A MULTICHANNEL SUBBAND GSVD BASED APPROACH FOR SPEECH ENHANCEMENT IN HEARING AIDS

Ann Spriet[1,2], *Marc Moonen*[1], *Jan Wouters*[2]

[1]K.U. Leuven, ESAT-SISTA
Kasteelpark Arenberg 10
B-3001 Leuven-Heverlee, Belgium
{spriet,moonen}@esat.kuleuven.ac.be

[2]K.U. Leuven, Lab. Exp. ORL
Kapucijnenvoer 33
B-3000 Leuven, Belgium
jan.wouters@uz.kuleuven.ac.be

ABSTRACT

The major goal of a noise reduction algorithm for hearing aid applications is to improve speech intelligibility. In [1], a multichannel noise reduction technique, based on a Generalized Singular Value Decomposition (GSVD) has been proposed. The GSVD based filter minimizes the Mean Square Error (MSE) between the desired signal portion in the received signals and a filtered sum of the received microphone signals. In this paper, we propose a subband implementation of the GSVD based filter. It is shown that - in case of coloured signals - the subband implementation improves intelligibility more than the fullband approach. In addition, a significant complexity reduction is achieved.

1. INTRODUCTION

Noise reduction algorithms are crucial to improve the speech intelligibility for hearing impaired people in background noise and/or reverberation. Since the disturbing signal is often also speech like, multi-microphone algorithms are preferred to single microphone procedures. The most common adaptive multi-microphone noise reduction technique is the generalized sidelobe canceller (GSC) [2]. The GSC assumes the position of the desired source, the microphone characteristics and positions to be known. Deviations from these assumptions are detrimental to its performance. In [1], a multichannel signal enhancement technique based on a Generalized Singular Value Decomposition (GSVD) has been proposed. The GSVD based optimal filter minimizes the Mean Square Error (MSE) between the desired signal portion in the received signals and the filtered sum of the received microphone signals. Since no assumptions are made about the location of the desired speaker, the microphone characteristics and positions, this algorithm is shown to be more robust than the GSC [1][3].

In this paper, we propose a subband implementation of the GSVD based filter. Section 2 briefly reviews the GSVD based optimal filter. The subband implementation is motivated and described in Section 3. Section 4 and Section 5 compare the computational cost and the performance of the fullband and subband approach. It is shown that the subband implementation improves intelligibility more than the fullband approach. In addition, a significant complexity reduction is achieved.

Ann Spriet is a Research Assistant supported by the Fonds voor Wetenschappelijk Onderzoek (FWO) - Vlaanderen. This research work was carried out at the ESAT laboratory and Lab. Exp. ORL of the K. U. Leuven, in the framework of the Concerted Research Action GOA-MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology) of the Flemish Government, IUAP P4-02 (1997-2001) 'Modeling, Identification, Simulation and Control of Complex Systems' and FWO Research Project nr. G.0233.1 ('Signal processing and automatic patient fitting for advanced auditory prostheses'). The scientific responsibility is assumed by its authors.

2. GSVD BASED OPTIMAL FILTER

Consider an array of M microphones and let $u_i[k] = d_i[k] + n_i[k]$ be the k-th signal sample at the i-th microphone, where $d_i[k]$ is the desired speech portion and $n_i[k]$ the environment plus internal noise. Define L as the filter length per channel. Construct the vector $\mathbf{u}_k = [\mathbf{u}_{1,k}^T \quad \mathbf{u}_{2,k}^T \cdots \mathbf{u}_{M,k}^T]^T$ with $\mathbf{u}_{i,k} = [u_i[k] \quad u_i[k-1] \cdots u_i[k-L+1]]^T$, and similarly the vectors $\mathbf{d}_k$ and $\mathbf{n}_k$. The optimal filter $\mathbf{W} \in \mathcal{C}^{ML\times ML}$ produces speech estimates $\hat{\mathbf{d}}_k = \mathbf{y}_k = \mathbf{W}^H\mathbf{u}_k$ such that the MSE diag $\left\{\varepsilon\left\{\mathbf{e}_k\mathbf{e}_k^H\right\}\right\}$ with $\mathbf{e}_k = \mathbf{d}_k - \mathbf{W}^H\mathbf{u}_k$ is minimized, and is given as:

$$\mathbf{W} = \varepsilon\{\mathbf{u}_k\mathbf{u}_k^H\}^{-1}\varepsilon\{\mathbf{u}_k\mathbf{d}_k^H\}. \tag{1}$$

This filter is approximated at time instant k by means of a GSVD of an input data matrix $\mathbf{U}_k \in \mathcal{C}^{p\times ML}$ and a noise data matrix $\mathbf{N}_k \in \mathcal{C}^{q\times ML}$, collecting respectively p speech + noise and q noise only signal vectors (corresponding to speech pauses) up to time instant k:

$$\left\{\begin{array}{lcl} \mathbf{U}_k &=& \mathbf{Q}_\mathbf{U}.\mathrm{diag}\{\sigma_i\}.\mathbf{X}^H \\ \mathbf{N}_k &=& \mathbf{Q}_\mathbf{N}.\mathrm{diag}\{\eta_i\}.\mathbf{X}^H \end{array}\right. \tag{2}$$

with $\mathbf{Q}_\mathbf{U}, \mathbf{Q}_\mathbf{N}$ orthogonal matrices, $\mathbf{X}$ an invertible matrix and $\frac{\sigma_i}{\eta_i}$ the generalized singular values:

$$\mathbf{W}_k^* \simeq \mathbf{X}^{-H}\mathrm{diag}\{\max\left(1-\frac{p}{q}\frac{\eta_i^2}{\sigma_i^2},0\right)\}\mathbf{X}^H. \quad (3)$$

Filtering $\mathbf{u}_k$ with the l-th column of $\mathbf{W}_k$ yields the estimate of the l-th element of $\mathbf{d}_k$,

$$\hat{d}_i[k-\Delta_l] = \mathbf{W}_k(:,l)^H\mathbf{u}_k, \quad (4)$$

with $i = \mathrm{mod}(l-1,L)+1$ and $\Delta_l = \mathrm{rem}(l-1,L)$. In practice, the GSVD is not calculated from scratch at each time k but updated recursively [1].

3. SUBBAND GSVD APPROACH

3.1. Improved intelligibility

In hearing aid applications, the major purpose of a noise reduction algorithm is to improve the speech intelligibility, rather than to maximize Signal-to-Noise Ratio (SNR) or to minimize MSE. The quality of a noise reduction algorithm is therefore assessed in terms of speech intelligibility weighted measures [4] such as the intelligibility weighted SNR, denoted as $\mathrm{SNR}_{\mathrm{intellig}}$:

$$\mathrm{SNR}_{\mathrm{intellig}} = \sum_i I_i \mathrm{SNR}_i, \quad (5)$$

where the band importance function I_i expresses the importance of the i-th one-third octave band for intelligibility [5] and SNR_i equals the SNR in dB in the i-th one-third octave band.
These measures, based on the articulation index, reflect the fact that distinct frequency bands contribute independently to intelligibility. Since the interferers and desired source are generally both speech like, minimizing the fullband MSE will weight the MSE more at the high energy low frequencies, resulting in a smaller improvement and larger distortion at higher frequencies (see also Section 5). However, at low frequencies, the achievable performance of a small-sized array is limited. In [6], a prewhitening operation is proposed to improve the intelligibility obtained with a GSC. Such prewhitening is intrinsically present in a subband implementation. Maximization of intelligibility can thus better be achieved by minimizing the MSE in separate subbands.

3.2. Uniform subband GSVD

The concept of subband GSVD is illustrated in Figure 1. The M microphone signals are each split into K frequency bands by a (nearly) Perfect Reconstruction (PR), oversampled uniform DFT modulated filterbank [7].These filterbanks can be implemented efficiently using a polyphase decomposition of the prototype filter H_0 and an FFT. Since the fullband signals are real and half of the DFT modulated analysis bank filters are the complex conjugate of the other half, only half of the subband signals need to be processed.

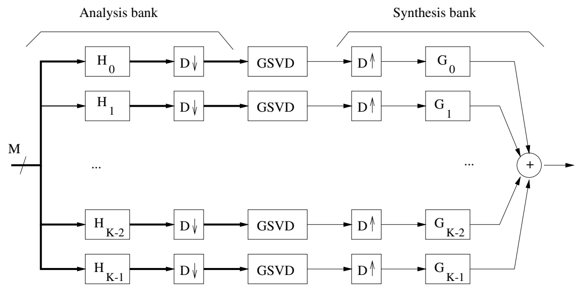

Figure 1: Concept of subband GSVD.

The optimal GSVD filter operates on the subband signals at the decimated sampling frequency $\frac{f_s}{D}$. To avoid audible interband aliasing, the downsampling factor D is kept smaller than the number of subbands K. In [8], it is shown that for a system identification setup, the aliasing level in the subbands corresponds to a nonlinear distortion and thus presents a lower bound for the minimum MSE in each subband. This indicates that -if the MSE is minimized in each subband independently (and thus no fullband adaptation criterion is used) - the amount of aliasing determines the performance and should thus be kept small.
A disadvantage of using a subband approach is the extra delay caused by the filterbank. In hearing aid applications, the total processing delay should be smaller than 20 ms in order not to degrade intelligibility due to asynchronism with lip reading [9]. The delay Δ caused by the filterbank equals [7]

$$\Delta = D\lceil\frac{L_p}{D}\rceil + D - 1 \quad (6)$$

samples at f_s, with L_p the length of the prototype filter. In the sequel, a sampling frequency $f_s = 16000$ Hz is used. To achieve a large stopband attenuation with a reasonable L_p and thus delay Δ, so-called nearly PR filterbanks [7] are used. Note that the perfect reconstruction property would be lost anyway due to the filtering operation on the subband signals.
In the sequel, two nearly PR oversampled DFT modulated filter banks are used: a 16 subband filter bank with $D = 12$, prototype filter length $L_p = 120$ and delay $\Delta = 8.2$ ms and a 32 subband filter bank with $D = 20$, $L_p = 240$ and $\Delta = 16.2$ ms.

4. COMPUTATIONAL COMPLEXITY

4.1. Fullband GSVD

Although approximate recursive GSVD algorithms are available [1], the computational complexity of a fullband GSVD based approach is still too high for hearing aid applications. The cost of a real, recursive square-root free GSVD implementation [1] equals

$$f_s\left(\frac{17.5}{d_g} + \frac{4}{d_f}\right)L^2M^2 \quad (7)$$

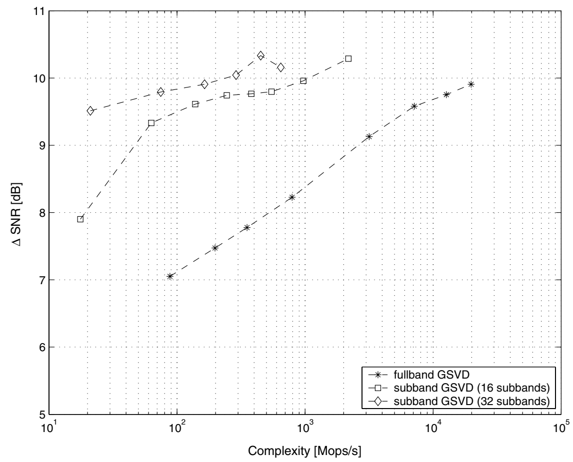

Figure 2: Improvement in SNR of fullband and subband GSVD as a function of complexity (Mops/s) for $M = 4$, $d_f = d_g = 1$.

operations (multiplications and additions) per second (ops/s). The factors d_g and d_f indicate the number of samples between respectively two GSVD and filter updates. They trade off convergence speed and cost [1].

4.2. Subband GSVD

The overall computational cost of the subband GSVD algorithm consists of the complexity of the filter bank and the complexity of the subband GSVD. When using a DFT modulated filterbank, only half the number of complex subband signals have to be processed. However, a complex recursive GSVD costs about 4 times as much as a real one. The subband GSVD filters thus cost about

$$2.K.\frac{f_s}{D}(\frac{17.5}{d_g} + \frac{4}{d_f}).L_{sub}^2 M^2 \tag{8}$$

ops/s, where L_{sub} equals the filter length per channel used in the subbands.
The implementation of a DFT modulated analysis or synthesis filter bank requires

$$(2L_p + 2K\log_2 K)\frac{f_s}{D} \tag{9}$$

ops/s. The total complexity of the filter bank implementation thus equals

$$(M+1)(2L_p + 2K\log_2 K)\frac{f_s}{D}. \tag{10}$$

In general, the cost of the filterbank will be negligible compared with the cost of the GSVDs. From (7) and (8), it follows that the subband GSVD will roughly be a factor $\frac{D^3}{2K}$ less complex if the same temporal window $L.D$, i.e. $L_{sub} = \frac{L_{full}}{D}$, is used. This amounts to roughly a factor 54 and 125, respectively, for the above given setups.

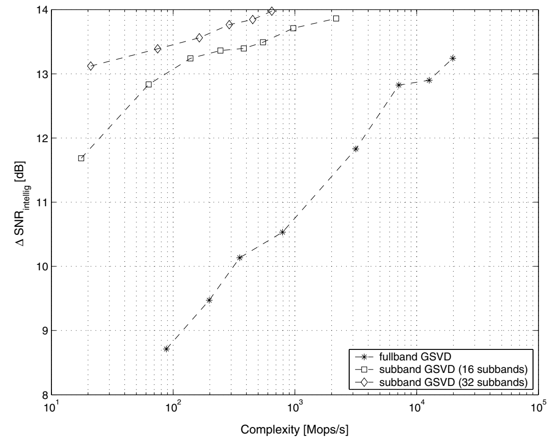

Figure 3: Improvement in $\text{SNR}_{\text{intellig}}$ of fullband and subband GSVD as a function of complexity (Mops/s) for $M = 4$, $d_f = d_g = 1$.

4.3. Comparison

Table 1 compares the complexity (expressed in Mops/s) of the recursive fullband and subband approach (16 and 32 subbands) with $d_f = d_g = 1$. The complexity of subband GSVD approaches the computational capacity available in current hearing aids (i.e. about 10 Mops/s). A further decrease in complexity can be obtained by reducing the number of filter and GSVD updates $\frac{1}{d_f}$ and $\frac{1}{d_g}$.

5. EXPERIMENTAL RESULTS

The performance of the fullband and the subband GSVD algorithm is evaluated based on recordings in a realistic environment. A linear endfire array with 4 omnidirectional microphones (Knowles EM-4368) and microphone interspacing $d = 0.02$ m has been mounted on a dummy head in an office room. The reverberation time T_{60dB} is about 700 ms for a speech weighted noise. The desired and interfering source are positioned at a distance of 1 meter from the head, at an angle of 0° and 90°, respectively, relative to the microphone array axis.

		fullband	subband	
			$K = 16$, $D = 12$	$K = 32$, $D = 20$
$M = 2$	$L.D \approx 24$	792.6	16.2	6.3
	$L.D \approx 48$	3170.0	60.2	19.5
$M = 3$	$L.D \approx 24$	1783.3	35.0	12.5
$M = 4$	$L.D \approx 24$	3170.0	61.2	20.8

Table 1: Complexity (in Mops/s) of the recursive fullband and subband GSVD for different number of microphones M and different temporal windows $L.D$, $d_f = d_g = 1$.

The desired and noise signal are uncorrelated, stationary and speech like. They both have a level of 70 dB SPL at the center of the head.

Figure 2 and Figure 3 plot the improvement in SNR, ΔSNR, and in $\mathrm{SNR}_{\mathrm{intellig}}$, $\Delta\mathrm{SNR}_{\mathrm{intellig}}$, obtained by varying the filter length L_{full} of the fullband GSVD and the filter length per subband L_{sub} in the subband GSVD, as a function of complexity (expressed in Mops/s). For the same complexity, a significant performance gain is achieved with subband GSVD. The gain in $\mathrm{SNR}_{\mathrm{intellig}}$ (2 to 4 dB) is larger than the gain in SNR (1 to 2.5 dB). This indicates that - for coloured signals such as speech - the subband approach indeed tends to provide improved intelligibility compared to the fullband method.

Figure 4 plots the Power Transfer Function (PTF) of the desired and interfering signal obtained with the fullband and the 16 subband approach for a temporal window $LD = 24$. The improvement in SNR, ΔSNR, equals 9.1 dB and 9.3 dB for the fullband and subband GSVD, respectively, the improvement in $\mathrm{SNR}_{\mathrm{intellig}}$ equals 11.8 dB and 12.8 dB, respectively. In the fullband implementation, there is more distortion and less noise reduction at the higher frequencies compared to the subband approach. Since the desired and interfering signal are both speech like, a fullband MSE or SNR criterion pays much more attention to the high energy low frequencies and tends to ignore the MSE and SNR at the higher frequencies. However, the noise reduction at higher frequencies also contributes to intelligibility. This especially occurs if the filter length L is small (small number of degrees of freedom), which explains the larger gain in $\Delta\mathrm{SNR}_{\mathrm{intellig}}$ offered by the subband GSVD at low complexity than at high complexity in Figure 3. The subband approach does not have these effects thanks to the intrinsic prewhitening.

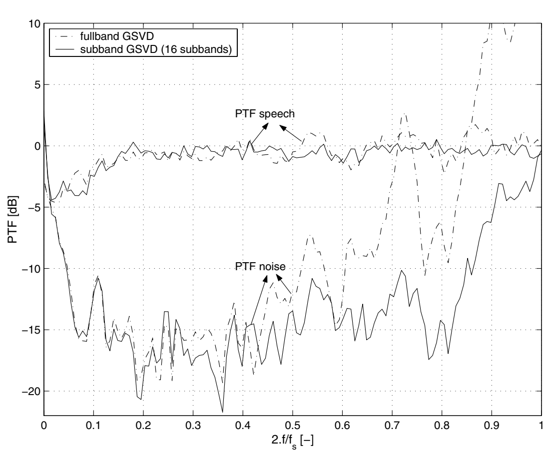

Figure 4: Power Transfer Function of speech and noise signal obtained with fullband and 16 subband GSVD for $M = 4$ and a temporal window $LD = 24$.

6. CONCLUSIONS

We have presented a subband implementation of a multichannel GSVD based noise reduction technique. It is shown that a subband implementation offers an improvement in intelligibility at a significantly lower complexity compared to the fullband approach. Hearing aid applications benefit from these two advantages.

7. REFERENCES

[1] S. Doclo and M. Moonen, *GSVD-Based Optimal Filtering for Multi-Microphone Speech Enhancement*, chapter 6 in “Microphone Arrays: Signal Processing Techniques and Applications” (Brandstein, M. S. and Ward, D. B., Eds.), pp. 111–132, Springer-Verlag, May 2001.

[2] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. AP*, vol. 30, pp. 27–34, Jan. 1982.

[3] A. Spriet, M. Moonen, and J. Wouters, “Robustness analysis of GSVD based optimal filtering and generalized sidelobe canceller for hearing aid applications,” to appear in *Proc. of WASPAA*, New Paltz NY, USA, Oct. 2001.

[4] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, “Intelligibility-weighted measures of speech-to-interference ratio and speech system performance,” *J. Acoust. Soc. Amer.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[5] Acoustical Society of America, “ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index,” June 1997.

[6] M. J. Link and K. M. Buckley, “Prewhitening for intelligibility gain in hearing aid arrays,” *J. Acoust. Soc. Amer.*, vol. 93, no. 4, pp. 2139–2140, Apr. 1993.

[7] K. Eneman and M. Moonen, “Para-Unitary Filter Bank Design for Oversampled Subband Systems,” Tech. Rep. ESAT-SISTA/TR 1997-119, ESAT, K.U.Leuven, Belgium, Dec. 1997.

[8] S. Weiss and R. W. Stewart, “Performance Limitations of Subband Adaptive Filter,” in *Proc. of EUSIPCO*, Rhodos, Greece, 1998, pp. 1245–1248.

[9] M. A. Stone and C. J. Moore, “Tolerable Hearing Aid Delays. I. Estimation of Limits Imposed by the Auditory Path Alone Using Simulated Hearing Losses,” *Ear and Hearing*, vol. 20, pp. 182–191, 1999.