A SPATIALLY CONSTRAINED SUBBAND BEAMFORMING ALGORITHM FOR SPEECH ENHANCEMENT

Per Cornelius, Zohra Yermeche, Nedelko Grbić and Ingvar Claesson

Blekinge Institute of Technology
School of Engineering
372 25 Ronneby, Sweden. E-mail:pco@bth.se

ABSTRACT

This paper discusses speech enhancement in an enclosed environment such as communication in a motorcycle helmet. A new constrained subband adaptive beamformer is proposed, which uses the concept of an earlier proposed calibrated beamformer mainly developed for a hands-free in-car environment. The highly non-stationary nature of the disturbing sound field encountered in an motorcycle helmet and the fact that the source is situated in the extreme nearfield of the array, causes the beamformer to produce an unwanted fluctuation in the output level. The spatially constrained beamformer proposed in this paper makes sure that the output maintains a constant gain, as long as the corresponding source originates from the desired location.

1. INTRODUCTION

An efficient approach to improve speech enhancement/noise suppression is to additionally make use of spatial information. The use of microphone arrays have been studies for many acoustical applications such as hands-free in-car communication, teleconferencing, speech-recognition and hearing aids [1]. The source of interest may be corrupted by interfering signals, echoes or reverberation from the environment, or from other speakers and from ambient noise sources. These environments are generally very difficult to describe by *a priori* model, whereby sequences of calibration signals can be used effectively for the design of the beamformers [2].

Recently, a new calibrated adaptive frequency domain beamformer was proposed which is based on the principle of a soft constraint RLS type of algorithm, formed from calibration data [3]. This constraint may also be precalculated from free-field assumptions as it is done in [4], but the benefit from using calibration data is that the acoustical environment, such as information about reverberation and microphone misplacement are taken into account in the model. The algorithm has been shown to produce good results in different environments.

An unwanted gain fluctuation of the output may appear which originates from the recursive updating process of the least square solution. This becomes significant mainly when the signal-to-noise-ratio is changing rapidly. The algorithm make use of the the second order statistics of the calibration data combined with the actually observed realtime data. When the source from the desired position increases its signal power, the algorithm compensates by decreasing the level of the weights, which in turn give rise to a decreased output signal power.

In this paper we propose a method which make use of the information from the calibration signal, and continuously adjusts the level such that the source of interest is processed with a constant gain.

Simulation in a real motorcycle environments is presented. Results show that the proposed method significantly reduces these unwanted gain fluctuations.

2. PROBLEM FORMULATION

Consider a scenario where the desired speech source is located in the near field of a microphone array in a fix position and the noise sources may change position with time. Assume there are I elements in the microphone array. In general, the sampled signal received by the microphone element i can be represented by

$$x_i[n] = s_i[n] + n_i[n] + \sum_{d=1}^{D} v_{id}[n], \quad i = 1, 2, \ldots, I \quad (1)$$

where $s_i[n], n_i[n]$, and $v_{id}[n], d = 1, \ldots, D$, are the source signal, the mixtures of the coherent and incoherent noise sources, and D number of interfering directional sources. The output of the beamformer is given by

$$y[n] = \sum_{i=1}^{I} w_i[n] * x_i[n] \quad (2)$$

where '$*$' denotes convolution and $w_i[n]$ denotes the beamformer filters.

The computational complexity of the convolution operation is reduced by using the frequency domain formulation of the filtering operations, which corresponds to a multiplication with I number of complex frequency domain representation weigths, $w_i^{(f)}$ for each frequency. For a specific frequency, f, the output is given by

$$y^{(f)}[n] = \sum_{i=1}^{I} w_i^{(f)} x_i^{(f)}[n] \tag{3}$$

where the signals, $x_i^{(f)}[n]$ and $y^{(f)}[n]$, are narrow band, time domain signals, containing essentially components at the frequency f.

A multichannel uniform over-sampled analysis DFT filter bank is employed to decompose each of the I microphone input signals into K numbers of subbands with a decimation factor $\frac{K}{2}$. Likewise, a synthesis filter bank is used to reconstruct the subband output signals into fullband representation. Both filter banks are designed with the metholology described in [5], where transformation and reconstruction aliasing effects are minimized. An illustration of the subband beamformer is shown in figure 1.

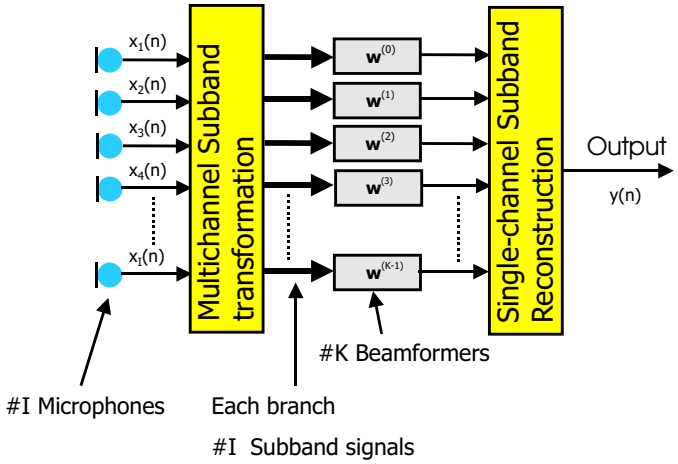

Fig. 1. *Structure of the subband beamformer*

2.1. Subband beamformer

The soft constraint subband beamformer proposed by Grbić [3], is based on a calibration and an operational phase. The calibration phase consist of collecting data from the source of interest in a quit environment. The operational phase use stored second order statistical information calculated in the first phase in combination with the present data to continuously calculate the optimal weights.

The optimal weight vector, derived from the least square solution is formulated in the frequency domain,

$$\mathbf{w}_{ls}^{(k)} = [w_1^{(k)} \quad w_2^{(k)} \quad \dots w_I^{(k)}]^T \tag{4}$$

where index k indicates the subband number, and it is recursively calculated at time instant n according to

$$\mathbf{w}_{ls}^{(k)}[n] = \left[\hat{\mathbf{R}}^{(k)}[n]\right]^{-1} \hat{\mathbf{r}}_s^{(k)}(N) \tag{5}$$

where $\hat{\mathbf{R}}^{(k)}[n]$ is a combined correlation matrix estimate

$$\hat{\mathbf{R}}^{(k)}[n] = \hat{\mathbf{R}}_{ss}^{(k)}(N) + \hat{\mathbf{R}}_{xx}^{(k)}[n]. \tag{6}$$

From N data samples, collected in the first calibration phase the correlation matrices are precalculated from

$$\hat{\mathbf{R}}_{ss}^{(k)}(N) = \frac{1}{N} \sum_{m=0}^{N-1} \mathbf{s}^{(k)}[m] \mathbf{s}^{(k)}[m]^H \tag{7}$$

$$\hat{\mathbf{r}}_s^{(k)}(N) = \frac{1}{N} \sum_{m=0}^{N-1} \mathbf{s}^{(k)}[m] s^{(k)}[m]^* \tag{8}$$

where

$$\mathbf{s}^{(k)}[m] = [s_1^{(k)}[m] \quad s_2^{(k)}[m] \quad \dots \quad s_I^{(k)}[m]]^T \tag{9}$$

and where each signal, $s_i^{(k)}[m]$, is the i's microphone received data when only the source signal of interest is active, for subband k.

The observed data correlation matrix is given by

$$\hat{\mathbf{R}}_{xx}^{(k)}[n] = \sum_{l=0}^{n-1} \lambda^{n-1-l} \mathbf{x}^{(k)}[l] \mathbf{x}^{(k)}[l]^H \tag{10}$$

where $\mathbf{x}^{(k)}[l]$ is the input vector for subband k, at time instant l and where λ is a weighting factor.

In the originally proposed calibration subband beamformer, the inverse of (6) is effectively updated, recursively for every time instant.

2.2. Spatially constrained beamformer

The observed correlation matrix estimate, may be viewed as a combination of two matrices

$$\hat{\mathbf{R}}_{xx}^{(k)}[n] = \hat{\mathbf{R}}_{cc}^{(k)}[n] + \hat{\mathbf{R}}_{nn}^{(k)}[n] \tag{11}$$

where $\hat{\mathbf{R}}_{nn}^{(k)}[n]$ corresponds to the noise plus interference correlation matrix at time instant n and $\hat{\mathbf{R}}_{cc}^{(k)}[n]$ corresponds to the correlation matrix from a signal source positioned at the same spatial location as the calibrated signal. The optimal weight vector from (5) may then be rewritten as

$$\mathbf{w}_{ls}^{(k)}[n] = [\hat{\mathbf{R}}_{ss}^{(k)}(N) + \hat{\mathbf{R}}_{cc}^{(k)}[n] + \hat{\mathbf{R}}_{nn}^{(k)}[n]]^{-1} \hat{\mathbf{r}}_s^{(k)}(N). \tag{12}$$

By observing the above expression we conclude that the power level of the weight vector may fluctuate, depending on the power level from the source of interest and the noise sources. If the SNR is high, the variation will mainly depend on the correlation matrix of the source of interest. This phenomena corresponds to a fluctuation of the output gain from the beamformer positioned at the desired spatial location.

Assuming a free field propagation, for a subband k with a certain frequency f, the array data vector received from a point d in space, at time n, may be described by

$$\mathbf{s}^{(k)}[n] = \mathbf{a}_d^{(f)}(\tau) s^{(k)}[n] \tag{13}$$

where $s[n]$ is the source of interest and the array response vector

$$\mathbf{a}_d^{(f)}(\tau) = [\beta_1 e^{i2\pi f \tau_1} \quad \beta_2 e^{i2\pi f \tau_2} \quad \dots \quad \beta_I e^{i2\pi f \tau_I}]^T \tag{14}$$

represents the propagation channel between the signal source and the array, where β_i is the channel attenuation, τ_i is the propagation time delay from point d to element i in the array [6].

Sources located at point d with the propagation time delay τ from the array should pass the array unaltered. By multiplying with a scalar function

$$q^{(k)}[n] \mathbf{w}_{ls}^{(k)}[n]^H \mathbf{s}^{(k)}[n] = s^{(k)}[n]. \tag{15}$$

defined as

$$q^{(k)}[n] = \frac{1}{\mathbf{w}_{ls}^{(k)H}[n] \mathbf{a}_d^{(f)}(\tau)} \tag{16}$$

the beamformer will produce a constant gain from signals originating from this point in space.

Since the correlation vector $\hat{\mathbf{r}}_s^{(k)}(N)$, calculated from the calibration data, acts as an estimate of the response vector, where microphone imperfections are also taken into account, we use this information instead of $\mathbf{a}_d^{(f)}(\tau)$. The weights are finally updated according to

$$\mathbf{w}_{new}^{(k)}[n] = \frac{\mathbf{w}_{ls}^{(k)}[n]}{\left| \mathbf{w}_{ls}^{(k)H}[n] \hat{\mathbf{r}}_s^{(k)}(N) \right|} P^{(k)}(N) \tag{17}$$

where we have introduced a scalar, $P^{(k)}(N)$ to compensate for the power of the desired signal from the calibration phase

$$P^{(k)}(N) = \frac{1}{N} \sum_{m=0}^{N-1} s^{(k)}[m]^* s^{(k)}[m] \tag{18}$$

for each subband k.

3. SIMULATION AND RESULTS

3.1. Conditions

In subsequent sections an evaluation of the proposed approach is presented. We use a uniform over-sampled analysis DFT filter bank, and compare the original proposed adaptive beamformer with the proposed method.

The array, consisting of sex omnidirectional microphones, were mounted inside a full face motorcycle helmet in front of the mouth onto the face shield. The space between the microphones were approximately 5 cm. The data were gathered on a portable multichannel digital audio tape recorder with a sample rate of 12 kHz. The input signal were bandlimited to the frequency band between 300-3400Hz.

In order to gather the calibration signal, an utterance of speech from the driver, with the helmet on and the windshield open, were collected before the engine was turn on.

3.2. Results

To clearly illustrate the effects of the unwanted gain fluctuation of the existing calibrated beamformer a sequence corrupted with ambient motorcycle helmet noise with a high SNR, depicted in figure 2 (a), is used for this evaluation. At lower SNR's, i.e. driving at higher speeds with the motorcycle, the existing beamformers source power fluctuation decrease and become almost neglecteble. When the driver have to stop, the opposite situation with a high SNR occurs and the fluctuation are defacto evident. The span of SNR for different speeds are presented in table 1 with corresponding noise suppression from the beamformers.

Velocity	Input sig.	Prop. beamf		Orig. beamf	
	SNR	SNR	Diff.	SNR	Diff.
50	15.8	19.9	4.1	22.8	7.0
70	8.1	11.3	3.2	16.8	8.7
90	1.5	8.4	6.9	15.2	13.7
110	-1.2	1.6	2.8	9.5	10.7
150	-11.3	-1.4	9.9	4.4	15.7
km/h	dB	dB	dB	dB	dB

Table 1. *The SNR of the the input signal, original beamformer and the proposed beamformer for 50,70,90,110 and 150 km/h are presented. The SNR differences between the input and the outputs are shown for clearness.*

The simulation were performed with 64 subbands and the constant λ was set to 0.9. By comparing figure 2(a) and 2(b) we clearly see the effect of the gain fluctuation from the original beamformer algorithm. Figure 2(c) shows the result from the proposed method, and it can be seen that these

fluctuations are cancelled. Two arrows are placed in figure 2(b) to show where these fluctuations are most noticeable.

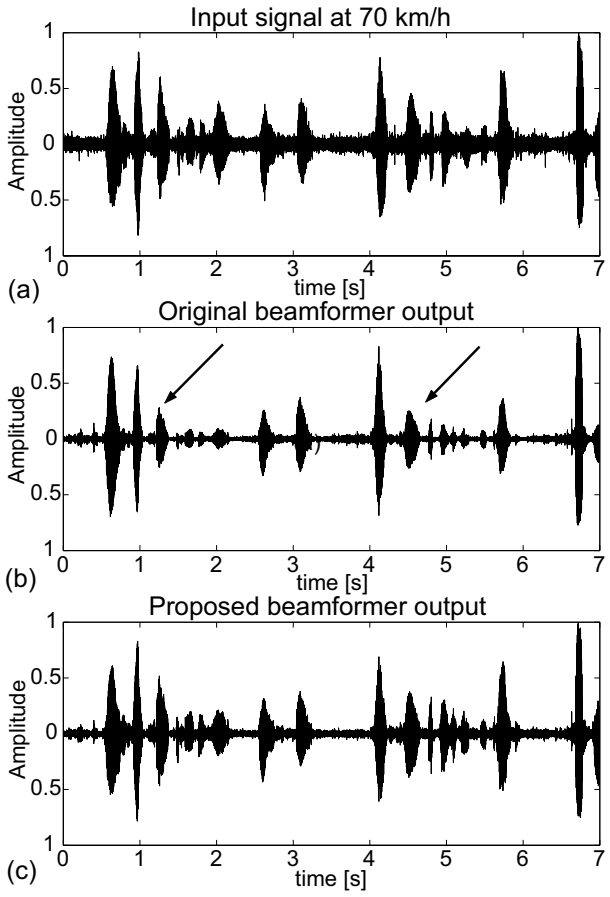

Fig. 2. *The figure shows a time sequence of 7 seconds where (a) is the input signal, (b) shows the resulting output from the original beamformer. The arrows points where the effect of gain fluctuations are most noticible, and (c) shows the resulting output from the proposed algorithm*

The calibration signals power spectrum density (PSD) effects the output PSD of the original beamformer. When the SNR decrease, the spectrum where the power is high from the calibration signal, become higher and vice versa. With the proposed method the spectrum from the desired location are passed unaltered. Figure 3(a) shows the PSD of the speech and noise at 70 km/h for the input signal (solid line), the original (doted line) and proposed beamformer (dashed line). Figure 3(b) shows the PSD of the speech and (c) the noise at 70 km/h. It can be seen that the output of the proposed method follows the spectrum of the input speech signal while the original beamformer are lower at low and high frequencies.

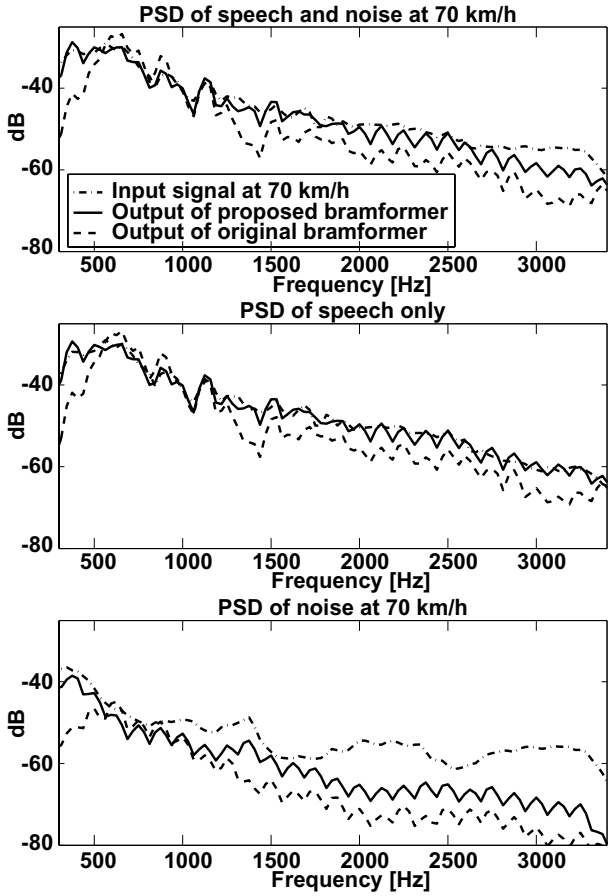

Fig. 3. *The PSD for each of the signals presented in figure 2(a-c) are showed in (a), (b) and (c) shows the PSD of the speech resp. noise at 70km/h from the PSD presented in (a). The output power level of the beamformers are adjusted to have the same speech power level as the input.*

4. CONCLUSIONS

A new spatially constrained subband adaptive beamformer used in a motorcycle environment has been presented. The proposed spatial constraints acts only on sources originating from a desired spatial location and thus it preserves the ability to attenuate sources from all other locations. Results from a real conversation with a person driving a motorcycle is presented and the results show no audible effects of output level fluctuation.

5. REFERENCES

[1] M. S. Brandstein and D. B. Ward, “Mircophone arrays: Techniques and Applications, ” Springer Verlag, 2001.

[2] S. Nordholm, I. Claesson, and M. Dahl, “Adaptive microphone array employing calibration signals: An ana-

lytical evaluation," *IEEE trans. Speech and Audio Processing*, vol. 7, pp. 241-252, Maj 1999.

[3] N. Grbić, "Optimal and Adaptive Subband Beamforming - Principles and Applications", PhD thesis, Blekinge Institute of Technology, ISBN 91-7295-002-1, Jun. 2001.

[4] N. Grbić and S. Nordholm, "Soft constrained subband beamforming for hands-free speech enhancement," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. I, pp. 885-888, 2002.

[5] J. M. de Haan, N. Grbić, I. Claesson, and S. Nordholm, "Design of oversampled uniform dft filter banks with delay specifications using quadratic optimization," *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. VI, pp. 3633-3636, May 2001.

[6] D. Johnson and D. Dudgeon," Array Signal Processing - Concepts and Techniques," Prentice Hall, 1993