THEORETICAL ANALYSIS OF ADAPTIVE NOISE REDUCTION ALGORITHMS FOR HEARING AIDS

Jean Baptiste Maj[1,2], *Marc Moonen*[1] *and Jan Wouters*[2]
[1]ESAT-SISTA KULeuven, Kasteelpark Arenberg 10, 3001 Leuven Belgium
[2]Lab.Exp.ORL KULeuven, Kapucijnenvoer 33, 3000 Leuven Belgium
e-mail: Jean-Baptiste.Maj@uz.kuleuven.ac.be

ABSTRACT

This paper presents the theoretical analysis of 2 different adaptive noise reduction algorithms for twin-microphone hearing aids. A first noise reduction algorithm is based on a beamformer technique [1] and a second is based on optimal filtering and singular value decomposition (SVD) [2]. On the one hand, it is shown that the SVD-based technique gives a robust solution against unmatched microphone characteristics. On the other hand, the beamformer technique has a better robustness against voice activity detector (VAD) errors. [1]

1 INTRODUCTION

In speech processing, it is generally assumed that the recorded signal equals $\mathbf{u} = \mathbf{s} + \mathbf{n}$, where $\mathbf{s}$ is the speech part and $\mathbf{n}$ is the noise part. Furthermore, the speech signal has two distinct signal conditions, leading to periods where only background noise is present and periods where speech and noise are present. This paper presents an evaluation of 2 adaptive noise reduction algorithms, a beamformer technique [1] and a SVD-based optimal filtering technique [2]. The noise reduction algorithms estimate the statistic of the noise during noise periods and subtract the noise from the speech plus noise signal during speech-and-noise periods. To discriminate between these two periods, voice activity detection algorithms are used. The strategies of the noise reduction algorithms are shown *Figure 1 and 2*.
The hearing aid contains two omnidirectional microphones which are used to create a software directional microphone. The software directional microphone parameters are the interport distance d, internal delay τ and the weight factor for the back port is $\beta(f) = a.e^{-j2\pi f\tau}$. The delay τ and the weight a have been chosen to give a hypercardioid spatial characteristic in anechoic conditions. The signals of the software directional microphone and the rear omnidirectional microphone are used as inputs to the noise reduction algorithms.

[1]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 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. The authors also thank GN ReSound for the dual microphone hearing aid prototypes.

2 TWO-STAGE ADAPTIVE BEAMFORMER

The two-stage adaptive beamformer (A2B) has two different signal processing stages (Figure 1). A first, where a filter $\mathbf{w}_1^{A2B}$ is fixed to give a specific look direction to the two-stage adaptive beamformer. In practice, this filter is trained in anechoic conditions with the direction of the desired signal at 0^o. A second filter filter stage $\mathbf{w}_2^{A2B}$ implements adaptive noise cancellation (ANC) and attempts to model noise during noise periods, and subtracts noise from speech plus noise when speech is present. The sum and subtraction (middle part of figure 1) improves the noise reference of the ANC. The additional delays actually allow to have non-causal filters.

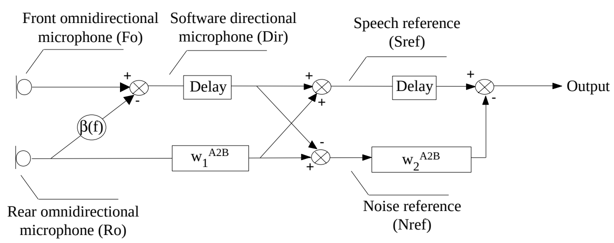

Figure 1: Scheme of the two-stage adaptive beamformer strategy.

3 SVD-BASED TECHNIQUE

In the single-microphone case, the SVD-based technique reconstructs the speech signal $\mathbf{s}_k$ from noisy data by means of a linear filter $\mathbf{W}_k$ using $\tilde{\mathbf{s}}_k = \mathbf{W}_k.\mathbf{u}_k$. In the sequel, $P_{x,x} = \mathcal{E}\{\mathbf{x}.\mathbf{x}^*\}$ is the spectral power density (PSD) of signal $\mathbf{x}$ and $P_{x,y} = \mathcal{E}\{\mathbf{x}.\mathbf{y}^*\}$ the cross-PSD of signals $\mathbf{x}$ and $\mathbf{y}$. The asterisk denotes complex conjugation and $\mathcal{E}\{.\}$ the expectation. Using a Minimum Mean Square Error criterion, the optimal filter $\mathbf{W}_{WF}$ is equal to:

$$\mathbf{W}_{WF} = P_{u,u}^{-1}.P_{u,s} \tag{1}$$

Assuming that the noise signal $\mathbf{n}$ is short-term stationarity and statistically independent of the speech signal $\mathbf{s}$ ($P_{n,s} = 0$), $\mathbf{W}_{WF}$ becomes:

$$\mathbf{W}_{WF} = P_{u,u}^{-1}.P_{s,s} = P_{u,u}^{-1}.(P_{u,u} - P_{n,n}) \tag{2}$$

In a multiple microphone application, $\mathbf{u} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_M]^T$ where $\mathbf{u}_i$ is a vector containing successive time samples of microphone signal i ($\mathbf{n}$ is similarly defined). In our case, the microphone inputs are the software directional microphone (Dir) and the rear omnidirectional microphone (Ro) (see figure 2). $P_{u,u}$ and $P_{n,n}$ become:

$$P_{u,u} = \begin{bmatrix} P^{speech}_{Dir,Dir} + P^{noise}_{Dir,Dir} & P^{speech}_{Dir,Ro} + P^{noise}_{Dir,Ro} \\ P^{speech*}_{Dir,Ro} + P^{noise*}_{Dir,Ro} & P^{speech}_{Ro,Ro} + P^{noise}_{Ro,Ro} \end{bmatrix}^T \quad (3)$$

$$P_{n,n} = \begin{bmatrix} P^{noise}_{Dir,Dir} & P^{noise}_{Dir,Ro} \\ P^{noise*}_{Dir,Ro} & P^{noise}_{Ro,Ro} \end{bmatrix}^T \quad (4)$$

The computation of the optimal filter $\mathbf{W}_{WF}$ provides estimators $\mathbf{w}$ for the different signals $\tilde{\mathbf{s}}_k$, $[\tilde{\mathbf{s}}_1\ \ \tilde{\mathbf{s}}_2]^T = \mathbf{W}_k^T.[\mathbf{u}_1\ \ \mathbf{u}_2]^T$. Maj *et al.* [3] have shown that using the first column of $\mathbf{W}_{WF}$ gives a good estimate of $\mathbf{s}_1$, $[\mathbf{w}_1^{SVD}\ \ \mathbf{w}_2^{SVD}].[\mathbf{u}_1\ \ \mathbf{u}_2]^T = \tilde{\mathbf{s}}_1$. The strategy of the SVD-based technique is depicted in figure 2.

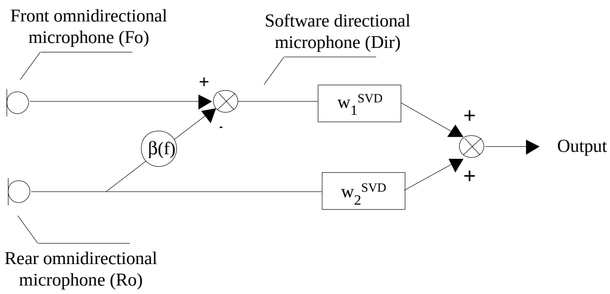

Figure 2: Scheme of the SVD-based technique strategy

4 THEORETICAL ANALYSIS

The theoretical analysis can be performed by using the complex coherence function (CCF) between the two omnidirectional microphones (Fo and Ro) [4] and the PSD at the output of the noise reduction algorithms. The CCF between two omnidirectional microphones is expressed by:

$$\Gamma_{Fo,Ro}(f) = \frac{P_{Fo,Ro}(f)}{\sqrt{P_{Fo,Fo}(f).P_{Ro,Ro}(f)}} \quad (5)$$

$\Gamma_{Fo,Ro}(f) = exp(j.2.\pi.f.cos(\theta).d/c)$ with a source located at angle θ. The CCF for a speech source at the endfire angle 0^o and with one noise source at broadside 90^o become:

$$\Gamma^{speech}_{Fo,Ro}(f) = exp(j.2.\pi.f.d/c) \quad \Gamma^{noise}_{Fo,Ro}(f) = \frac{1}{1+\rho(f)} \quad (6)$$

Where $\rho(f)$ is the sensor-to-environmental noise ratio, d the distance between the two microphones and c the velocity of the sound in air ($c \approx 340m/s$). In the next paragraphs, the PSD at the output of the software directional microphone and the noise reduction algorithms are defined. It is assumed the PSD of the received signals at the microphone inputs is the same ($P^{noise}_{Ro,Ro}(f) = P^{noise}_{Fo,Fo}(f) = P^{noise}_{In,In}(f)$ and $P^{speech}_{Ro,Ro}(f) = P^{speech}_{Fo,Fo}(f) = P^{speech}_{In,In}(f)$). Finally, $\alpha(f)$ is defined as the signal-to-noise ratio at the omnidirectional microphone $P^{noise}_{In,In}(f) = \alpha(f).P^{speech}_{In,In}(f)$. The performances of the noise reduction algorithms are function of the distance (d) between the two microphones, the angle (θ) of the speech and the noise sources, the sensor-to-environmental noise ratio ($\rho(f)$) and the signal-to-noise ratio ($\alpha(f)$).

It is also feasible to evaluate the performance in the case of unmatched microphones characteristics (in gain and phase) and VAD errors. To study the unmatched microphone case, a deviation in gain ($gain$) and in phase ($phase$) is brought to the rear omnidirectional microphone (Ro) such as $Ro_{dev}(f) = Ro(f).gain.e^{j.phase.\pi/180}$ The VAD discriminates noise periods from speech-and-noise periods. When VAD errors are present, the statistic of the speech (or the noise) signal is corrupted by the noise (or the speech) signal. The errors of the VAD are modeled by:

$$P^{noise}_{in,in}(f) = (1-coef).P^{noise}_{in,in}(f) + coef.P^{speech}_{in,in}(f) \quad (7)$$

$$P^{speech}_{in,in}(f) = (1-coef).P^{speech}_{in,in}(f) + coef.P^{noise}_{in,in}(f) \quad (8)$$

where $coef$ denotes the degree of the corruption. When $coef = 0$, there is a perfect voice activity detection.

4.1 Directional Microphone

The PSD and cross-spectral density as a function of the CCF at the output of the software directional microphone (Dir) (see figure 1 and 2) are:

$$\begin{aligned} P_{Dir,Dir}(f) &= P_{in,in}(f).(1 - 2.Re(\beta^*(f).\Gamma_{Fo,Ro}(f)) \\ &\quad + \beta(f).\beta^*(f)) \\ P_{Dir,Ro}(f) &= P_{in,in}(f).(\Gamma_{Fo,Ro}(f) - \beta(f)) \end{aligned} \quad (9)$$

4.2 Two-Stage Adaptive Beamformer

The first filter $\mathbf{w}_1^{A2B}$ is kept fixed, under the assumption that the speaker is always in front of the listener. In fact, a specific look direction is given to the two-stage adaptive beamformer, namely the direction of the desired signal, e.g. at 0^o. The filter equals $\mathbf{w}_1^{A2B}(f) = P^{speech}_{Dir,Ro}(f)/P^{speech}_{Ro,Ro}(f)$

$$\begin{aligned} P_{Sref,Sref} &= P_{Dir,Dir}(f) + |\mathbf{w}_1(f)|^2.P_{Ro,Ro}(f) \\ &\quad + 2.Re(\mathbf{w}_1^*(f).P_{Dir,Ro}(f)) \end{aligned} \quad (10)$$

$$\begin{aligned} P_{Nref,Nref}(f) &= P_{Dir,Dir}(f) + |\mathbf{w}_1(f)|^2.P_{Ro,Ro}(f) \\ &\quad - 2.Re(\mathbf{w}_1^*(f).P_{Dir,Ro}(f)) \end{aligned} \quad (11)$$

$$\begin{aligned} P_{Sref,Nref}(f) &= P_{Dir,Dir}(f) - |\mathbf{w}_1(f)|^2.P_{Ro,Ro}(f) \\ &\quad + 2.Im(\mathbf{w}_1^*(f).P_{Dir,Ro}(f)) \end{aligned} \quad (12)$$

The second filter is adapted during noise periods and equals: $\mathbf{w}_2^{A2B}(f) = P^{noise}_{Sref,Nref}(f)/P^{noise}_{Nref,Nref}(f)$.

$$\begin{aligned} P_{Out,Out}(f) &= P_{Sref,Sref(f)} + |\mathbf{w}_2^{A2B}(f)|^2.P_{Nref,Nref}(f) \\ &\quad - 2.Re(\mathbf{w}_2^{A2B*}(f).P_{Sref,Nref}(f)) \end{aligned} \quad (13)$$

4.3 SVD-Based Technique

From paragraph 3, the two-channel estimator $\mathbf{w} = [\mathbf{w}_1^{SVD}\ \ \mathbf{w}_2^{SVD}]^T$ is given by:

$$\begin{aligned} \mathbf{w}_1^{SVD} &= 1/Det.(P^{speech}_{Dir,Dir}.(P^{speech}_{Ro,Ro} + P^{noise}_{Ro,Ro}) \\ &\quad - P^{speech}_{Dir,Ro}.(P^{speech*}_{Dir,Ro} + P^{noise*}_{Dir,Ro})) \end{aligned} \quad (14)$$

$$\begin{aligned} \mathbf{w}_2^{SVD} &= 1/Det.(P^{speech}_{Dir,Ro}.(P^{speech}_{Dir,Dir} + P^{noise}_{Dir,Dir}) \\ &\quad - P^{speech}_{Dir,Dir}.(P^{speech}_{Dir,Ro} + P^{noise}_{Dir,Ro})) \end{aligned} \quad (15)$$

where Det corresponds to the determinant of the matrix $P_{u,u}$. The PSD at the output of the SVD-based technique corresponds to:

$$\begin{aligned} P_{Out,Out}(f) &= |\mathbf{w}_1^{SVD}(f)|^2.P_{Dir,Dir} + |\mathbf{w}_2^{SVD}(f)|^2.P_{Ro,Ro} \\ &\quad + 2.Re(\mathbf{w}_1^{SVD}(f).P_{Dir,Ro}(f).\mathbf{w}_2(f)^{SVD*} \end{aligned} \quad (16)$$

5 PERFORMANCE

To evaluate the theoretical performance of the noise reduction algorithms, performance measures, namely *noise reduction* $NR(f) = P^{noise}_{Out,Out}(f)/P^{noise}_{In,In}(f)$ and *speech conservation* $SC(f) = P^{speech}_{Out,Out}(f)/P^{speech}_{In,In}(f)$ in the frequency domain (f) are used. To evaluate the improvement of the speech intelligibility, a performance metric G_{AI} has been developed, which is based on an averaged intelligibility gain. The improvement of the speech intelligibility is estimated between the input, the omnidirectional microphone (Fo) in our case, and the output of the noise reduction algorithm:

$$G_{AI} = SNR_{weigthed,output} - SNR_{weigthed,input} \qquad (17)$$

where $SNR_{weigthed} = \sum_{i=1}^{k} I_i.(SC_i - NR_i)$. $SC(f)$ and $NR(f)$ are decomposed in k-th third octave bands and for each frequency band, I_i weights are applied as defined in the speech intelligibility index [5]. The G_{AI} does not give information about the level difference between the input speech signal and the output speech signal of the strategy. An average spectrum level difference measure D is introduced:

$$D = \frac{2}{Nfft} \sum_{k=1}^{\frac{Nfft}{2}} 10.log_{10}|P^{speech}_{k,output} - P^{speech}_{k,Dir}| \qquad (18)$$

where $Nfft$ is the FFT length (=256), $P^{speech}_{k,output}$ and $P^{speech}_{k,input}$ the PSD of the input and the output of the speech signal respectively. $D = 0$ corresponds to no spectral level difference.

6 RESULTS

Simulations have been carried out by varying the different parameters such as the distance between the two microphones (d), the sensor-to-environmental noise ratio $(\rho(f))$ and the signal-to-noise ratio $(\alpha(f))$. These experiments have been performed under the assumption that the two omnidirectional microphones (Fo and Ro) are matched ($gain = 1$ and $phase = 0$) and that no VAD errors occur($coef = 0$). It appears there are no significant differences (in G_{AI} and D) between the behaviour of the two-stage adaptive beamformer and the SVD-based technique. The longer the distance between the two microphones, the better the G_{AI}. The higher the sensor-to-environmental noise ratio, the worse the G_{AI}. Finally, the higher the signal-to-noise ratio $\rho(f)$ the better the G_{AI}. The value of the performance measure D is always around 0dB during these experiments. This means that the two-stage adaptive beamformer and the SVD-based technique do not bring in spectral level difference of the speech signal during the noise reduction processing.

Figures 3 and 4 show the behaviour of the noise reduction algorithms with phase and gain deviations ($d = 2cm$, $\rho = -45dB$ and $\alpha = 0dB$). The phase deviation decreases the G_{AI} but has no impact on D for both noise reduction algorithms. With a gain deviation, the performance of the G_{AI} of the two-stage adaptive beamformer decreases a lot, on the other hand, the SVD-based technique technique seems to be robust against gain deviations.

Figure 5 shows the $SC(f)$ and the $NR(f)$ measures of the beamformer technique as a function of frequency for gain deviations. When the gain equals 0.8 or 1.2, the $NR(f)$ decreases. Furthermore, an additional distortion on the $SC(f)$ at the low frequencies is brought when the gain is 1.2.

Figure 6 shows the effect of the VAD errors for both algorithms with no deviation in phase and gain ($d = 2cm$, $\rho = -45dB$ and $\alpha = 0dB$) as a function of $coef$. These experiments have been carried out for 4 different positions of the speaker and the noise statistic is corrupted by the speech signal (see equation 7). The G_{AI} performance of the SVD-based technique is independent of the speaker position but drops rapidly in function of the VAD errors. For the two-stage adaptive beamformer, the G_{AI} performance depends on the speaker position. Indeed, when the speaker is at the angle 0^o, the performance of the beamformer technique is not affected. However, when the speaker is not at the angle 0^o, the performance of the beamformer also decreases as a function of the VAD errors but not as drastically as for the SVD-approach. If the speaker is not positioned at the look direction (angle 0^o) of the beamformer, a leakage of the speech signal into the noise reference is obtained (*figure 7*). With VAD errors, the estimate of the noise statistics by $\mathbf{w}_2^{A2B}$ are corrupted by the statistic of the speech signal and brings a cancellation of the speech signal at the output of the beamformer. When the speech statistic is corrupted (see equation 8), it is found that there is no influence on the beamformer and the SVD-based techniques performances.

Finally, a last experiment combines a gain deviation ($gain =$

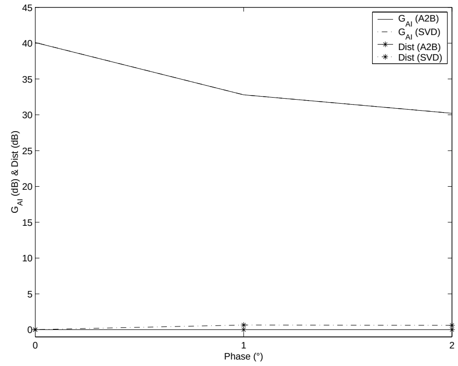

Figure 3: Influence of the phase deviation of the microphones on the improvement of the speech intelligibility G_{AI} and the distortion $Dist$ of the noise reduction algorithm. ($\alpha = 0$dB, $\rho = -45$dB, $d = 2cm$).

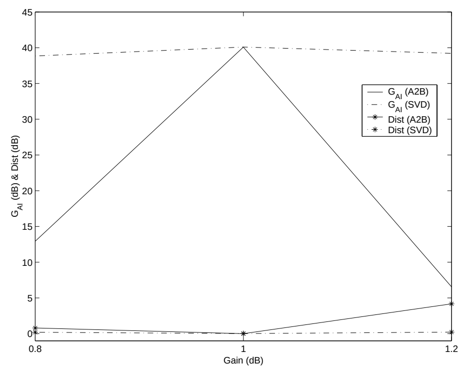

Figure 4: Influence of the gain deviation of the microphones on the improvement of the speech intelligibility G_{AI} and the distortion $Dist$ of the noise reduction algorithm. ($\alpha = 0$dB, $\rho = -45$dB, $d = 2cm$).

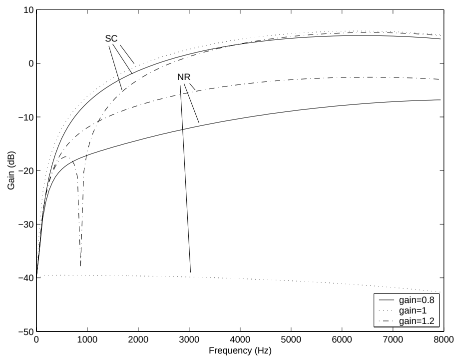

Figure 5: Influence of the gain deviation of the microphones on the noise reduction $NR(f)$ and speech conservation $SC(f)$ of the A2B ($\alpha = 0$dB, $\rho = 45$dB, $d = 2cm$).

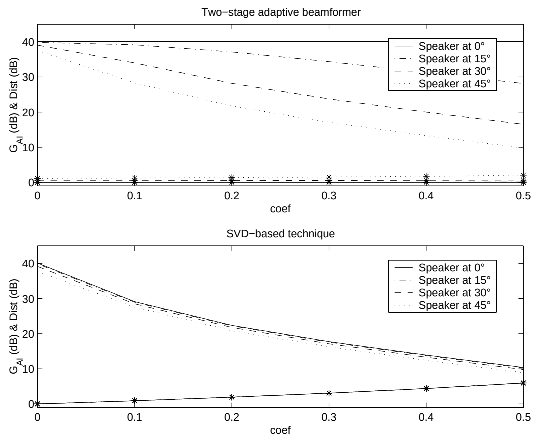

Figure 6: Influence of unperfect VAD on the improvement of the speech intelligibility G_{AI} and the distortion $Dist$ of the noise reduction algorithm. $(gain = 1)$.

0.8) of Ro with VAD errors (*figure 8*). The SVD-based technique keeps the same performance as in the last experiment, unlike the beamformer technique. Indeed, the performance drops significantly, and the beamformer has roughly the same behaviour as the SVD-based technique. Furthermore, the higher the angle of the speaker, the worse the performance of the beamformer. For all these cases, the spectral level difference of the speech signal D is practically not affected.

7 CONCLUSION

In this work we theoretically evaluated 2 noise reduction processing strategies for application in dual-microphone hearing aids. The SVD-based technique is very robust against gain variations of the microphones, unlike the two-stage adaptive beamformer. Moreover, the adaptive beamformer approach works with assumptions about the look direction, the direction of visual contact. The necessity of a robust voice activity detection is important and enhances considerably the effectiveness of the SVD noise reduction technique.

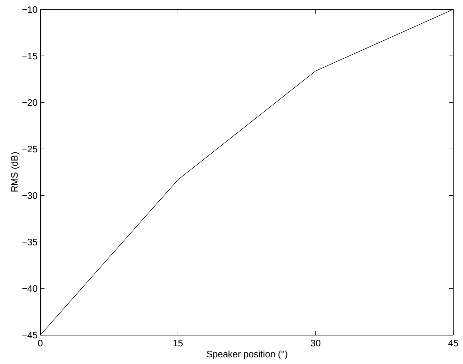

Figure 7: Level (dB) of the speech signal in the noise reference of the two-stage adaptive beamformer in function of the speaker position.

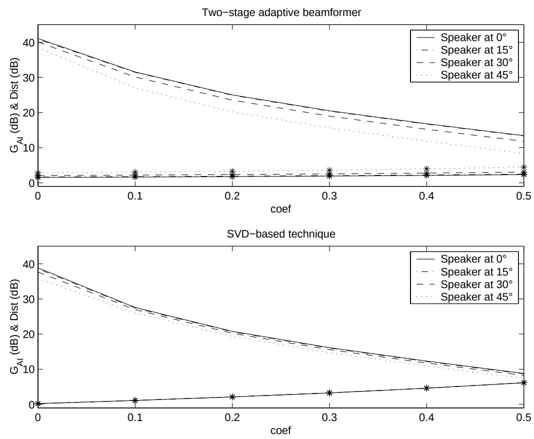

Figure 8: Influence of unperfect VAD on the improvement of the speech intelligibility G_{AI} and the distortion $Dist$ of the noise reduction algorithm. $(gain = 0.8)$.

8 REFERENCES

[1] Maj JB, Wouters J, Moonen M. *Noise Reduction Results of an Adaptive Filtering Technique for Behind-The-Ear Hearing Aids with 2 Microphones.* Ear & Hear. (submitted).

[2] Doclo S, Moonen M. *Microphone Arrays: Signal Processing Techniques and Apllications.* Springer-Verlag. Chap.6, pp111-132, 2001.

[3] Maj JB, Moonen M, Wouters J. *SVD-based optimal filtering for noise reduction in hearing aids using two microphones.* Applied Signal Processing. (In revision).

[4] Spriet, A. Wouters, J. Moonen, M.: *Robustness Analysis of GSVD Based Optimal Filtering and Generalized Sidelobe Canceller for Hearing Aid Applications.* WASPAA, New Paltz, 21-24 October 2001.

[5] American National Standard. *Methods for Calculation of the Speech Intelligibility Index.* Acoust. Soc. Am. ANSI S3.3-1997, 1997.