Adaptive Microphone Array Employing Calibration Signals. An Analytical Evaluation.

Sven Nordholm Ingvar Claesson

September 4, 1995

Abstract

This report evaluates an adaptive microphone array which facilitates a simple built-in calibration to the environment and instrumentation. The method has been suggested for use in hands-free mobile telephones for both speech enhancement and acoustic echo-cancellation as well as speech recognition. The scheme has several advantages such as a simple calibration procedure, directed jammer suppression, versatile robust beamforming and reduced target signal distortion. The analysis employs noncausal Wiener filters yielding compact and effective theoretical suppression limits.

Contents

Chapter 1

Introduction

Methods to increase the signal-to-noise ratio when using mobile telephones in hands-free operation are spectral subtraction [1, 2, 3], temporal filtering (high-pass) and array techniques [4]. We have previously presented [5] an adaptive microphone array solution based on a spatial filtering.

In the European car industry DRIVE-II project, one of the main priorities has been to replace many of the hand-controlled functions in a car by voice control. We have also evaluated array noise reduction for speech recognition in this application [6, 7].

The method analysed in this report employs a calibrated adaptive microphone array. The method was evaluated, and outperformed the spatial filtering approach both in mobile telephony and speech enhancement in cars [8, 9]. In car applications it is necessary to consider the near field in an enclosure, and the use of low cost, low precision, microphones and equipment. This aspect raises new problems which are difficult to describe in detail with a priori modeling. A simple beamformer with reduced inherent theoretical modeling is likely to outperform an ingenious array based on erroneous a priori information. The scheme presented in this report calibrates the array to the speaker and jammer locations, the microphone elements and their positions and lobe gains, amplifiers. The array is also calibrated for the acoustic situation in the car. The main idea behind this development is the calibration which provides the adaptive algorithm with excellent "desired" and "undesired" signals.

A feature of vital interest both for telephony and for speech recognition system is speech distortion. The discrepancy between the total transfer function in training and operation should be small. This can be controlled in the proposed signal beamformer by means of amplification of training signals. A vital problem for hands-free mobile telephony is to avoid acoustic feedback and maintain good speech quality. This means that the filtering must suppress the loudspeaker as well as the background noise and room reverberation without causing severe speech distortion.

In the analytical evaluation below, point sources and free field are considered. We do not try to model the interior of a car. The main subject for this study was to find the parameters that affect the suppression of the hands-free loudspeaker and the distortion of the speaker. The model is, however, also well-suited to include real measured transfer functions between the hands-free loudspeaker and the microphone elements. In this way it is possible to include a fairly good model of the car situation and also make theoretical calculations in this environment. This

will, however, be discussed in a later report.

Chapter 2

Working Scheme for the Adaptive Beamformer

The performance and theory of most adaptive arrays rely on high-precision channel matching [5, 10, 11]. This demands a careful calibration of the array elements, which can be an expensive and difficult task, in particular for broadband arrays. Another aspect is that the microphone elements that are likely to be used are of standard quality with a considerable spread in performance. This motivates a new type of adaptive beamformer which incorporates the calibration phase.

The main idea is to record calibration sequences from both jammer and target in the real situation through the actual system, with all its imperfections, and when no car noise is present. The recordings are gathered in a memory and are later used as training signals in the adaptive phase. This approach gives an inherent calibration where it is also possible to average and weigh interesting frequency bands, microphones and spatial points. This approach does not rely on any geometric a priori information on array element similarities or accurate positioning. We obtain a system that is tailored for the actual situation.

2.1 Description of the Calibration Phase

The adaptive beamformer can be calibrated on site in a parked car by using the existing handsfree loudspeaker and letting the human speaker read a representative sequence either directly or via a loudspeaker in the correct position. The sequences are gathered into a memory, see Fig. 2.1. This means that the total channels from the (loud)speakers to A/D converters are included.

The environmental noise level in a parked car, is very low providing a good signal-to-noise ratio during recording. If the situation and equipment can be regarded as time invariant, the array system possesses calibration signals, which can be combined to form inputs and a suitable desired signal for an adaptive system.

The microphone elements and placement can be chosen arbitrarily. In order to obtain a more robust system the (loud)speaker position can be varied during calibration in the vicinity of the speakers' normal position with varying power. By adding the recorded signals, weighted average training signals are obtained. These signals, gathered in the memory are later used as training input signals, and are also combined to form a desired signal during adaptation.

2.2 Operation/Adapting Phase

In the adapting phase, while the telephone is in use, the speaker is silent (approximately half the time) and the memory signals are utilized instead, see Fig.2.2. The signals are combined to form a desired signal for the adaptive filters, and are also added to the incoming microphone signals, which otherwise contain environmental noise only. This produces memory speech signals plus jammer and noise at the lower beamformer, and a known desired signal which has passed through the same electronic equipment when no noise was present. This is a conventional situation for adaptive filters, which now possess all the information needed to adapt to the correct filter coefficients in the least square sense. For this purpose we used a normalized LMS algorithm, which converges within fractions of a second and behaves very robustly in this application. The coefficients are hence updated only while the speaker is silent.

The coefficients are copied and used in the filtering upper beamformer. Two beamformers must be used since the calibration signals are added at the filter inputs in the lower beamformer. When the speaker is active the adaptation is switched off. This is done to avoid echo-effects and also to yield a more robust system in the sense that the adaptive filters can not operate on the real speech signal. Note that the speech signal is filtered through the upper beamformer only.

2.3 Choice of Reference Signal

The construction of the reference signal is a vital factor in determining speech quality. As previously explained, the calibrations signals are affected by room reverberation, the microphones,amplifiers etc.. This implies that the sequences gathered from the speakers position are affected. In order to achieve high speech quality from the beamforming scheme, it is vital that room reverberation is limited, and that undesired signals are rejected from the speech signal. The task for the adaptive filters is to minimise the difference between $y(t)$ and the reference signal $y_r(t)$ in a least-squares sense, implying that if the reference signal is chosen in an improper way it will seriously affect $y(t)$. Assuming that the channels from the speakers position are time invariant, the task for the beamformer is to perform a blind restoration of the original sequence, and at the same time, suppress any background noise. It is thus essential that the reference signal $y_r(t)$ closely represents the original sequence, optionally with some spectral weighting.

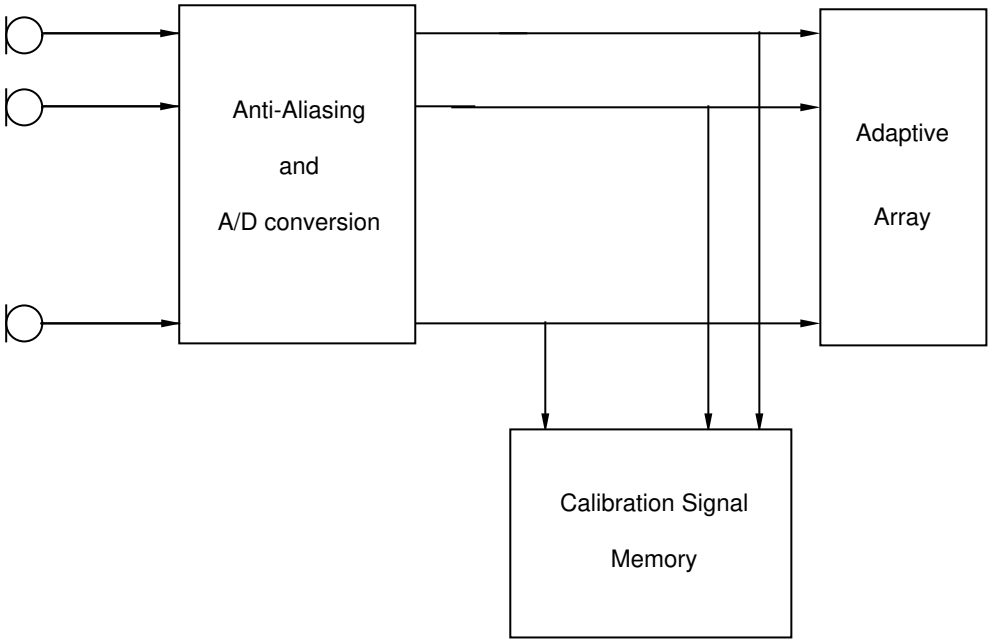

Figure 2.1: Data Gathering/Recording

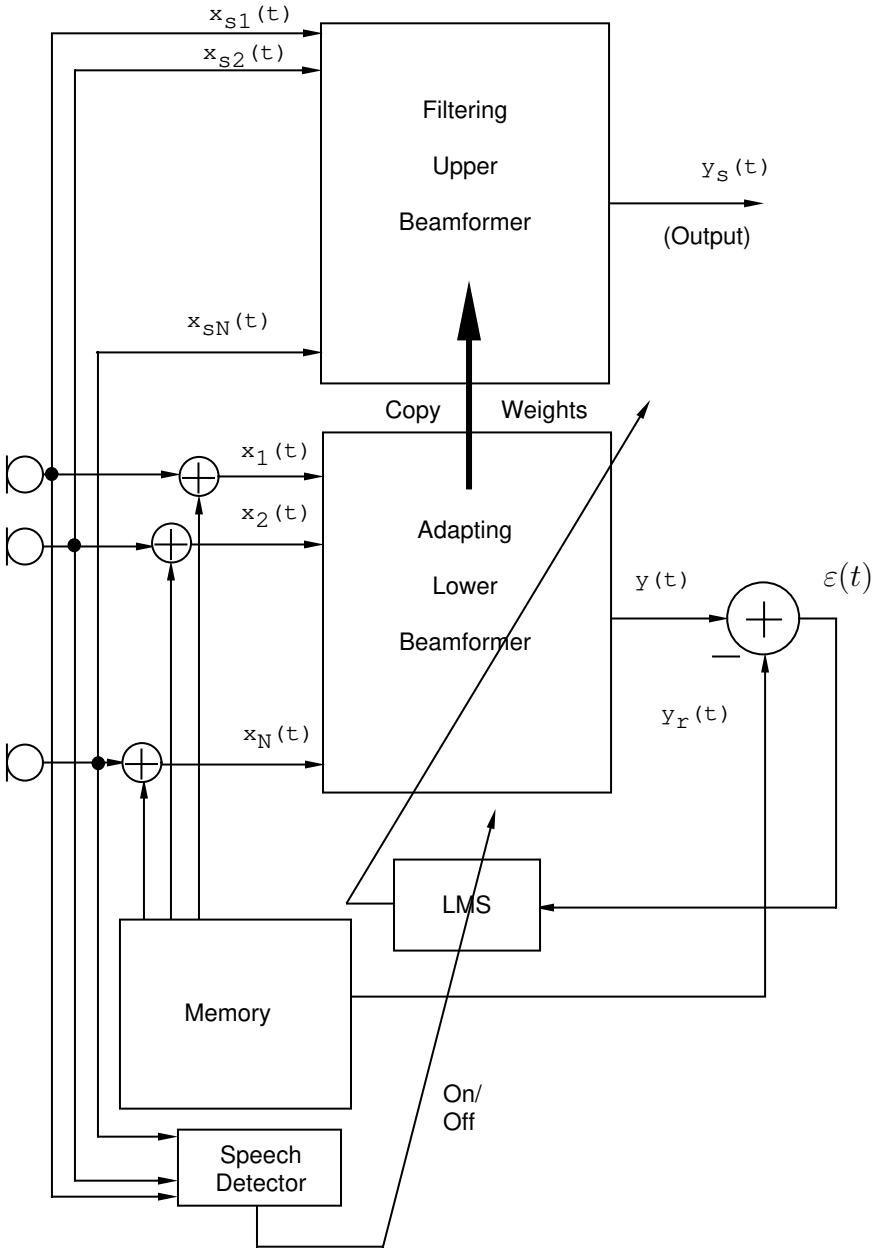

Figure 2.2: Adaptive Microphone Array/Operation phase

Chapter 3

Signal Model

The signal model is general in the sense that microphone elements and sources can be placed arbitrarily with any spectral content. The M different point signal sources $s_m(t),\ m = 1...M$ with spectral densities $R_{s_m s_m}(\omega)$ are assumed to be mutually uncorrelated, i.e. the cross power spectral density $R_{s_l s_m}(\omega)$ is zero if $l \neq m$. All sources impinge on an array of N microphone elements, each corrupted with mutually uncorrelated noise $n_l(t)$. The transfer function between source no. m and an array element no. n is denoted $G_{m,n}(\omega)$ and is either measured, or modeled. In the model, a spherical source in a free field and homogeneous medium has been assumed.

In the evaluation the situation is the adapting phase assuming that the desired signal $y_r(t)$ in Fig. 2.2 is combined from the recorded memory signals from the calibration phase. The input signals, $x_n(t)$ in Fig. 2.2 are

$$x_n(t) = \sum_{m=1}^{M} s_m(t) * g_{m,n}(t) + n_n(t) \quad n = 1....N \tag{3.1}$$

where $*$ denotes convolution. The desired signal to the adaptive scheme $y_r(t) = s_1(t) * f(t)$ is a filtered version of the target signal. This filter can be formed more generally as a fix beamformer, or formed from an electrical signal taken just before the loudspeaker where one is used. In this way the filter $f(t)$ can be used for weighing the target signal both in the spatial and frequency domains. Alternatively, another high quality microphone may be used which is placed close to the mouth during the calibration phase.

In the sequel we assume that all input signals are sampled and correctly band-limited. Hence, the power spectral density matrix for the discrete-time signals is given by

$$\mathbf{R_{xx}}(\Omega) = \sum_{m=1}^{M} R_{s_m}(\Omega)\mathbf{G}_{s_m}(\Omega)\mathbf{G}_{s_m}^H(\Omega) + \mathbf{R_{nn}}(\Omega) \tag{3.2}$$

where $\mathbf{G}_{s_m}(\Omega)$ denotes a column vector with dimension N containing all the corresponding digitised signal frequency transfer functions from a source m to all elements. The noncausal Wiener solution minimises the error spectral density with respect to the row filter vector $\mathbf{W}(\Omega)$ and can be found by expressing the orthogonality between the error and the inputs

$$\mathbf{R}_{\varepsilon\mathbf{x}}(\Omega) = \mathbf{R}_{y_r\mathbf{x}}(\Omega) - \mathbf{W}(\Omega)\mathbf{R_{xx}}(\Omega) = 0 \tag{3.3}$$

yielding

$$\mathbf{W}_{opt}(\Omega) = \mathbf{R}_{y_r\mathbf{x}}(\Omega)\mathbf{R}_{\mathbf{xx}}^{-1}(\Omega) \tag{3.4}$$

The corresponding minimum error power spectral density is given by

$$\mathbf{R}_{\varepsilon\varepsilon,opt}(\Omega) = \mathbf{R}_{y_r y_r}(\Omega) - \mathbf{R}_{y_r\mathbf{x}}(\Omega)\mathbf{R}_{\mathbf{xx}}^{-1}(\Omega)\mathbf{R}_{\mathbf{x}y_r}(\Omega) \tag{3.5}$$

By expressing the Wiener solution, the optimum performance for the adaptive beamformer can be studied.

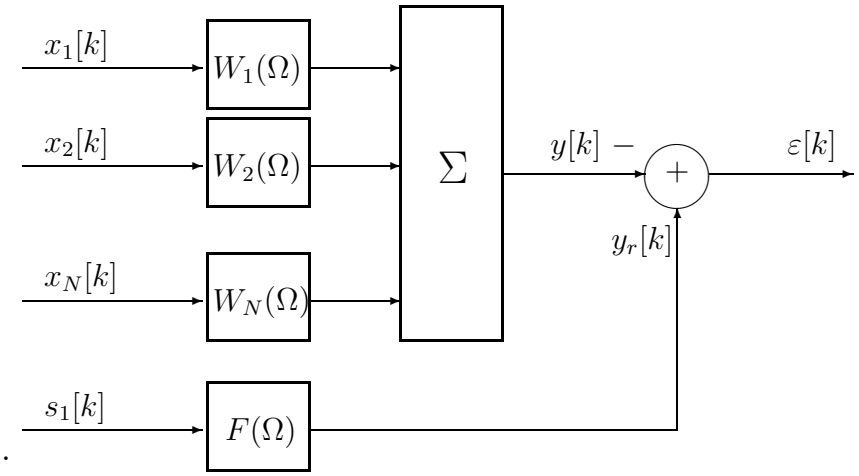

Figure 3.1: Figure describing the beamforming scheme

The power spectral density matrix $\mathbf{R}_{\mathbf{xx}}$ consists of target signal and jamming signals corrupted with noise. We can separate $\mathbf{R}_{\mathbf{xx}}$ in two parts, one with desired signals and one with undesired

$$\mathbf{R}_{\mathbf{xx}}(\Omega) = R_{s_1}(\Omega)\mathbf{G}_{s_1}(\Omega)\mathbf{G}_{s_1}^H(\Omega) + \mathbf{R}_{\nu\nu}(\Omega) \tag{3.6}$$

The matrix $\mathbf{R}_{\nu\nu}$ consists of all undesired signals, including measurement noise. Hence, it is assumed to be non-singular, and the matrix inversion lemma can be employed in order to express $\mathbf{R}_{\mathbf{xx}}^{-1}$ in closed form. Using this expression in the eq. (3.4) and finding an expression for $\mathbf{R}_{\mathbf{x}y_r}$ yields a compact formula for the solution.

$$\mathbf{W}_{opt}(\Omega) = \frac{R_{s_1}(\Omega)F(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{R}_{\nu\nu}^{-1}(\Omega)}{R_{s_1}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{R}_{\nu\nu}^{-1}(\Omega)\mathbf{G}_{s_1}(\Omega) + 1} \tag{3.7}$$

The corresponding expression for the error power spectral density is

$$R_{\varepsilon\varepsilon,opt}(\Omega) = \frac{R_{s_1}(\Omega)|F(\Omega)|^2}{R_{s_1}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{R}_{\nu\nu}{}^{-1}(\Omega)\mathbf{G}_{s_1}(\Omega) + 1} \tag{3.8}$$

The array output signal $y(k)$ is closely related to the desired signal $y_r(k)$, i.e. $F(\Omega)$.

Once the optimum filters $\mathbf{W}_{opt}(\Omega)$ are found for a given situation, we can investigate the total transfer functions from any spatial point, provided the transfer function is known from the point to each element. In particular, the transfer function from each source to the output of the array can be determined. The total

transfer function $H_m(\Omega)$ expresses how each signal is affected by the optimum beamformer and is given by

$$H_m(\Omega) = \mathbf{W}_{opt}(\Omega)\mathbf{G}_{s_m}(\Omega) = \frac{F(\Omega)R_{s_1}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{R}_{\nu\nu}^{-1}(\Omega)\mathbf{G}_{s_m}(\Omega)}{R_{s_1}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{R}_{\nu\nu}^{-1}(\Omega)\mathbf{G}_{s_1}(\Omega) + 1} \tag{3.9}$$

3.1 Two Directional Sources

The normal situation for acoustic echo-cancellation for hands-free mobile telephones is a single target (speaker) and a dominating jammer (hands-free loudspeaker). In this situation, it is possible to derive somewhat more explicit expressions on the transfer functions.

These expressions can be used to evaluate parameters and features that affect the echo suppression, such as the number of microphones, signal levels, loudspeaker placement and microphone configuration. The construction of the desired signal $y_r[k]$ is crucial, since it provides two important quantities, an output which matches the original speech and a frequency weighting of the input signal in order to obtain better speech quality.

A single jamming signal together with noise constitutes only two terms in the matrix $\mathbf{R}_{\nu\nu}$ representing all unwanted signals

$$\mathbf{R}_{\nu\nu}(\Omega) = R_{s_2}(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega) + \sigma_n^2\mathbf{I} \tag{3.10}$$

In order to derive a more explicit expression of 3.9 the matrix inversion lemma is again used to yield the inverse of the matrix $\mathbf{R}_{\nu\nu}(\Omega)$

$$\mathbf{R}_{\nu\nu}^{-1}(\Omega) = \frac{1}{\sigma_n^2}\left(\mathbf{I} - \frac{R_{s_2}(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega) + \sigma_n^2}\right) \tag{3.11}$$

Inserting this expression in eq. (3.9) finally yields

$$H_m(\Omega) = \frac{F(\Omega)\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_m}(\Omega) - \frac{R_{s_2}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_m}(\Omega)}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega) + \sigma_n^2}\right)}{\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{R_{s_2}(\Omega)|\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)|^2}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega) + \sigma_n^2}\right) + 1} \tag{3.12}$$

This expression is used below to investigate different parameters that affect the total transfer function from source to array output. By evaluating the expression from the target direction we can quantify the impact on the target signal, and analogously, an evaluation from the jammer direction yields the echo suppression of the loudspeaker. A similar evaluation can also be performed for realistic situations in a car using measured transfer functions.

Chapter 4

Target Distortion and Jammer Suppression

This evaluation assumes a simple model, point sources and spherical wave propagation in a homogeneous medium. A simple situation can still give guidelines for different parameter choices, such as element placing, and quality and spread in microphones and loudspeakers in a more complex environment.

From the simple model we obtain transfer functions from a point source to each element

$$G_{m,l}(\omega) = \frac{e^{-j\omega\tau_{m,l}}}{r_{m,l}} \tag{4.1}$$

where $\tau_{m,l}$ is the delay of the signal and $r_{m,l}$ is the distance from source to array element. In the far-field $r_{m,l}$ is much larger than the array aperture and a linear array simplifies the array response vector to

$$\mathbf{G}_{s_m}(\omega,\theta) = \frac{e^{-j\omega\tau_{bulk}}}{r_m}\left(1, e^{-j\omega\frac{d\sin(\theta)}{c}} \cdots e^{-j\omega(N-1)\frac{d\sin(\theta)}{c}}\right)^T \tag{4.2}$$

where $e^{-j\omega\tau_{bulk}}$ is an initial bulk delay, d the distance between adjacent elements and c the sound propagation velocity in the medium. Below, we again work only with sampled and bandlimited signals.

4.1 Target Direction

The digital target direction total transfer function is given by inserting $\mathbf{G}_{s_1}$ in eq. (3.12). We obtain

$$H_1(\Omega) = \frac{F(\Omega)\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{R_{s_2}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_1}(\Omega)}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)}{\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{R_{s_2}(\Omega)|\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)|^2}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)+1} \tag{4.3}$$

In this direction

$$\mathrm{SINR}_{eq}(\Omega) = \frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{R_{s_2}(\Omega)|\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)|^2}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right) \tag{4.4}$$

can be viewed as an Equivalent Signal-to-Interference-Noise Ratio, $\mathrm{SINR}_{eq}(\Omega)$ simplifying $H_1(\Omega)$ to the conventional Wiener expression

$$H_1(\Omega) = \frac{F(\Omega)\mathrm{SINR}_{eq}(\Omega)}{1+\mathrm{SINR}_{eq}(\Omega)} = \frac{F(\Omega)}{1+\frac{1}{\mathrm{SINR}_{eq}(\Omega)}} \tag{4.5}$$

The expression $\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)$ can be interpreted as a scalar product and can be rewritten as $\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega) = ||\mathbf{G}_{s_1}(\Omega)||||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)$ where $\rho(\Omega)$ is a measure of closeness between $\mathbf{G}_{s_1}(\Omega)$ and $\mathbf{G}_{s_2}(\Omega)$. From Schwarz inequality it follows that $|\rho(\Omega)| \leq 1$. An alternative expression of SINR_{eq} is obtained by using the relation above

$$\mathrm{SINR}_{eq}(\Omega) = \frac{R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||^2}{\sigma_n^2}\left(\frac{R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2(1-|\rho(\Omega)|^2)+\sigma_n^2}{R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2+\sigma_n^2}\right)$$

The maximum effective signal-to-noise ratio is obtained if $\rho(\Omega) = 0$ i.e. the vectors $\mathbf{G}_{s_1}(\Omega)$ and $\mathbf{G}_{s_2}(\Omega)$ are orthogonal.

We also define a Signal to Noise Ratio at each element given as the target signal power to the measurement noise

$$\mathrm{SNR} = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_1}(\Omega)d\Omega}{\sigma_n^2} \tag{4.6}$$

and a Signal to Interference Ratio is also defined

$$\mathrm{SIR} = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_1}(\Omega)d\Omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_2}(\Omega)d\Omega} \tag{4.7}$$

4.1.1 Examples

In order to obtain low distortion of the target signal, SINR_{eq} must be large. A parameter of interest is the correlation coefficient $\rho(\Omega)$ which should be kept small, i.e. the target and jammer array responses should be orthogonal and the target signal-to-noise ratio large, see Fig. 4.1. However, when the array responses are similar $\mathbf{G}_{s_1}(\Omega) \approx \mathbf{G}_{s_2}(\Omega)$ over some frequency range, the distortion of the target signal might become large and highly dependent on the jammer spectral density $R_{s_2}(\Omega)$.

When the noise level is small compared to $R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2$, an approximative expression with close array responses is given by

$$H_1(\Omega) \approx \frac{F(\Omega)}{1+\frac{R_{s_2}(\Omega)}{R_{s_1}(\Omega)}} \tag{4.8}$$

In Fig. 4.1 the target and jammer have identical spectral densities, and are approximately flat within 600 Hz - 3600 Hz. When the signal to noise ratio is improved, see Fig. 4.2, the jammer affects the target signal less, provided they are well separated in directions. Employing more array elements gives also a possibility to separate the target and jammer for smaller angular differences, compare figures 4.1 and4.3. When the jammer is stronger, see Fig. 4.4, the transfer function is determined by the target to jammer ratio for small angular differences, as predicted by Eq. (4.8).

Finally, Fig. 4.5 illustrates the effect of spatial aliasing when the array elements are too far apart.

4.2 Jammer Direction

In the jammer direction the transfer function is given by Eq. (3.12) with G_{s_2} inserted yielding

$$H_2(\Omega) = \frac{F(\Omega)\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega) - \frac{R_{s_2}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)}{\frac{R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{R_{s_2}(\Omega)|\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)|^2}{R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)+1} \tag{4.9}$$

This transfer function can also be expressed using the SINR_{eq} and inner product interpretation

$$H_2(\Omega) = \frac{F(\Omega)R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{(R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2+\sigma_n^2)\,(\text{SINR}_{eq}(\Omega)+1)} \tag{4.10}$$

which is close to zero when target and jammer are well separated

4.2.1 Examples

In Figs. 4.6 to 4.10, the target and jammer have identical spectral densities and are approximately flat within 600 Hz - 3600 Hz. We observe that it is essential that the scalar product between $\mathbf{G}_{s_2}(\Omega)$ and $\mathbf{G}_{s_1}(\Omega)$ is kept small, i.e. the two sources must be well separated, see Fig. 4.6, or more elements must be used, see Fig. 4.7.

Other essential parameters are the noise level at the array elements, see Fig. 4.8, and the strength of the jamming signal, see Fig. 4.9. Figure 4.10 again shows the effect of spatial undersampling. Observe that the function $\mathbf{G}_{s_2}(\Omega)$ is actually sampled twice i.e. it is periodic both in Ω and θ [12].

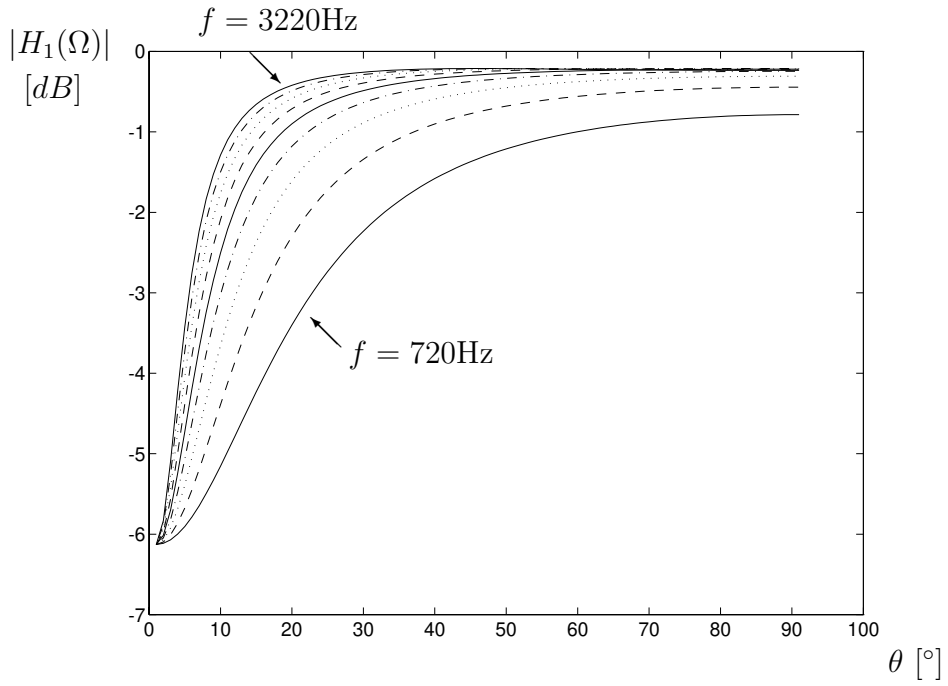

Figure 4.1: Target transfer function $H_1(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.05m, SNR=10dB, SIR=0dB

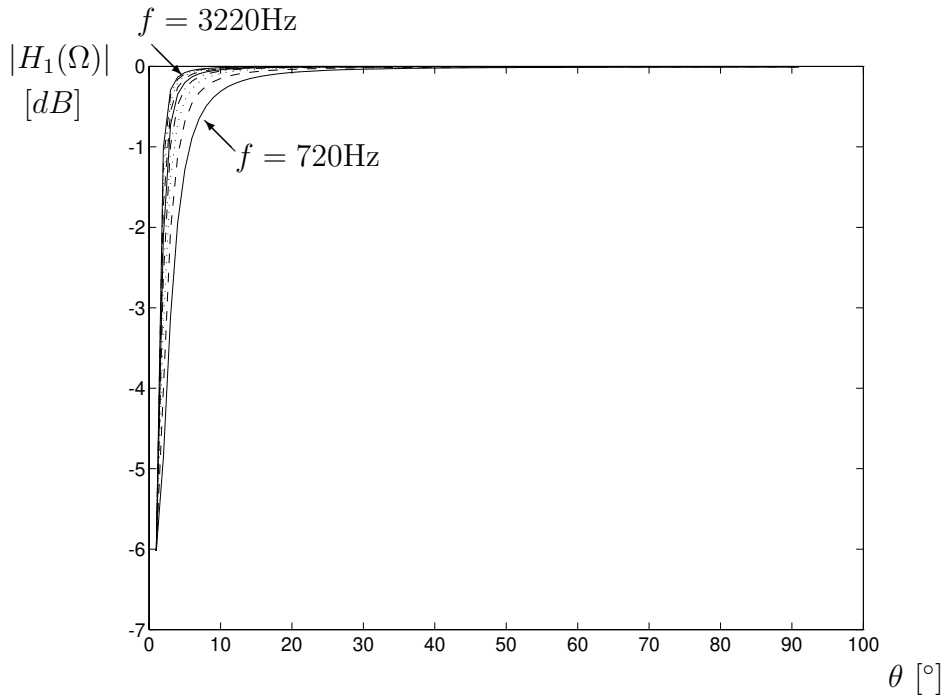

Figure 4.2: Target transfer function $H_1(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.05m, SNR=30dB, SIR=0dB

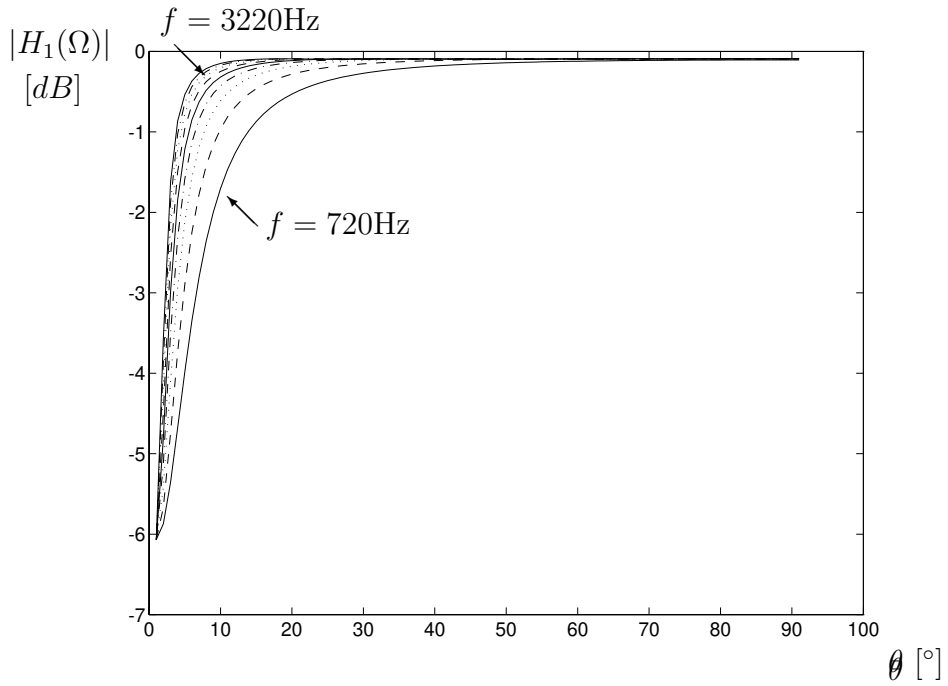

Figure 4.3: Target transfer function $H_1(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=7, d=0.05m, SNR=10dB, SIR=0dB

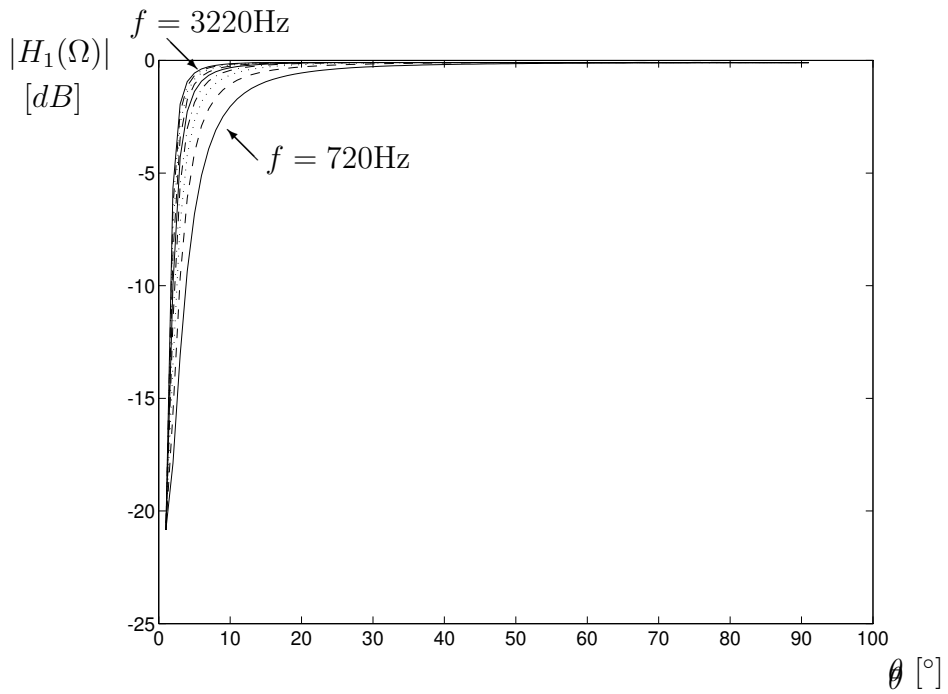

Figure 4.4: Target transfer function $H_1(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=7, d=0.05m, SNR=10dB, SIR=-10dB

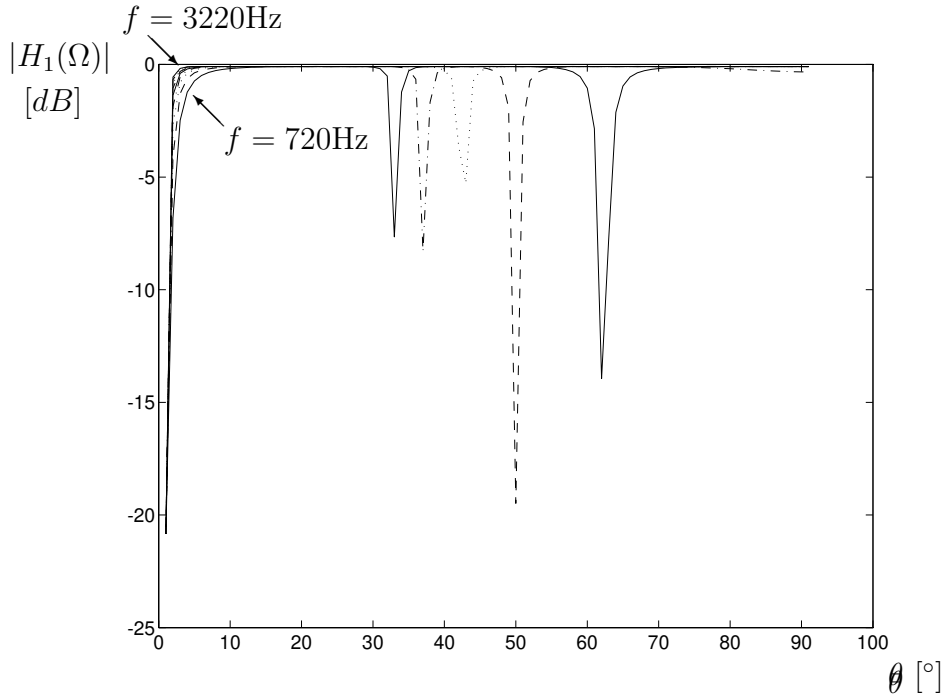

Figure 4.5: Target transfer function $H_1(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=7, d=0.2m, SNR=10dB, SIR=-10dB

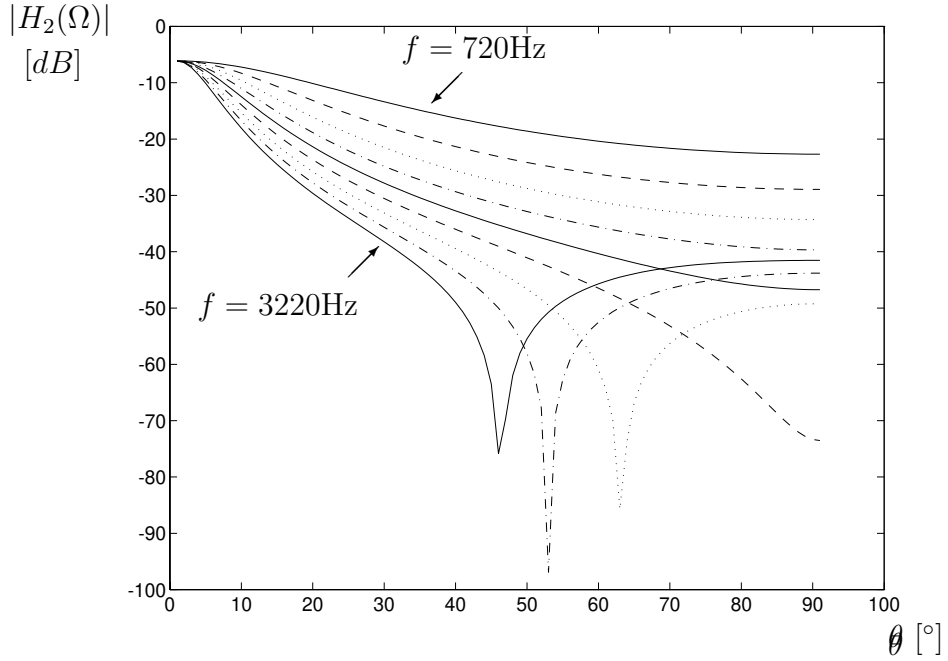

Figure 4.6: Jammer transfer function $H_2(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.05m, SNR=10dB, SIR=0dB

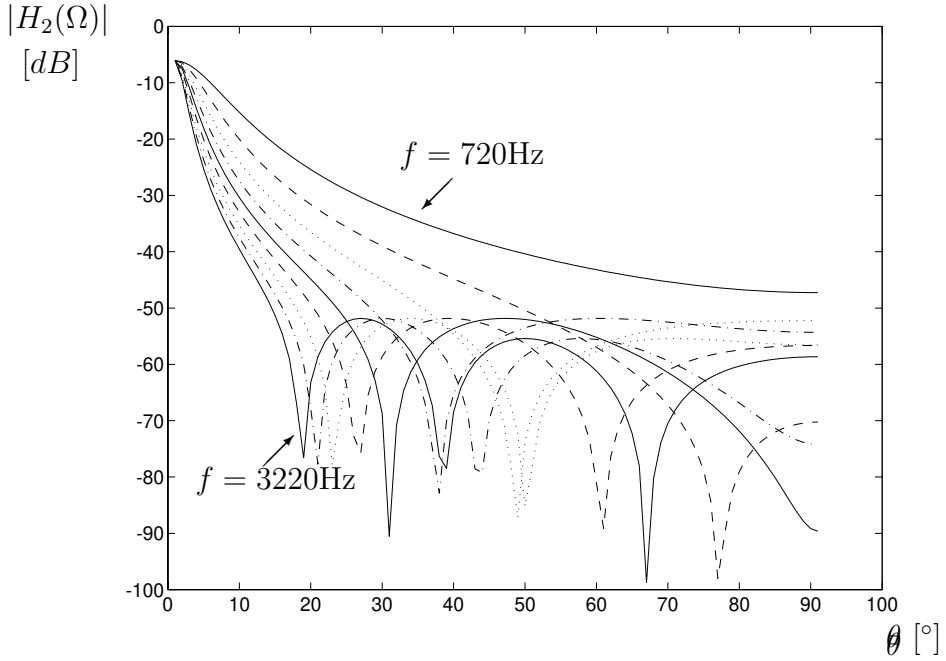

Figure 4.7: Jammer transfer function $H_2(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=7, d=0.05m, SNR=10dB, SIR=0dB

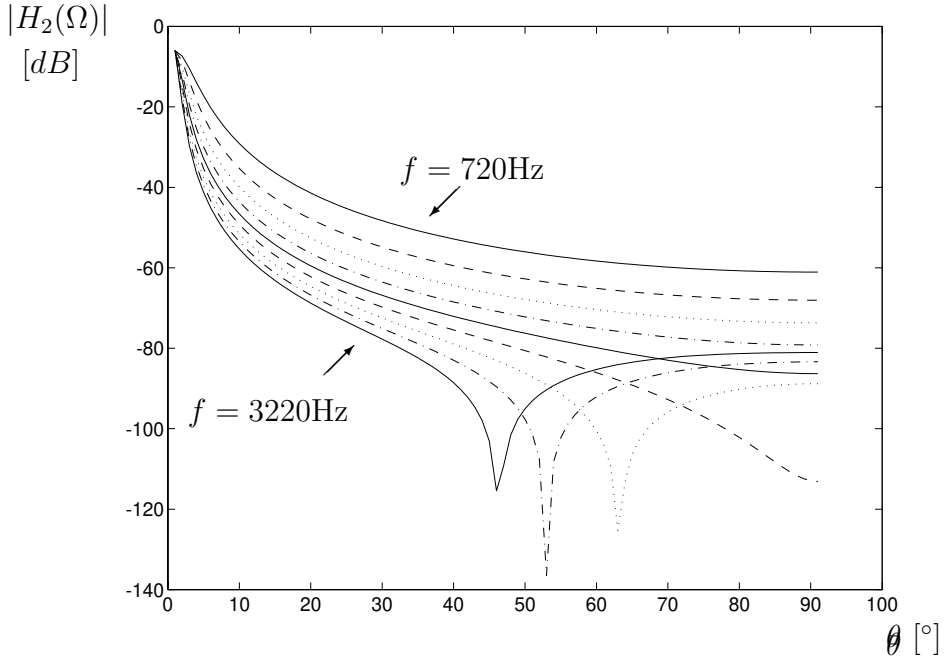

Figure 4.8: Jammer transfer function $H_2(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.05m, SNR=30dB, SIR=0dB

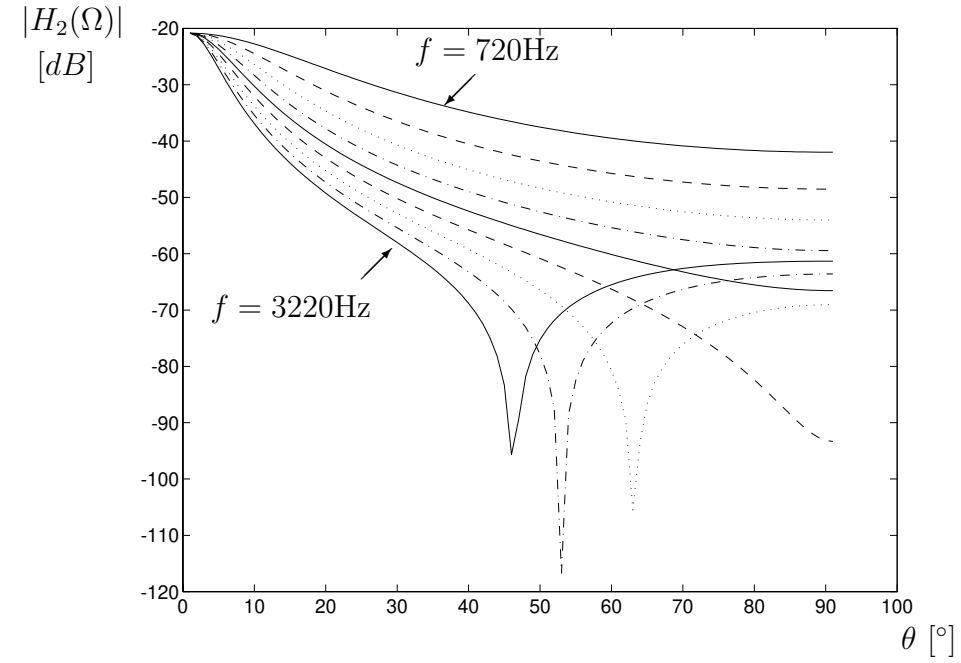

Figure 4.9: Jammer transfer function $H_2(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.05m, SNR=30dB, SIR=-10dB

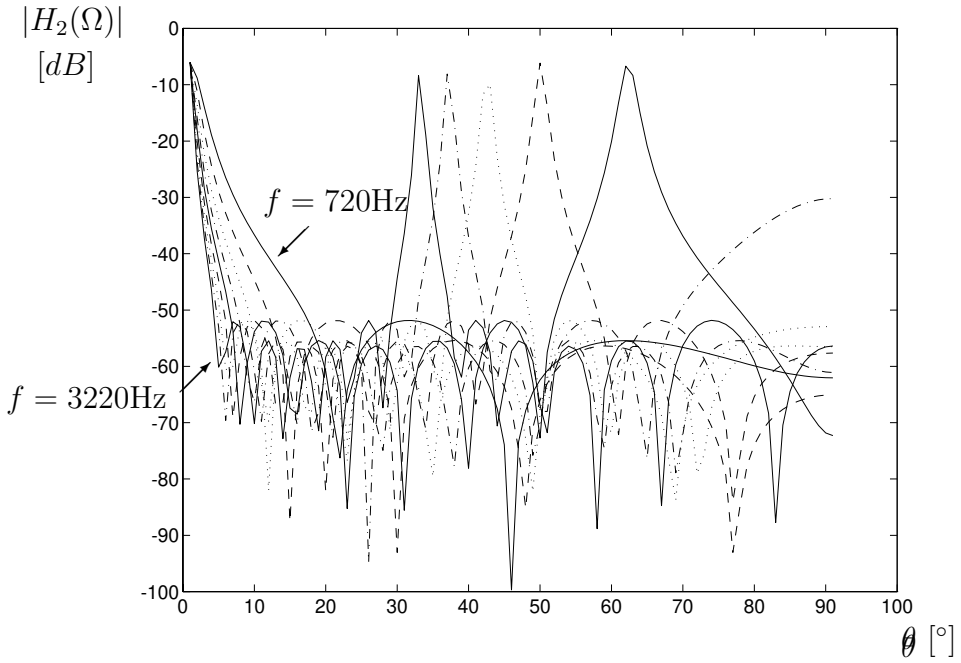

Figure 4.10: Jammer transfer function $H_2(\Omega)$ for varying jamming directions at 10 different frequencies varying from 720 Hz to 3220 Hz. Linear array with N=3, d=0.2m, SNR=30dB, SIR=-10dB

Chapter 5

Varying Signal Levels during Adaptation

In the adaptive mode, see Fig. 5.1, we also want to vary the signal levels of the gathered calibration signals in order to control the desired performance of the beamformer. This mean that level controls for the target calibration signal as well as for the jamming calibration signal are included. These levels can be adjusted independently.

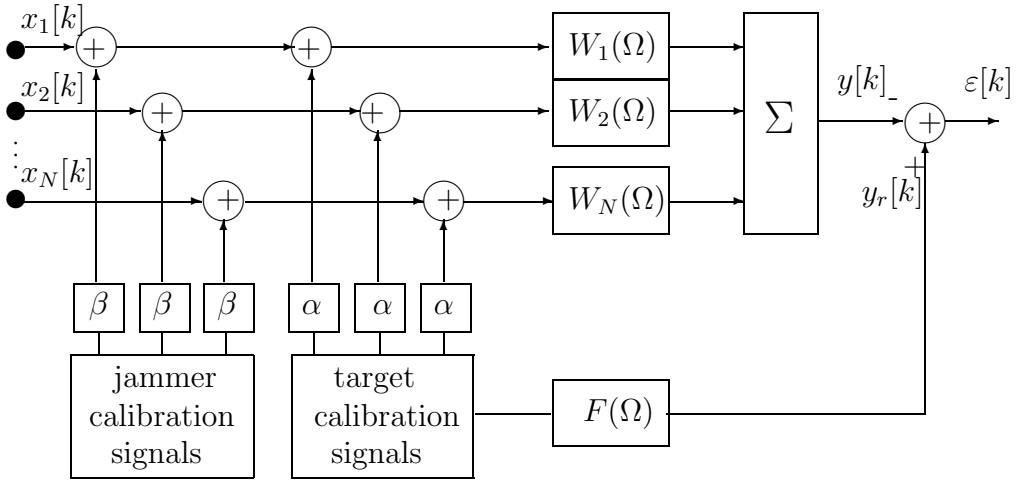

Figure 5.1: A description of the adaptive mode

5.1 Target Distortion

The memory target signal is amplified a factor α, implying that the spectrum $R_{s_1}(\Omega)$ is multiplied by α^2 and the calibrated jamming signal is analogously am-

plified with a factor β. This is inserted in eq. (4.3) yielding

$$H_1(\Omega) = \frac{F(\Omega)\frac{\alpha^2 R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{\beta^2 R_{s_2}(\Omega)\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_1}(\Omega)}{\beta^2 R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)}{\frac{\alpha^2 R_{s_1}(\Omega)}{\sigma_n^2}\left(\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_1}(\Omega) - \frac{\beta^2 R_{s_2}(\Omega)|\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)|^2}{\beta^2 R_{s_2}(\Omega)\mathbf{G}_{s_2}^H(\Omega)\mathbf{G}_{s_2}(\Omega)+\sigma_n^2}\right)+1} \tag{5.1}$$

and the $\mathrm{SINR}_{eq}(\Omega)$ is given by

$$\mathrm{SINR}_{eq}(\Omega) = \frac{\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||^2}{\sigma_n^2}\left(\frac{\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2(1-|\rho(\Omega)|^2)+\sigma_n^2}{\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2+\sigma_n^2}\right) \tag{5.2}$$

The transfer function $H_1(\Omega)$ can be used to calculate the target signal distortion ratio, which we define as

$$D = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_1}(\Omega)|H_1(\Omega)-F(\Omega)|^2 d\Omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_1}(\Omega)d\Omega} \tag{5.3}$$

Observe that the actual real target signal, jammer signals and background noise are not affected by these amplifications, since they are only performed on the memory signals in order to control the total transfer function.

Since

$$H_1(\Omega) = \frac{F(\Omega)}{1+\frac{1}{\mathrm{SINR}_{eq}(\Omega)}} \tag{5.4}$$

larger SINR_{eq} gives reduced target distortion.

The distortion is strongly dependent on the amplitude of the calibration signal, see Figs. 5.2 and 5.3. These figures present the distortion ratio D versus the amplifying parameter α for different jamming directions. If α is small, $\mathrm{SINR}_{eq}(\Omega)$ will also be small causing $H_1(\Omega)$, to deviate more from $F(\Omega)$ which in turn yields high distortion. For larger α, the distortion is lower.

In Fig. 5.3 more microphone elements are used, which give better source resolution and smaller $\rho(\Omega)$. The actual dependence is given by a sind-function(for a plane wave and a linear array case). This can easily be verified by evaluating $\mathbf{G}_{s_1}^H(\Omega)\mathbf{G}_{s_2}(\Omega)$.

In Figs. 5.4 and 5.5 the distortion function D is presented for varying memory jammer amplification β. A small β i.e. when $\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2$ is of the same order as σ_n^2, makes the target distortion almost independent of the jamming signal. For a large β the main influence will be via $\rho(\Omega)$. For well-separated sources, an expression for the equivalent signal-to-noise-ratio is

$$\mathrm{SINR}_{eq}(\Omega) \approx \frac{\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||^2}{\sigma_n^2}(1-|\rho(\Omega)|^2) \tag{5.5}$$

A good approximation with close target and jammer will be

$$\mathrm{SINR}_{eq}(\Omega) \approx \frac{\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||^2}{\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2+\sigma_n^2} \tag{5.6}$$

and the distortion depends directly on the target to jammer ratio.

5.2 Jammer Suppression

The suppression of a jammer for varying $\alpha$and β has also been investigated in relation to the jammer level at a single array element. We define the suppression ratio as

$$S = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_2}(\Omega)|H_2(\Omega)|^2 d\Omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi} R_{s_2}(\Omega) d\Omega} \tag{5.7}$$

The jammer transfer function is given by

$$H_2(\Omega) = \frac{F(\Omega)\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{(\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2 + \sigma_n^2)\,(\mathrm{SINR}_{eq}(\Omega)+1)} \tag{5.8}$$

corresponding to eq. (4.10) with α and β included.

In Figs. 5.6 and 5.7, the suppression ratio S is presented for different jamming directions with α as parameter. As the figures show, the true jammer suppression is strongly dependent on the amplitude of the calibration memory target signal. The interaction of Eq. (5.8) is, however, rather involved since the parameter α affects both nominator and denominator.

For small α, i.e. when $\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||^2$, the jammer transfer function $H_2(\Omega)$ will be small, i.e. good suppression is obtained and the transfer function can be approximated by

$$H_2(\Omega) \approx \frac{F(\Omega)\alpha^2 R_{s_1}(\Omega)||\mathbf{G}_{s_1}(\Omega)||||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2 + \sigma_n^2} \tag{5.9}$$

For large α:s, the transfer function can be approximated by

$$H_2(\Omega) \approx \frac{F(\Omega)||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{(\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2 + \sigma_n^2)\left(\frac{||\mathbf{G}_{s_1}(\Omega)||}{\sigma_n^2}(1-|\rho(\Omega)|^2)\right)} \tag{5.10}$$

In Figs. 5.8 and 5.9, the strength of the calibration jammer is varied. A small β implies that the transfer function can be approximated by

$$H_2(\Omega) \approx \frac{F(\Omega)||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{||\mathbf{G}_{s_1}(\Omega)||} \tag{5.11}$$

Here $\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2$ is assumed to be of the same strength or weaker than σ_n^2, and accordingly, $H_2(\Omega)$ will not be dependent on the calibration memory jammer or target strength. The direction is the dominating parameter. This is the classic situation for a data independent beamformer, and this is the performance the beamformer will show when there is only target signal and uncorrelated noise at each array element.

When investigating larger β:s, the calibration memory jammer rules behaviour. A good approximation of the true jammer transfer function is

$$H_2(\Omega) \approx \frac{F(\Omega)||\mathbf{G}_{s_2}(\Omega)||\rho(\Omega)}{(\beta^2 R_{s_2}(\Omega)||\mathbf{G}_{s_2}(\Omega)||^2)\left(\frac{||\mathbf{G}_{s_1}(\Omega)||}{\sigma_n^2}(1-|\rho(\Omega)|^2)\right)} \tag{5.12}$$

The suppression ratio is proportional to β^4, see Figs. 5.8 and 5.9.

5.2.1 Multipath Jammer Situation

In the above examples, the jamming signal was assumed to arrive as a plane wave. We conclude with some examples with multipath jamming, which are more closely related to a realistic situation.

The jammer transfer function is formed by a sum of plane wave array response vectors

$$\mathbf{G}_{s_2}(\omega) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{d}(\omega,\theta_k) \tag{5.13}$$

where

$$\mathbf{d}(\omega,\theta_k) = \left(1, e^{-j\omega\frac{d\sin(\theta_k)}{c}} \cdots e^{-j\omega(M-1)\frac{d\sin(\theta_k)}{c}}\right)^T$$

Some examples have been calculated for two different multipath situations. In the first situation the jamming signal angle θ_k was more concentrated and closer to the target signal, see Figs. 5.10 and 5.11. In the second situation, θ_k:s were more spread out and further away from the target, see Figs. 5.12 and 5.13.

We observe that the ability to suppress a multipath jammer is as good as for a single plane wave. The reason is that since the jamming signal originates from a single source only, it only increases the rank of the power spectral density matrix by one. This means that the adaptive beamformer can treat the signal as if it comes from "one" direction.

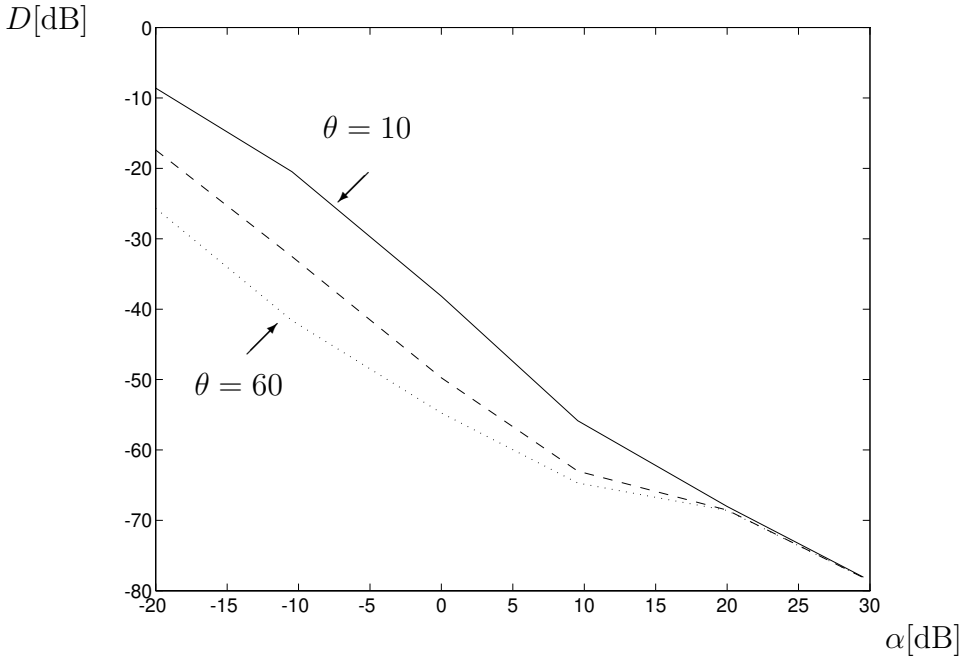

Figure 5.2: Target distortion D versus α. When $\alpha = 1$ (0 dB) it corresponds to SNR=30dB and SIR=0dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 45, 60]$.

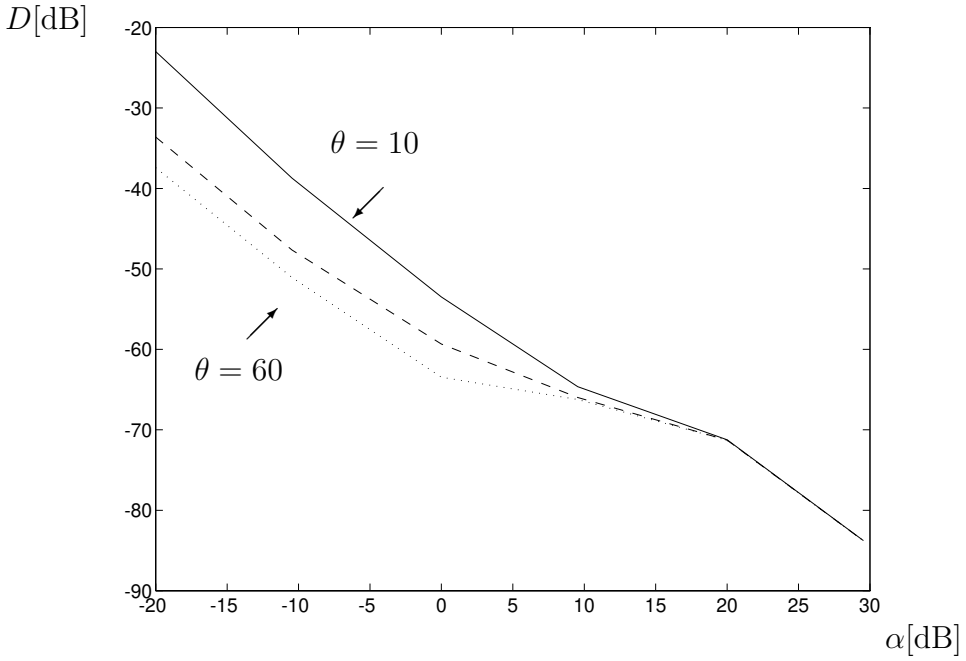

Figure 5.3: Target distortion D versus α. When $\alpha = 1$ (0 dB) it corresponds to SNR=30dB and SIR=0dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 45, 60]$.

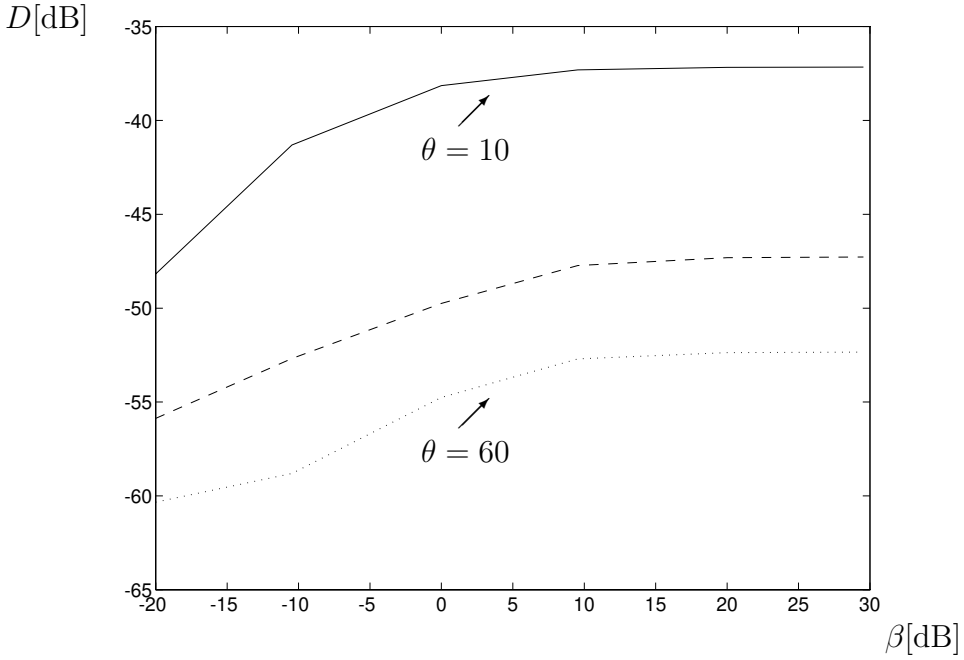

Figure 5.4: Target distortion D versus β. When $\beta = 1$ (0dB) it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 45, 60]$.

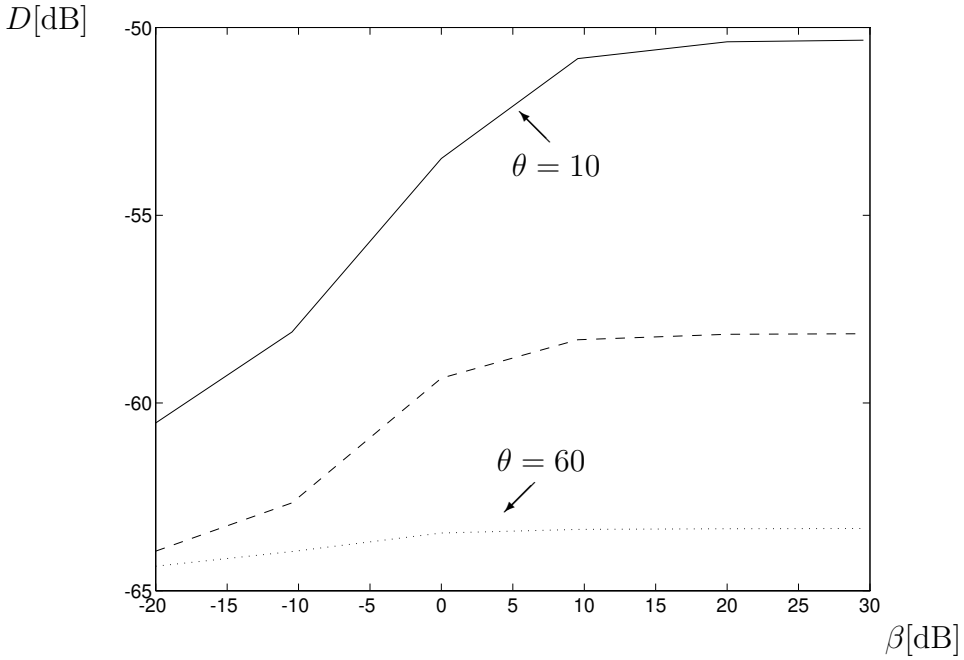

Figure 5.5: Target distortion D versus β. When $\beta = 1$ (0dB) it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=7. Jamming directions $[10, 45, 60]$.

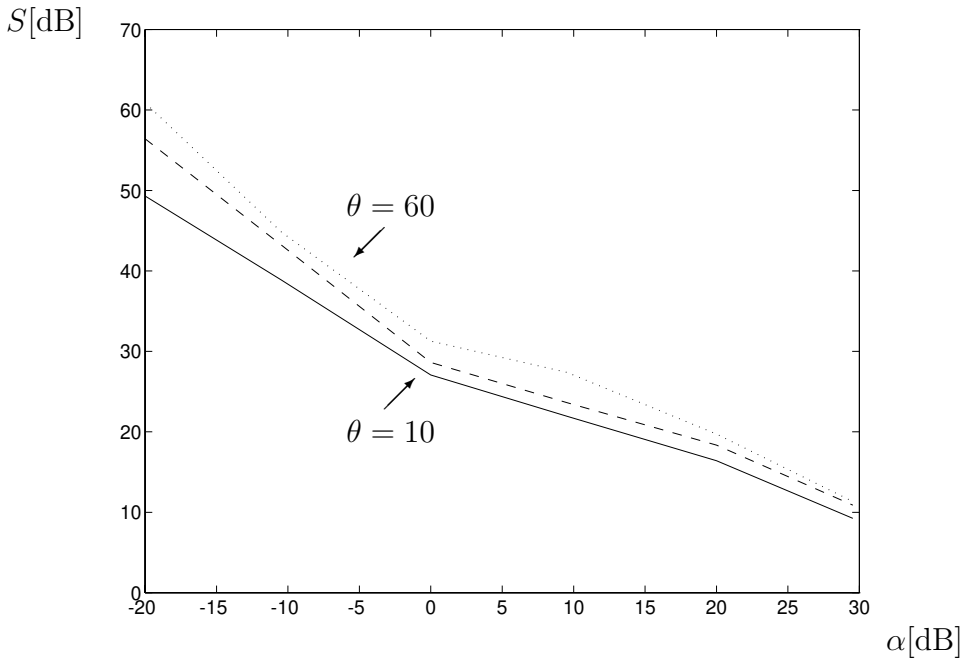

Figure 5.6: Jammer suppression S versus α. When $\alpha = 1$ (0 dB) it corresponds to SNR=30dB and SIR=0dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 45, 60]$.

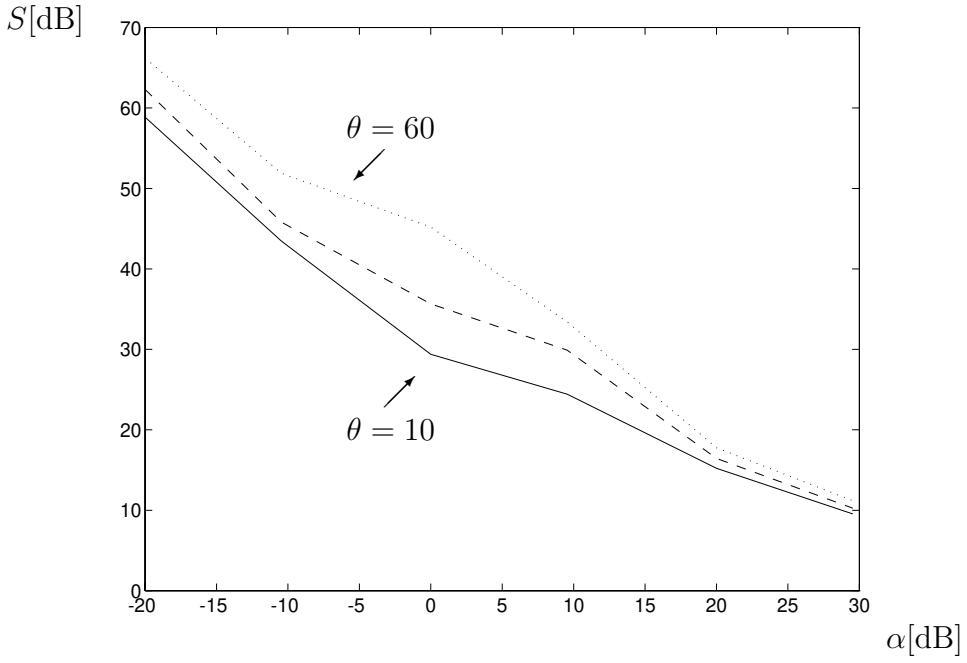

Figure 5.7: Jammer suppression S versus α. When $\alpha = 1$ (0 dB) it corresponds to SNR=30dB and SIR=0dB. Linear array with d=0.05 m, N=7. Jamming directions $[10, 45, 60]$.

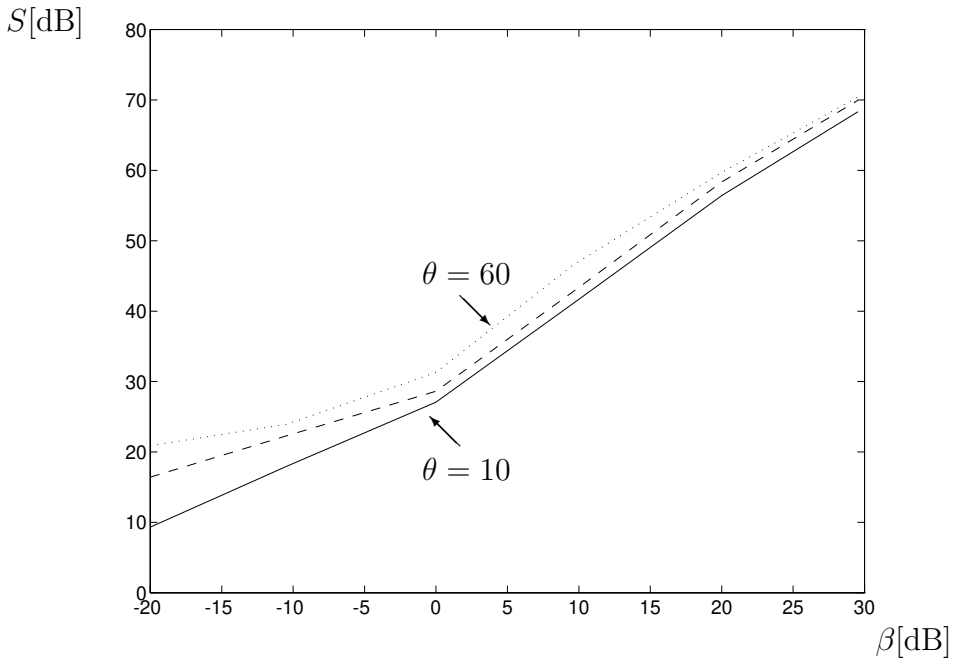

Figure 5.8: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 45, 60]$.

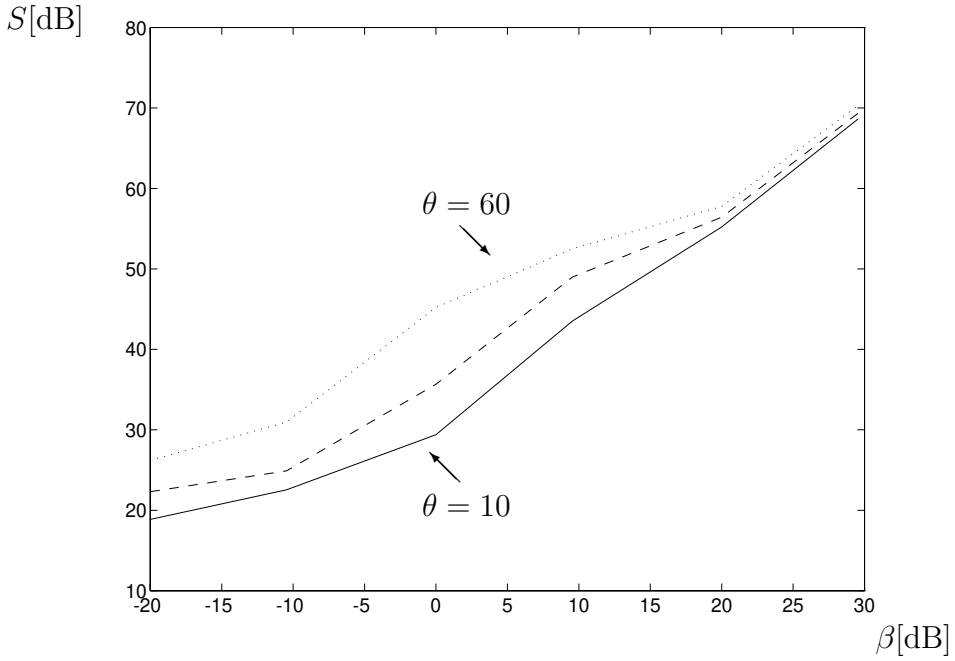

Figure 5.9: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=7. Jamming directions $[10, 45, 60]$.

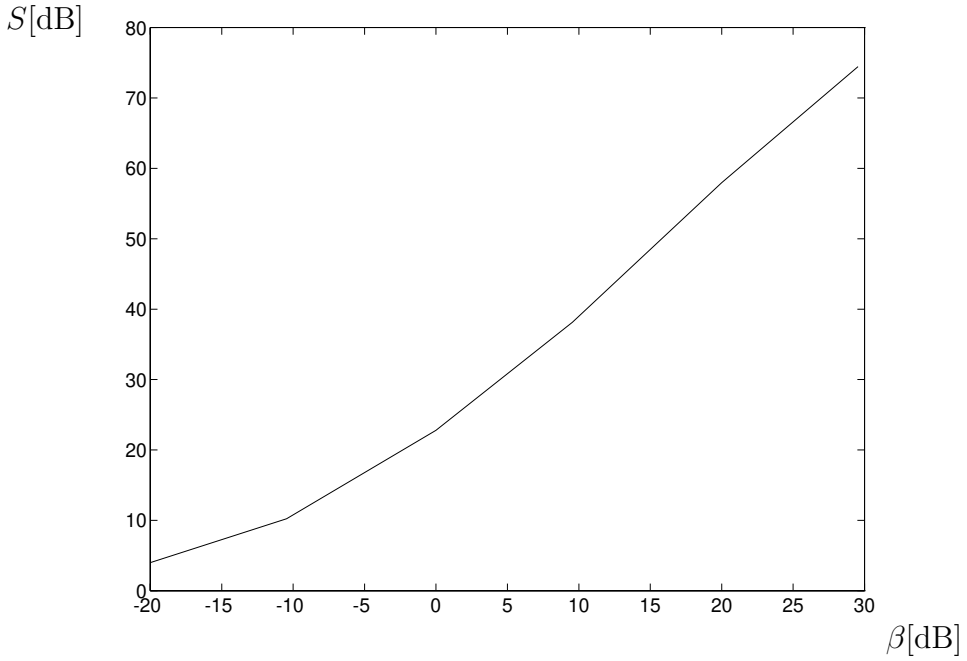

Figure 5.10: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=3. Jamming directions $[9, 10, 11, 12, 13]$.

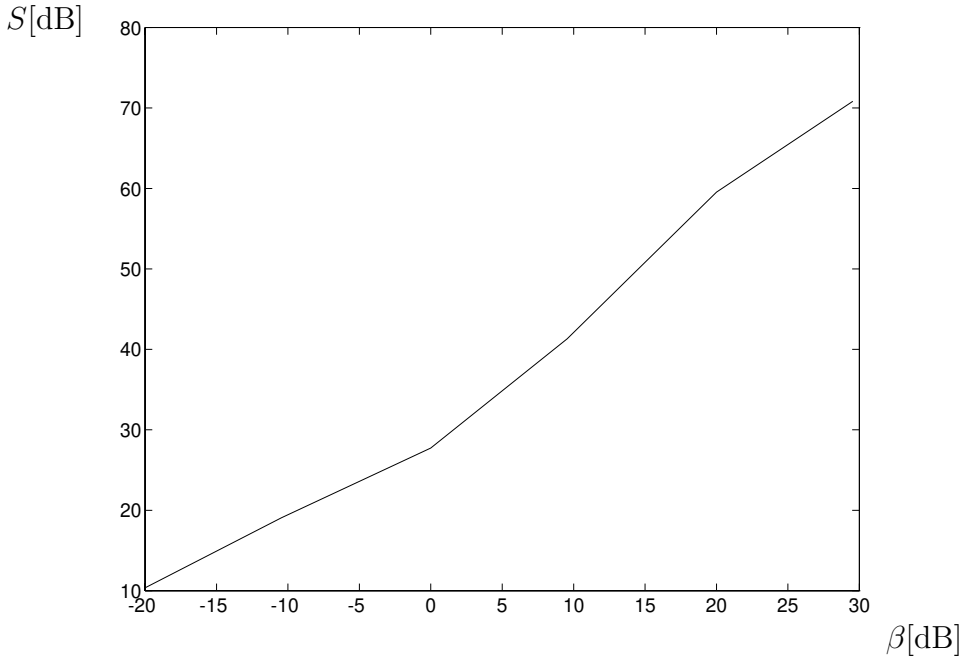

Figure 5.11: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=7. Jamming directions $[9, 10, 11, 12, 13]$.

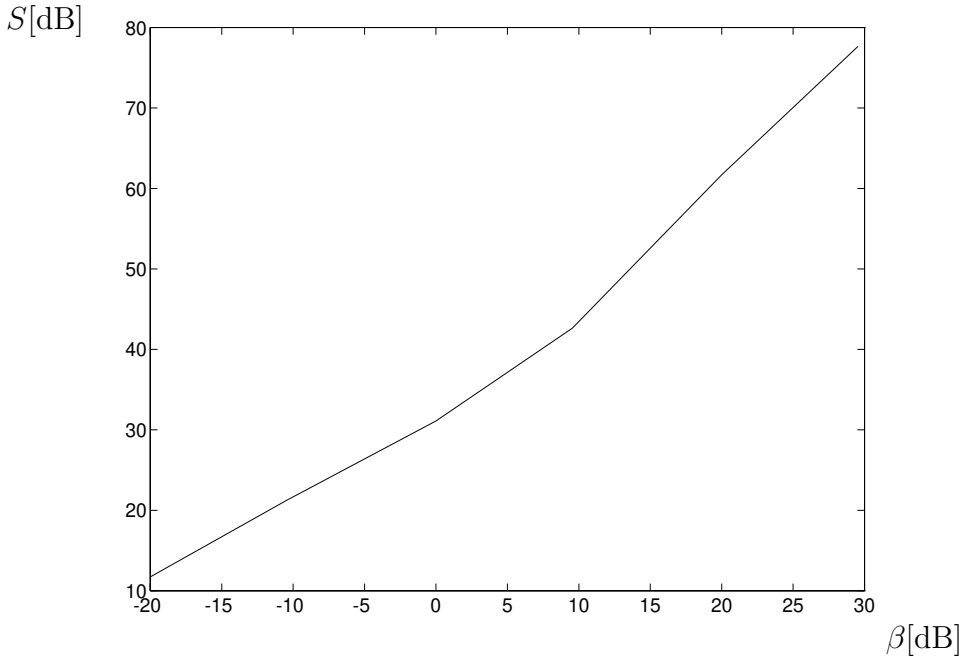

Figure 5.12: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB, SNR=30dB. Linear array with d=0.05 m, N=3. Jamming directions $[10, 30, 45, 60, 80]$.

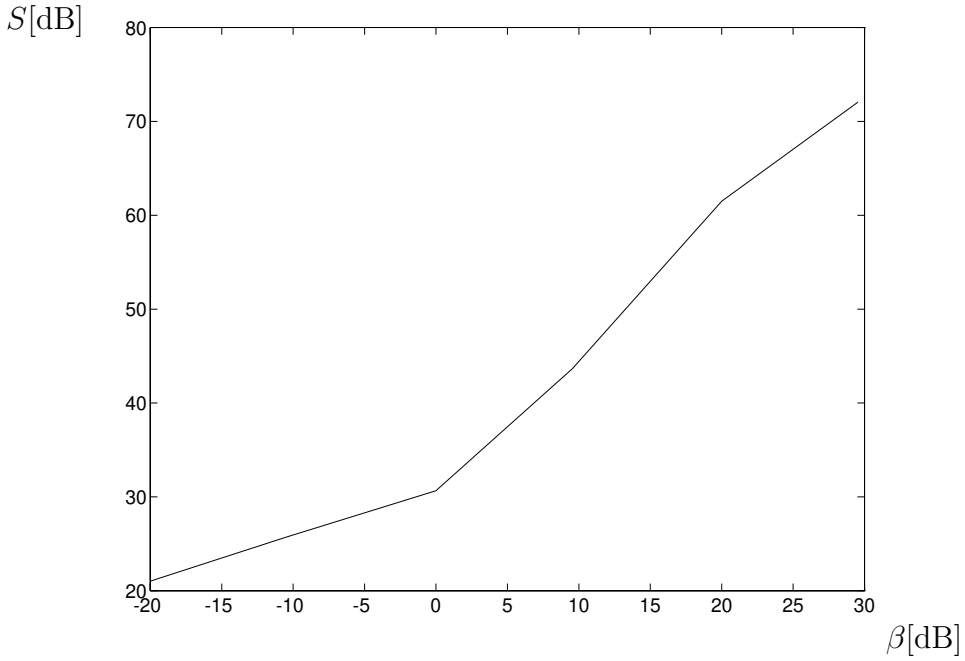

Figure 5.13: Jammer suppression S versus β. When $\beta = 1$ it corresponds to SIR=0dB,SNR=30dB. Linear array with d=0.05 m, N=7. Jamming directions $[10, 30, 45, 60, 80]$.

Chapter 6

Conclusions and Future Work

A cumbersome part of the microphone array realizations is the calibration of the microphones and analog channels at the inputs. A self-calibrating realization has thus been developed. This beamformer has been studied from an analytical point of view by using noncausal Wiener solutions. The study has gained insight into which parameters affect the jammer suppression and give target signal distortion.

An interesting application is when the jammer is a hands-free loudspeaker. Placing and design of the hands-free loudspeaker provide different performance.

The study have shown that in order to achieve good echo suppression and small target distortion, the target signal and jammer signals should be well-separated in spatial domain in order to make the inner product between their transfer functions small.

A method to improve spatial resolution at low frequencies is to consider a sub-band beamformer with unequal element spacing in different frequency bands.

Further work is also needed on this basic beamformer type. We will employ measured transfer functions, and study finite length FIR filters, as well as combine the beamformer with a conventional echo canceller. The design and placing of microphones and loudspeakers in a real environment will also be studied.

Bibliography

[1] J. R. Deller jr, J. G. Proakis, J. H. L. Hansen
"Discrete-Time Processing of Speech Signals"
Macmillan, 1993

[2] S. F. Boll
"Suppression of Acoustic Noise in Speech using Spectral Subtraction"
IEEE Trans. on Acoustics, Speech and Signal Processing, vol 27, Apr. 1979.

[3] J. Yang
"Frequency Domain Noise Suppression Approaches in Mobile Telephone System"
Proc. ICASSP-1993, pp II-363

[4] Y. Kaneda, J. Ohga
"Adaptive microphone-Array System for Noise Reduction"
IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 34, no. 6, Dec. 1986,

[5] S. Nordholm, I. Claesson, B. Bengtsson
"Adaptive Array Noise Suppression of Handsfree Speaker Input in Cars"
IEEE Transactions on Vehicular Technology, vol 42, Nov. 1993

[6] S. Nordebo, B. Bengtsson, I. Claesson, S. Nordholm, A. Roxström, M. Blomberg and K. Elenius
"Noise Reduction Using an Adaptive Microphone Array for Speech Recognition in a Car"
Proc. RVK93, RadioVetenskaplig Konferens, Lund, April 1993

[7] S. Nordebo, S. Nordholm, B. Bengtsson, I. Claesson
" Noise Reduction Using an Adaptive Microphone Array in a Car-A Speech Recognition Evaluation"
Proc. 1993 Workshop on applications of signal processing to audio and acoustics, Mohonk Mountain New York, Oct 1993.

[8] S. Nordebo, I. Claesson, S. Nordholm
"An Adaptive Microphone Array Employing Calibration Signals Recorded On-Site"
Proc. ICSPAT94, Dallas, Oct. 1994

[9] I. Claesson, S. Nordebo, S. Nordholm, M. Dahl
"An "in situ" Calibrated Adaptive Microphone Array
Submitted to Journ. Acoust. Soc. Am., March 1995

[10] I. Claesson, S. Nordholm
"A Spatial Filtering Approach to Robust Adaptive Beamforming"
IEEE Transactions on Antennas and Propagation, vol. 40, no. 9., Sept 1992

[11] B. D. van Veen, K. M. Buckley
"Beamforming: A Versatile Approach to Spatial Filtering"
IEEE ASSP Magazine, April 1988

[12] Nordholm S., Claesson I., Nordebo S.
"Adaptive Beamforming: Spatial Filter Designed Blocking Matrix",
IEEE Journal of Oceanic Engineering, vol 19, no. 4 Oct 1994