3, pp. 1486–1494, 2009.
- [2] 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: Application to auditory prostheses," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 1, pp. 63–72, Jan. 2013
- [3] N.Li and P.C.Loizou, "Factors influencing intelligibility of binary-masked speech: Implications for noise reduction," *Journal of the Acoustical Society of America*, vol. 123, pp. 1673–1682, 2008.
- [4] D. Wang, U. Kjems, M. S. Pedersen, J. B. Boldt, and T. Lunner, "Speech intelligibility in background noise with ideal binary time-frequency masking," *J. Acoust. Soc. Amer.*, vol. 125, no. 4, pp. 2336–2347, 2009.
- [5] M. C. Anzalone, L. Calandruccio, K. A. Doherty, and L. H. Carney, "Determination of the potential benefit of time-frequency gain manipulation," *Ear Hearing*, vol. 27, no. 5, pp. 480–492, 2006.
- [6] E.W.Healy, S.E.Yoho, Y.Wang, and D.Wang, "An algorithm to improve speech recognition in noise for hearing-impaired listeners," *J.Acoust.Soc. Amer.*, vol. 134, no. 4, pp. 3029–3038, 2013.
- [7] Y. Hu and P. C. Loizou, "A comparative intelligibility study of singlemicrophone noise reduction algorithms," *J. Acoust. Soc. Amer.*, vol. 122, pp. 1777–1786, 2007.
- [8] U. Kjems, J. B. Boldt, M. S. Pedersen, T. Lunner, and D. Wang, "Role of mask pattern in intelligibility of ideal binary-masked noisy speech," *J. Acoust. Soc. Amer.*, vol. 126, pp. 1415–1426, 2009.
- [9] D. Wang, U. Kjems, M. S. Pedersen, J. B. Boldt, and T. Lunner, "Speech perception of noise with binary gains," *J. Acoust. Soc. Amer.*, vol. 124, no. 4, pp. 2303–2307, 2008.
- [10] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, pp. 588–601, 2007.
- [11] 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.
- [12] R. J. van Hoesel and G. M. Clark, "Evaluation of a portable two microphone adaptive beamforming speech processor with cochlear implant patients," *J. Acoust. Soc. Amer.*, vol. 97, no. 4, pp. 2498–2503, 1995.
- [13] J. Wouters and J. Vanden Berghe, "Speech recognition in noise for cochlear implantees with a two microphone monaural adaptive noise reduction system," *Ear Hearing*, vol. 22, pp. 420–430, 2001.



- [14] A. Spriet, L. Van Deun, K. Eftaxiadis, J. Laneau, M. Moonen, B. van Dijk, A. van Wieringen, and J. Wouters, "Speech understanding in background noise with the two-microphone adaptive beamformer beam in the nucleus freedom cochlear implant system," *Ear Hearing*, vol. 28, pp. 62–72, 2007.
- [15] S. Kochkin, "Consumers rate improvements sought in hearing instruments," *Hearing Review*, vol. 9, pp. 18-22, 2002.
- [16] Romain Serizel, Marc Moonen, Jan Wouters, and Soren Holdt Jensen "A Zone-of-Quiet Based Approach to Integrated Active Noise Control and Noise Reduction for Speech Enhancement in Hearing Aids" *IEEE Trans. Audio, Speech and language Process.*, vol. 20, no. 6, Aug. 2012.
- [17] I. Tashev, *Sound Capture and Processing: Practical Approaches*. New York: Wiley, 2009.
- [18] Yilmaz, "O., & Rickard, S. (2004). Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52(7).
- [19] H. Dillon, *Hearing Aids*, 2nd Ed. ed. Stuttgart, Germany: Thieme, 2012.
- [20] Jan Wouters, Bram Cornelis and Marc Moonen "Speech Intelligibility Improvements with hearing aids using bilateral and binaural adaptive multichannel Wiener filtering based noise reduction" *J. Acoust. Soc. Am.*, Vol. 131, No. 6, June 2012.
- [21] K. Kokkinakis, B. Behnam, Y. Hu, and D. R. Friedland, "Single and multiple microphone noise reduction strategies in cochlear implants," *Trends Amplification*, vol. 16, no. 2, pp. 102–116, 2012.
- [22] S.J. Mauger, P. W. Dawson, and A. A. Hersbach, "Perceptually optimized gain function for cochlear implant signal-to-noise ratio based noise reduction," *J. Acoust. Soc. Amer.*, vol. 131, no. 1, pp. 327–336, 2012.
- [23] N. J. Versfeld, L. Daalder, J. M. Festen, and T. Houtgast, "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," *J. Acoust. Soc. Amer.*, vol. 107, no. 3, pp. 1671–1684, 2000.