NTNU
Innovation and Creativity

Hearing Aid for Social Situations

Emil Wiik Larsen
Espen Oldervoll Moberg

Master of Science in Electronics
Submission date: June 2007
Supervisor: Odd Kr. Pettersen, IET
Co-supervisor: Svein Sørsdal, SINTEF

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem Description

One of the most problematic social situations for people with hearing loss is communicating in a room with a lot of people present and talking at the same time. They often hear the spoken words, but they can not understand what is being said. If the use of a microphone array can increase the speech intelligibility and make it easier to understand speech in a noisy environment, most people with hearing loss agree that the device not necessary have to be invisible.

The task is to design a prototype microphone array for testing beamforming algorithms in a real environment. The processing must be done in real-time and the system must be fully portable. The evaluation of the beamformer is done in three different stages; verification of system design in an anechoic chamber, rhyme word testing in a simulated noisy environment and finally a subjective evaluation of the microphone array in a real environment.

Assignment given: 15. January 2007
Supervisor: Odd Kr. Pettersen, IET

Preface

This master thesis was given by SINTEF ICT. It was written in the spring 2007 at the Department for Electronics and Telecommunication at the Norwegian University of Science and Technology (NTNU).

We would like to use this opportunity to thank our supervisors Odd Kr. Ø. Pettersen, Svein Sørsdal and Trym Holter. We would also like to thank Øyvind Lervik for lending of equipment and Magne Hallstein Johnsen for providing the DSP-kit. A special thanks goes out to all the people who participated in the system testing and evaluation.

Abstract

Conventional hearing aids perform badly in environments with reverberation and noise. In this paper the use of microphone arrays as hearing aids to increase directivity and signal-to-noise ratio (SNR) in a noisy environment are evaluated. A portable microphone array prototype is constructed to test beamforming algorithms in a real environment. Delay and sum beamforming, sub-band beamforming and an experimental type of binaural beamforming is implemented in real-time using the digital signal processor ADSP-BF533. Results from testing showed that a four microphone array using sub-band beamforming outperforms delay and sum beamforming using the same number of microphones. The results also showed that it is possible to obtain binaural impression of the array output and source localization using the proposed binaural technique called beamspreading.

Contents

1 Introduction

One of the most common problem among people with hearing loss is the reduced ability to understand speech in environments where reverberations and interference is present. The cocktail party problem is the problem to distinguish speakers and understand words said when the background noise is of the same character as the wanted speech. The main problem with conventional single channel hearing aids is that they amplify both speech and this background noise equally. A study performed in the United States revealed that the two top reasons for not wearing hearing aids was that they did not provide any benefits and that the hearing aids did not work in background noise. 29,6% of the users said they felt the benefits were minimal or non-existing and 25.3% reported that the hearing aid did not work in difficult listening situations[8]. Background noise was reported to be either annoying, distracting or unacceptable. Conventional hearing aids fail when they are needed the most. This problem has motivated the use of microphone arrays as hearing aids.

The main benefit of using microphone arrays as hearing aids is the possibility to selectively attenuate unwanted noise based on the direction of arrival and leave the wanted signal unchanged. With the use of a multi microphone arrays, it is possible to achieve higher speech intelligibility and enhance the performance in a noisy environment.

2 Microphone arrays

Microphone array processing involves the use of multiple microphones. The microphones can be placed in many different geometries. In this project the microphones will be placed on a linear array in a broadside setup. Microphone arrays are capable of enhancing directionality and signal-to-noise ratio(SNR). These are benefits which are suitable for hearing aids. Microphone arrays achieve directionality using beamforming.

Beamforming is a technique that takes advantages of the fact that the distance from the source to the microphones in an array is different. The difference in propagation paths means that the signal recorded at each microphone is a replica of each other, but that the signals are phase shifted. For a linear array with uniformly spaced microphones as shown in figure 2.1, the phase shifts can easily be calculated.

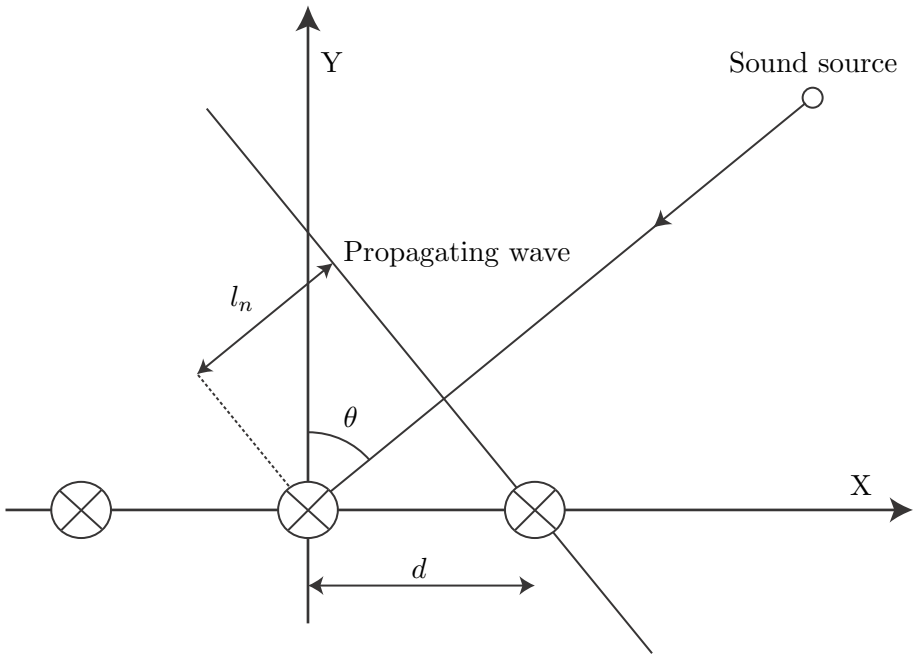


Figure 2.1: A linear array with uniformly spaced microphones with incoming wavefront from a far-field source

A microphone array is said to perform a spatial filtering. It leaves signals coming from a desired direction unchanged, and reduce noise and reverberation in all other directions. Beamforming algorithms are used to determine the delays needed for each microphone to implement a desired shaping of the array directivity pattern. How this is done is described in the section about delay and sum array 2.2.1.

The optimal solution for a broadband signal as speech, would be a frequency invariant array response, but a limitation with conventional beamforming techniques is the performance at low frequencies. If the microphones in the array are too closely spaced, the phase differences between the microphones at low frequencies will be negligible because of large wavelengths and no directivity can be achieved.

2.1 Array properties

2.1.1 Directivity pattern

Directivity patterns or beam patterns show the variation of the intensity received by a microphone as a function of incident angle and frequency. In other words, they show amplification and attenuation of a system for given frequencies and angles. One particular type of pattern is considered in this paper; that of a discrete sensor array. If the microphones have identical frequency response and are linearly equally spaced, the pattern is given by

$$D(f,\theta) = \sum_{N=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j\frac{2\pi f}{c} nd\sin\theta}, \tag{2.1}$$

where f is frequency, θ is incident angle, N is the number of microphones, $w_n(f)$ is the complex weight for element n, c is the speed of sound and d is distance between the microphones. This equation is valid for far field considerations [11].

Directivity patterns can be measured directly utilizing methods like the one described in section 6.4

The directivity of a delay and sum beamformer is dependent on frequency, incident angle of a source, number of microphones, inter-microphone distance and the effective length of the array. The effective length of the array is defined as $L = Nd$. The effective length of an array is the portion of the propagating wave that the array samples. The real physical length of the array is given by $d(N-1)$. Variations in either of these parameters influences the response of the array. How each parameter influences the response will now be described.

To show how the response is influenced by the number of microphones in the array, the frequency and the effective length of the array is set to a fixed value and the number of microphones is varied. The directivity patterns in the figure 2.2 are calculated using equation 2.1.When the number of microphones is increased the sidelobes become more attenuated as shown in the figure. For a fixed length of the array, this means that the spatial sampling per meter is increased and the sidelobes is suppressed more.

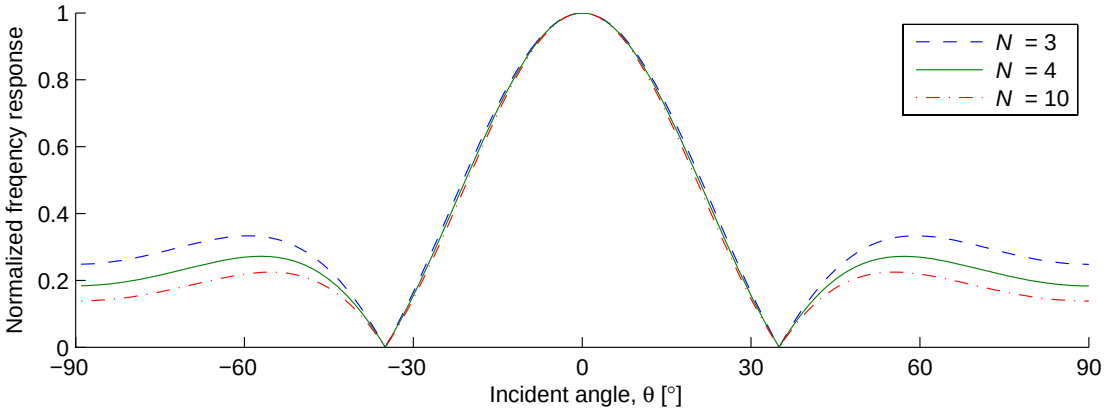


Figure 2.2: Array response for varying number of microphones, $f = 5$kHz and $L = 12$cm.

To show this effect different effective lengths of the array, the frequency and number of microphones is set to a fixed value and the effective length L is varied as shown in figure 2.3. As the inter-microphone distance is increased the effective length is increased and this affects the width of the main beam. The beam width is inversely proportional with the product fL.[11] This means that if L is fixed, the beam width will become narrower as the frequency increases. On the other hand, if N is fixed and $L = Nd$ the beam width can be varied by adjusting fd.

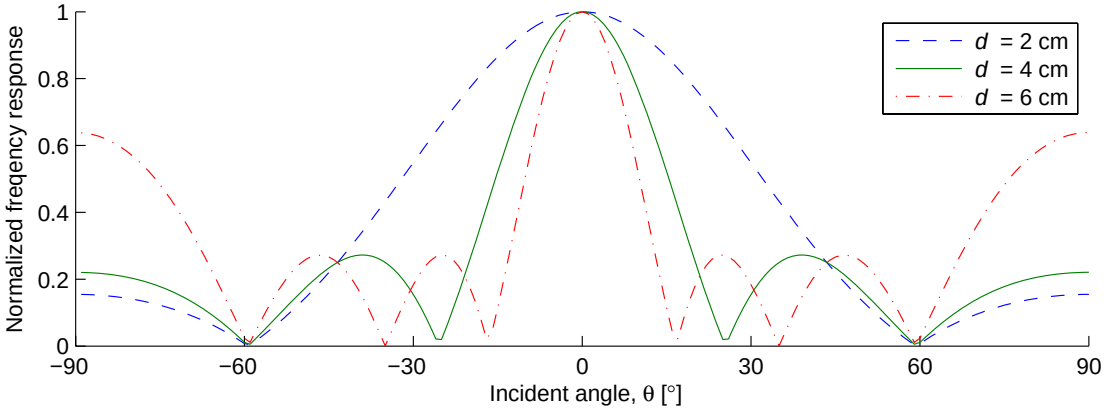


Figure 2.3: Array response for varying effective length L, $f = 5000$Hz and $N = 4$.

When the inter-microphone distance in figure 2.3 is set to $d = 6$ cm, the sidelobe level is approaching the level of the main lobe. This phenomenon is called spatial aliasing.

2.1.2 Spatial aliasing

In terms of temporal sampling, the well known Nyquist frequency gives the minimum sampling frequency needed to avoid aliasing. The microphones in an array perform a spatial sampling of the propagation wave and a similar sampling criterion exists to avoid spatial aliasing. In practice, spatial aliasing will cause the array to pick up interference from directions where there is no desired signal. For temporal sampling, the Nyquist sampling theorem states that to avoid aliasing the signal must be sampled at a rate given by

$$f_s = \frac{1}{T_s} \geq 2f_{max}, \tag{2.2}$$

where f_{max} is the highest frequency component in the signal. The same rules apply for spatial sampling and the requirement to avoid spatial aliasing is given by

$$f_{x_s} = \frac{1}{d} \geq 2f_{x_{max}}, \tag{2.3}$$

where f_{x_s} is the spatial sampling frequency given by sample per meter along the x-axis, and $f_{x_{max}}$ is the highest frequency component in the angular spectrum of the signal. Based on equation 2.3 and knowledge of the geometry of a linear array with uniformly spaced array, the requirement to avoid spatial aliasing is found to be

$$d < \frac{\lambda_{min}}{2}, \tag{2.4}$$

where λ_{min} is the minimum wavelength in the desired signal and d is the inter-microphone distance[11]. This means that when the frequency of the signal increases, the inter-microphone distance must decrease to avoid spatial aliasing. For a broadband signal as speech, this has its implications for the array processing. When the array has a fixed dimension with d and N fixed, the frequency response will be optimized for a given frequency range. If the distance between the microphones is high, it will sample low frequencies well, but if the inter-microphone distance is small, the array will have a better response for higher frequencies. The distance between the microphones is an important property when designing the array.

2.1.3 Far field assumption

In this project, it is assumed that the array is operating in the far field. This far-field assumption is done to simplify the calculations of the array weighting functions. A source is said to be in the far field when the spherical wave front of the propagating wave appears planar at the array. This means that the curvature of the wavefront can

be neglected, making the calculations of the propagation path differences simpler. A source is said to be in the far field of a linear array if

$$|r| > \frac{2L^2}{\lambda}, \tag{2.5}$$

where r is the distance from the array to the source, L is the effective length of the array and λ is the wavelength of the signal [13]. This means that when the frequency of the wave gets higher the wavelength will become smaller, and the far-field distance will increase. For example, an array that has a effective length of 12cm receiving a signal with the highest frequency component at 9kHz, the far field distance is 75cm. The far-field assumption is valid since the hearing aid is planned to be used in a cocktail party setting where the source would be a person talking from a distance equal to or greater than the far-field distance.

2.1.4 Directivity index

Directivity index is a common measure of the performance of arrays. It is defined [14] for linear arrays as

$$DI(f, \theta_0) = 10 \log_{10} \left(\frac{|D(f, \theta_0)|^2}{\frac{1}{2} \int_0^\pi |D(f, \theta)|^2 \sin\theta \, d\theta} \right), \tag{2.6}$$

where $D(f, \theta)$ is the directivity pattern as function of frequency f and incident angle θ, and θ_0 is the look direction of the array. The nominator in the fraction represents the received intensity in the steering angle and the denominator represents the average received intensity for all angles. The directivity index is thus a measure of how well the array suppresses noise from other angles than the desired direction.

2.2 Array algorithms

2.2.1 Delay and Sum Array

The delay-sum beamforming (DSB) is the classic beamforming technique. In DSB the frequency response is fixed and known at all times, and the main beam can be steered to a desired direction. The advantage of delay-sum beamforming technique is that it is easy to realize and the required processing is relative simple. This technique is best suited for narrow band signal and is not optimal for a broadband signal as speech, but it will produce directivity within the speech frequency range. The general formula for delay and sum is given by

$$y(t) = \sum_{n=1}^{N} w_n x_n (t - \tau_n) \tag{2.7}$$

where $y(t)$ is the array output, w_n is the amplitude weighting of each microphone, $x_n(t)$ is the output of microphone n and τ_n is the delay applied to each microphone output. As the name implies, the time domain signal from each microphone are first delayed by τ_n seconds, and then summed to give the array output. If the microphones are uniformly spaced on a linear array like in figure 2.1, the delay for the n^{th} microphone for a source in the far-field is given by

$$\tau_n = \frac{(n-1)d\sin\phi}{c} \tag{2.8}$$

When the source is in the far field perpendicular to the broadside array, the sound wave will be planar at the array and no delaying is needed before adding the signals. The delay is maximum when $\cos\phi = 1$, i.e. when the source is 90° left or right of the array. By calculating and applying the time delays given by equation 2.8, the main beam can be steered to a desired direction.

2.2.2 Sub-band Array

The directivity pattern of a uniformly spaced microphone array is dependent on the frequency range, the inter-microphone distance and the number of microphones in the array. For a fixed numbers of microphones and geometry of the array, the array response will only remain constant for narrow band signals. A sub-band array beamformer performs the same processing as a DSB. The difference is that a sub-band array divides the array response into two or more sub-bands to give a broadband response.

A simple method to achieve beamforming is to construct the array as a group of sub-arrays. Each sub-array is an uniformly spaced linear array, but with different inter-microphone distances. The sub-band arrays are often designed in a interconnected manner, such that the microphones may be used in multiple sub-arrays.

Figure 2.4 shows a two-band sub-array where microphone signals x_1, x_2 and x_3 form the high frequency array and x_1, x_3 and x_4 form the low frequency band. Two of the microphones are used in both sub-arrays.

With the different inter-microphone distances, each sub-band is designed to give the desired response characteristics for a given frequency range. Each sub-array is restricted to a given frequency range by bandpass filtering and the sub-bands are summed to give the final response. Well designed bandpass filters between the adjacent sub-bands is an important parameter in the design process.

2.3 Binaural beamforming

The use of microphone arrays can improve signal-to-noise ratio. However, the advantages come at the price of spatial hearing. Normally the brain takes advantage of the time and

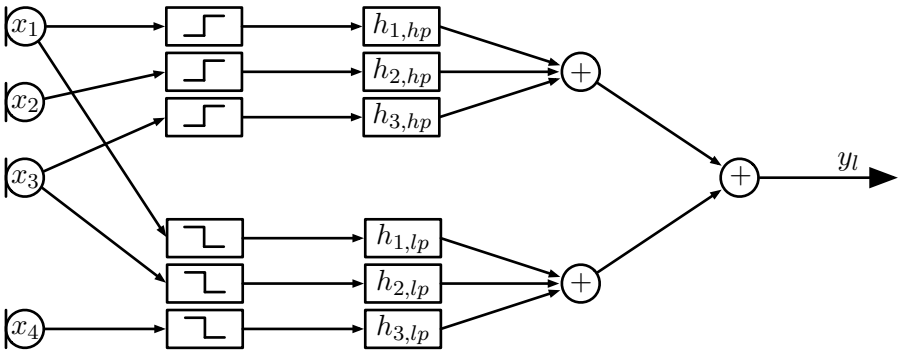


Figure 2.4: Subband filtering

amplitude differences of the sound received by the two ears. It uses these cues for source localization and distinction of sound sources. When microphones are put other places than close to each ear, these cues are lost, and the cocktail party problem arises as the microphones pick up noise from all directions and the brain is not able to distinguish the sources. Microphone arrays can help out in this distinguishing, but it leaves the brain's capabilities unused. Using two traditional hearing aids have proved beneficial over one single [6] as the binaural cues are preserved.

The cues can also to some extent be preserved using microphone beamforming. By using the same physical array, two different beamforming settings can be applied simultaneously. This paper uses a technique that steers the array in different directions for the each of the two ears. This will be called beam spreading, as the main beams of the directivity pattern will be spread out around one focus angle. This will preserve the binaural cues best in higher frequencies where the interaural intensity differences normally are most effective [2]. Sources to the left of the person sound more in the left ear than in the right, resembling the normal experience. The facts that most beamformers work best in higher frequencies and that hearing impaired persons in most cases have problems with high frequencies makes this method an adequate proposition. To be able to process the signal going to each ear differently, two sets of filters are used. This is indicated later on in figure 3.1.

3 Digital signal processing

The beamforming in this paper is done using digital signal processing. The sampled microphone signals are processed digitally to perform the algorithms described in the previous section before the conversion back to analog sound signals. The main processing in these algorithms is done by means of digital filters.

Filtering means changing a signal. The properties changeable are among others delaying and frequency weighting. FIR filters are created in time domain as impulse responses. The time domain equation for a FIR filter is

$$y(n) = \sum_{k=0}^{M-1} h_k x(n-k), \tag{3.1}$$

where $x(n)$ is the sampled input signal, $y(n)$ is the output and h_k are the M filter coefficients of the filter. The equation says that the output is a weighted sum of the M previous samples of the input. This makes the filter stable, linear and time invariant. A well designed FIR filter can perform all the above mentioned filtering given that the number of coefficients is high enough. In real-time applications memory and real-time constraints determine the maximum filter size.

A large amount of beamforming techniques can be performed using only filters and summing. Using a processing as indicated in figure 3.1, one has a very versatile platform for testing beamforming algorithms.

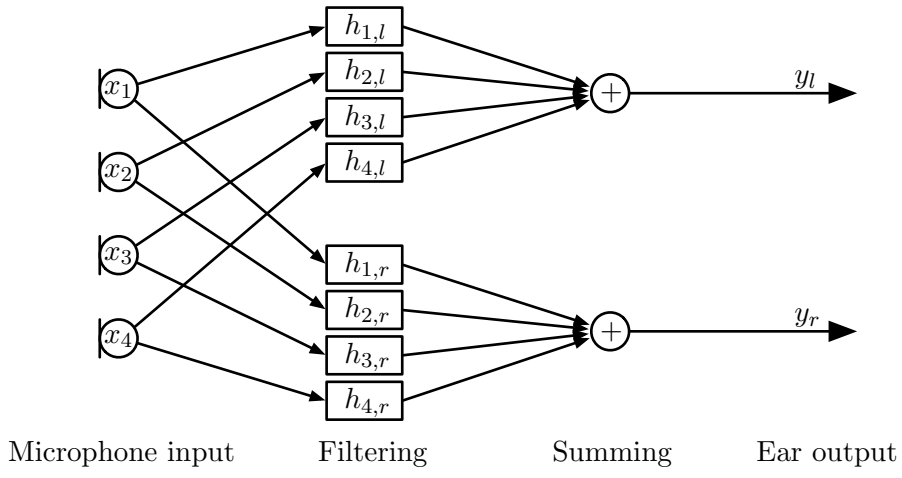


Figure 3.1: Signal processing path

The signals x from the microphones are filtered and summed before sent to the outputs

y. As there are two outputs, one for each ear, two sets of filters are used, to be able to filter the signals differently for each ear.

All filters used in this project are passive, i.e. not adaptive. That means that filter coefficients are not updated automatically or algorithmically, only in the case of interaction from the user.

3.1 Time delaying

An important part of the beam steering algorithms is time delaying. The foundation for delay-sum beamforming is time delaying. The following subsections discuss properties and challenges of digital time delaying.

3.1.1 Integer delaying

The simplest way to implement time delay using FIR filters is to right shift the filter coefficients a given number of positions, e.g. right shifting the filter $b = [1, 0, 0, 0]$ two positions to $b_{delayed} = [0, 0, 1, 0]$ introduces a time delay of $\tau = 2/f_s$, where f_s is the sampling frequency. This type of delaying is in this paper called integer delaying as it is only possible to delay in integer steps n,

$$\tau_{integer} \quad = \quad n/f_s, \, n \in \mathbb{N}. \tag{3.2}$$

As seen from equation 3.2 the delay depends on the sampling rate. Increasing this sampling rate increases the accuracy of the delaying. In delay-and-sum array processing the idea is to time delay signals to make them match in phase. For high frequencies, inaccuracies in time delay introduce a significant mismatch in phase, as the period of the wave is in the same range as the time delay quantization error.

Figure 3.2 shows maximum absolute phase error that the quantization of delay time can introduce as a function of frequency. It is calculated as

$$|\phi_{err,max}(f)| \quad = \quad 0.5 \cdot 360 f \frac{1}{f_s}. \tag{3.3}$$

As frequency gets closer to the Nyquist frequency $f_s/2$, the maximum absolute phase error increases towards 90 degrees. One can observe from the figure that a sample rate of $f_s = 48$ kHz introduces less potential phase error than a smaller sample rate, e.g. $f_s = 16$ kHz. Within the spectral area of human speech ($f \leq 7$kHz, see section 4.1), the maximum error is about 30° with $f_s = 48$kHz, as indicated by the filled box in the figure, but at $f_s = 16$kHz it is almost 80°. The effects of these errors in a microphone array can be evaluated considering the summation of sine waves. When summing two sine waves with a phase difference of $\Delta\phi$, one gets the following result:

$$x_{sum} = \sin(\omega t) + \sin(\omega t + \Delta\phi) = 2\cos(\frac{\Delta\phi}{2})\sin(\omega t + \frac{\Delta\phi}{2}). \tag{3.4}$$

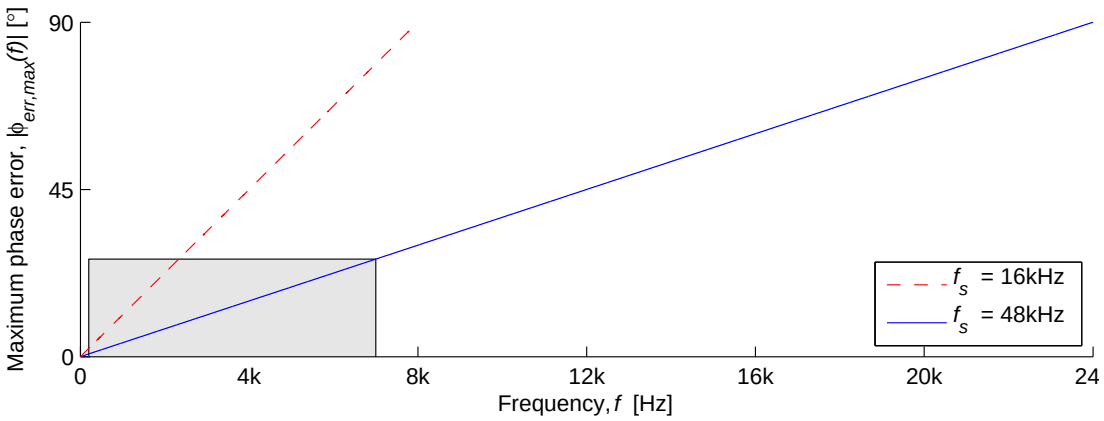


Figure 3.2: Maximum quantization error introduced by rounding off the delay time

Inserting

$$C = 2\cos(\frac{\Delta\phi}{2}) \tag{3.5}$$

results in

$$x_{sum} = C\sin(\omega t + \frac{\Delta\phi}{2}), \tag{3.6}$$

which is a new sine wave with a phase $\frac{\Delta\phi}{2}$ and a new amplitude C. For a two-microphone array where one of the signals has a phase shift due to either delay quantization or inaccurate microphone positioning, the attenuation is given as

$$A_{dB} = 20\log_{10} 2 - 20\log_{10} C. \tag{3.7}$$

This attenuation affects the sound level of sound meant to be in phase, i.e. sound arriving at the array from the desired steering angle.

When summing N microphone signals with different phases $\phi_1, \phi_2, \cdots, \phi_N$, the attenuation becomes:

$$A_{dB} = 20\log_{10} N - 20\log_{10}\sqrt{2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\cos(\phi_i - \phi_j) + N}. \tag{3.8}$$

In stead of having gain in desired direction of $20\log_{10} N$ as wanted, the signal is attenuated with A_{dB}. Assuming worst case scenario, letting half of the microphone signals have phase errors $\phi_{err} = -\phi_{err,max}$ and the other half $\phi_{err} = \phi_{err,max}$, the attenuation can be expressed as a combination of equation 3.3 and 3.8 and showed graphically as shown in figure 3.3 for an array of $N = 4$ microphones.

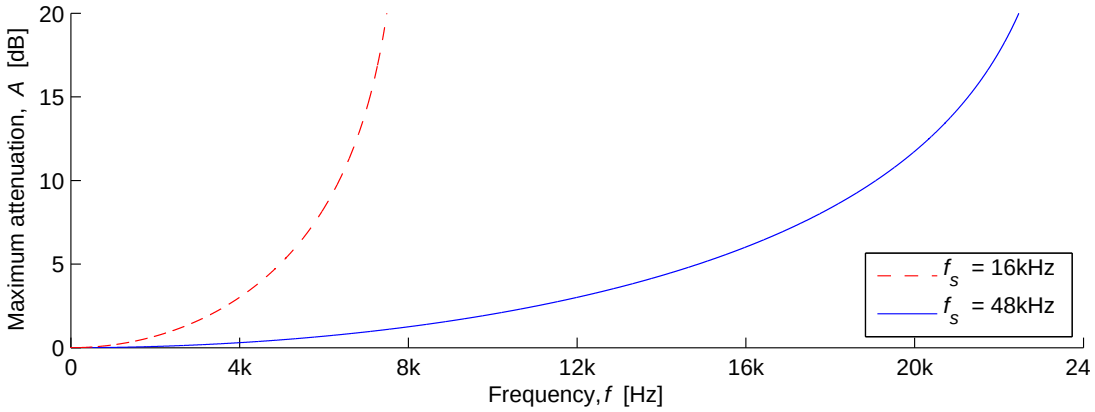


Figure 3.3: Worst case scenario unwanted attenuation caused by quantization error in time delay calculation

It is observable from the figure that the worst case attenuation increases exponentially and gets very high as the frequency closes to the Nyquist frequency. This worst case scenario is not very likely to happen and for a sampling rate $f_s = 48$kHz the attenuation does not exceed 1dB within the speech area ($f \leq 7$kHz), which can be considered acceptable. Using a sample rate $f_s = 16$kHz on the other hand results in severe inaccuracies in large parts of the speech area, and measures has to be done to avoid this. Alternative and more accurate methods exist for calculating delay and one of them is described in the following section.

3.1.2 Delay using Lagrange interpolation

To avoid quantization errors in delaying, fractional delaying using Lagrange interpolation [9] is used. The delay is split up in two parts, one integer part which is calculated as described in section 3.1.1 and one fractional part which is calculated as

$$h(n) = \prod_{k=0,k\neq n}^{N} \frac{D-k}{n-k} \text{ for } n = 0, 1, 2, \ldots, N, \tag{3.9}$$

where $h(n)$ are the $N+1$ filter coefficients and D is the fractional delay between 0 and 1.

Preliminary test listening and measuring comparing arrays using the two delay-types do not reveal any difference in experience or directivity, but the Lagrange-method is used to assure accuracy, since once calculated it does not require any additional computational costs in the real-time processing.

3.1.3 Filter size and delays

Delays might be introduced when using FIR filters even in cases where they are not desired. Group delay depends on the derivative of the phase and is defined as

$$\tau_{grp}(\omega) = -\frac{d\phi(\omega)}{d\omega}, \tag{3.10}$$

and group delay peaks grow large for frequency areas with steep frequency responses, such as low pass or high pass transition areas. Minimum phase filters have minimum delay in areas where the response is flat, but the above mentioned peaks distort the signal by smearing some frequencies in time. For frequencies around 3kHz this is audible if they are more than about 1-1.5ms [1]. Linear phase filters delay all frequency components equally, avoiding the problem of group delay distortion. The problem here is the large total delay introduced, which depends on the filter size N_{taps} and sampling rate f_s as follows:

$$\tau_{fir,linphase} = \frac{N_{taps} - 1}{2F_s}. \tag{3.11}$$

This is tested together with the entire system in section 6.3.

A compromise between the two filter designs may be used to keep clear of both notable group delay distortion and notable total delay, e.g. when assembling a complex filter consisting of multiple parts, some parts can be made minimum phase and some linear phase.

When working with array signals it is important to keep signals in phase and make sure the same delay is applied to all the microphone signals in all active frequency areas. This suggests linear phase filters for channel dependent filtering and the liberty for minimum phase filters for general purpose filtering.

3.2 General purpose filtering

In addition to steer the array beam according to the algorithms in section 2.2, some general filtering is done. High pass filtering is done to reduce low frequency noise, and low pass filtering dampens the higher frequencies where no speech information lies and spatial aliasing begins to occur. An smoothed equalizing filter is implemented to flatten the frequency response of the signal that presented to the user.

Figure 3.4 shows how the different components of the filter chain form part of each filter h_n.

These general purpose filters are implemented as simple linear phase filters. The tap sizes of these filters are kept as small as possible, to keep the total system delay short. The delays these filters introduce are preferred over the smaller but frequency dependent delays that their counterpart minimum phase filters introduce. They are joined with the

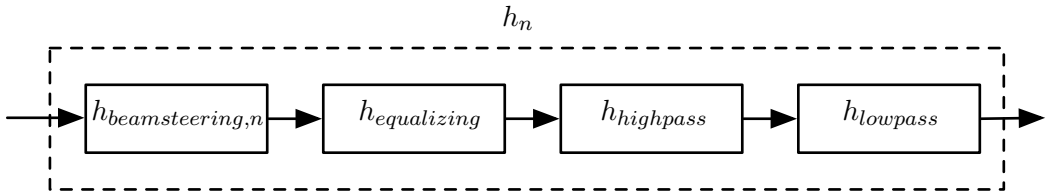


Figure 3.4: Filter Composition.

beam steering filters using convolution to produce a filter with a greater number of coefficients. The final throughput is then given as:

$$h_n = h_{beamsteering,n} * h_{equalizing} * h_{highpass} * h_{lowpass}. \quad (3.12)$$

3.2.1 Equalizing filter

The first filter block after the beamsteering part is the equalizing filter, created using inverse filtering techniques. The impulse response through the entire system is measured using WinMLS software as described in section 6.2. This response is inverted in the frequency domain, smoothed and bandpass filtered to result in a small linear phase filter. The smoothing is done to prevent narrow peaks and dips in the frequency response to cause severe dips and peaks in the inversed response, as they may be caused by small, changing uncertainties in the measuring. The large dips on the other hand are taken account of. This filter's main task is to equalize a dip around 3kHz caused by the reflection of the body of the user, and the increase of gain in the area above this frequency, as described in section 4. The bandpass part of the filter is implemented to prevent the bandpass characteristics of the measured response to turn in to a bandstop filter when inverting it. The very lowest and highest frequency components of the measured response are naturally dampened in all parts of the signal chain as all the components are designed to pass along sound in the audible frequency area. If not handled well this dampening turns out as increase of amplitude in areas where no sound information lies, i.e. the result is high and low frequent noise.

3.2.2 Band pass filtering

The next step is a low pass filter, used to prevent the presence of spatial aliasing in higher frequencies where no speech information lies. Implementing low pass filtering of the sound given to the user removes the frequency area where spatial aliasing is occurring. This can increase the overall directivity, but also reduce the overall sound quality. Even though most important speech information resides in the frequency area below 7 kHz, music and other ambient sounds that might be of interest contain frequency components outside this area, and the quality loss introduced by such filtering can be

felt as annoying to users. Based on this and preliminary listening, a low pass filter with cutoff frequency of 9kHz is selected as a compromise between directivity and full band sound quality

The last part is the high pass filter. As it is necessary with a big filter size to create a steep high pass filter with a low cutoff frequency, it is created fairly gentle, with a dampening of less than 10dB in the stop band. This dampens general low frequent rumble, and gives more clarity to human speech signals.

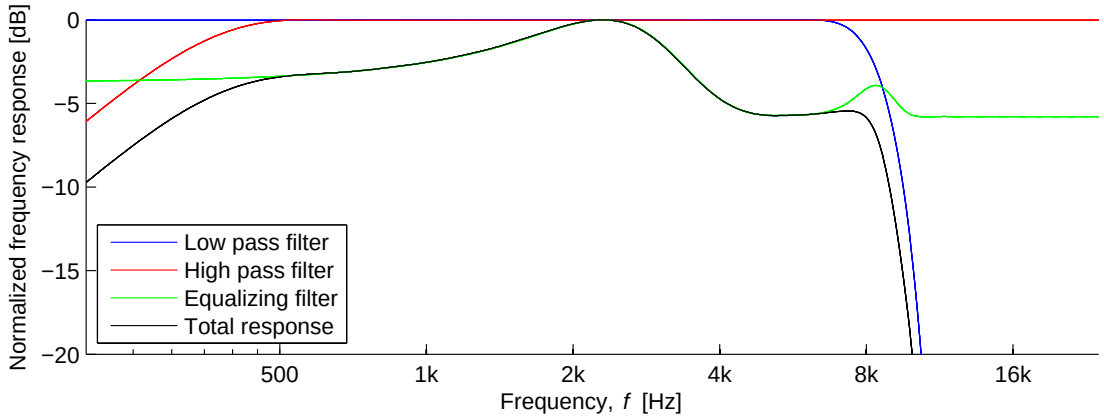


Figure 3.5: Frequency response of the different general purpose filters used and the total output

Figure 3.5 shows the frequency responses for the above mentioned filters together with a final throughput, h_n, combining all the filters.

3.2.3 Sub-band filtering

Sub-band processing requires each adjacent bands to be separated by high and low pass filters. For a two-band sub-band array, one low-pass filter is needed for the lower band and one high-pass filter for the higher band. Since the different filtered signals are to be joined again by summing in the end of the filter chain as suggested in figure 2.4, the transition between the bands has to meet certain requirements. When joining the bands again, the output frequency should be flat. This is done by setting the cutoff frequency equal for the two filters and selecting an appropriate filter algorithm and filter size. For example, both Chebychev and Parks-McClellan filters with small filter sizes fulfill this requirement.

Figure 3.6 shows the summing of two sub-band filters, where it is observable that the sum plotted in black has a flat response. An equivalent real-life measurement of band summation is given in section 6.2.

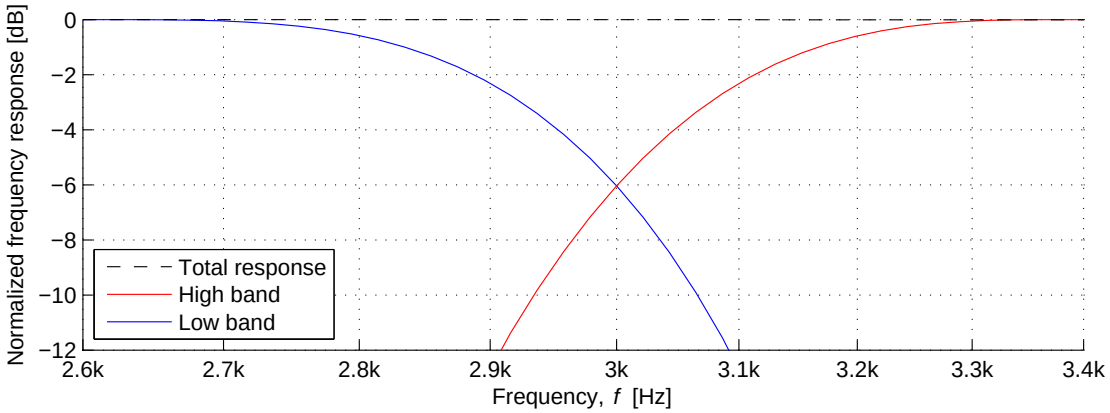


Figure 3.6: Frequency response of two filters used in sub-band filtering, together with the summation of the two.

The choice of crossover frequency depends on the geometry of the array to use, and the spatial aliasing limit given in equation 2.4 defines the upper limits of any given microphone basis.

Another issue to consider when designing sub band crossover filters is the group delay in the crossover section between the bands. As group delay according to equation 3.10 depends on the derivative of the phase, a steep filter will introduce a considerable delay in this area. The choices are delaying the rest of the signal to maintain equal group delay in all frequencies utilizing a linear phase filter, or using a minimum phase filter that will have the delay in just the transition area. The first option leads to a large delay, which together with other filters in the signal chain may compromise any real-time constraint set for this project. The second option has low delay in most parts of the spectrum, but has a phase distortion in the transition band that may worsen the overall sound quality.

Figure 3.7 shows group delay for two different filters. The blue line is the result of adding two sub-bands created using linear phase filters and the red is created using two minimum phase filters. They both have 128 filter coefficients, and the differences discussed above are visible as the linear phase filters have a longer delay all over, but the minimum phase filter produce a frequency varying delay. As they both comply with the constraints for real-time processing defined in section 4.4 and for group delay distortion, they can be used in the project.

An evaluation of different sub-band filters are done empirically and subjectively in sections 6.4.5 and 7.1, respectively.

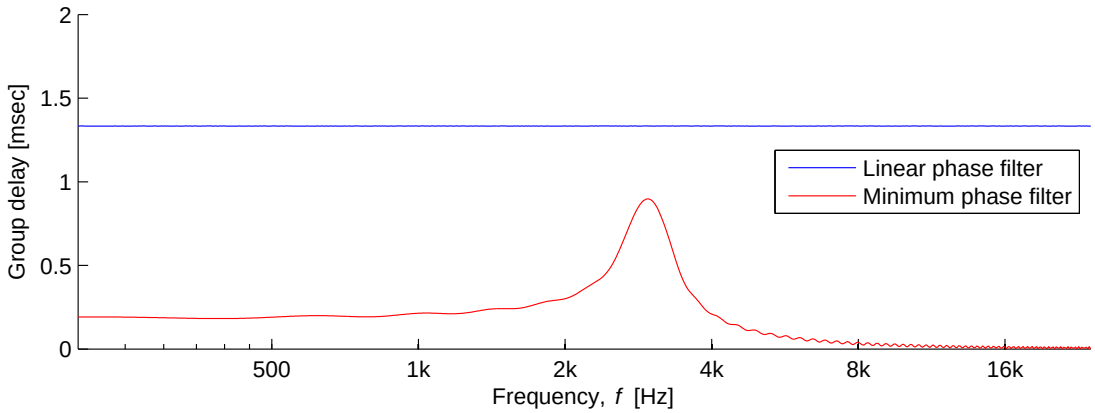


Figure 3.7: Group delay plotted as function of frequency for a linear phase filter and a minimum phase filter

4 Acoustic considerations

4.1 Speech signals

Speech is used to communicate information from a speaker to a listener. Human speech is a continuous signal that has a fundamental frequency in the range of 100-400 Hz. A male speaker has an average fundamental frequency of about 100 Hz and the average for female speakers is 200 Hz.[7] Speech contains frequency components in the range from about 80 Hz to 10 kHz. Speech contains most energy in low frequencies, but the high frequencies are important for speech intelligibility.

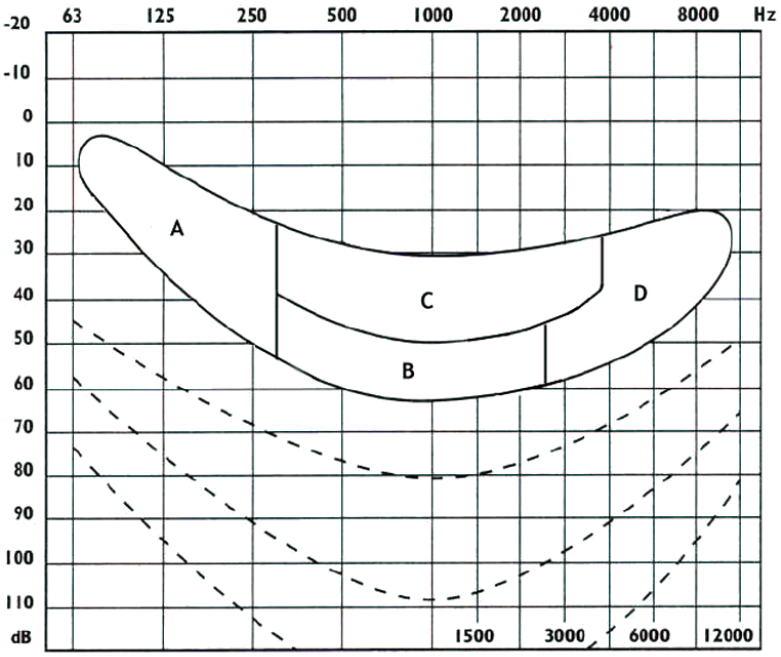


Figure 4.1: Diagram showing dynamic and frequency range of speech signals. A: fundamental frequency area, B: vowels, C: voiced consonants, D: unvoiced consonants

Figure 4.1 shows the placement of the different components of speech in an audiogram. Voiced consonants are defined as the letters b, d, g, j, m, n, r, v and unvoiced consonants are defined as the letters f, h, k, p, s, t. Two things are important to notice from the figure. The consonants are higher pitched than vowels and consonants are spoken more softly than vowels. Consonants may be as much as 27 dB lower in amplitude than the vowels[7]. These factors are important in our ability to understand speech. The great majority of people with hearing loss, suffer from loss in the higher frequencies. This is the problem among people who suffer from sensorineural hearing loss due to aging. Another common problem is noise-induced hearing loss (NIHL). This is also a

sensorineural hearing loss that begins at the higher frequencies in the range 3000 to 6000 Hz and develops gradually as a result of repeatedly exposure to excessive sound levels[12]. A difference between aging and noise-induced hearing loss is that with NIHL the hearing often recovers at about 8000 Hz while hearing loss due to aging do not show this effect.

People with hearing loss in higher frequencies tend to hear the word, but cannot understand what is being said. Consonants convey most of the word information; they are much more important to speech intelligibility than vowels. In a noisy environment like a cocktail party consonants tend to get masked by other voices and noise, thus making it hard for people with a hearing loss to understand speech.

4.2 Array placement

Several things should be considered when selecting the placement of the microphone array. The frequency responses changes when placed differently on the user and effects like body shadowing and reflection affects the choice. Certain placements may also cause more noise than others.

4.2.1 Self noise from body

Different placements on the body will cause the array to pick up different unwanted noise. An array placed on the chest may pick up noise from clothing or vibration noise from the chest, while an array placed on the head or neck may pick up more of the users own voice. Either way this can have a masking effect on the desired signal from the array and could cause loss of intelligibility. The distance from the mouth to the array will either way influence how much of the user's own voice is picked up, suggesting chest placement over a neck or forehead placement.

4.2.2 Body shadowing

If the microphone array is placed on the chest of the wearer, the body will have a shadowing effect on the sound coming for behind the person. The body block incoming sound waves and attenuate the sound received by the microphones. This effect is most prominent for mid and high frequencies. For low frequencies, the sound pressure amplitude is almost the same on both sides of the body. This applies when λ is much larger than the size of the body. For higher frequencies, the shadowing effect of the body is significant as the size of the body is large compared to the wavelength.

Body shadowing is a desirable effect as the default steering of the array is forward and the shadowing helps attenuating sound from unwanted angles, and suggests that chest placement is preferable over head placement.

4.3 Body reflections

The physical depth of the array casing influences the frequency response of the array. Sound reflections from the body can cause destructive interference with the direct sound, and this interference depends on the distance between the microphones and the body, l_1. Even if the microphones are built into a casing, sound waves can propagate through this casing, as it mostly contains air. As the array algorithms used in this paper at some point sum up the signals from the microphones, not only interference from one microphone occurs, but from all the others too.

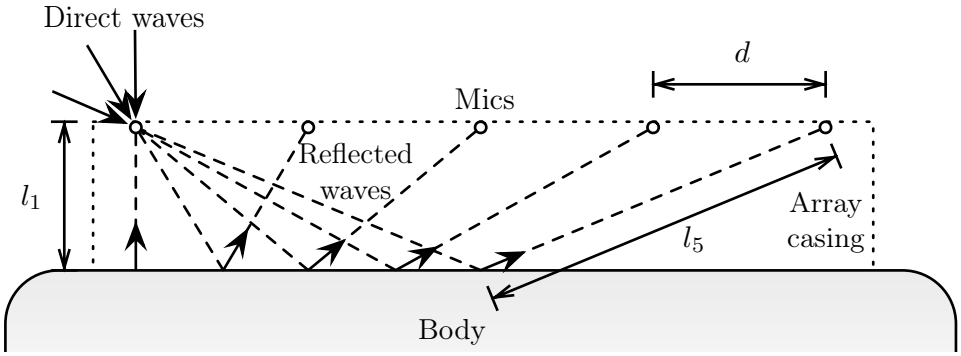


Figure 4.2: Reflections from the body interfere with the direct sound in multiple microphones.

Figure 4.2 shows a microphone array on a human body seen from above. The distance l_1 between the microphones and the body is indicated. Sound entering the leftmost microphone (solid lines) can enter one of the others after being reflected in the body of the user following the paths indicated by the dashed lines. The traveled distance for the reflection is between $2l_1$ and $2l_5$.

A sound wave coming directly from the front of the array, is reflected back to the same microphone. If the wavelength of the wave is twice the traveled distance from the direct impact to the reflection impact, the reflection will partially cancel out the direct sound, as it will be 180° phase shifted. This traveled distance is twice the distance l_1 between the microphones and the body. Frequencies with this wavelength will thus be partially canceled out, and causing a dip in the frequency response around this frequency,

$$f_d = \frac{c}{4l_1}, \tag{4.1}$$

where c is the speed of sound and l_1 is the distance between the microphones and the body. The casing built for this paper leaves the microphones in a distance $l_1 = 2.5$cm from the body, which sets the center frequency of this dip to $f_d = 3440$Hz.

Sound waves coming from other angles than $\theta = 0°$, are not necessarily reflected back to the same microphone, but as the microphone signals are summed up by the array algorithms, a reflection entering a second microphone might interfere with the direct sound entering in the first one. The frequency of the dip will change as the traveled distance is changed for this kind of reflection, making the frequency response angular dependent. For instance sound coming in from an angle $\theta = 30°$ will reflect back to the adjacent microphone having traveled a distance of $2l_1/\cos\theta = 5.86$cm resulting the center frequency for the dip $f_d = 2.9$kHz. In general, considering the array as continuous the dip frequency is given as

$$f_d = \cos\theta \frac{c}{4l_1}. \tag{4.2}$$

The dip can be filtered out for a given angle, but as the response is different from other angles, this filtering will distort sound coming from other angles. In frequencies above the dip, the response will in be periodically gained, as the reflection is creating a comb filter effect [4], which is difficult to filter away.

The distance l_1 can be changed to change the dip frequency f_d. Measurements show that moving the array away from the body results in a lower f_d. To lower it enough for it to stay out of the speech frequency area one would have to move it very far out from the body, something that is not very practically nor advisable. As an example, the distance needed to move the dip down to 340 Hz is 25cm. The reflections do not differ any when measuring the frequency response on a Head and Torso Simulator with or without clothing. Mounting the microphones in a casing significantly dampens the reflections and decreases the negative effects of it compared to when mounting the microphones only on a plate as seen in figure 6.5 in section 6.2.

4.4 Real-time contraint

Real time is for audio defined as the ability for a system to process the signals without introducing a delay of more than 20 ms. This is an important constraint that the array processing must fulfill. The sound from the array is in this project delivered to the ears through ear plugs and all other sounds are blocked. This means that if the system introduces a delay of more than 20 ms the sound delivered to the user will be out of sync with the lips of the person talking.

4.5 Noise analysis in real environment

To get a better understanding of the acoustical parameters in a real cocktail party setting a preliminary analysis is done of two cafeterias, representing a real cocktail party setting. In each cafeteria the sound pressure level (SPL) is measured. Both the tests are performed with a lot of people present. The a-weighted SPL is measured and both the

max value and average value is recorded. The measured values are used as a guideline to set the gain needed in the microphone pre amplifier, see section 5.2.

Figure 4.3: The two cafeterias used for measuring SPL

5 System design

A main part of this project is to build a working microphone array for testing beamforming algorithms in a real environment. The prototype of the microphone array can be seen is picture 5.1. The finalized system consists of different modules. These modules are the microphone array, microphone amplifier, DSP-kit and the headset with attachable earplugs. The array is hung around the neck, the microphone amplifier is mounted at the hip with a clip and the DSP-kit is worn on the back. This makes the system fully portable. In the following sections each module will be described.

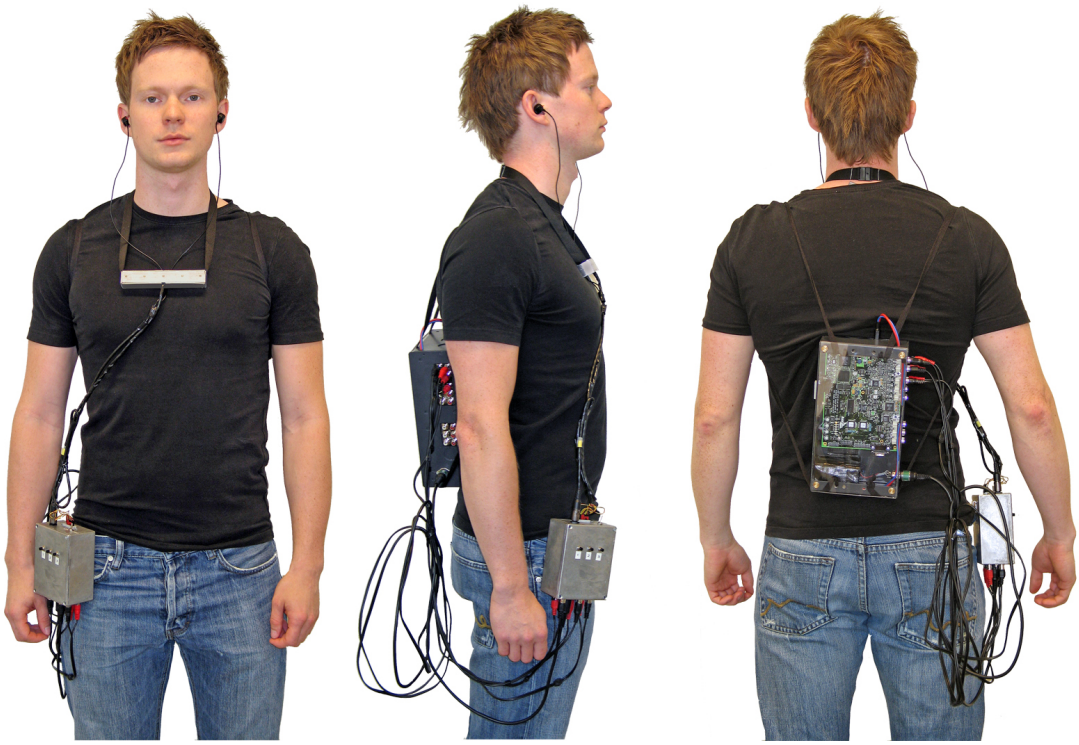


Figure 5.1: Finalized system worn on body

5.1 Microphone array

The microphones used in the array are five high performance FG-23629 microphones from Knowles Acoustics. The FG microphone is one of the smallest microphone on the marked and it is used in conventional hearing aids. The microphone is omni directional, has a low vibration sensitivity and is resistant to mechanical shock. It has a flat frequency response within the range of audible sound, and very low inter microphone phase variance. All these properties make it an ideal microphone for fitting in an array.

The microphone has three terminals. These are ground, output and voltage supply, and are connected with black, yellow and red cables, respectively, to a flat three pin connector as shown in figure 5.2.

Figure 5.2: Headset and microphone connectors.

The effective length of the array should be as large as possible to maximize the performance for lower frequencies. The obvious limitation of the size of the body of the wearer applies, but also the number of microphones/channels available is a limiting factor when wanting to build a broad array. The number of microphones dictates the distance between the microphones at any given array length, $d = L/N$, and to avoid spatial aliasing for high frequencies this distance has to be kept small. Using equation 2.4 a spacing of 3.05cm avoids spatial aliasing up to 5.6kHz. However, as the aliasing limit depends on the steering angle, this frequency is higher with moderate steering angles. Simulations show that when steering the beam straight ahead, no spatial aliasing occurs for frequencies up to 9kHz.

To conclude, given the number of microphone channels, the spacing between the microphones of 3.05cm is chosen as a compromise between a wide effective length favoring directivity in low frequencies, and a short inter distance favoring no spatial aliasing in high frequencies.

Figure 5.3 shows the constructed microphone array and the ear plugs used in this project. The 5 microphones are named A-E and are used in two different setups. Microphones A, B, C and D are uniformly spaced and form the standard delay and sum setup. Microphones A, B, C and E constitute the sub-band setup, where the A, B and C form the high frequency array with microphone spacing of 3.05cm and A, C and E form the low frequency array with inter distance of 6.10cm. The sub-band array will in theory perform better at low frequencies, due to the extended effective length of the array when using A, B and E.

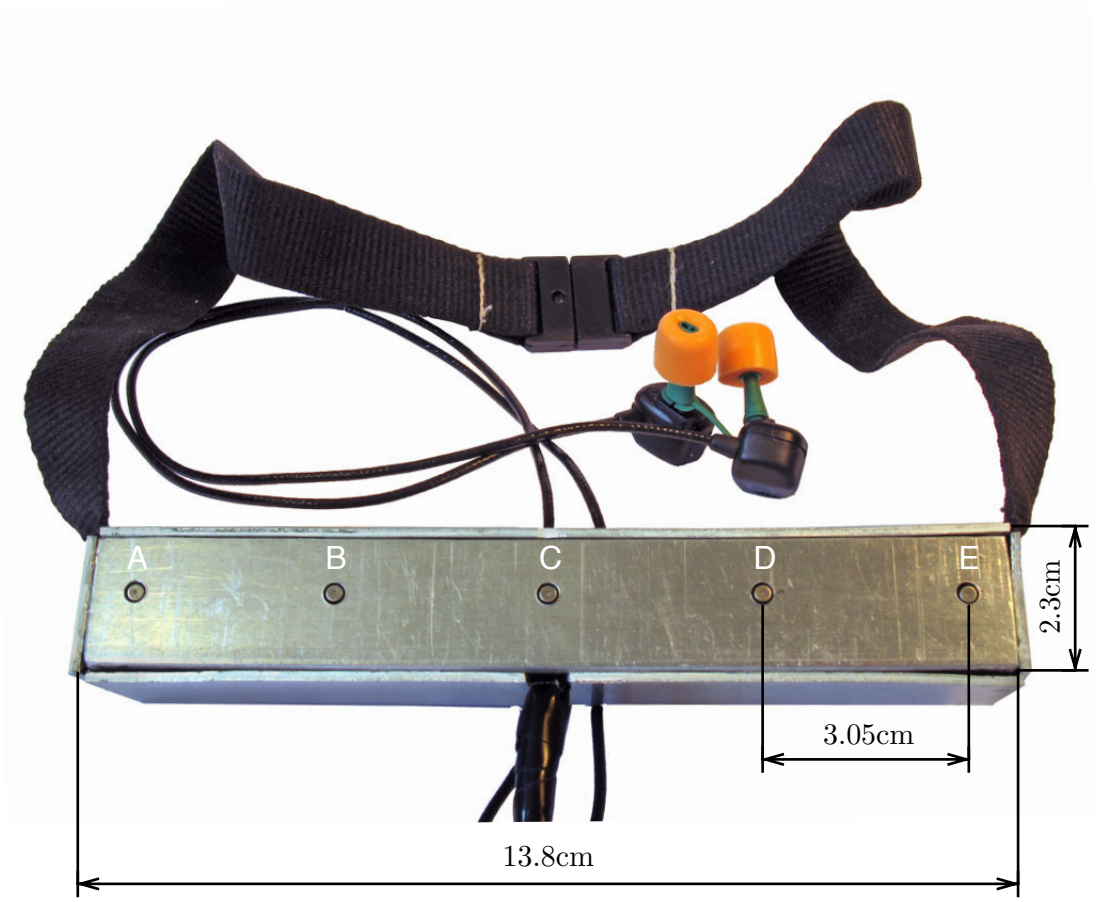


Figure 5.3: The microphone array with microphone distance and casing size indicated

5.2 Microphone preamplifier

The microphone array consists five microphones where four of them are active at in each configuration and therefore a four channel amplifier circuit is needed. To amplify the signals a two channel Burr-Brown OPA2134 operational amplifier is chosen. The OPA2134 is an ultra-low distortion, low noise operational amplifier specified for audio applications. It offers a wide output swing, to within 1V of the rails, which allows increased headroom. The op amps can be operated from ±2.5V to ±18V power supplies. Two of these operational amplifiers give the four channels needed to amplify the signals from the microphones.

The operational amplifiers are powered by a single supply rail given by a 9V battery. This gives the amplifier circuit a dynamic range of 7V which is enough to drive the inputs of the DSP-kit that has a dynamic range of 6.16V. To achieve a functional operational

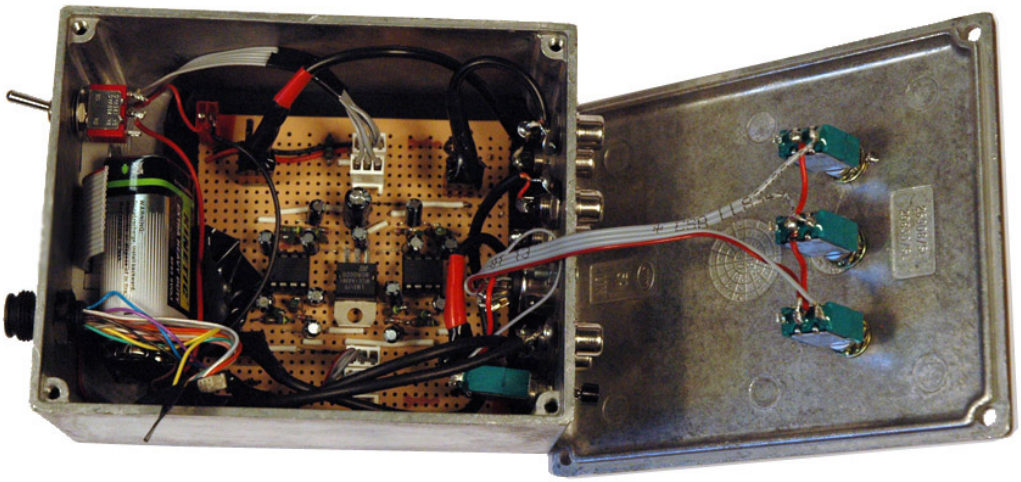


Figure 5.4: Preamplifier circuit mounted in amplification box.

amplifier using a single supply, a voltage divider consisting of the two 47kΩ resistors shown in figure 5.5 is used to give the op-amp a virtual ground at 4.5V. Amplification is chosen so that the entire dynamical range can pass along the signal chain undistorted. Maximum expected sound pressure level based on measurements in section 4.5 is 94 dB. With an extra 13 dB headroom the dynamic range is set to include levels up to 107 dB. The sensitivity of the microphones is -53 dB referenced to 1V/0.1 Pa, which means that for a sound pressure level of 74dB (0.1Pa), the output from the microphones is given as

$$V_{mic,74dB} = 1\text{V} \cdot 10^{\frac{S}{20}} = 1\text{V} \cdot 10^{\frac{-53}{20}} = 2.2\text{mV}, \tag{5.1}$$

The level at 107dB will then be 98mV, and to take advantage of the entire dynamic area of the DSP-kit line level input of 2.18V_{RMS}, the gain is set to

$$G = \frac{2.18\text{V}}{98\text{mV}} = 22. \tag{5.2}$$

The op-amp gain is set by selecting appropriate values for R_1 and R_2. The gain is given by

$$G = \frac{R_1 + R_2}{R_1}, \tag{5.3}$$

which means that selecting values $R_1 = 2.2\text{k}\Omega$ and $R_2 = 47\text{k}\Omega$ the resulting gain remains as

$$G = \frac{47k\Omega + 2.2k\Omega}{2.2k\Omega} = 22.3636. \tag{5.4}$$

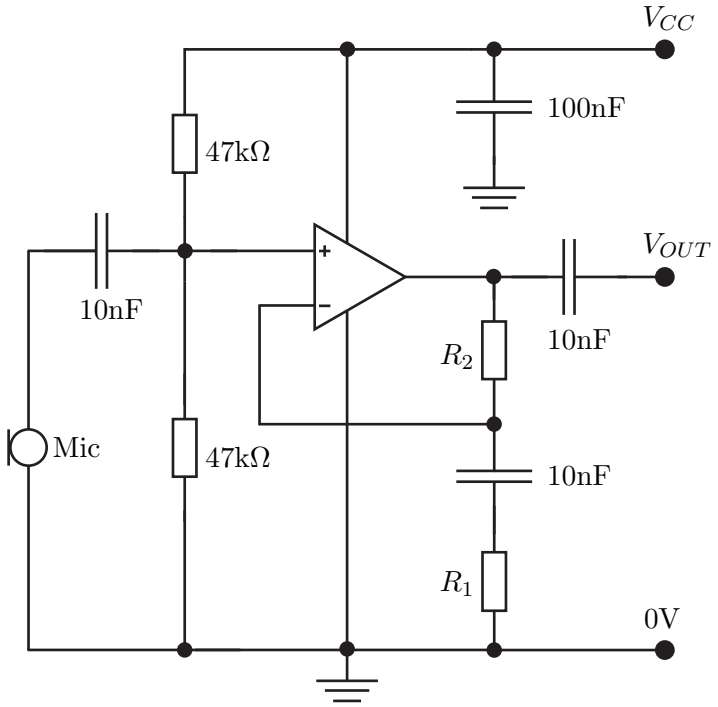


Figure 5.5: Circuit schematic for preamplifier

The amplifier circuit's input and output are decoupled using two 10nF capacitors to remove the DC-offsets from the microphones and the voltage division. In addition a 10nF is connected between R_1 and R_2 to smooth out any ripples and finally a 100nF capacitor is connected between V_{CC} and ground to smooth out any sudden voltage spikes and protect the components from excess currents on disconnection of power supply.

5.3 DSP-kit

The hardware unit chosen for realizing the digital signal processing of this project is a DSP BF 533 from Analog Devices, delivered on an evaluation board BF 533 EZ-KIT with 4 analog inputs and 6 outputs.

There are many possible ways to implement a beamformer in hardware. The most natural choices for this purpose are PICs - general microprocessors, DSPs - digital signal processors - and FPGAs - field programmable gate arrays. The two first choices are programmed in assembly or higher level programming languages such as C, and the instructions are executed one after the other, in a serial manner. The last choice is programmed in a hardware descriptive language and the program just defines how the

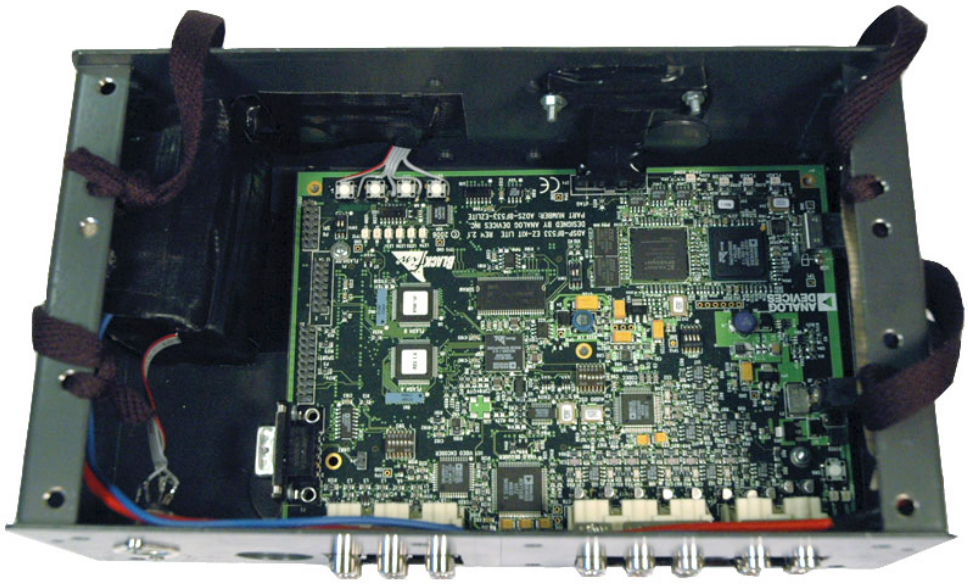


Figure 5.6: DSP-kit mounted in casing

built-in ports and blocks of the chip are to be connected, and in operation all connections are already done and the signal processing is done in a parallel manner. Traditionally DSPs have been the most used choice when developing signal processing equipment such as the one intended for this project, but FPGAs have become increasingly popular recently because of it's ability to handle large amounts of data in parallel. Still, more work has been put into DSP development over the years, and there exists a large number of well developed algorithms and modules for basic and advanced signal processing made available by both the producers of the processors and by third party developers.

The DSP evaluation kit selected for this project is chosen because of these already available code modules, and also because of it's development software and it's hardware specifications.The software that ships with the evaluation board has functionality for designing signal chains graphically and inserting blocks of code as modules in these chains, through the designing tool Visual Audio. This significantly reduces the number of lines of code needed to be written. The hardware of the evaluation board stands out in the crowd of kits with it's number of on board digital-analog and analog-digital converters. Most development kits on the market only have two analog inputs, but as this one has four, data can be processed from four microphones without the necessity of connecting and configuring external analog-digital converters.

The different configurations programmed on the DSP are called presets. These presets are stored in the evaluation board's memory, and when loaded, behavioral parameters change. Such parameters include filter coefficients and signal routing. The presets are

defined using a mathematical scripting software, Matlab, compiled together with the rest of the code and loaded onto the DSP.

Code is written to create the ability of cycling through and loading these presets during use. Push buttons create run time interrupts that pauses the normal running of the signal processing, change the parameters affected by the preset, and returns to normal running, all without notable gaps in the signal flow. This absence of gaps is verified empirically and the verification is more deeply described in section 6.3. On board LED-lights on the evaluation board are lit to indicate which preset is active and can be seen through the Plexiglas lid of the box.

Having presets means that several settings can be programmed into the DSP. There are two main reasons for including this feature; the first reason is for test purposes. Different presets are made available for the test person to choose from, and the freedom of being able to change parameters on-the-fly gives the test person the opportunity to compare and evaluate different settings without the need of reprogramming. The second reason is for adjustability in a real use situation. The user might want to be able to adjust volume, desired beam width or angle, or even select a broad band preset for music listening.

5.4 Head set

The headset used is taken from a hearing protection gear developed by Nacre. The headset consists of two small earpieces with two microphones and one loudspeaker in each. In this project, only the loudspeaker part will be used to present the processed sound to the user. Three different sizes earplugs can be attached to each ear piece to give maximum fit and damping. The three sizes available are small (red), green (medium) and blue (large). The earplugs are developed by SINTEF for Nacre. The headset is driven directly by the line level outputs from the DSP-card and no extra amplification of the signal is needed.

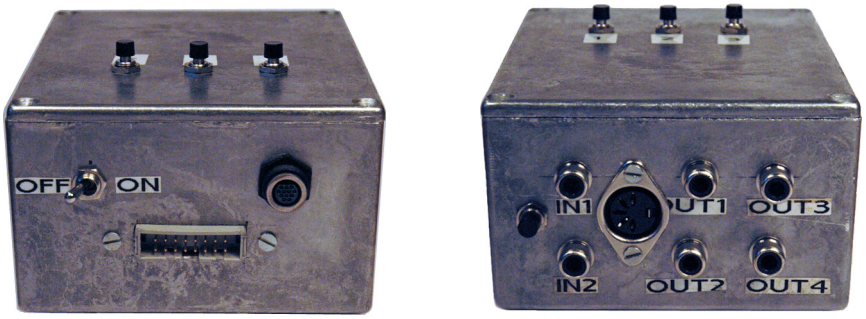
5.5 Interconnecting modules

To create a portable system, both the DSP-kit and the microphone preamplifier circuit are mounted in metal boxes for protection against noise and direct contact. The modules are connected together with cables.

As described in chapter 5.1, the microphones are terminated in a flat connector with three conductors. To get four of these connected to the amplifier box a custom connector is created using a 16 channel bus connector. Picture 5.7 (a) shows the connector fitted in the center of the side of the amplifier box.

Four standard phono connectors are mounted on the side of the casing and are use to connect the amplified microphone signals to DSP-kit. In this way, stock phono cables can

be used to connect the amplifier-box and the DSP-kit. The phono connectors are visible on the right side of DSP-kit box in figure 5.8. In the same way, the processed audio signals from the DSP-kit is sent back to the amplifier-box using two phono connectors mounted on the left hand side of the amplifier casing as shown in picture 5.7 (b). The connectors are labeled OUT1 to OUT4 and IN1 to IN2 as shown in picture 5.7 (b).



(a) (b)

Figure 5.7: Connectors on the amplifier-box

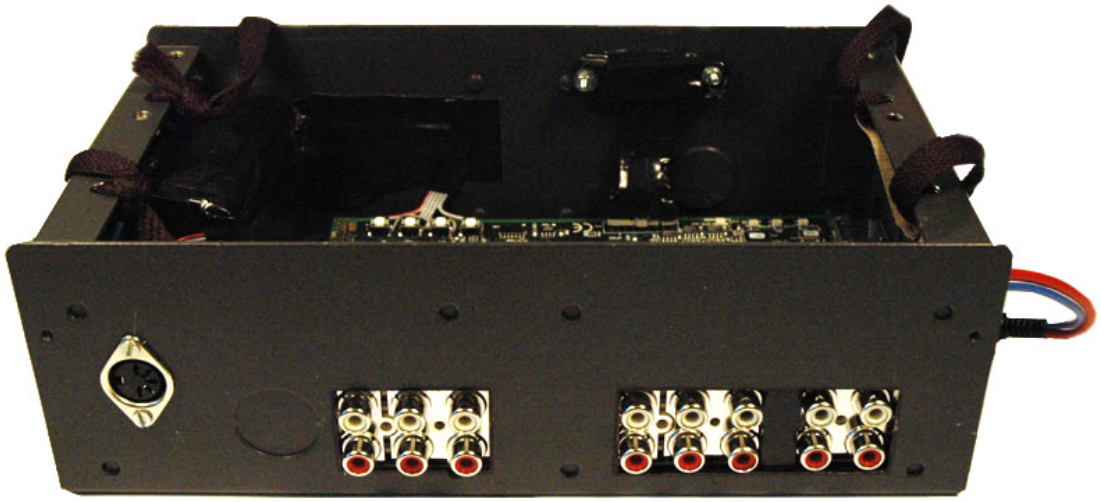


Figure 5.8: Connectors on the DSP-box

To connect the headset described in chapter 5.4, a custom connector made by AMGAB is used. This connector is shown to the right in picture 5.7 a).

5.6 User interface

To give the user a possibility to change between different parameters in real time during testing, the buttons situated on the DSP-kit are used to update the filter coefficient in the FIR filters running on the DSP. Since the DSP-kit is placed inside a box, external buttons have to be made. This is done by soldering wires directly onto the button leads on the DSP-kit and connected to a chassis mounted five pin DIN connector. The connection between the DSP-box and the amplifier-box is done by a single five pin DIN cable. The DIN-connector is used because is has the numbers of pins needed for the four buttons and premade cables in different lengths can be used. External buttons on the amplifier box is fitted to remotely change the presets.

Figure 5.9: Preset button control

Three buttons mounted topside on the amplifier box will give the user the ability to change between three presets as described in section 5.3. There is also a fourth button for changing between different groups of three presets. This button can be seen in picture 5.7 (b) and is placed far left on the side of the amplifier box.

5.7 Power supply

To make the system portable, both the microphone amplifier and the DSP-kit have to be powered by batteries. This section will describe how this is solved.

5.7.1 Microphone preamplifier power supply

The current drain for the microphones and microphone amplifiers where measured to be 33mA with high level sound input on the microphones. A standard 9 V alkaline battery is sufficient to deliver this amount of current and is also chosen because of its small size. The battery is directly connected to the operational amplifiers. There is no need to regulate the supply voltage to the op-amps because they will still operate when the battery is near depletion. The voltage supplied to the microphones need to be regulated from 9V to the 1.25V they need as supply voltage. This is done by a simple regulating circuit using a LM317T voltage regulator connected as shown in figure (5.10).

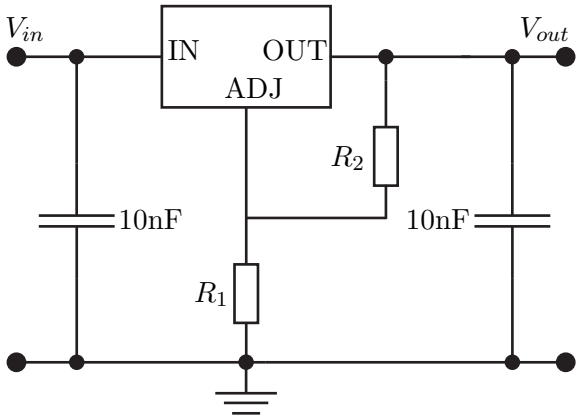


Figure 5.10: The voltage regulator ciruit.

The LM317T ouput voltage is given by the equation

$$V_{out} = 1.25 \cdot (1 + \frac{R_2}{R_1}). \tag{5.5}$$

The LM317T always supplies 1.25 V between the ADJ and OUT pin. By connecting the ADJ pin directly to ground and thereby short circuiting R_2, V_{out} is become 1.25V which is the voltage needed to supply the microphones. A common value of R_1 is in the range 100-1000Ω, and for this regulator R_1 is chosen to be 220Ω.

5.7.2 DSP-kit power supply

The DSP-kit needs more current than the microphone amplifier. With all the cables connected and algorithm performing delay and sum running on the DSP, the current is measured to be 1.2A. An alkaline battery will not be able to deliver this type of current. To solve this problem, six rechargeable nickel metal hydride batteries (NiMH) are soldered together in series to deliver 7.2V at normal capacity. This voltage will vary from about 8.4V when fully charged, to 6V when near depletion. The DSP-kit

has an internal regulator, so there is no needs to regulate the voltage from the battery pack.

To recharge the batteries, the battery pack is connected directly to a power supply. The power supply is adjusted to give the right voltage and a steady current. The minimum voltage needed to get a full charge varies with temperature. A NiMH battery need at least 1.41V per cell at 20 ℃. The voltage is set to 12V which is more than enough for the six cell battery pack. The capacity C of the NiMH batteries used is 4100mAh. A safe and easy way to charge the NiMH batteries is to charge at $\frac{C}{10}$ where C is the capacity of the batteries. The batteries used to power the DSP-kit do not come with any available technical information and the coulometric charging efficiency is therefore assumed to be between 60-70%. A typical value for the coulometric charging efficiency for NiMH batteries is 66% and used in the calculation to find the charging time. The simple equation for charging time is given by

$$T_{charge} = \frac{C}{A_{charge} \cdot Charge_{eff}} \tag{5.6}$$

where T_{charge} is the charging time, A_{charge} is the charging current and $Charge_{eff}$ is the coulometric charging efficiency. This means that if equation 5.6 is used and the charging current is set to 410mA, the battery is fully charged after 15 hours. When the batteries are fully charged, the DSP-kit has a battery lifetime of at least three hours.

To verify the designed system, the system is thoroughly tested to ensure correct operation. The methods and results used are described in the following section.

6 System tests and verification

6.1 Amplification of microphone signals

As described in section 5.2, the output voltage needed is calculated to be 22.4 times the input voltage. To verify that all the channels are amplified correctly, a signal with $1V_{pp}$ amplitude is sent through each channel and the output amplitude is read from an oscilloscope. By dividing the output amplitude with the input amplitude the gain is found.

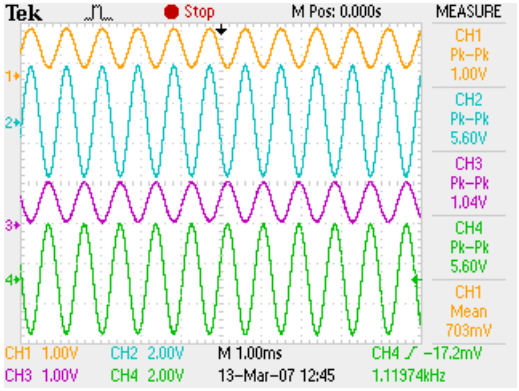


Figure 6.1: Screen shot from oscillator showing amplification

Figure 6.1 shows how the amplification is verified for two channels at the time using an oscilloscope. In this preliminary measurement the amplification is set to 5.7 and measured to be 5.6. The measured value has a deviation of only 1.8% of the calculated value. The final amplification is set to 22 and the maximum deviation is measured to be 2%. All four channels are confirmed to be satisfactorily close to the calculated value. The decoupling of DC-components are also verified to be decoupled in this test.

6.2 Microphone array frequency response

A frequency analysis is done in an anechoic chamber to ensure that the microphone signals are in phase and operates correctly. The analysis is done by the software WinMLS 2004, level 6c. WinMLS generates a sine sweep which is sent to a speaker inside the anechoic chamber. The sine sweep is picked up by the microphone array and routed back to the measuring computer outside the chamber. Based on the returned signal, WinMLS can calculate both the magnitude and phase response of the microphones.

Part of the system testing is to evaluate the variation of the frequency response for different placements of the array on a Head And Torso Simulator as discussed in sec-

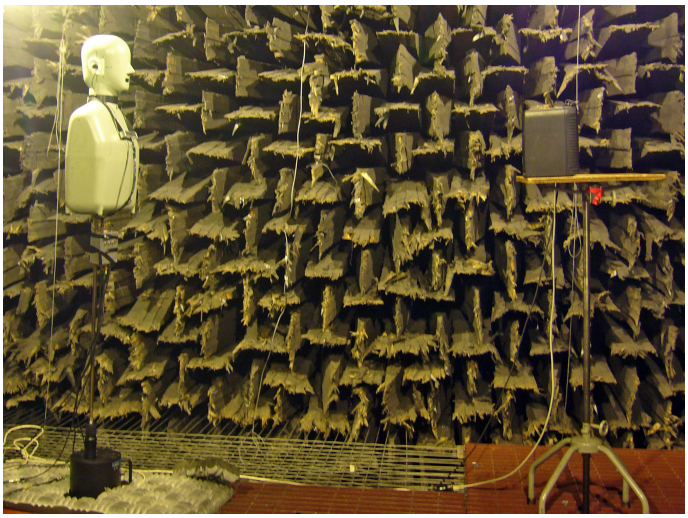


Figure 6.2: HATS and speaker seen in anechoic chambe

tion 4.2. How the frequency response varies with different positions can be seen in figure 6.3. There are pros and cons with the different placements. Placing the array on the head gives the user ability to steer the beam by turning the head, but is not cosmetically pleasing. The user must either wear the array mounted on glasses or a headband. The frequency response for the head placement results in a frequency response with multiple dips which is not optimal. A simple and discrete solution is to place the array on the chest with a strap around the neck. The chest placement gives a relatively flat frequency response with a dip at about 3kHz. The neck placement has a similar response as the chest placement, but the dip is 10dB deeper and is therefore excluded from further considerations. The chest placement is chosen because of the frequency response and easy fitting on a human body.

To demonstrate the effect of different array distances between the array and the body as described in 4.3 a plot of three different distances are shown in Figure 6.3. A distance of 2.2 cm gives a dip in the frequency response at about 3kHz. When the distance is increased to 3.5 cm the dip appears at 2100Hz. This is because the wave has to travel further than with a distance of 2.2 cm and the negative interference affects frequencies with longer wavelengths. With a distance of 5.8 cm the interference creates a wide dip in the frequency response spanning from 900 to 1900Hz. The dip frequencies are a somewhat lower than anticipated from the calculations in section 4.3, but the trends are clear as the dip frequency lowers as the distance increases.

The array constructed in this project is 2.5 cm in depth when it is fitted in the casing. Figure 6.5 shows the effect of fitting the array in a casing. The end result is that the dip is not that prominent which is a wanted effect.

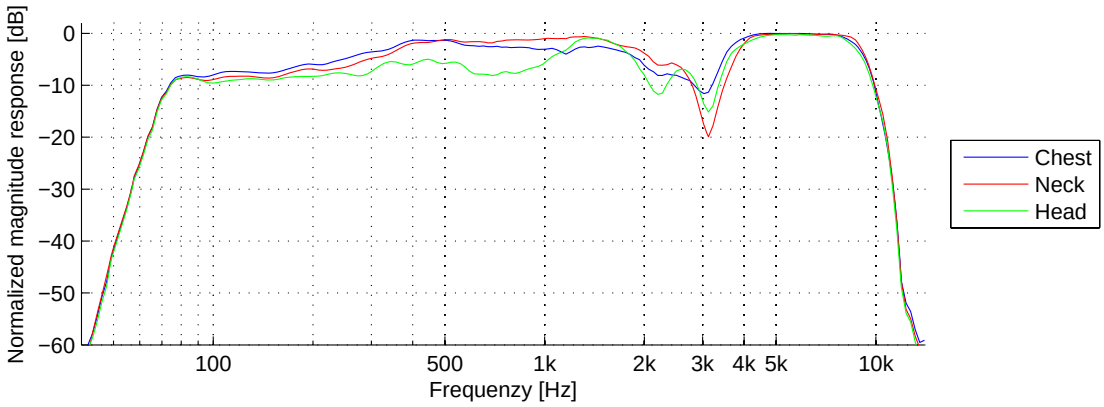


Figure 6.3: Frequency response for different placements of the array in vertical direction

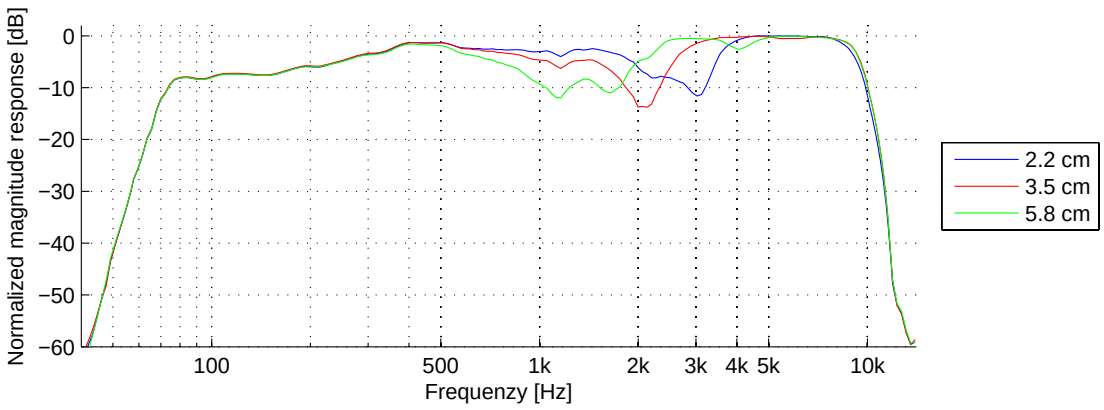


Figure 6.4: Frequency response for different distances from microphone to body of the wearer

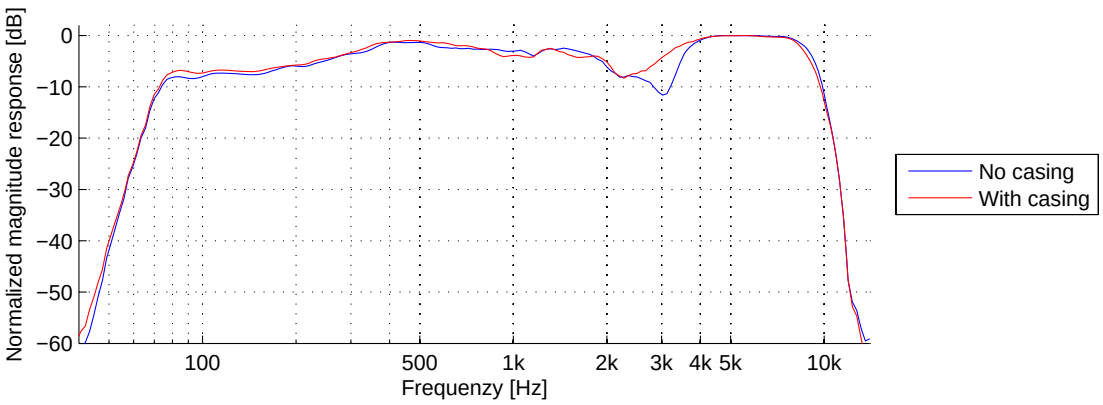


Figure 6.5: Frequency response with and without casing

6.3 System delay

The delay introduced by converting the input signals from analog to digital, processing them and converting them back to analog signals is measured in order to make sure the system acts according to given real-time constraints described in section 4.4.

In addition to frequency analysis, WinMLS can be used to measure system delay. A test signal is generated by the software, sent through the computer's sound card into the analog inputs of the DSP-kit and the output from the kit is sent back to the sound card and registered by the software. The sound card itself introduces some delay itself, but this can be measured and subtracted from the measured system delay in order to find the delay introduced by the system itself.

The signal chain is as shown in figure 6.6.

Figure 6.7 shows the maximum delays introduced by FIR filters of different tap sizes. The solid dots in the figure shows real measured values of the delay, and the dotted line is estimated system delay based on these measurements. A real-time constraint of 20 ms implies that the maximum tap size within this constraint is 800 taps. If the filters are linear phase filters only half the delay will be introduced, e.g. 10ms for a tap size 800 filter, see equation (3.11). If they are minimum phase filter even smaller delays are expected, but in order to fulfil the real time-constraint for arbitrary filters, the above mentioned tap size should not be exceeded.

A last test is performed to verify that there is no gap in the signal flow introduced by the changing between the filter coefficients described in section 5.3. A gap of silence in the signal flow sounds the same way as a Dirac impulse, i.e. a click with frequency components from the entire spectrum. This must be avoided since the array is going to be tested in a real situation where the user can change between different presets with

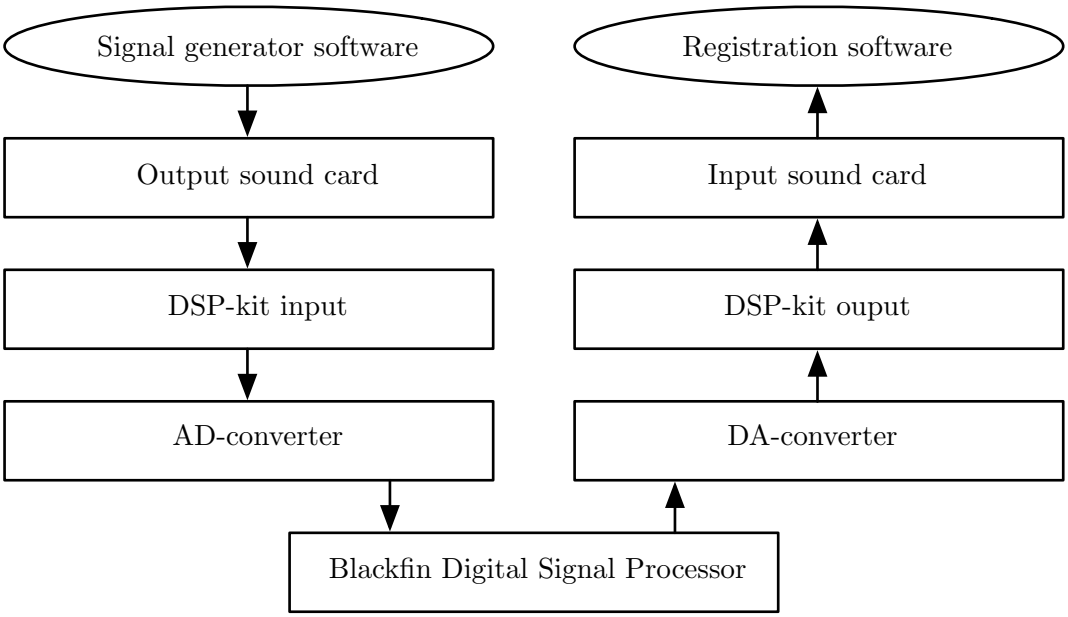


Figure 6.6: Signal path for delay measuring

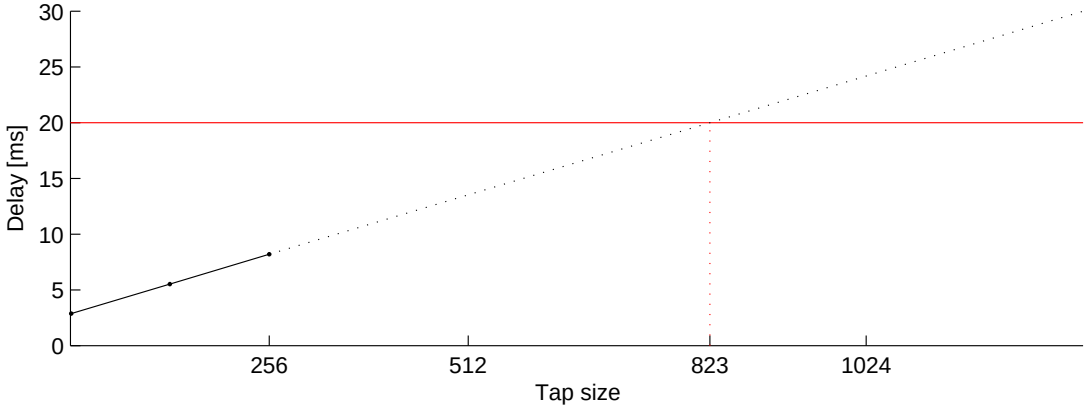


Figure 6.7: Maximum introduced delays by FIR filters compared to their tap sizes

the push of a button.

A sine wave is generated and is registrated in a wave recorder. The signal goes through all the blocks of figure 6.6 and is recorded in the final block. While recording, the filter coefficients are changed to observe the effects of this in the continuity of the signal. The recorded wave is analyzed visually for discontinuities.

No delay caused by the loading of filter coefficients is observed.

6.4 Beam pattern measurements

The purpose of these measurements is to empirically verify theoretic beam responses from microphone arrays. The frequency responses of the sound arriving the array from different angles are measured. Depending on the incident angle of the incoming sound wave the array will attenuate the signal differently.

The following equipment is used in these test:

- Matlab 2006b, The Mathworks.
- WinMLS 2004 Level 6, Morset Sound Development. Controllable from Matlab.
- Nor265 Turntable, Norsonic. Controllable from Matlab.
- Head and Torso Simulator type 4128C (HATS), Brüel & Kjær.

The microphone array is mounted on a head and torso simulator, HATS, and placed on a computer controllable turntable. A loudspeaker is set 1.5 meters away from the array, pointing towards it. WinMLS plays back a sine sweep through the loudspeaker and measures the response of the array. The HATS is rotated 3 degrees counter clockwise by the turntable and another measurement is done, and the procedure is repeated until a full 360 degrees is done. This test is done with different array configurations and algorithms.

The results of these tests are given in the subsequent sections. Both measured responses and theoretical responses are presented. The measured ones are calculated with basis of the measurements described in section 6.4 with the array placed on the chest of a head-and-torso-simulator, and the theoretical ones are derived from equation 2.1.

6.4.1 Comparison theory/measurements

In this section the performance of an array of 4 microphones is compared to its theoretical counterpart. The inner distance between the microphones is set to $d = 3.05$cm and the delay-and-sum algorithm has a steering angle $\phi = 0°$. The theoretical plots are calculated using free field, far field assumptions, while the measured ones are measured with the array mounted on the HATS-body.

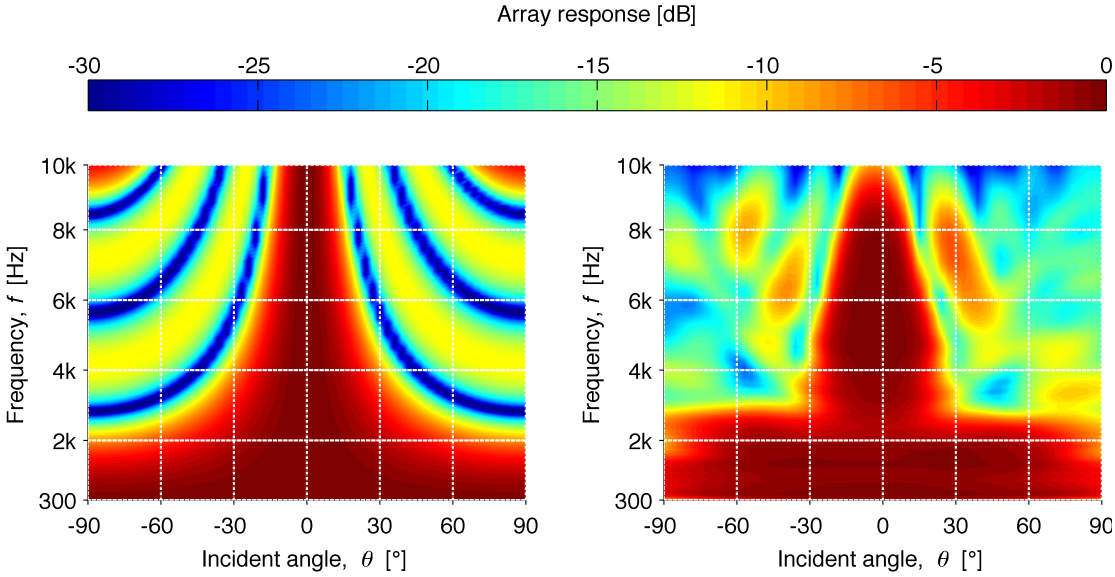


Figure 6.8: Theoretical (a) and measured (b) array response from delay-and-sum array with steering angle $\phi = 0°$.

Figure 6.8 shows calculated and measured beam patterns for this array configuration. The plot shows the magnitude frequency response for angles between $-90°$ and $-90°$. The color indicates the response in dB and all the responses next to each other form a color pattern indicating directivity for the multiple angles and frequencies. The frequency scale is here presented in a linear manner to show the technical details of the directivity pattern. The red (or dark) parts symmetrically surrounding $0°$ is called the main lobe, and represents the angular / frequency area from where sound is most audible. Comparing (a) and (b) it is observable that the beam width of the implemented array is approximately the same as that of the theoretical for high frequencies. The fact that the measured response (b) is measured with the array mounted on a HATS can explain the differences between the plots in the frequency area around 2-3kHz. The reflections from the HATS body partially cancel out direct sound in this frequency area as described in section 4.3, and lower the directivity here. Even though this cancellation ought to dampen the response, the equalization filter described in section 3.2.1 rises the entire area to maintain flat frequency response in forward direction ($\theta = 0°$), and a broader lobe is seen.

The low pass filter used in the array implementation is observable in figure 6.8 (b) as the magnitude starts decreasing around the cutoff frequency $f_x = 9$kHz. This is not considered when calculating the theoretical responses and the brightening of color is thus not present in (a).

The response in the suppressed area outside the main lobe is comparable in the two plots, and even though the smooth wave shapes of destructive interference seen in the theoretical response are not as distinct in the measured one, the tendencies are similar. The small asymmetric features in the measured response can be assumed caused by the fact that the microphone array is constructed in a slight asymmetric way, i.e. the four leftmost of the 5 microphones are utilized for the current array configuration.

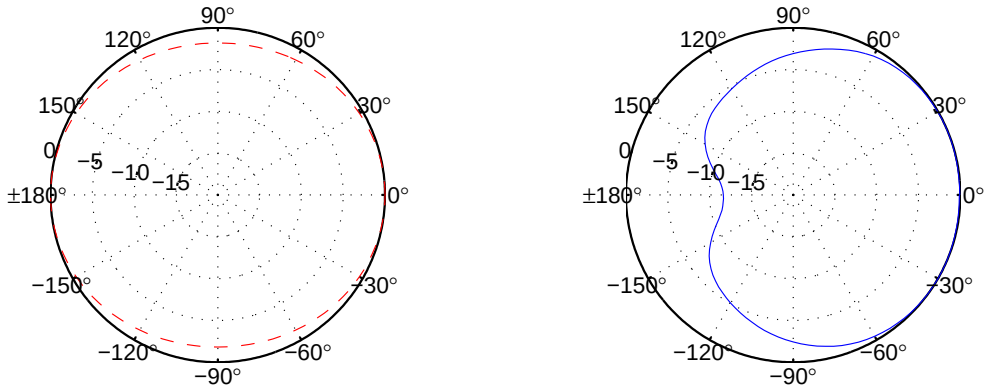


Figure 6.9: Theoretical (a) and measured (b) array responses for $f = 1$kHz from array configuration with steering angle $\phi = 0°$.

The body shadowing suppresses sound coming from behind the user. Figure 6.9 shows a polar representation of the directivity pattern in $f = 1$kHz. At this frequency the effective length of the array is too small to theoretically suppress sound from unwanted directions as observable in (a), but the measured response when worn by the HATS (b) shows an attenuation of 5-10dB in large parts of the area behind the user. Even though this is not due to the array properties of the system and the main lobe remains very wide, it is an desirable effect of placing the microphones on the chest of the user.

Figure 6.10 shows the polar array pattern for the same array at 4kHz. At this frequency the array suppresses sound from the sides well, and the beam pattern in forward direction is very similar for the theoretical (a) and the measured (b) response. The shadowing from the body is also more significant at these frequencies, and one can see from the figure that almost all sound coming from behind is attenuated 15dB or more. The asymmetries before mentioned are also viewable here as the side lobes at each side are not exactly equal.

Directivity index, DI, for the same array is plotted as a function of frequency in figure 6.11, calculated using both the theoretical and measured directivity patterns. The measured array scores better than the theoretical one over the entire examined frequency

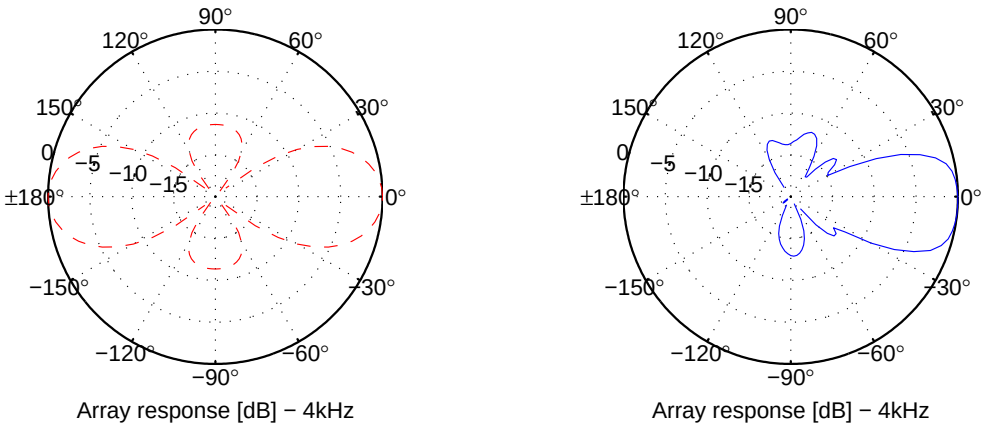


Figure 6.10: Theoretical (a) and measured (b) array responses for $f = 4$kHz from array configuration with steering angle $\phi = 0°$.

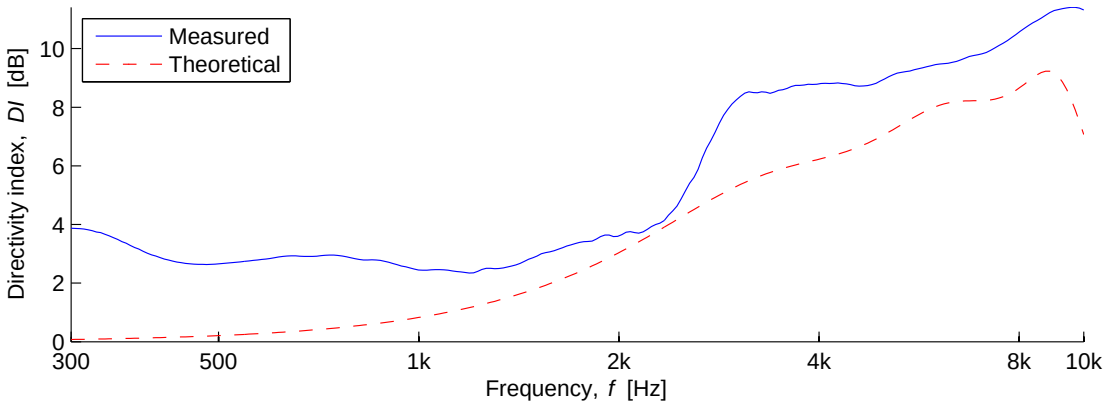


Figure 6.11: Theoretical and measured directivity index, for a linear array of four microphones with $d = 3.05$cm and a steering angle $\phi = 0°$.

range,and it is observable that even for low frequencies the directivity is better than the theory. This is assumed to be caused by body shadowing. In the area between 2 and 3kHz the gradient of the measured DI is lower than for the theoretical one, and despite the fact that the measured one is with body shadowing, it is almost as low as the theoretical. This is because of the reflections from the body described in section 4.2. For frequencies from around 3kHz and above, the DI rises significantly as here both the shadowing and the beam forming are working most efficiently.

6.4.2 Beam steering

In this section the same array is examined, this time with a steering angle $\phi = -30°$. All other parameters are the same, $N = 4$, $d = 3.05$cm and the algorithm is delay-and-sum. Also here comparisons with theoretically expected responses are done, in addition to comparisons with the responses from last section with steering angle $\phi = 0°$.

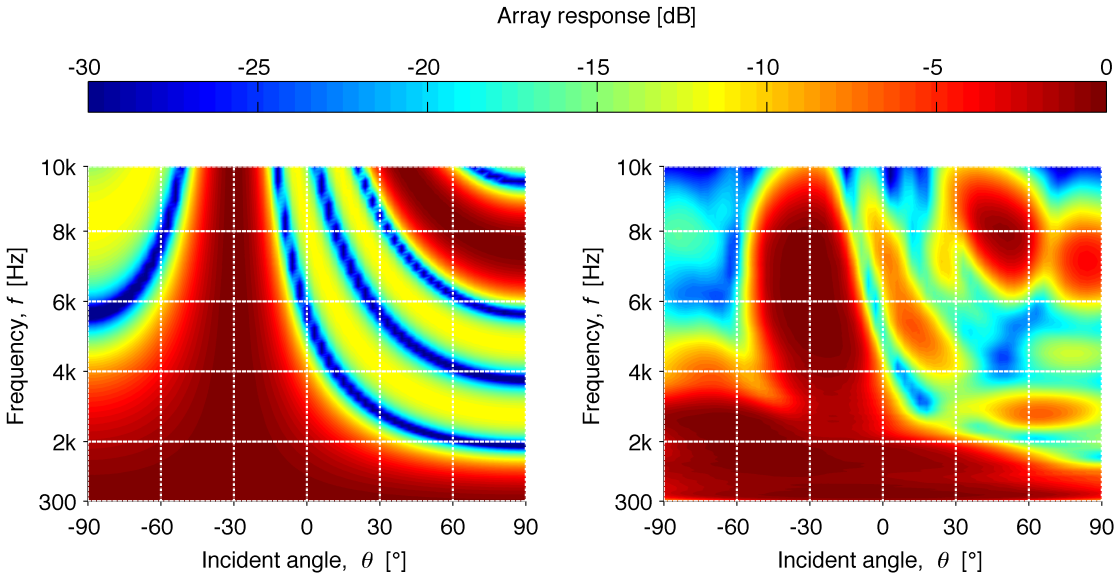


Figure 6.12: Theoretical (a) and measured (b) array response from narrow-band configuration with steering angle $\phi = -30°$.

Figure 6.12 shows theoretical (a) and measured (b) array response as function of frequency and incident angle. The steering angle $\phi = -30°$ is observable, as the main beam, i.e. the biggest red / dark area, has its center at $-30°$. This means that the array is steerable only by means of changing parameters / filter coefficients. It shows that with this configuration at least for frequencies from 2.5-3kHz and above, most of the sound entering the array is that coming from sources with an angle of 30° to the left

for the user. The measured pattern (b) resembles very much that of the theoretical; in addition to the angle of the main beam being the same, it is also observable that sound from the right is better attenuated for lower frequencies than from the left. Even if the main beam is wide at these frequencies (2-3kHz), it is a desirable effect when using this steering for beam spreading as described in 2.3, as the left ear will mainly be receiving sound coming from the left, and the right ear from the right. Another effect appearing in this steered version of the array is the side lobe observable in both plots in the figure. They appear as areas of intensities comparable to the main lobe for frequencies from 7kHz and higher, starting 30° to the right.

6.4.3 Beam spreading

This section shows how beam spreading is used to create binaural output to the user. The array configuration is the same as in previous sections, with $N = 4$ microphones, inner distance $d = 3.05$cm between the microphones. The beam spreading is set to $\phi = \pm 30°$, i.e. the main beam for the left output is steered 30° to the left and 30° to the right for the right ear.

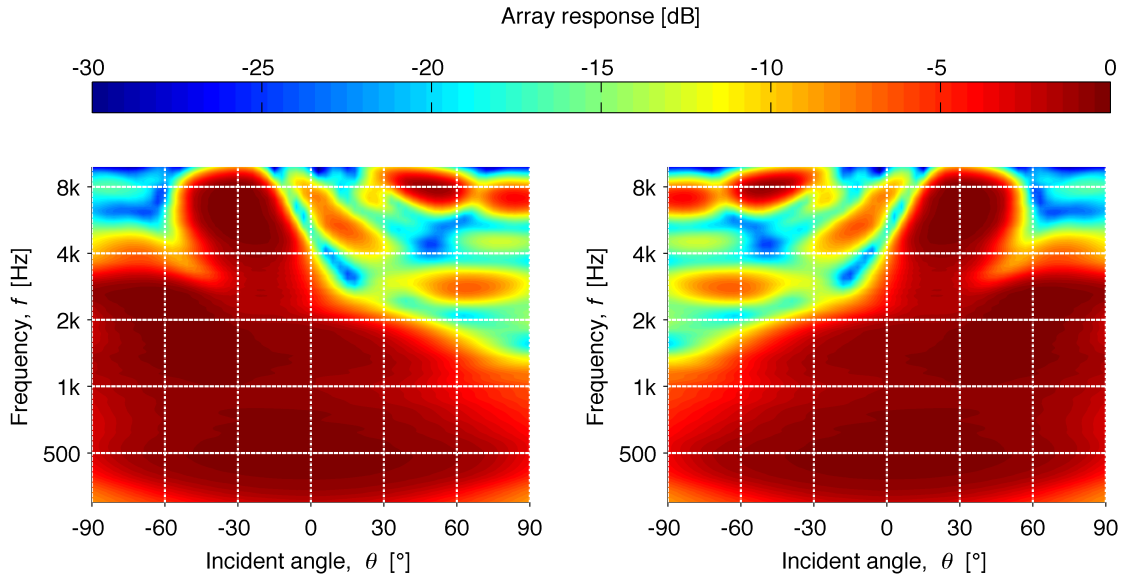


Figure 6.13: Array response for left ear (a) and right ear (b) with beam spreading $\phi = \pm 30°$.

Figure 6.13 shows how each ear is presented with different beam patterns. The plots are copies of that in figure 6.12 (b), but repeated to visually present the beam spreading. The frequency scale is logarithmic to emphasize speech frequencies.

6.4.4 Equalizing filter

The equalizing filter is designed to gain frequencies around 3kHz where the reflections from the body dampen the array signal. This section studies the effects of this filter implementation.

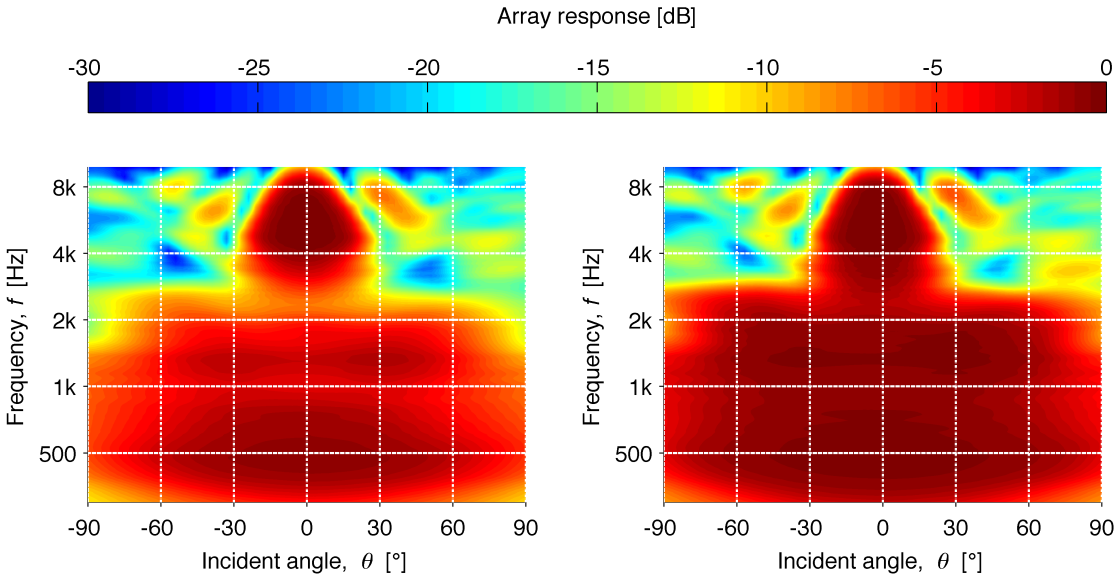


Figure 6.14: Measured array response from array without (a) and with (b) equalizing filter

Figure 6.14 shows the beam pattern for the array without (a) and with (b) the equalizing filter. A logarithmic frequency scale is used to emphasize speech frequencies. The goal of the equalizing filter is achieved as the color nearly change along the 0° vertical axis, i.e. the frequency response in $\theta = 0°$ is flat. The beam width, on the other hand, is not improved. For frequencies between 1.5 and 3kHz it is observable that the intensity increases and the resulting beam ends up wider in (b) than in (a).

6.4.5 Sub band

The sub band implementation measurements are presented here. This is the implementation where two sub arrays of 3 microphones each are processed in two separate frequency bands. The distance between the microphones are $d_l = 6.1$cm and $d_h = 3.05$cm for the low frequency band and high frequency band, respectively.

Figure 6.15 shows the directivity pattern for the single-band array (a) from above for comparison with the pattern for the sub-band array (b). The sub-band version has a

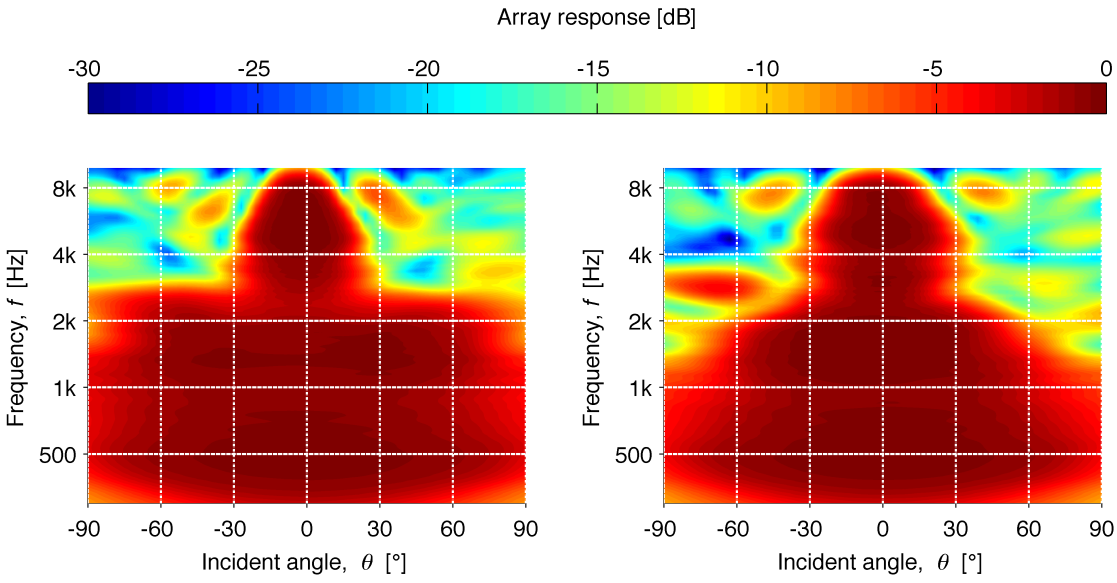


Figure 6.15: Measured array response from single-band configuration (a) and sub-band configuration (b) with steering angle $\phi = 0°$. Crossover frequency f_c for (b) is 3khz.

narrower beam between around 1kHz and 3kHz. This is due to the extended effective length of the array; the three microphones forming the lower band have a distance of $d_l = 6.10$cm and set the effective length of the array to $L = Nd = 18.3$cm. This narrower beam suggests a better directivity in these lower frequencies where the single-band array does not perform very well. The expanding of effective length requires the payment of number of microphones per frequency band. This is observable in frequencies above 4kHz where the main beam is wider for the sub-band array than for the single-band version. Despite this, the beam remains fairly narrow, and the attenuation outside this main beam looks to be comparable to that of the single-band array.

The improvements in the frequency area 1-3kHz are also observable in figure 6.16. Here the directivity index is plotted as function of frequency for the sub-band array and the single-band array. The red dashed line representing the sub-band array is higher then the blue solid single-band line in this area indicating that the narrow beam from figure 6.15 results in better directivity. For frequencies above 3kHz the directivity for the sub-band array is slightly lower than for the single-band version, a result which concurs with the larger beam width observed in figure 6.15 (b).

In later sections it will be shown that the improved directivity presented here has desirable qualities that make the sub-band array also score well in subjective tests and evaluations.

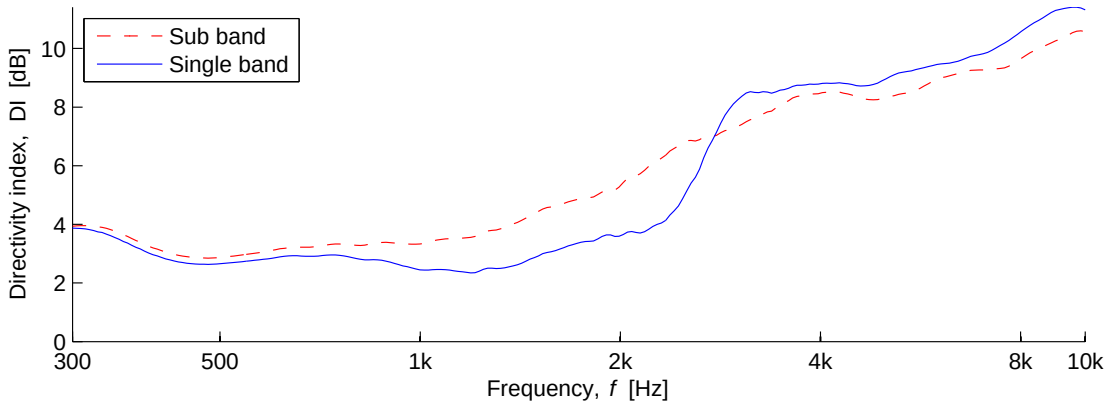
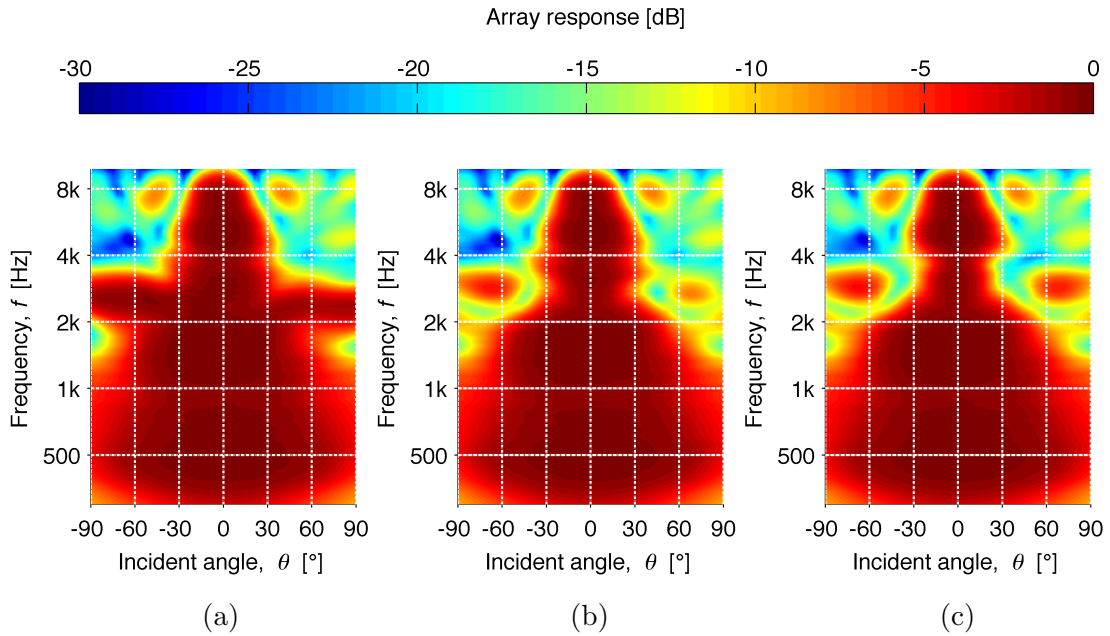


Figure 6.16: Directivity index plottet for a single-band array configuration and a sub-band array configuration.



(a) (b) (c)

Figure 6.17: Measured array response from sub-band array with crossover frequencies 2kHz (a), 3kHz (b) and 4kHz (c).

The crossover frequency between the bands in the sub-band array is chosen to keep the beam width constant and avoiding spatial aliasing in the upper end of the bands. Spatial aliasing limit is calculated using equation 2.4, which suggests 2.8kHz as the upper limit for the lower band. When steering the beam straight forward or with moderate angles close to 0°, the resulting sidelobes are not very significant. Figure 6.17 shows the beam patterns for sub-band configurations with crossover frequencies set to 2, 3 and 4kHz. In (a) the crossover frequency of 2kHz makes the upper band active in a too low frequent area. The three microphones with distance $d = 3.05$cm result in a wide beam here compared to the other configurations in this plot. Setting the crossover frequency to 3kHz (b) leaves the area between 2 and 3kHz for the lower band where the distance between the microphones is 6.1cm and the resulting beam is narrower. The sidelobes are starting to become visible around ±60° at just below 3kHz, but are still considered more acceptable than the wide main beam in (a). In (c) the crossover frequency is 4kHz and the lower band is used all the way up to this frequency. It results in a narrower beam further up the frequency scale, but the side lobes are also more significant here.

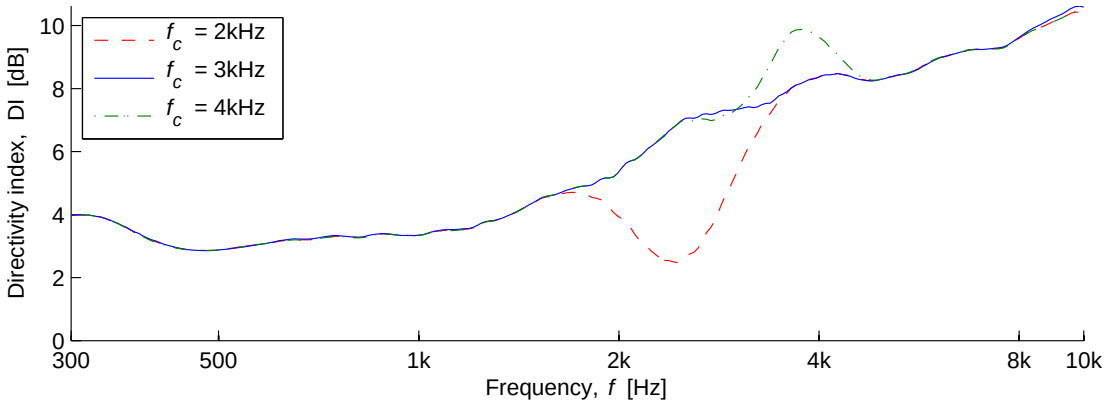


Figure 6.18: Measured directivity index, for sub-band array with crossover frequency between the bands of 2, 3 and 4kHz.

Figure 6.18 shows that the crossover frequency of 4kHz gives a better directivity index than the other settings. The figure shows the measured directivity indices for the three tested crossover settings for the sub-band array, 2, 3 and 4kHz. As anticipated the wide main beam in the frequency range 2-3kHz in figure 6.17 (a) results in low directivity in this area for the configuration with $f_c = 2$kHz. The $f_c = 4$kHz version on the other hand scores better than the others despite the spatial aliasing that starts right below the crossover frequency, this because of the narrower beam.

The reason for dividing an array into sub-bands and sub-arrays is to intend to make the beam width less dependent of frequency, or at best making it constant with respect to frequency. That means that even if the 4kHz crossover frequency results in a bet-

ter directivity index, it is not necessarily better in use than for instance the one with $f_c = 3$kHz. The beam width is more varying and sound from angles like 30° may have undesired frequency alterations.

7 Experiments

7.1 Rhyme test

There are a number of tests designed to establish the intelligibility and quality of an audio system. Two of these are the diagnostic rhyme test (DRT) and the modified rhyme test (MRT) [3].

The MRT consists of a word list for testing statistical intelligibility . The MRT uses 50 six-word lists of rhyming and similar sounding English words. Each of these words are constructed from a consonant-vowel-consonant sound sequence, and the words only differ in the initial or final consonant sound. During testing, the listener is shown six different words, and is asked to identify the word spoken by the talker. It is common to use a carrier sentence to make the listener aware that the word is coming.

Similar to the MRT, the DRT uses monosyllabic English words that are constructed in a consonant-vowel-consonant sequence. Unlike the modified rhyme test, no carrier sentences are used. Instead the listener is shown a pair of words differing only in their initial consonants, and asked to identify the word presented by the talker. The speech material consists of 96 pair of rhyming words.

Because of the material for these tests are written in English, the tests also require that both the talker and the listener speak English as their first language. The test used in this project is similar to the modified rhyme test, but consists of 50 four-word lists of rhyming or similar sounding Norwegian words. The list of sentences with alternative answers is given in appendix A.

The test is done to measure speech intelligibility when using different array settings. The test persons are forced to choose between four words based on which of them they think are read to them. The words are tangled into carrier sentences to prepare the test person for the important word, but the sentences are made in such a manner that it is not possible to guess the correct answer from any semantic context. The four options given to the test persons rhyme in such a way that only one syllable differs.

The tests are performed in a synthesized noisy environment. 16 loudspeakers are used to simulate 16 people talking, of which only one of them is of interest and the 15 others are considered noise. The measure in this test is how loud this one person has to speak to be understood in the crowd. This speaker level is adjusted throughout the test to find a threshold, and the final level is considered the result of each run. If the test person understands the spoken word and answers correctly, the voice is attenuated 1dB for the next word. If the word is not understood, and the wrong answer is given, the test voice is gained up 2dB befor the next word. The noise level is kept constant for all sentences. A total of 15 persons participated in the testing. Each test consists of 20 sentences randomly chosen from the 50 different sentences.

To perform the tests a graphical interface is developed in Matlab. This makes it easier

Figure 7.1: A test person answering a rhyme test on the screen

Figure 7.2: Rhyme test showing a test person sitting centered between speakers with the microphone array worn on his chest.

to perform the test and analyze the results. The graphical test interface as seen by the user is shown in figure 7.3. The interface consists of a text field at the top of the window, four alternative-buttons, a play sound button and a next button.

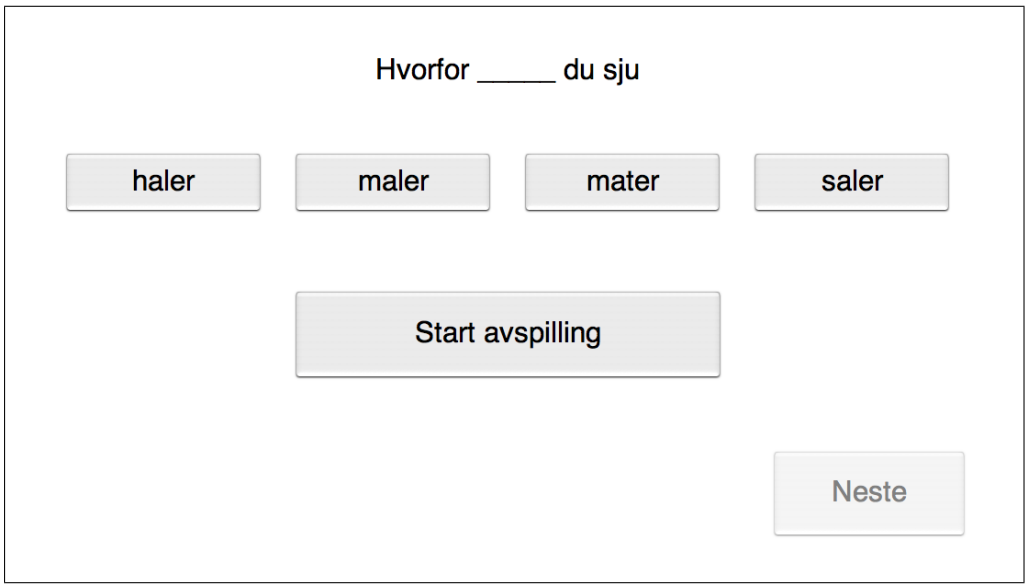


Figure 7.3: Rhyme test user interface

When the user presses the next button, the text field is updated with a randomized carrier sentence with one of the words missing. The four alternative buttons are updated with the four alternatives belonging to the carrier sentence. The placement of the alternatives on the four buttons is also randomized. When the listener has read both the carrier sentence and the four alternatives, the play button should be pressed. This action plays back the prerecorded test material producing speech noise from 15 speakers and the rhyme test sentence from the speaker directly in front of the listener. When the playback is finished, the listener must choose one of the four alternatives. The listener will be given feedback in the text field if the answer is either right or wrong. The user will then press next and a new sentence is presented.

As shown in picture 7.1 and 7.2 the listener is placed in a chair surrounded by 16 speakers. The speakers are placed on a circle with radius 2.2 meters. The speakers are symmetrically placed with 22.5 degrees separation.

Figure 7.4 shows an example of a rhyme test run. The current speaker level is plotted for each trial or test sentence. Right and wrong answers are indicated for each trial by plus-signs and dots, respectively, and it is observable that the speaker level rises or drops after each trial according to the answer. The adjusted value after the last trial is called the final value. It is indicated by a square in the figure and is considered the threshold of needed speaker level. The plot is a real-life example from one of the trials forming

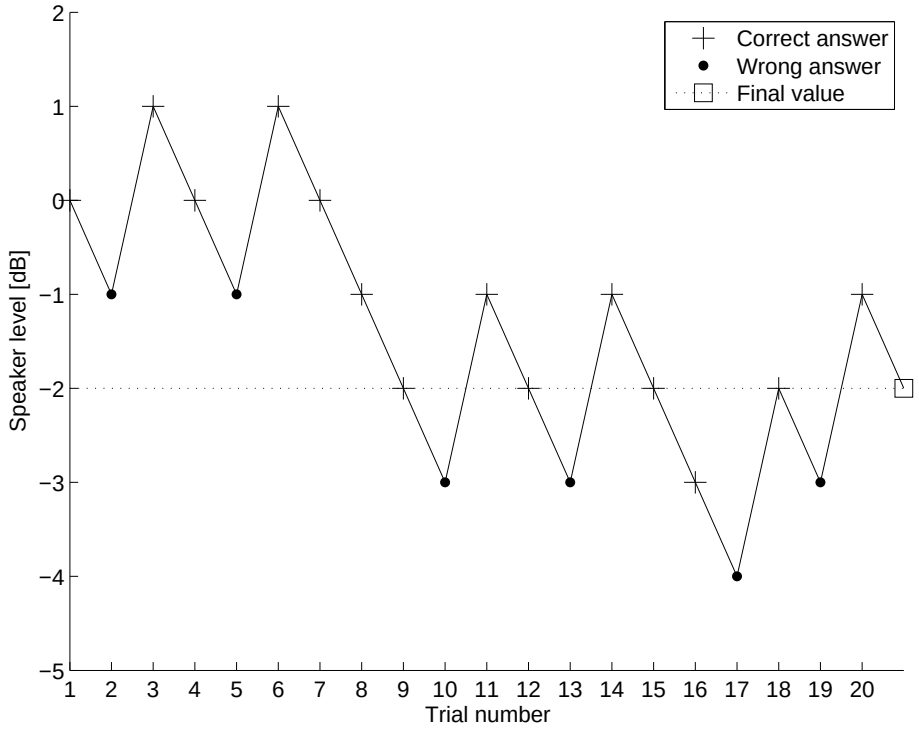


Figure 7.4: Example of a typical rhyme test run, showing the varying speaker level of the test sentences as function of trial number and answers.

part of the results in the section 8.1. It is a typical run where after some initial getting-used-to the level stabilizes around the threshold. The percentage of correct answers, the word articulation, in this example is 70%, which is very close to the expected and recommended value [10] for a 2-up-1-down algorithm like this,

$$p_{exp} = \sqrt{1/2} = 70.7\%. \tag{7.1}$$

The speaker levels indicated in the figure are decibel values referred relatively to a diffuse noise field, adjusted to have a starting point not very far from the expected final value, to ensure rapid approximation to a steady level.

The purpose of performing these rhyme tests is to compare intelligibility when using different array settings. The settings tested are chosen to be able to compare different specific parameters. The following subsections present the groups of presets to be tested utilizing rhyme tests.

7.1.1 Beam Spreading

The idea of bream spreading is as described in section 2.3 to preserve some of the binaural cues that the human brain utilizes to localize and differentiate sounds spatially. The rhyme test can reveal if the cues can help the understanding of word despite the fact that the beams in some cases are steered in other directions than towards the interesting speaker. The settings compared are

Straight, $\phi = 0°$: Both the left ear beam and the right ear beam are steered straight forward,

Wide, $\phi = \pm 30°$: Left ear beam is steered $30°$ to the left and right ear beam is steered $30°$ to the right, and

Extra Wide, $\phi = \pm 60°$: Left ear beam is steered $60°$ to the left and right ear beam is steered $60°$ to the right.

7.1.2 Sub band filter type

The second test compares various filters for sub-band band-division as described in section 3.2.3, to evaluate the effects of group delay distortion introduced by minimum phase filters. The filters chosen for testing are:

Minimum phase filter: Minimum delay, but some phase distortion.

Linear phase filter: No phase distortion, but longer total delay.

Compromise filter: A compromise between the two latter filters.

7.1.3 Sub-band crossover frequency

Changing the crossover frequency changes the frequency response and directivity pattern of the system, as shown in section 6.4.5. A rhyme test with different crossover frequencies will show how these changes affect speech intelligibility. The same frequencies evaluated in setion 6.4.5 are tested here. The sub-band configurations in this test has the beams for both ears steered straight ahead. The settings compared are:

$f_c = 2$**kHz:** A low frequency band division.

$f_c = 3$**kHz:** A sub-band division with crossover frequency chosen close to the limit for spatial aliasing.

$f_c = 4$**kHz:** Crossover frequency set to cause some spatial aliasing.

7.1.4 Equalizing filter

The dip in the frequency response caused by body reflections described in section 4.3 is taken care of for forward direction sound by an equalizing filter. This test evaluates the effects of using this, in measures of speech intelligibility. The settings used for comparison are:

Equalizing filter disabled: The signals coming from the microphones are directly processed by the beam-steering algorithm.

Equalizing filter enabled: The microphone signals are filtered with the equalizing filter to negate the effects of body reflections in forward direction.

7.1.5 Normal hearing comparison

To evaluate the absolute improvement in intelligibility increase caused by using the built microphone array, a series of tests are done without any array. The test persons perform the tests as normal, but are using their normal hearing only. A comparison is made between the following setups:

With array: An sub-band array is steered straight forward.

Without array: Normal binaural hearing only is used.

7.2 Subjective tests in real environments

The entire system is tested in noisy environments and the performance of the different array at different settings are evaluated and compared subjectively by test persons.

A set of 3 predefined array presets are made available for the test persons and they may choose between these as they wish during the test period. The evaluation is to be done according to the following criteria:

- Speech intelligibility
- Suppression of noise/unwanted sound
- Sound quality of wanted sound
- Sound quality of attenuated sound
- Source localization
- Overall impression

For each criteria the test person may give a score in the range of 1 to 5, where 5 is best and 1 is worst. This method of quality testing follows a standard method of subjective measuring called Mean Opinion Score, MOS [5]. In addition to giving scores the test person may make comments on all of the above measures.

Figure 7.5 shows the evaluation form given to the test persons. The form is given to the test persons with only the three preset identifiers filled out.

Four test persons take part in this testing, and they all perform three different tests with different array settings. They are given a review of the different criteria they are to evaluate and a demonstration of how the equipment works. First they get time to get used to the equipment and a chance to experiment at free will with the three presets made available to them. The testing is done in different environments to give the user an impression of how the array behaves, but the main testing is done in cafeterias with a lot of people present, where the test persons participate in normal social activities. The test persons are observed throughout the test period.

The collection of presets available to the test person at any given time is composed in such a way that the presets only differ in one way, i.e. only one parameter is changed as the test person changes preset. Three different collections of presets form the three test and are described in the subsequent sections.

7.2.1 Beam Spreading

The first collection of presets is equal to the one used for the beamspreading comparison using rhyme tests in section 7.1.1. As for the rhyme test in addition to steer the array straight ahead, two presets use beam spreading to steer the main lobe of the array

Name	Location	Date and time

Description of environment

Preset:	Score					Comments
	1	2	3	4	5	
Taleforståelighet						
Demping av lyd fra sidene						
Lydkvalitet forfra						
Lydkvalitet fra sidene						
Lokalisering av kilder						
Generelt inntrykk						

Preset:	Score					Comments
	1	2	3	4	5	
Taleforståelighet						
Demping av lyd fra sidene						
Lydkvalitet forfra						
Lydkvalitet fra sidene						
Lokalisering av kilder						
Generelt inntrykk						

Preset:	Score					Comments
	1	2	3	4	5	
Taleforståelighet						
Demping av lyd fra sidene						
Lydkvalitet forfra						
Lydkvalitet fra sidene						
Lokalisering av kilder						
Generelt inntrykk						

Figure 7.5: Evaluation form for subjective tests in real environments

separately for the two ears. This is done to preserve some of the binaural cues that the human brain utilizes to localize and differentiate sounds spatially, and to keep a more spatial feeling in contrast to the monaural in-a-box-feeling. This test tests how well this works when worn by a user in a real life situation. A standard delay-and-sum algorithm is used, the differences between the presets presented to the test person are as follows:

Straight, $\phi = 0°$: Both the left ear beam and the right ear beam are steered straight forward,

Wide, $\phi = \pm 30°$: Left ear beam is steered 30° to the left and right ear beam is steered 30° to the right, and

Extra Wide, $\phi = \pm 60°$: Left ear beam is steered 60° to the left and right ear beam is steered 60° to the right.

7.2.2 Single microphone vs array

This test compares a setup where only one microphone is activated with two two array-presets. This is done to be able to confirm the difference between using a single microphone and array beamforming. The settings used for the array-presets are sub-band algorithms with the beam steered straight ahead.

One single omni directional microphone

Sub-band array with equalizing filter

Sub-band array without equalizing filter

7.2.3 Sub-band vs single-band array

To compare sub-band array configurations with single-band configurations, two single-band configurations of three microphones are designed. They use a basis of $d = 3.05$cm and $d = 6.10$cm inter microphone distance, respectively. The sub-band combines the two by dividing them into frequency bands.

Sub-band array: A sub-band array configuration with $d_h = 3.05$cm and $d_l = 6.10$cm inter microphone distance for the high and low band, respectively.

Single-band small basis array: A single band array configuration of 3 microphones with inter distance of $d = 3.05$cm.

Single-band wide basis array: A single band array configuration of 3 microphones with inter distance of $d = 3.05$cm.

8 Results and discussion

8.1 Rhyme test

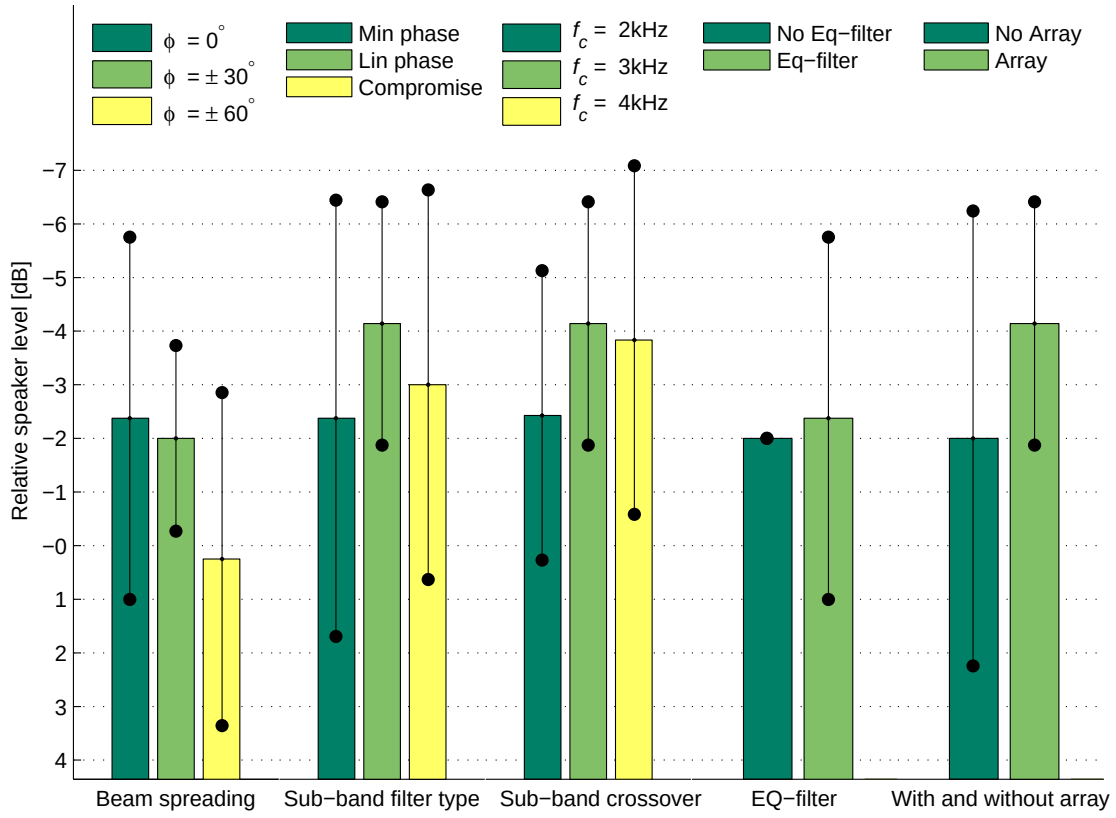


Figure 8.1: Threshold SNR from rhyme word tests

The results from the rhyme tests are shown in figure 8.1. Each bar represents a preset within a group. Each bar shows the value of the last speaker level in each test. A taller bar means better score, as a lower speaker level is needed to compensate for the noise from other directions. This means that all the values in the figure are directly comparable, but the results are grouped to compare different parameters. The black lines enclosing the top of each bar indicate the standard deviation of the values, and is included to give a measure of the certainty of the test results.

8.1.1 Beam Spreading

The beamspreading group forms a comparison of different values of the beamspreading angle, $|\phi|$. Presets with angles set to $\phi = 0°$, $\phi = \pm 30°$ and $\phi = \pm 60°$ are compared.

The algorithm used for these three presets is normal single-band delay and sum, and the equalization filter is turned on. It is observable that the preset steered straight forward, $\phi = 0°$, scores the highest. The $\phi = \pm 30°$ version is very close, but the widest spreading, $\phi = \pm 60°$, ends up with score more than 2dB below the others. The standard deviation bars indicate that the variances are rather big here, but still the tendencies suggest that a narrow beam spreading is better. The difference in score between $\phi = 0°$ and $\phi = \pm 30°$ is so small that if the thirty degrees version can provide additional benefits, like better source localization or more comfortable spatial experience, this version can be considered better than the straight ahead version.

8.1.2 Sub-band filter type

The second group in figure 8.1 shows a comparison of the three crossover filter types listed in section 7.1.2. The filters are used to divide the sub-bands in a sub-band array setup. The crossover frequency f_c is set to 3kHz, the equalization filter is turned on and the beam is steered to straight ahead. The first preset uses minimum phase filters to reduce total delay in the system, the second uses big linear phase filters to create an accurate crossover with no phase-interrupts in the crossover section, while the third is a compromise between the two. As the difference between these filters is a tradeoff between phase distortion and total delay, it is as expected the linear phase version that comes out best as total delay does not matter when listening to prerecorded sound. The standard deviations in this test are also quite big, but it seems that the distorted phase in the transition area causes a worsening of speech intelligibility.

8.1.3 Sub-band crossover frequency

The third group labeled sub-band crossover in figure 8.1 changes the crossover frequency between the two sub-bands to 2kHz, 3kHz and 4kHz, respectively. The three presets use the liner phase filter used in the winning preset from last paragraph to divide the sub-bands in a sub-band array setup. As the preset with $f_c = 3$kHz is equal to the one labeled "Lin phase" in group two, the same results are used for comparison. The preset with $f_c = 2$kHz scores the lowest. This is in accordance with the results from section 6.4.5, where it is shown that a broader lobe in important frequencies causes a loss of directivity. The versions with $f_c = 3$kHz and $f_c = 4$kHz score almost identically, with the 3kHz just ahead. The variance in the test says that it is not possible to throw away neither of the two, but as the shown in 6.4.5 the more constant beam width of the 3kHz version might be preferable.

8.1.4 Equalizing filter

The forth comparison done with rhyme tests is turning the equalizing filter off and on. The equalizing filter is implemented to maintain a flat frequency response in forward direction, but as mentioned in section 6.4.4 no improved directivity is expected to be gained, since noise from the sides also will be filtered equally. As they score almost equally in this test, the equalizing filter can be said not to affect the directivity. All the tests performed with the equalizing filter turned off resulted in the same score, hence the standard deviation of zero is indicated in the figure.

8.1.5 Normal hearing comparison

In the rightmost group a comparison is done between the winner of the array presets and a situation of not using a microphone array at all, to show the difference between normal hearing and the array aided hearing. The array used in this comparison is the sub-band array with linear phase crossover filter with $f_c = 3$kHz, beam steering angle $\phi = 0^\circ$ and equalization filter turned on, as this setup has shown best results in the other tests. The 2dB improvement indicated by the bars shows that the use of microphone arrays actually helps understanding speech in noisy environments.

To obtain more information about the performance of the microphone array and to get an subjective opinion of the how the system performs, a subjective listening test is performed. The results from this test is described in the next section.

8.2 Subjective tests in real environments

The figures in the subsequent sections show the Mean Opinion Scores (MOS) from the subjective testing of the array in a real environment. As described in chapter 7.2 the users are asked to evaluate the array considering six different criteria. These six criteria are shown along the horizontal axis in the figures and the Mean Opinion Score is shown along the vertical axis. Within each criteria there are three bars. These bars corresponds to the three presets within each preset collection, i.e. each single plot compares one single collection of three presets. Standard deviation is indicated as black terminated lines enclosing the top of each bar. These are included to give an overview of the variance in the responses.

8.2.1 Beamspreading

Figure 8.2 shows the results from the beamspreading group. The presets available to the user are described in section 7.2.1. The first preset is normal delay and sum with the main beam steered straight ahead, the second preset has the main beam split up and

steered 30° the sides and sent to each ear, and the third preset has the main beam split up and spread out 60° to each side.

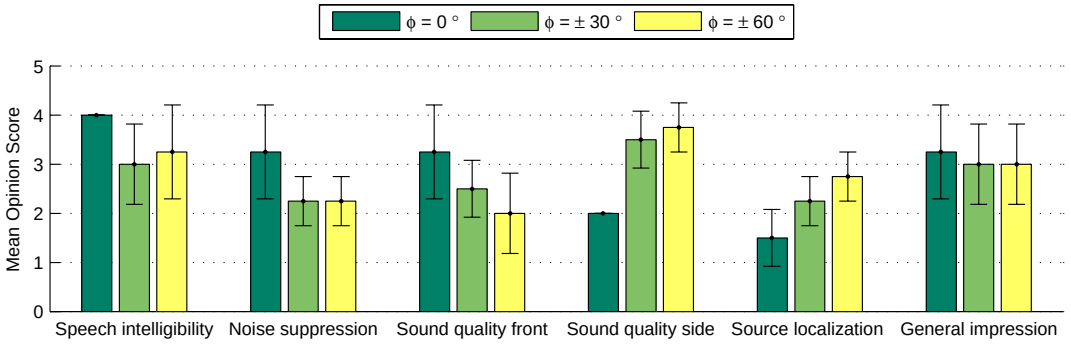


Figure 8.2: Mean opinion scores from testing beamspreading in a real environment.

A few things can be noticed from the results. Speech intelligibility scores highest when the beam is steered straight ahead, $\phi = 0°$. As the beam is spread further and further outwards, the sound quality in front decreases, but in return the sound quality from the sides increases. This is as expected, since the beam is steered to the sides, the sound coming from those directions should remain unchanged. Another interesting observation is that when the beam is spread out, the ability to localize sounds is enhanced.

The idea behind beamspreading is to combine beamforming with natural binaural impression. The concept seem to be working, but there seems to be a trade-off between source localization and a comfortable sound impression on one hand and noise suppression and speech intelligibility on the other hand. This result is further backed up by comments given by the users during the test. ±30° beamspreading gives users a normal and comfortable sound impression, but less noise suppression and speech intelligibility. ±60° beamspreading gives even better source localization, but people find this setting unnatural and too excessively steered to each ear. The binaural impression is lost. One user comment on this test suggest that a wide spreading, e.g. ±60°, could feel comfortable to use when communicating to a group of people, or in general while not speaking to one specific person. The opposite case when there is only one speaker of interest, straight-forward-steering of the lobes could be the most comfortable. This suggests that a in a final product a choice between these settings could be made available to the user, e.g. in form of a push button.

8.2.2 Single microphone vs array

The next test compares two sub-band array settings and a setting with only one active microphone. The test also gives the user the possibility to evaluate the equalizing filter as one of the array-presets is with and one is without. The first preset in figure 8.3 is

the one microphone preset, the second is a sub-band array preset with equalizing filter activated and the third is the same array setup, but without equalizing.

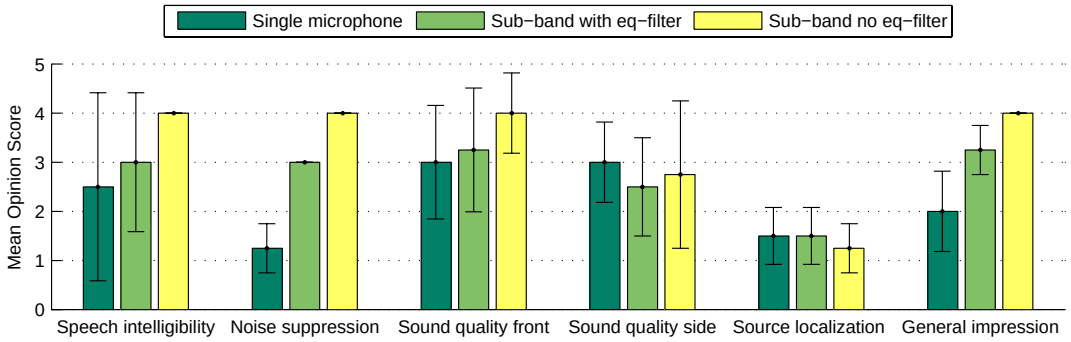


Figure 8.3: Mean opinion scores from testing sub-band array versus single microphone in a real environment.

A single omni directional microphone has no directivity. As seen in figure 8.3, this is confirmed with the first preset getting a low noise suppression score. Both the sub-band settings get high noise suppression score as anticipated, but the one without equalizing scores highest.

The two first presets get high scores in perceived sound quality from the look direction. This could be because they both have a flat frequency response in this direction and this should result in good sound quality. On the contrary the sub-band setting with out equalizing scores higher even though there is a dip in the frequency response about 3kHz as described in section 4.3. Another interesting observation is that the same subband setting also scores highest in speech intelligibility and general impression and outperforms the sub-band setting with equalizing filter in all criteria. This is further backed up by user feedback; the general impression is that preset without equalizing filter has the best ability to suppress sound from and from the sides, which gives this setting the best speech intelligibility and sound quality. The sound is said to feel more natural and less noisy witch results in a better general impression, which suggests that the equalizing filter does not fulfill the intentions of giving the user a more normal hearing experience.

As presumed, all the presets in this test have low source localization since the sound output is monophonic.

8.2.3 Sub-band array vs single-band array

The last test consists of a sub-band array and two single-band configurations. The two single-band configurations has inter-microphone distance 3.05cm and 6.10cm, respectively. The sub-band beamformer is a combination of the two delay and sum presets, i.e. the low frequency band uses the 6.10cm basis and the high frequency band uses

the 3.05cm basis. The test is done to compare sub-band beamforming against basic delay and sum beamformers. Because of the limited input channels on the DSP-kit, the delay and sum beamformers can only have a three microphone basis to be able to swap between sub-band and delay and sum beamforming.

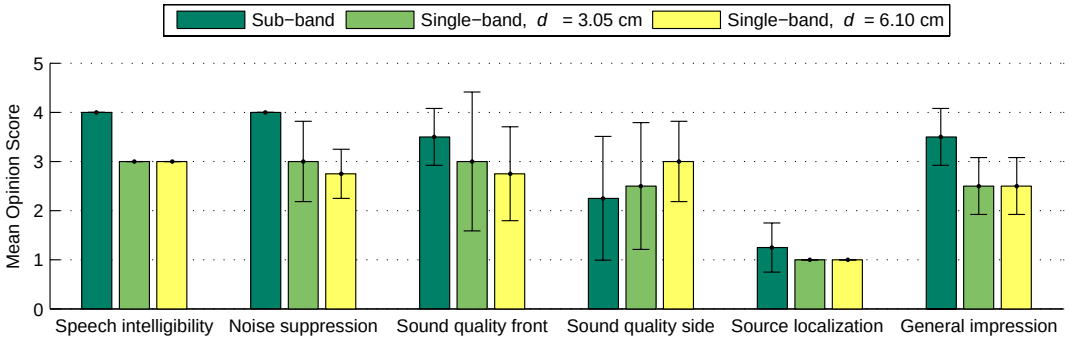


Figure 8.4: Mean opinion scores from testing sub-band array versus single band array.

As seen in figure 8.4 the sub-band beamformer outperforms the other presets in almost all criteria. The low standard deviation shows general agreement of this. The difference between the two single-band beamformers is not significant, but the 3.05cm basis array which has better frequency response in high frequencies gets a marginal better score in noise suppression and sound quality in front. Once again, no source localization is achieved.

The scores within each collection are not comparable with scores from the other collections. When the test persons are asked to evaluate a presets with another preset in a different group, their response is that this is not possible. Even though the different criteria within each collection are not directly comparable, the preset without the equalizing filter in the second collection gets best feedback in terms of noise suppression, voice quality and the sound seems focused straight ahead. On the contrary, the common opinion is that the beamspreading preset with a $\pm 30^{\circ}$ spreading gives a comfortable and natural sound impression at the sacrifice of some of the noise suppression and speech intelligibility.

9 Conclusion

The main goal of building a microphone array test platform for testing different beamforming algorithms in a real environment is completely achieved. The system is portable and fully functional.

In this paper it has been shown that it is possible to achieve directivity close to theoretical values in the mid and high frequency range with four microphones and a 9cm wide array. By using a sub-band array, the directivity is further improved.

Even though the delay and sum algorithm do not offer much directivity for low frequencies, the directivity achieved is in the mid and high frequency range where hearing loss is most common. The use of microphone arrays increases signal-to-noise ratio in these frequency bands and can help people with hearing loss to understand what is being said.

The proposal of using an equalizing filter to overcome body reflections and give a normal feeling sound to the user was proved not adequate for the task, as users experienced it as more noisy and unnatural than when not using it.

The results from testing beamspreading in a real setting revealed that a spreading of $\pm 30^\circ$ gave the best perceived binaural sound. People found the perceived sound quality in this setting the most natural to listen to. Tests show that there is a trade-off between binaural hearing on one side and directivity and speech intelligibility on the other side. It is therefore suggested to make a compromise between steering the beam straight ahead and a beamspreading of $\pm 30^\circ$ to give the user the best of two worlds.

References

[1] J. Blauert and P. Laws. Group delay distortions in electroacoustical systems. *The Journal of the Acoustical Society of America*, 63(5):1478–1483, 1978.

[2] M. Brandstein and D. Ward, editors. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer-Verlag, New York, 2001.

[3] J.R. Deller Jr, J.G. Proakis, and J.H. Hansen. *Discrete Time Processing of Speech Signals*. Prentice Hall PTR Upper Saddle River, NJ, USA, 1993.

[4] Alton F. Everest. *Master Handbook of Acoustics*. McGraw-Hill/TAB Electronics, September 2000.

[5] Jerry D. Gibson, Toby Berger, Tom Lookabaugh, Dave Lindbergh, and Richard L. Baker. *Digital compression for multimedia: principles and standards*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998.

[6] D.B. Hawkins and W.S. Yacullo. Signal-to-Noise Ratio Advantage of Binaural Hearing Aids and Directional Microphones under Different Levels of Reverberation, 1984.

[7] Ralph Jones. Speech intelligibility papers. 2007. `http://www.meyersound.com/support/papers/speech/`.

[8] Sergei Kochkin. "why my hearing aids are in the drawer": The consumers' perspective. *The Hearing Journal*, 53(2):34–41, 2000.

[9] T.I. Laakso, M. Valimaki, V. anf Karjalainen, and U.K. Laine. Splitting the unit delay [fir/all pass filters design]. *Signal Processing Magazine, IEEE*, 13(1):30–60, Jan 1996.

[10] Neil A. Macmillan and C. Douglas Creelman. *Detection Theory - A User's Guide*. Lawrence Erlbaum Associates, Inc, 2005.

[11] I. A. McCowan. *Robust Speech Recognition using Microphone Arrays*. PhD thesis, Queensland University of Technology, Australia, 2001.

[12] ACOEM Noise and Hearing Conservation Committee. Noise-induced hearing loss. *Occupational and Environmental Medicine*, 45(6):579–581, 2003.

[13] B. D. Steinberg. *Principle of Aperture and Array System Design*. John Wiley & Sons, Inc., New York, 1976.

[14] Harry L. Van Trees. *Optimum Array Processing*. John Wiley & Sons, Inc., New York, 2002.

A Rhyme test sentences

Table A.1: List over sentences used in the rhyme tests

Boka tror at billene vil	helle	hulle	hyle	hylle
De tenker at .....tør holde	baren	broen	broren	troen
Han roper at hodet må	bære	kjæle	lære	skjære
Hunden mener at rillene kan	lære	kjæle	skjære	bære
Jeg lærer at selen kan	be	se	sy	så
Posten frykter at tanten bør	gå	se	sy	så
Vi forteller at veien burde	banne	lande	vinne	vanne
Hvorfor treffer bilen	byer	sjøer	skjeer	skyer
Når spiser blomstene	feer	sjøer	skjeer	skyer
Hvor ser de	ni	sju	ski	ti
Når mottar dere	sjøer	skjeer	skyer	løer
Hvorfor .....du sju	haler	maler	mater	saler
Hvor krever hunden	gaver	haver	maver	paver
Når kjører huset	lakk	lam	land	lass
Når elsker vi	bad	bar	barn	garn
Aldri synger faren en	bille	fille	pille	rille
Helst skriver gaven en	måke	måne	måte	tåke
Idag ser havet en	kasse	mappe	masse	matte
I morgen leser kisten en	kiste	lisse	liste	niste
Neste uke kjører kjolen en	labb	lapp	lakk	last
Om sommeren har læreren en	kanne	kappe	kasse	masse
Til middag bærer været en	ball	pall	hall	stall
Den faste .....synger aldri	bilen	filen	pilen	silen
Det gale skjegget .....denne uka	raser	reiser	riser	roser
Den fine øya .....i morgen	rager	raker	raller	raser
Det fjerne skinnet .....ofte	faller	raker	raller	raser
Den flate snøen .....om høsten	lekker	rekker	renner	revner
Det frie preget .....om våren	frir	glir	rir	rår
Den friske fuglen .....sjelden	rir	rår	ror	trår
Den fulle tavla .....til frokost	hører	kjører	røper	rører
Den første dokken .....til middag	bygger	rigger	rygger	tygger
Blide tanker .....ukentlig	firer	flirer	forer	fyrer
Gode .....grubler overalt	biller	filler	piller	riller
Rare .....jubler minst	kjeder	kjeler	kjever	reder
Til middag rugger været en	dram	pram	tram	trapp
Til frokost .....tjæren nitten riller	lugger	pugger	rugger	vugger
Om høsten .....listen en masse	koser	loser	moser	roser
Denne uka skjønner fienden en	gran	klan	kran	plan
Alltid .....fela en dram	lader	lager	lakker	lapper

Våkne briller . . . . . ukentlig	brer	grer	rer	trer
Gode filler . . . . . fort	skraper	skratter	sprader	spraker
Slemme . . . . . skrubber plutselig	kjeler	seler	sjefer	sjeler
Vakre paver . . . . . sikkert	brøler	føler	nøler	søler
Sikkert vrikker . . . . . en tang	haken	hanen	haren	hasen
Tregt . . . . . hunden en purk	slenger	slynger	sprenger	trenger
Uten . . . . . skinner luften sjelden	land	rand	tang	tann
Uten lakk svikter . . . . . farlig	bukten	fukten	lukten	lykten
På . . . . . lyder ferskenen lykkelig	bordet	gjerdet	hjula	jorda
Den kalde . . . . . brenner presist	nesen	nepen	neven	nisen
Hvordan . . . . . gutten glass	føler	kjøler	siler	søler