A REAL TIME IMPLEMENTATION AND AN EVALUATION OF AN OPTIMAL FILTERING TECHNIQUE FOR NOISE REDUCTION IN DUAL MICROPHONE HEARING AIDS

Jean Baptiste Maj[1,2], *Liesbeth Royackers*[1] *and Jan Wouters*[1]

[1]Lab.Exp.ORL KULeuven, Kapucijnenvoer 33, 3000 Leuven
[2]ESAT-SISTA KULeuven, Kasteelpark Arenberg 10, 3001 Leuven
`Jean-Baptiste.Maj@uz.kuleuven.ac.be`

ABSTRACT

A real time implementation and an evaluation of a Singular Value Decomposition (SVD) based optimal filtering technique [1] for noise reduction in a dual microphone BTE hearing aid is presented. A method to improve the performance of a Voice Activity Detector (VAD) is described and evaluated physically. This method is used in the real time implementation of the optimal filtering technique. A perceptual evaluation by normal hearing subjects is carried out for single and multiple jammer sound sources with speech weighted noise. The SVD-based technique can perform as well as an adaptive beamformer [2] strategy in a single noise scenario (i.e. the ideal scenario for the latter technique), and, can outperform the beamformer technique in a multiple noise sources scenario. [1]

1. INTRODUCTION

Noise reduction strategies are important in hearing aid devices to improve speech intelligibility in a noisy background [3]. Modern digital hearing aids using dual-microphone configurations in a single behind-the-ear (BTE) hearing aid allow to process more complex noise reduction algorithms. Recently, adaptive noise reduction algorithms have been developed and implemented in hearing aids. These algorithms can adapt to changing jammer sound directions and can track moving noise sources. In this study, an adaptive procedure using a SVD-based optimal filtering technique is evaluated perceptually. This strategy was assessed theoretically and physically in previous studies [1, 4, 5]. The optimal filtering strategy works without assumptions about the desired target direction, however, this strategy needs a robust VAD. In this paper, the SVD-based optimal filtering technique is presented and the real time implementation is described. Furthermore, a method to improve the performance of the VAD is introduced. A physical evaluation allows to assess the latter method. Finally, a perceptual evaluation with subjects is carried out by measuring the SNR-improvements of the SVD-based technique, and comparing these to the results obtained with an adaptive beamformer technique [2].

[1]The authors would like to acknowledge Marc Moonen[2] for his scientific contribution. They consider him to be a co-author of this paper, however he cannot be listed due to conference restrictions. This study is supported by the Fund for Scientific Research - Flanders (Belgium) through the FWO projects 3.0168.95 ("Signal processing for improved speech intelligibility of hearing impaired"), G.0233.01 ("Signal processing and automatic patient fitting for advanced auditory prostheses"), and Cochlear (IWT project 20540), and was partially funded by the Belgian State, Prime Minister's Office - Federal Office for Scientific, Technical and Cultural Affairs - IUAP P4-02 (Modeling, Identification, Simulation and Control of Complex Systems) and the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government. The scientific responsibility is assumed by its authors.

2. SVD-BASED OPTIMAL FILTERING TECHNIQUE

The SVD-based optimal filtering technique considered here, in general reconstructs a speech signal $\mathbf{s}_k$ from noisy data $\mathbf{u}_k = \mathbf{s}_k + \mathbf{n}_k$ by means of an optimal filter $\mathbf{W}_{WF} \in \mathbb{R}^{N\times N}$ using $\hat{\mathbf{s}}_k = \mathbf{W}_{WF}^T\mathbf{u}_k$ at time k. Using a Minimum Mean Square Error-criterion (MMSE), the optimal filter $\mathbf{W}_{WF}$ is equal to:

$$\mathbf{W}_{WF} = \mathcal{E}\{\mathbf{u}_k.\mathbf{u}_k^T\}^{-1}.(\mathcal{E}\{\mathbf{u}_k.\mathbf{u}_k^T\} - \mathcal{E}\{\mathbf{n}_k.\mathbf{n}_k^T\}) \quad (1)$$

Doclo and Moonen [1] use an interesting and useful simplification in formula (1), where $\mathbf{W}_{WF}$ is derived from the GSVD *(generalized singular value decomposition)* of the data matrices $\mathbf{U}_k \in \mathbb{R}^{p\times N}$ and $\mathbf{N}_k \in \mathbb{R}^{q\times N}$ (with p and q typically larger than N), such that $\mathcal{E}\{\mathbf{u}_k.\mathbf{u}_k^T\} \Rightarrow (\mathbf{U}_k^T.\mathbf{U}_k)/p$ and $\mathcal{E}\{\mathbf{n}_{k'}.\mathbf{n}_{k'}^T\} \Rightarrow (\mathbf{N}_k^T.\mathbf{N}_k)/q$. $\mathbf{u}_k$ is collected during *speech-and-noise periods*, while $\mathbf{n}_k$ is collected during *noise periods*. The GSVD of the matrices $\mathbf{U}_k$ and $\mathbf{N}_k$ is defined as

$$\begin{cases} \mathbf{U}_k = \mathbf{Y}.diag\{\sigma_i\}.\mathbf{X}^T \\ \mathbf{N}_k = \mathbf{V}.diag\{\eta_i\}.\mathbf{X}^T \end{cases} \quad (2)$$

where $\mathbf{Y} \in \mathbb{R}^{p\times N}$ and $\mathbf{V} \in \mathbb{R}^{q\times N}$ are orthogonal matrices, $\mathbf{X} \in \mathbb{R}^{N\times N}$ is an invertible matrix and $\frac{\sigma_i}{\eta_i}$ are the generalized singular values. By substituting the above formulas in (1), we obtain:

$$\mathbf{W}_{WF} = \mathbf{X}^{-T}.diag\left\{1 - \frac{p}{q}\frac{\eta_i^2}{\sigma_i^2}\right\}.\mathbf{X}^T \quad (3)$$

By using a time constrained estimator, the energy of the signal distortion ϵ_s^2 is minimized under the constraint that the residual noise energy ϵ_n^2 stays under a threshold α [1].

$$\underset{\mathbf{W}_{WF}}{Min} \quad \epsilon_s^2 \quad subject \quad to \quad \epsilon_n^2 \le \alpha \quad where \quad 0 \le \alpha \le 1 \quad (4)$$

Thus, the filter $\mathbf{W}_{WF}$ becomes:

$$\mathbf{W}_{WF} = \mathbf{X}^{-T}.diag\left\{\frac{q.\sigma_i^2 - p.\eta_i^2}{q.\sigma_i^2 + (\mu - 1)p.\eta_i^2}\right\}.\mathbf{X}^T \quad (5)$$

The speech distortion parameter $\mu \in [0, \infty]$ allows a trade-off between signal distortion and noise reduction. If $\mu = 1$ the original MMSE solution is obtained. More emphasis is put on the signal distortion when $\mu < 1$ at the expense of decreasing the

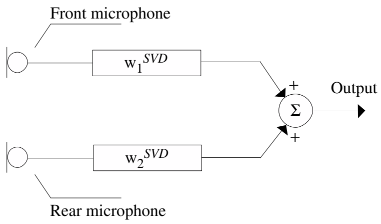

Fig. 1. Representation of the SVD-based optimal filtering technique.

noise reduction performance. The residual noise level is reduced when $\mu > 1$ at the expense of increasing speech distortion. With $\mu \to \infty$, all the emphasis is put on the noise reduction without taking into account of the signal distortion. In a two microphone application, the vector $\mathbf{u}_k \in \mathbb{R}^{MN}$ takes the form:

$$\mathbf{u}_k = \left[\begin{array}{cc} \mathbf{u}_{1k} & \mathbf{u}_{2k} \end{array} \right] \tag{6}$$

with

$$\mathbf{u}_{jk} = \left[\begin{array}{cccc} u_j(k) & u_j(k-1) & \ldots & u_j(k-N+1) \end{array} \right]^T \tag{7}$$

where the *j* refers to the *j*-th microphone. The vector $\mathbf{n}_k$ is similarly defined. The computation of the optimal filter $\mathbf{W}_{WF}$ results in a $(2 \times N)-$taps estimator $\mathbf{w}_{WF}$ for the signal $\tilde{\mathbf{s}}_k$.

$$\tilde{\mathbf{s}}_k = \left[\begin{array}{l} \tilde{s}(k) \\ \tilde{s}(k+1) \\ \vdots \\ \tilde{s}(k+p-1) \end{array} \right] = \mathbf{U}_k.\mathbf{w}_{WF} \tag{8}$$

where $\tilde{\mathbf{s}}_k$ is an estimate for the (delayed version of the) speech part of either front microphone or rear microphone depending on the choice for $\mathbf{w}_{WF}$, which is one column of $\mathbf{W}_{WF}$. Maj et al. [5] showed that using the middle column of $\mathbf{W}_{WF}$ in the front microphone part, a good estimate of $\tilde{\mathbf{s}}_k$ is obtained. This filter $\mathbf{w}_{WF}$ (see figure 1) is as a two-channel filter, where each microphone was filtered with a N-taps filter $\mathbf{w}_j^{SVD}$. In our experiments N will be 15.

$$\mathbf{w}_{WF} = \left[\begin{array}{c} \mathbf{w}_1^{SVD} \\ \mathbf{w}_2^{SVD} \end{array} \right] \tag{9}$$

3. REAL TIME IMPLEMENTATION

The real time implementation of the SVD-based technique is illustrated in figure 2. Four steps are necessary to compute the filter coefficients in real time:

• <u>*Step* 1</u> : The VAD discriminates the *speech-and-noise periods* from the *noise periods* of the noisy speech signals. The VAD used in this study is based on the log-energy of the signal [2]. The log-energy of the signal is computed with an overlap method on 128 samples. The decision of the VAD is taken from the computation of two thresholds namely, *Tspeech* and *Tnoise*. *Tspeech* and *Tnoise* are computed from the statistics of the signal (the mean and the variance). The function *Signal* equals the log-energy when the energy of the signal increases, and drops with an exponential curve when the energy dropps. A function *Offset* preserves the *VAD*=1 during a number of samples when a *noise period* is detected. In this way, a *speech-and-noise period* is still identified when there is a silence in a word or a sentence. With these different thresholds, the VAD works as follows:

- if *Signal* $>$ *Tspeech*, a *speech-and-noise period* is detected, *VAD*=1.

- if *Tnoise* $>$ *Signal* and *Offset*=1, a *noise period* is detected but *VAD*=1.

- if *Tnoise* $>$ *Signal* and *Offset*=0, a *noise period* is detected, *VAD*=0.

• <u>*Step* 2</u> : Classification errors between the *speech-and-noise periods* and the *noise periods* occur with the VAD. If the *speech-and-noise periods* are wrongly classified, speech-and-noise vectors are added to the noise matrix ($\mathbf{N}_k$). In this case, the factor $F = 1 - \eta_i^2/\sigma_i^2$ of the filter $\mathbf{W}_{WF}$ tends to be small ($\sigma_i^2 \to \eta_i^2$), resulting in signal cancellation. Since F varies in time, the gradient G of this factor can be measured during the processing:

$$G = \frac{\delta(1/N.\sum_{i=1}^{N}(1 - \eta_i^2/\sigma_i^2))}{\delta t} \tag{10}$$

If the gradient G is below a given threshold β, this means that the VAD detects *speech-and-noise periods* instead of *noise periods*. Then, a correction is made to the VAD and the decision made in *Step 1* is modified. Otherwise, when $G > \beta$, the decision made in *Step* 1 is kept valid.

• <u>*Step* 3</u> : A recursive technique is used to approximate the SVD-based optimal filtering technique. This technique is based on a Jacobi-type GSVD-updating algorithm [6]. Recursive GSVD-updating algorithms use the decomposition of the GSVD at time $k-1$ to compute the decomposition at time k. The equation 2 at time $k-1$ can be rewritten as:

$$\left\{ \begin{array}{lcl} \mathbf{U}_{k-1} & = & \mathbf{Y}_{k-1} \cdot \mathbf{R}_{U,k-1} \cdot \mathbf{X}_{k-1}^T \\ \mathbf{N}_{k-1} & = & \mathbf{V}_{k-1} \cdot \mathbf{R}_{N,k-1} \cdot \mathbf{X}_{k-1}^T \end{array} \right. \tag{11}$$

where $\mathbf{R}_{U,k-1} \in \mathbb{R}^{N \times N}$ and $\mathbf{R}_{N,k-1} \in \mathbb{R}^{N \times N}$ are upper triangular matrices having parallel rows and $\mathbf{X}_{k-1} \in \mathbb{R}^{N \times N}$ is an orthogonal matrix. For the computation, only $\mathbf{R}_{U,k-1}$, $\mathbf{R}_{N,k-1}$ and $\mathbf{X}_{k-1}$ are stored. When a new data vector $\mathbf{u}_k$ (speech-and-noise) or $\mathbf{n}_k$ (noise) is present at time k, the GSVD of $\mathbf{U}_k$ and $\mathbf{N}_k$ need to be recomputed as

$$\mathbf{U}_k = \left[\begin{array}{c} \lambda_s \cdot \mathbf{U}_{k-1} \\ \mathbf{u}_k \end{array} \right] \quad or \quad \mathbf{N}_k = \left[\begin{array}{c} \lambda_n \cdot \mathbf{N}_{k-1} \\ \mathbf{n}_k \end{array} \right] \tag{12}$$

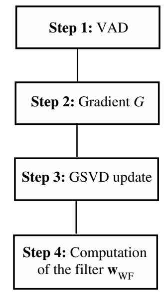

Fig. 2. Real time implementation of the SVD-based optimal filtering technique.

where λ_s and λ_n are exponential weighting factors for speech and noise matrix, respectively. For details on the updating scheme, the reader is referred to [6].

• $\underline{Step\ 4}$: This step consists of computing the optimal filter $\mathbf{w}_{WF,k}$ after the update of the recursive GSVD-updating algorithm. Substituting formulae (11) into (1), the equation can be rewritten as:

$$\mathbf{W}_{WF,k} = \mathbf{X}_k.\mathbf{R}_{U,k}^{-1}. \\ diag\left\{ \frac{(1-\lambda_n^2).(\mathbf{R}_{U,k}^{ii})^2 - (1-\lambda_s^2).(\mathbf{R}_{N,k}^{ii})^2}{(1-\lambda_n^2).(\mathbf{R}_{U,k}^{ii})^2 + (\mu-1).(1-\lambda_s^2).(\mathbf{R}_{N,k}^{ii})^2} \right\}. \\ \mathbf{R}_{U,k}.\mathbf{X}_k^T \tag{13}$$

The factor p/q is replaced by $(1-\lambda_n^2)/(1-\lambda_s^2)$. Only one column (the $i-th$ column, $\mathbf{w}_{WF,k}^i$ of $\mathbf{W}_{WF,k}$) is computed as the solution of the linear set by a back-substitution method. In our experiments, the speech distortion parameter μ is set to 1.75.

4. METHODS

4.1. Hearing aids

The hearing aid was a prototype based on a Cochlear Nucleus behind-the-ear headset housing. One hardware directional microphone (Microtronic 6001), as front microphone, and one omnidirectional microphone (Knowles FG-3452), as rear microphone, were mounted in an endfire array configuration. The hardware directional microphone had a cardioid spatial characteristic (null at 180^o) in anechoic conditions. The distance between the front entry port and the back entry port of the hardware directional microphone was 1cm. The distance between the front entry port of the hardware directional microphone and the omnidirectional microphone was 2.5cm.

4.2. Physical evaluation

In general, several signals are available to the VAD, such as the signal of the omnidirectional microphones, the directional microphone or even the output of the noise reduction technique. In this study, the behaviour of the VAD is evaluated when the VAD is connected to these different signals. When the VAD algorithm is connected to the omnidirectional microphone or the directional microphone, the signals are directly available. When the VAD is connected to the output of the strategy, the signals are only available after a first update of the adaptive filters. The SVD-based technique needs at least a *noise period* and a *speech-and-noise period*. To solve this problem of initialization, the VAD is connected first to the directional microphone and when several samples are classified as *speech-and-noise periods* or *noise periods*, the optimal filters are updated. Only then, the VAD algorithm is connected to the output of the SVD-based strategy. The performance of the VAD is evaluated by calculating the percentage correctly detected samples by the VAD algorithm for *speech-and-noise periods* and *noise periods* of the signals. The percentage (Per) is calculated as:

$$Per = \frac{SN_{RealTime} \times 100}{SN_{Perfect}} \qquad Per = \frac{N_{RealTime} \times 100}{N_{Perfect}} \tag{14}$$

where $N_{Perfect}$ and $SN_{Perfect}$ are the number of samples, which are known to be classified as *noise periods* (N) or *speech-and-noise periods* (SN) by the 'perfect' VAD. $N_{Realtime}$ and $SN_{Realtime}$ are the number of samples which are correctly classified as *noise periods* or *speech-and-noise periods* by the real time VAD. The signals of the speech signals (0^o) and the noise signal (90^o) are recorded when the hearing aid is positioned on a dummy head. The signals are recorded during 90 seconds. In the calculation, the first 20 seconds of the signals are not taken in account. This is the time needed to the noise reduction algorithm to converge.

4.3. Perceptual evaluation

The perceptual evaluation was performed with ten normal hearing listeners by measuring the Speech Reception Threshold (SRT) of sentences in a stationary speech weighted noise, with an adaptive procedure [7]. The tests of the omnidirectional microphone and the adaptive beamformer [2] were carried out in two different noise scenarios in a moderately reverberant room ($T_{60} = 0.76s$). A first, where the speech source was at an angle of 0^o (in front of the mannequin) and the noise source at 90^o, and a second, where the speech source was at 45^o and three independent noise sources were at 90^o/180^o/270^o. The distance between the loudspeakers and the center of the mannequin was 1 meter. The SVD-based technique was compared to an adaptive beamformer technique, which was known to give significant improvements in speech intelligibility [2].

5. RESULTS

5.1. Physical evaluation

Figure 3 shows the results of the percentage (Per) correctly detected samples by the VAD algorithm for *speech-and-noise periods* and *noise periods* in a stationary speech weighted noise. The VAD algorithm detected correctly the *noise-only periods* when it was connected to the omnidirectional microphone, the directional microphone or the output of the noise reduction strategy ($Per > 90$). The detection performance for the *speech-and-noise periods* was clearly a function of the signal to which the VAD was connected. The performance of the VAD dropped significantly when it was linked to the omnidirectional or directional microphone for a SNR below 5dB. When the VAD used the output signal of the SVD-based technique, the percentage of well-detected samples stayed above 90% for a SNR above -5dB. At a SNR of -10dB, the scores were about 90% with the optimal filtering technique. Connecting the VAD to the output of the noise reduction algorithm revealed the best performance. In this study, the VAD was connected to the output of the noise reduction strategy for the real time implementation.

5.2. Perceptual evaluation

Figure 4 shows the SRT-improvements (in dB) of the two noise reduction algorithms (SVD-based optimal filtering technique versus adaptive beamformer [2]) relative to the omnidirectional microphone, for both jammer sound scenarios. To compare the performance of the noise reduction techniques between each other, a statistical analysis (a paired comparison) was performed for the two noise scenarios. In the single jammer sound scenario, important SRT-improvements were obtained, 15.8dB and 15.1dB, for the adaptive beamformer and the optimal filtering technique respectively. There were no significant differences between both strategies (p=0.103). This means that the SVD-based technique

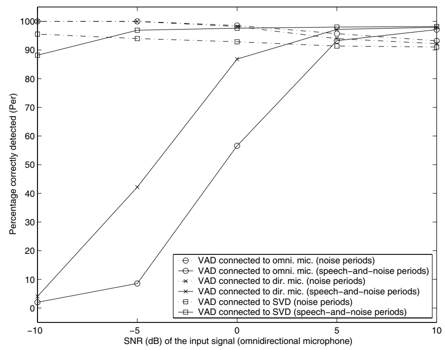

Fig. 3. Performance of the VAD when it is connected to the omnidirectional microphone, the directional microphone, the output of the SVD-based technique.

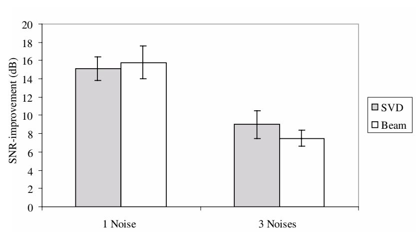

Fig. 4. SRT-improvements (in dB) of the SVD-based optimal filtering technique (SVD) and the adaptive beamformer (Beam) relative to the omnidirectional microphone for both jammer sound scenarios

can perform as well as the adaptive beamformer when the noise scenario is optimal for the latter technique. Indeed, the desired target (speech at 0^o) was in the look direction of the beamformer (angle 0^o). In the multiple noise scenario, the SVD-based technique was significantly better than the adaptive beamformer when a stationary speech weighted noise was present (p=0.005). SRT-improvements of 7.5dB and 9.0dB were obtained with the adaptive beamformer and the optimal filtering technique, respectively. The difference between the two strategies (1.5dB) is important for hearing-aid users. In critical listening conditions (close to 50% of speech understood by the listener) an improvement of 1dB in SNR corresponds to an increase of speech understanding of about 15 per cent in every day speech communication [3].

On one hand, the SVD-based optimal filtering technique works without assumptions about the desired target direction, however, this strategy needs a robust VAD. On the other hand, the adaptive beamformer works with assumptions about the desired target direction and the characteristics of the microphones. When these assumptions are violated, it leads to a leakage of the speech signal into the noise reference. If then the VAD misclassifies the *speech-and-noise periods*, the adaptive filter takes in account the statistics of the desired signal and subsequent target cancellation.

In the multiple jammer sound scenario, the noise reduction strategies did not achieve the same performance as the single jammer sound scenario. The SRT-improvements decreased by about 8dB. Theoretically, a signal processing strategy comprising N microphones can potentially separate up to N statistically independent sources. More specifically, a configuration with two microphones is optimal for the cancellation of one jammer sound. The directional microphone is important in adverse listening conditions. In a diffuse listening environment (the jammer sources are not located in well defined directions), the adaptive effect of the noise reduction strategies falls back to the effect of the directional microphone [2].

SVD-based procedures are known to have a high computational complexity, but, recent studies showed that the complexity problem can be controlled, making this approach attractive for practical systems. Recently, a LMS approach was found to have approximately the same cost of calculation as the adaptive beamformer [8].

6. CONCLUSIONS

A real time implementation and an evaluation of a Singular Value Decomposition (SVD) based optimal filtering technique for noise reduction in a dual microphone BTE hearing aid is presented. Connecting the VAD to the output of the noise reduction algorithm reveals a good performance for discriminating the *speech-and-noise periods* from the *noise periods*. Perceptual measurements showed that the optimal filtering technique is more robust than the adaptive beamformer in a multiple noise source scenarios and could perform as well as the latter technique in a single jammer sound scene.

7. REFERENCES

[1] S Doclo and M Moonen, "Gsvd-based optimal filtering for single and multiple speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, 2002.

[2] J B Maj, J Wouters, and M Moonen, "Noise reduction results of an adaptive filtering technique for dual-microphone behind-the-ear hearing aids," *Ear and Hearing*, vol. In revision, 2003.

[3] R Plomp, "Noise, amplification, and compression: considerations of three main issues in hearing aid design," *Ear and Hearing*, vol. 15, no. 1, pp. 2–12, 1994.

[4] J B Maj, M Moonen, and J Wouters, "Theoretical analysis of adaptive noise reduction algorithms for hearing aids," *European Signal Processing Conference (EUSIPCO)*, vol. September 3-6, pp. Toulouse, France, 2002.

[5] J B Maj, J Wouters, and M Moonen, "Svd-based optimal filtering technique for noise redcution in hearing aids using two microphones," *Journal on Applied Signal Processing*, vol. 4, pp. 432–443, 2002.

[6] M Moonen, P VanDooren, and J Vandewalle, "A singular value decomposition updating algorithm for subspace tracking," *SIAM Journal of Matrix Anal. Application*, vol. 13, no. 4, pp. 1015–1038, 1992.

[7] N Versfeld, L Daalder, J M Festen, and T Houtgast, "Extension of sentence materials for the measurement of the speech reception threshold," *Journal of the Acoustical Society of America*, vol. 107, no. 3, pp. 1671–1684, 2000.

[8] A Spriet, M Moonen, and J Wouters, "Spatially preprocessed, speech distortion weighted multi-channel wiener filtering for noise reduction," *Submitted*, 2003.