PRESERVATION OF INTERAURAL TIME DELAY FOR BINAURAL HEARING AIDS THROUGH MULTI-CHANNEL WIENER FILTERING BASED NOISE REDUCTION

Thomas J. Klasen, Marc Moonen

Department of Electrical Engineering
Katholieke Universiteit Leuven, Belgium
{tklasen,moonen}@esat.kuleuven.ac.be

Tim Van den Bogaert, Jan Wouters

Laboratory of Experimental Otorhinolaryngology
Katholieke Universiteit Leuven, Belgium
{tim.vandenbogaert,jan.wouters}@uz.kuleuven.ac.be

ABSTRACT

This paper presents a binaural extension of a monaural multi-channel noise reduction algorithm for hearing aids based on Wiener filtering. The algorithm provides the hearing aid user with a binaural output. In addition to significantly suppressing the noise interference, the algorithm preserves the interaural time delay (ITD) cues of the received speech, thus allowing the user to correctly localize the speech source.

1. INTRODUCTION

Hearing impaired persons often localize sounds better without their hearing aids than with their hearing aids [1]. This is not surprising, since noise reduction algorithms currently used in hearing aids are not designed to preserve speech localization cues [2]. The inability to correctly localize sounds puts the hearing aid user at a disadvantage as well as at risk. The sooner the user can localize the speech, the sooner the user can begin to exploit visual cues. Generally, visual cues lead to large improvements in intelligibility for hearing impaired persons [3]. Moreover, in certain situations, such as traffic, incorrectly localizing sounds could endanger the user. This paper focuses specifically on interaural time delay (ITD) cues, which help listeners localize sounds horizontally [4]. ITD is the time delay in the arrival of the sound signal between the left and right ear.

Most noise reduction algorithms, such as the generalized sidelobe canceler (GSC), make the assumption that the location of the speech source is known [5]. When this assumption is valid, the output speech sounds as if it is coming from the actual speech source. However, in practice this assumption is often violated. In this case, the ITD cues of the processed speech differ from those of the unprocessed speech; accordingly the processed speech appears to be coming from another direction than the unprocessed speech.

Unfortunately, many monaural noise reduction algorithms incorporate a beamformer. This type of algorithm estimates the speech component in the output of the beamformer, instead of in the individual channels, leaving one with a monaural output regardless of a possible binaural input.

In [6], a binaural adaptive noise reduction algorithm is proposed. This algorithm takes a microphone signal from each ear as inputs. The inputs are filtered by a high-pass and low-pass filter with the same cut-off frequency to create high and low frequency portions. The high frequency portion is adaptively processed and added to the delayed low frequency portion. Since interaural time delay (ITD) cues are contained in the low-frequency regions, as the cut-off frequency increases more ITD information will arrive undistorted to the user [4, 7]. The major draw back to this approach is that the low frequency portion containing the speech ITD cues also contains noise. Consequently, noise as well as speech ITD cues are passed from the input to the output unprocessed. Therefore, there is a trade-off between noise reduction and ITD preservation. As the cut-off frequency increases the preservation of the ITD cues improves at the cost of noise reduction.

In this paper we extend the monaural multi-channel Wiener filtering algorithm discussed in [5, 8, 9] to a binaural algorithm. This algorithm is well suited for binaural noise reduction because it makes no a priori assumptions (e.g. the location of the speech source), and is capable of estimating the speech components in all microphone channels [8].

2. SYSTEM MODEL

2.1. Listening scenario

Figure 1 shows a binaural hearing aid user in a typical listening scenario. The speaker speaks intermittently in the continuous background noise caused by the noise source. In this case, there is one microphone on each hearing aid. Nevertheless, we consider the case where there are M microphones on each hearing aid. We refer to the mth microphone of the left hearing aid and the mth microphone of the right hearing aid as the mth microphone pair. The received signals at time k, for k ranging from 1 to K, at the mth microphone pair are expressed in the equations below.

$$y_{L_m}[k] = h_{L_m}[k] \otimes s[k] + g_{L_m}[k] \otimes n[k] \quad (1)$$

$$y_{R_m}[k] = h_{R_m}[k] \otimes s[k] + g_{R_m}[k] \otimes n[k] \quad (2)$$

In (1) and (2), $s[k]$ and $n[k]$ represent the signals generated by the speaker and noise source respectively. The acoustic room impulse responses between the speaker and the mth microphone pair are $h_{L_m}[k]$ and $h_{R_m}[k]$. Similarly, the room responses between the noise source and the mth microphone pair are captured in the acoustic room impulse responses $g_{L_m}[k]$ and $g_{R_m}[k]$. The convolution, denoted by $\otimes$, of the speech signal, $s[k]$, and the left and right room impulse responses, $h_{L_m}[k]$ and $h_{R_m}[k]$, can be written as the speech component in the mth left microphone, $x_{L_m}[k]$, and the speech component in the mth right microphone, $x_{R_m}[k]$. Likewise, the noise components of the mth microphone pair, $v_{L_m}[k]$ and $v_{R_m}[k]$, are the convolution of the noise signal and the room impulse responses between the noise source and the mth microphone pair. Using the above definitions, the signals received at the mth microphone pair simplify to the equations below.

$$y_{L_m}[k] = x_{L_m}[k] + v_{L_m}[k] \quad (3)$$

$$y_{R_m}[k] = x_{R_m}[k] + v_{R_m}[k] \quad (4)$$

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the framework of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian Federal Science Policy Office IUAP P5/22 ('Dynamical Systems and Control: Computation, Identification and Modelling'), the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, Research Project FWO nr.G.0233.01 ('Signal processing and automatic patient fitting for advanced auditory prostheses') and IWT project 020540: 'Innovative Speech Processing Algorithms for Improved Performance of Cochlear Implants'. The scientific responsibility is assumed by its authors.

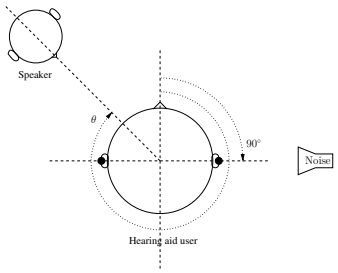


Fig. 1. Typical listening scenario

We make two standard assumptions that will be pertinent later. First, the speech signal is assumed to be statistically independent of the noise signal. Second, we assume that the noise is short-term stationarity.

2.2. Voice activity detection (VAD)

The signals received at the microphones of the left and right hearing aids contain either noise when speech is not present, or speech and noise. We assume in our system model that we have access to a perfect VAD algorithm. In other words, we can identify without error when there is only noise present, and when there is speech and noise present. For simplicity, let us call the time instants when there is only noise present $k^n$and when there is speech and noise present k^{sn}.

2.3. ITD calculation

As mentioned earlier, ITD cues are essential for the listener to localize sounds horizontally [4]. These cues are contained in the low-frequency regions of the signals received by the listener's ears [4]. We consider the ITD cues for the frequency region below 1500Hz. Above this frequency the cues become ambiguous, because the wavelength of the sound approaches the distance between one's ears [7]. In order to calculate the ITD between two signals, we use cross-correlation to determine the delay in samples between the two signals. Dividing this delay by the sampling frequency results in the ITD between the two signals.

3. BINAURAL MULTI-CHANNEL WIENER FILTERING

This algorithm is an extension of the multi-channel Wiener filtering technique discussed in [5, 8, 9]. The goal of this algorithm is to estimate the speech components of the mth microphone pair, $x_{L_m}[k]$ and $x_{R_m}[k]$, using all received microphone signals, $y_{L_{1:M}}[k]$ and $y_{R_{1:M}}[k]$. In order to estimate the speech components of the mth microphone pair, we design two Wiener filters that estimate the noise components in the mth microphone pair. The noise estimates of the mth microphone pair and therefore the output of the two Wiener filters are $\tilde{v}_{L_m}[k]$ and $\tilde{v}_{R_m}[k]$. To obtain the estimates of the speech components of the mth microphone pair, the noise components estimates are subtracted from the original signals received at the two microphones. The speech estimates are defined below.

$$\tilde{x}_{L_m}[k] = (x_{L_m}[k] + v_{L_m}[k]) - \tilde{v}_{L_m}[k] \tag{5}$$

$$\tilde{x}_{R_m}[k] = (x_{R_m}[k] + v_{R_m}[k]) - \tilde{v}_{R_m}[k] \tag{6}$$

The errors of the right and left estimates are written in (7) and (8).

$$e_{L_m}[k] = v_{L_m}[k] - \tilde{v}_{L_m}[k] \tag{7}$$

$$e_{R_m}[k] = v_{R_m}[k] - \tilde{v}_{R_m}[k] \tag{8}$$

The goal is to develop a left and right multi-channel Wiener filter that minimizes the error signals $e_{L_m}[k]$ and $e_{R_m}[k]$. See Figure 2 for a clear illustration.

Before going any further, a few definitions are necessary. We choose the filters $\mathbf{w}_{L_m}[k]$ and $\mathbf{w}_{R_m}[k]$ to be of length N. The filters are expressed in the following equations.

$$\mathbf{w}_{L_m}[k] = \left[w^0_{L_m}\ w^1_{L_m}\ \dots\ w^{N-1}_{L_m}\right]^T \tag{9}$$

$$\mathbf{w}_{R_m}[k] = \left[w^0_{R_m}\ w^1_{R_m}\ \dots\ w^{N-1}_{R_m}\right]^T \tag{10}$$

Next we create a stacked vector of the individual left and right microphone filters $\mathbf{w}_{L_m}[k]$ and $\mathbf{w}_{R_m}[k]$.

$$\mathbf{w}_L[k] = \begin{bmatrix} \mathbf{w}_{L_1}[k] \\ \mathbf{w}_{L_2}[k] \\ \vdots \\ \mathbf{w}_{L_M}[k] \end{bmatrix} \qquad \mathbf{w}_R[k] = \begin{bmatrix} \mathbf{w}_{R_1}[k] \\ \mathbf{w}_{R_2}[k] \\ \vdots \\ \mathbf{w}_{R_M}[k] \end{bmatrix} \tag{11}$$

$$\mathbf{w}_{Left}[k] = \begin{bmatrix} \mathbf{w}_L[k] \\ \mathbf{w}_R[k] \end{bmatrix} \tag{12}$$

The filter $\mathbf{w}_{Right}[k]$ is defined similarly. Both filters are vectors of length $2MN$. Correspondingly, we define the received microphone signals at mth microphone pair below.

$$\mathbf{y}_{L_m}[k] = [y_{L_m}[k]\ y_{L_m}[k-1]\ \dots\ y_{L_m}[k-N+1]]^T \tag{13}$$

$$\mathbf{y}_{R_m}[k] = [y_{R_m}[k]\ y_{R_m}[k-1]\ \dots\ y_{R_m}[k-N+1]]^T \tag{14}$$

Again we create a stacked vector of the individual left and right microphone inputs. Input vector $\mathbf{y}$ is of length $2MN$.

$$\mathbf{y}_L[k] = \begin{bmatrix} \mathbf{y}_{L_1}[k] \\ \mathbf{y}_{L_2}[k] \\ \vdots \\ \mathbf{y}_{L_M}[k] \end{bmatrix} \qquad \mathbf{y}_R[k] = \begin{bmatrix} \mathbf{y}_{R_1}[k] \\ \mathbf{y}_{R_2}[k] \\ \vdots \\ \mathbf{y}_{R_M}[k] \end{bmatrix} \tag{15}$$

$$\mathbf{y}[k] = \begin{bmatrix} \mathbf{y}_L[k] \\ \mathbf{y}_R[k] \end{bmatrix} \tag{16}$$

In this section we derive the left and right multi-channel Wiener filters in a statistical setting. Minimizing the following cost function,

$$E\left\{\left|\mathbf{y}^T[k]\,[\mathbf{w}_{Left}[k]\ \mathbf{w}_{Right}[k]] - [v_{L_m}[k]\ v_{R_m}[k]]\right|^2\right\} \tag{17}$$

minimizes the errors defined in (7) and (8). In (17), $E\{\cdot\}$is the expectation operator. The filters achieving the minimum of the cost function are the well known Wiener filters expressed below.

$$\left[\mathbf{w}_{WF_{Left}}[k]\ \mathbf{w}_{WF_{Right}}[k]\right] =$$

$$E\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\}^{-1} E\left\{\mathbf{y}[k]\,[v_{L_m}[k]\ v_{R_m}[k]]\right\} \tag{18}$$

Owing to (3) and (4), we can define $\mathbf{x}[k]$ and $\mathbf{v}[k]$, where $\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$. Recall that in the first assumption we assert that the speech signal and the noise signal are statistically independent. More specifically the following equation must hold.

$$E\left\{\mathbf{x}[k]\mathbf{v}^T[k]\right\} = 0 \tag{19}$$

Using the first assumption defined in (19) we can rewrite (18) by making the following substitution.

$$E\left\{\mathbf{y}[k]\,[v_{L_m}[k]\ v_{R_m}[k]]\right\} = E\left\{\mathbf{v}[k]\,[v_{L_m}[k]\ v_{R_m}[k]]\right\} \tag{20}$$

Unfortunately, in real life these statistical quantities are not available. Therefore we cannot calculate the left and right Wiener filters directly. Instead, we make a least squares approximation of the filters. This data based approach requires a few extra definitions. Using (16), we write the input matrix $\mathbf{Y}$, which is of size K by $2MN$.

$$\mathbf{Y}[k] = \begin{bmatrix} \mathbf{y}^T[k] \\ \mathbf{y}^T[k-1] \\ \vdots \\ \mathbf{y}^T[k-K+1] \end{bmatrix} \tag{21}$$

Analogously, the speech input matrix, $\mathbf{X}[k]$ and the noise input matrix, $\mathbf{V}[k]$, can be defined, where $\mathbf{Y}[k] = \mathbf{X}[k] + \mathbf{V}[k]$. Finally, we write the desired signals, $\mathbf{d}_L[k]$ and $\mathbf{d}_R[k]$, which are the unknown noise input vectors.

$$\mathbf{d}_L[k] = \quad \mathbf{v}_{L_m}[k] = \begin{bmatrix} v_{L_m}[k] \\ v_{L_m}[k-1] \\ \vdots \\ v_{L_m}[k-K+1] \end{bmatrix} \tag{22}$$

$$\mathbf{d}_R[k] = \quad \mathbf{v}_{R_m}[k] = \begin{bmatrix} v_{R_m}[k] \\ v_{R_m}[k-1] \\ \vdots \\ v_{R_m}[k-K+1] \end{bmatrix} \tag{23}$$

We define the desired matrix $\mathbf{D}[k]$ as $[\mathbf{d}_L[k]\ \mathbf{d}_R[k]]$. We can estimate $E\left\{\mathbf{y}[k]\mathbf{y}^T[k]\right\}$ by the matrix $\mathbf{Y}^T[k]\mathbf{Y}[k]$. In order to estimate $E\left\{\mathbf{v}[k]\left[v_{L_m}[k]\ v_{R_m}[k]\right]\right\}$ by $\mathbf{V}^T[k]\mathbf{D}[k]$, we must use the second assumption we made in our system model; since the input noise matrix $\mathbf{V}[k]$, and therefore the desired matrix $\mathbf{D}[k]$ are not known explicitly. The assumption is that the noise is short-term stationary. This means that $E\left\{\mathbf{v}[k]\left[v_{L_m}[k]\ v_{R_m}[k]\right]\right\}$ is the same whether it is calculated during noise only periods, k^n, or at all time instants, k. Assumption two is expressed below.

$$E\left\{\mathbf{v}[k]\left[v_{L_m}[k]\, v_{R_m}[k]\right]\right\} = E\left\{\mathbf{v}[k^n]\left[v_{L_m}[k^n]\, v_{R_m}[k^n]\right]\right\} \tag{24}$$

Invoking assumption two, $E\left\{\mathbf{v}[k]\left[v_{L_m}[k]\ v_{R_m}[k]\right]\right\}$ can be estimated by $\mathbf{V}^T[k^n]\mathbf{D}[k^n]\}$ at time instants where only noise is present. Therefore we can write the least squares approximation of the Wiener filter as,

$$\left[\mathbf{w}_{LS_{Left}}\ \mathbf{w}_{LS_{Right}}\right] = \left(\mathbf{Y}^T[k]\mathbf{Y}[k]\right)^{-1}\mathbf{V}^T[k^n]\mathbf{D}[k^n]. \tag{25}$$

This least squares approximation of the Wiener filter is what we use in practice.

4. PERFORMANCE

4.1. Experimental setup

The recordings used in the following experiments were made in an anechoic room. Two CANTA behind the ear (BTE) hearing aids were placed on a CORTEX MK2 artificial head. Mounted on each hearing aid were two omni-directional microphones. The speech and noise sources were placed one meter from the center of the dummy head. The sound level measured at the center of the dummy head was 70dB SPL. Speech and noise sources were recorded separately. All recordings were performed at a sampling frequency of 32kHz. HINT sentences and ICRA noise[1] were used for the speech and noise signals [11].

[1]ICRA 1: Unmodulated random Gaussian noise, male weighted (HP 100Hz 12dB/oct.) idealized speech spectrum [10]

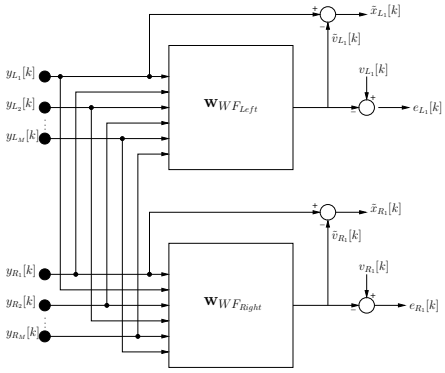


Fig. 2. Binaural multi-channel Wiener filter

The signals fed into the binaural algorithms were 10 seconds in length. The first half of the signal consisted of noise only. A short one and a half second sentence was spoken in the second half amidst the continuous background noise. The location of the speech source varied from 0 to 345 degrees in increments of 15 degrees. The noise source remained fixed throughout the simulations at 90 degrees. This situation is depicted in Figure 1. In the simulations only the front input from each ear was used. The algorithms estimated the speech component in the first microphone pair.

Simulations were run to compare the binaural multi-channel Wiener filtering algorithm and the binaural adaptive algorithm discussed in [6]. In the binaural multi-channel Wiener filtering algorithm the filter length, N, was fixed at 100. The filter length of the binaural adaptive algorithm was 201, and was adapted, during periods of noise only, by a normalized LMS algorithm. Cut-off frequencies of $500Hz$ and $1500Hz$ were simulated.

4.2. Results

The intelligibility weighted signal-to-noise-ratio (SNR_{INT}), defined in [12], is used to quantify the noise reduction performance.

$$SNR_{INT} = \sum_{j=1}^{J} w_j SNR_j \tag{26}$$

The weight, w_j, emphasizes the importance of the jth frequency band's overall contribution to intelligibility, and SNR_j is the signal-to-noise-ratio of the jth frequency band. The individual weights of the J frequency bands are given in [10].

Figure 3 shows the absolute difference between the input ITD and the output ITD of the speech component and the noise component. The noise reduction performance of the algorithms can be seen in Figure 4. Looking closely at Figure 3 we see that for the binaural multi-channel Wiener filtering algorithm there is no speech ITD error (except for a slight error when the speech source is located at 150 degrees). In other words, the speech ITD cues are preserved. Figures 3 and 4 clearly demonstrate the trade-off between ITD cue preservation and noise reduction that exists for the binaural adaptive algorithm. With a cut-off frequency of $500Hz$, the speech ITD cues are not preserved, but good noise reduction is achieved. Contrastingly, although noise reduction performance is poor when the cut-off frequency is $1500Hz$, both speech and noise ITD cues are preserved. Furthermore, the noise reduction performance of the binaural multi-channel Wiener filtering algorithm is similar to that of the binaural adaptive algorithm with a cut-off frequency of $500Hz$; yet the binaural multi-channel Wiener filtering algorithm preserves the speech ITD cues and the binaural adaptive

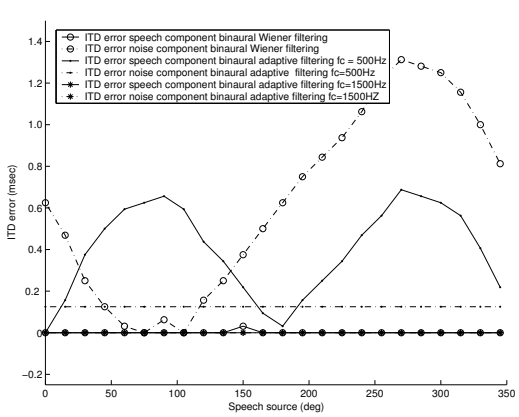


Fig. 3. ITD error: the absolute difference between the input ITD and output ITD of the first microphone pair

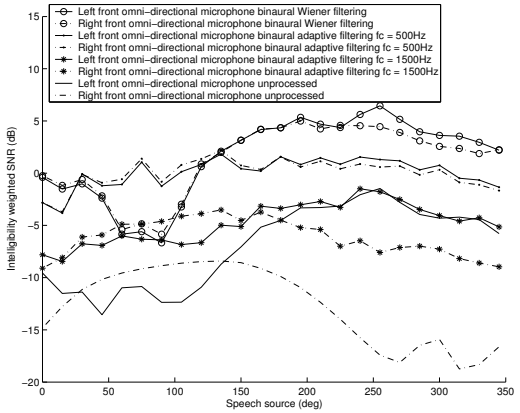


Fig. 4. Intelligibility weighted SNR for speech sources between 0 and 345 degrees with the noise source fixed at 90 degrees

algorithm does not. Clearly, the binaural multi-channel Wiener filtering algorithm is preferable to the binaural adaptive algorithm, since it does not sacrifice noise reduction in order to preserve the speech ITD cues.

It should be noted that the processing carried out by the multi-channel Wiener filtering does affect the ITD cues of the noise component. Typically, this will not cause any inconvenience to the hearing aid user, because the noise is sufficiently attenuated. Nevertheless, in some situations, such as traffic, a trade-off may exist between noise reduction and preservation of the noise ITD cues. Therefore, future research should focus on preserving the ability to localize both the speech and noise sources without sacrificing noise reduction.

Extensive studies have been carried out showing that monaural multi-channel Wiener filtering algorithms perform better than the generalized side-lobe canceler (GSC) [8, 13]. In situations where the assumptions of the GSC are violated, multi-channel Wiener filtering algorithms tend to be more robust [8, 14]. In addition, many variants of the multi-channel Wiener filtering algorithm have been developed to combat its complexity, making it possible to implement the algorithm in hearing aids [15, 9].

5. CONCLUSION

In conclusion, the binaural multi-channel Wiener filtering algorithm preserves the ITD cues without sacrificing noise reduction. As discussed above, the ITD cues of the speech component are exactly the same in the processed and unprocessed signal. Simultaneously, good noise reduction is achieved. Conversely, the binaural adaptive algorithm proposed in [6] sacrifices noise reduction in order to preserve ITD cues. This gives the binaural multi-channel Wiener filtering algorithm a clear advantage over the binaural adaptive algorithm.

6. REFERENCES

[1] J. Besing, J. Koehnke, P. Zurek, K. Kawakyu, and J. Lister, "Aided and unaided performance on a clinical test of sound localization," *J. Acoust. Soc. Amer.*, vol. 105, p. 1025, Feb. 1999.

[2] J. Desloge, W. Ravinowitz, and P. Zurek, "Microphone-Array Hearing Aids with Binaural Output-Part I: Fixed-Processing Systems," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 529–542, Nov. 1997.

[3] N. Erber, "Auditory-visual perception of speech," *J. Speech Hearing Dis.*, vol. 40, pp. 481–492, 1975.

[4] F. Wightman and D. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Amer.*, vol. 91, pp. 1648–1661, Mar. 1992.

[5] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction." Accepted: Signal Processing.

[6] D. Welker, J. Greenburg, J. Desloge, and P. Zurek, "Microphone-Array Hearing Aids with Binaural Output-Part II: A Two-Microphone Adaptive System," *IEEE Trans. Signal Processing*, vol. 5, pp. 543–551, Nov. 1997.

[7] W. Hartmann, "How We Localize Sound," *Physics Today*, pp. 24–29, Nov. 1999.

[8] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Single and Multi-Microphone Speech Enhancement," *IEEE Trans. Signal Processing*, vol. 50, pp. 2230–2244, Sep. 2002.

[9] G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," *Signal Processing*, vol. 83, Sep. 2003.

[10] A. S. of America, "American National Standard Methods for Calculation of the Speech Intelligibility Index," in *ANSI S3.5-1997*, 1997.

[11] M. Nilsson, S. Soli, and J. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Amer.*, vol. 95, pp. 1085–1096, 1994.

[12] J. Greenberg, P. Peterson, and Z. P.M., "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoust. Soc. Amer.*, vol. 94, pp. 3009–3010, Nov. 1993.

[13] S. Doclo and M. Moonen, "SVD-based optimal filtering with applications to noise reduction in speech signals," in *Proc. IEEE WASPAA*, pp. 17–20, Oct. 1999.

[14] A. Spriet, M. Moonen, and J. Wouters, "Robustness Analysis of GSVD based optimal Filtering and generalized Sidelobe Canceller for Hearing Aid Applications," in *Proc. IEEE WASPAA*, (New Paltz, New York), Oct. 2001.

[15] A. Spriet, M. Moonen, and J. Wouters, "A multichannel subband GSVD approach to speech enhancement," *European Transactions on Telecommunications*, vol. 13, pp. 149–158, Mar. 2002.