Sound Source Localitzation with the Eigenmike Microphone Array
-
Evaluation und Analysis

Manuel Brandner

Project Thesis
-
Supervisor: Dr. Alois Sontacchi
-
March 2014

institut für elektronische musik und akustik

Abstract

Directional detection of sound sources under defined ambience conditions using a spherical microphone array (Eigenmike) is examined. The used spatial detection algorithm correlates synthesized spherical wave spectra derived from theory with a set of concrete spherical spectra calculated from measured impulse responses. Thus, measurement signals were recorded with the 32 microphone equipped Eigenmike microphone array and two measurement sets were created for spatial sampling positions along an enclosing spherical surface with different radial distances. In order to simulate free field conditions and to compare the applicability of the proposed algorithm under real conditions the derived impulse responses are windowed adequately. Based on the Fourier transform of these resulting responses the calculation of the spherical wave spectra for specific source positions is possible. Under free-field conditions, the calculation of the synthesized spherical wave spectra of various spatial positions only depends on the structural properties of the microphone array and the position of the measured omnidirectional sound source. Correlation of measured and synthesized spherical wave spectra results in a data set with a maximum value for the sought direction of the sound source. Another aim of investigation is to understand the context between the size of the synthesized data - which serves as a lookup-table - and the directional accuracy. Within this thesis valuable information about functionality as well as the boundaries of the directional detection under defined spatial conditions with the spherical microphone array Eigenmike is given. The measurement is carried out at 612 equiangular source positions for the first measurement set and at 480 source positions for the second one. The results show a limited frequency resolution as expected, due to the arrangement of the microphones on the sphere. For the ideal case and a synthesis matrix including all measured source positions the algorithm yields for the full frequency range from 172 Hz up to the aliasing frequency (f_{alias} = 5.2kHz) an diminishing median deviation error. The accuracy is directly connected to the spatial sampling (microphone capsule spacing) and the size of the synthesis matrix.

Kurzfassung

Diese Arbeit untersucht eine Möglichkeit der Richtungsdetektion mit dem Eigenmike Mikrofonarray. Die Detektion erfolgt durch Korrelation synthetisierter sphärischer Wellenspektren für verschiedene Quellrichtungen mit einem aus Messdaten errechnetem sphärischen Wellenspektrum. Um die Berechnungen durchführen zu können, wurden zwei Messsets in zwei konstanten radialen Abständen entlang einer einhüllenden Kugeloberfläche gemessen. Um die Durchführbarkeit des Algorithmus in einer Freifeldsituation und dann für reale Bedingungen zu untersuchen wurden die Impulsantworten entsprechend gefenstert. Die fourier-transformierten Druckwerte aus den Impulsantworten werden dann zur Berechnung des sphärischen Wellenspektrums an einer bestimmten Quellposition verwendet. Die Synthesespektren werden durch die Bedingungen der Schallausbreitung, einbeziehen der baulichen Eigenschaften des Mikrofonarrays und der verwendeten Messanordnung über die Messsignale errechnet. Das Ergebnis der Korrelation der sphärischen Wellenspektren liefert normalisierte Werte in einem Vektor mit Information zur möglichen Quellrichtung. Hohe Werte treten für die Vektorindizes auf, die der detektierenden Quellrichtung am nächsten liegen. Vorzugsweise liefert der höchste Wert, die detektierte Quellrichtung. Ein Ziel des Projektes ist es den Einfluss der Größe des synthetisierten Datensets, welches einen "Lookup-Table" von sphärischen Spektren bildet, sowie die Genauigkeit zu untersuchen. Weiters wird Auskunft über die Durchführbarkeit der Richtungsdetektion mit dem Eigenmike und die realen Grenzen des Algorithmus für definierte räumliche Anordnungen gegeben. Es wurde für das erste Messdatenset an 612 Positionen und für das zweite Messdatenset an 480 Messpositionen im Abstand gleicher Winkel rund um das Mikrofonarray gemessen. Die Ergebnisse zeigen, wie zu erwarten war, durch die Anordnungen der Mikrofone an der Kugeloberfläche eine begrenzte Frequenzauflösung. Für die ideale Messanordnung und einer Synthesematrix, die alle Messpositionen beschreibt, liefert der Algorithmus im Frequenzbereich von 172Hz bis zur Aliasingfrequenz (f_{alias} = 5.2kHz) einen verschwindenden Mittelwert des Raumwinkelfehlers. Die Genauigkeit der Detektion hängt unmittelbar mit dem Sampling des Arrays und der Größe der Synthesematrix zusammen.

Danksagung

Besonderer Dank gebührt Alois Sontacchi für die geduldige Betreuung und die stets motivierenden Worte.

Außerdem möchte ich mich für die Hilfe von Mathias Frank, Franz Zotter und allen anderen Mitarbeiter des IEM bedanken, die sich immer wieder für lehrreiche Gespräche Zeit nahmen.

Contents

1 Theory **3**
- 1.1 Introduction . . . 3
- 1.2 Method . . . 4
  - 1.2.1 Synthesized Spectra . . . 5
  - 1.2.2 Measured Spectra . . . 6
  - 1.2.3 Matching profile of Spherical Wave Spectra . . . 7
  - 1.2.4 Examining the ASL Algorithm . . . 7

2 Measurement **9**
- 2.1 Measurement Setup . . . 9
- 2.2 Limitation due to Measurement Arrangement and Array Structure . . . 10

3 Evaluation and Analysis **12**
- 3.1 Reducing the Synthesis Matrix . . . 12
- 3.2 Dependence of Spherical Harmonics Order and Phase Information . . . 14
- 3.3 Vector Interpolation Approach . . . 15
- 3.4 D/R-Ratio - Direct-to-Reverberant Sound . . . 17
- 3.5 Dependence of the Radial Filter . . . 18

4 Summary and Conclusion **20**
- 4.1 Outlook . . . 21

Chapter 1

Theory

1.1 Introduction

Acoustic source localization (ASL) is a common subject in scientific literature and is feasible through different approaches. The angular information estimated by an source localization algorithm can be used as a steering indicator for microphone array beamforming [1]. A common used method for ASL is to calculate the sought direction through TDOA (time difference of arrival) which prerequisite is a spatially distributed array. Within this method the array spacing is directly related to the magnitude detection of TDOA and large spacings are common. A different approach for ASL is to use near-conincident microphone arrays (NCMA), where the spacing of the microphone capsules is reduced to a minimum. NCMA design emphasizes on spacing two up to more microphones as close as possible to the acoustic center of the array. The principle of determining the source direction here, is to use level differences between microphone signals.
Common Methods calculate the localization through the sound intensity vector approach [2] which calculates the sought direction in the frequency domain. Using eigenbeams or evaluating the sound pressure distributions via histogram estimates are further techniques for sound source localization. This paper presents an approach, also frequency dependent, but the evaluation of a source direction is made via spherical harmonics.

As a foundation for an ideal measurement setup we want the room influences to be zero or at least reduced to a minimum, later we want to have a look at the feasibility for an arbitrary room situation. The ideal setup means that only the direct sound is recorded without any influences of a diffuse sound field. If not possible adequate windowing is necessary. If this conditions are predominant, we can calculate the sound field around an microphone array

analytically with the spherical base solutions [3]

$$\begin{aligned} p(kr, \mathbf{\Theta}) &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [b_{nm} j_n(kr) + c_{nm} h_n^{(2)}(kr)] Y_{nm}(\mathbf{\Theta}) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [b_{nm} R_n^m(kr, \mathbf{\Theta}) + c_{nm} h_n^{(2)} S_n^m(kr, \mathbf{\Theta})]. \end{aligned} \tag{1.1}$$

The coefficients b_{nm} and c_{nm} can be called the wave spectrum of the incident and radiating field, respectively. $J_n(kr)$ are the spherical Bessel functions, and $h_n^{(2)}(kr)$ are the spherical Hankel functions of the second kind. $R_n^m(kr, \mathbf{\Theta})$ describes the regular incident field solution and $S_n^m(kr, \mathbf{\Theta})$ is the singular radiating field solution. Y_{nm} are the normalized spherical harmonics depending on the solid/room angle (unit vector). The indices n and m stand for the order and degree, respectively with the ranges $0 \leqslant n \leqslant N$, and $-n \leqslant m \leqslant n$. With the spherical base solutions the sound pressure distribution of an sound field can be fully described analytically. On the other hand of course measuring such a sound field with an microphone array is possible. If e.g. the analytically calculated pressure values from a sound source at a specific location are compared with the measured pressure values at this very position, they should show the same pressure contour around the sphere of the microphone array. The source direction can be determined by a maximum search of the correlation function obtained via correlation of a set of analytically derived wave spectra with measured wave spectra.

1.2 Method

This project presents a novel algorithm that correlates a synthesized spherical wave spectrum, which is calculated analytically, with a measured spectra in order to get the source direction towards an acoustic sound source. The measured wave spectrum of an source position can be calculate via the spherical harmonic transform (SHT, [4]). The transform of an arbitrary function $g(\vartheta, \varphi)$ into spherical harmonics coefficients γ_{nm} is described in (1.2) and (1.3). The correlation of a measured spectrum with a lookup-table of synthesized spectra should yield in the ideal case a matching profile for all possible source locations with a maximum at the position for the sought direction.

$$SHT_{nm}\{g(\vartheta, \varphi)\} = \gamma_{nm} \tag{1.2}$$

$$\gamma_{nm} = \int_{\varphi=0}^{2\pi} \int_{\vartheta=0}^{\pi} g(\vartheta, \varphi) Y_{nm}(\vartheta, \varphi) \sin(\vartheta) d\vartheta d\varphi \tag{1.3}$$

1.2.1 Synthesized Spectra

For the calculation of the pressure distribution from incident and radiating fields the spherical base solutions are the mathematical fundamentals (cf.(1.1)). After transforming the equation the spherical wave spectrum $A_{nm,s}$ can be stated as follows:

$$A_{nm,s} = j_n(kr_m)b_{nm} + h_n(kr_m)c_{nm}. \tag{1.4}$$

If a point source is examined at $(r_q, \mathbf{\Theta})$ the expression can be expanded:

$$b_{nm} = -ikh_n(kr_q)Y_{nm}(\mathbf{\Theta}) \tag{1.5}$$

$$A_{nm,s} = -ikj_n(kr_m)h_n(kr_q)Y_{nm}(\mathbf{\Theta}) + c_{nm}h_n(kr_m). \tag{1.6}$$

The spherical Bessel functions j_n and h_n the spherical Hankel functions are dependent on the order of the examined system. Equation (1.6) describes the synthesized spherical spectrum and holds two parts in the formula. The first part is the sound field of a point source from the source direction $(\mathbf{\Theta})$ decomposed in spherical harmonics. The distance r_q is the source radius starting from the microphone array center whereas r_m describes the distance from the center to the enclosing surface of the rigid sphere microphone array. Here only the far field conditions are taken into account. The second part in equation (1.6) describes the reflection of the incident sound field on the surface of the microphone array. The sound field decomposition from the center to the microphone surface can be analytically denoted via the coefficients c_{nm} and the propagation term h_n.
The derivation of the sound pressure yields the sound particle velocity. For the spectral terms the same is true, the derivation of the spherical wave spectrum of the pressure distribution yields the spherical wave spectrum of the sound particle velocity. On a rigid sphere we know that the normal component of the sound particle velocity has to be zero. The derivation of equation (1.4) will be set to zero. This steps allows us the calculation of the coefficients c_{nm} (cf.(1.7)). The derivation

$$\frac{i}{\rho_0 c} j_n^{'}(kr_m)b_{nm} + h_n^{'}(kr_m)c_{nm} = 0 \tag{1.7}$$

yields the coefficients c_{nm}

$$c_{nm} = -\frac{j_n^{'}(kr_m)}{h_n^{'}(kr_m)} b_{nm}. \tag{1.8}$$

Equation (1.4) will be expandend with the coefficients

$$A_{nm,s} = j_n(kr_m)b_{nm} - \frac{j_n^{'}(kr_m)h_n(kr_m)}{h_n^{'}(kr_m)} b_{nm} \tag{1.9}$$

and can be expressed as follows

$$A_{nm,s} = \frac{j_n(kr_m)h_n^{'}(kr_m) - j_n^{'}(kr_m)h_n(kr_m)}{h_n^{'}(kr_m)} b_{nm}. \tag{1.10}$$

The denominator (1.10) can be simplified via $1/i(kr)^2$, which is the Wronski determinant. With this simplification and equation (1.6) we can state the synthesis-vector.

$$A_{nm,s} = -\frac{h_n(kr_q)}{kr_{mic}^2 h_n^{'}(kr_m)} Y_{nm}(\mathbf{\Theta}) \tag{1.11}$$

The first multiplicand in (1.11) is the radial-filter term with the paramters frequency ($k = \omega/c$), radius from the center to the microphone surface r_m, and source radius r_q. The second multiplicand states the evaluation of the spherical harmonic decomposition of a source position from direction (φ, ϑ). To get an set of spherical wave spectra for every source direction (eq.1.12), it is necessary to build a $\mathbf{Y}_{nm}$-Matrix (eq.1.13) with all angle pairs $\{\varphi_l, \vartheta_l\}_{l=1\ldots L}$ at the respective source location,

$$\mathbf{A}_{nm,s} = -\frac{diag\left\{\mathbf{h_n}(kr_q)\right\}}{kr_{mic}^2 diag\left\{\mathbf{h}_n^{'}(kr_m)\right\}} \mathbf{Y}_{nm}(\mathbf{\Theta}_l)_{l=1\ldots L} \tag{1.12}$$

$$\mathbf{Y}_{nm}(\mathbf{\Theta}_l) = \begin{pmatrix} Y_{00}(\Theta_1) & Y_{00}(\Theta_2) & \cdots & Y_{00}(\Theta_l) \\ Y_{1-1}(\Theta_1) & Y_{1-1}(\Theta_2) & \cdots & Y_{1-1}(\Theta_l) \\ Y_{10}(\Theta_1) & Y_{10}(\Theta_2) & \cdots & Y_{10}(\Theta_l) \\ Y_{11}(\Theta_1) & Y_{11}(\Theta_2) & \cdots & Y_{11}(\Theta_l) \\ \vdots & \vdots & \ddots & \vdots \\ Y_{nm}(\Theta_1) & Y_{nm}(\Theta_2) & \cdots & Y_{nm}(\Theta_l) \end{pmatrix}^T . \tag{1.13}$$

1.2.2 Measured Spectra

An Investigation of a specific source position is expressed as in equation (1.1) for a specific frequency and yields with a transform a measured spherical spectrum. The vector $\mathbf{p}$ in equation (1.14) is already Fourier transformed and therefore frequency dependent. Because we want to investigate the spherical wave spectrum, equation (1.14) needs to be transformed to (1.15)

$$\begin{pmatrix} \mathbf{p}_{\varphi_{mic1},\vartheta_{mic1}}(f) \\ \mathbf{p}_{\varphi_{mic2},\vartheta_{mic2}}(f) \\ \vdots \\ \mathbf{p}_{\varphi_{mic32},\vartheta_{mic32}}(f) \end{pmatrix} = \mathbf{Y}_{nm}\mathbf{A}_{nm} \tag{1.14}$$

$$\mathbf{A}_{nm} = \mathbf{p}\mathbf{Y}_{nm}^{-1} \tag{1.15}$$

To prevent that the matrix inversion yields a singular solution, the Y_{nm} -matrix can be inverted via a pseudo inverse:

$$\mathbf{A}_{nm} = (\mathbf{Y}^T\mathbf{Y})^{-1}\,\mathbf{Y}^T\mathbf{p} \tag{1.16}$$

Nevertheless, adequate sampling is necessary to yield a condition number as small as possible[1]. Equation (1.16) results in a $(N+1)^2$ long vector and corresponds to the spherical wave spectrum calculated from the measured sound pressure distribution.

1.2.3 Matching profile of Spherical Wave Spectra

The feasibility of the algorithm is examined by the autocorrelation of the measured wave spectra for every measurement position. For this special case the angular deviation is zero. Further investigations are made on examining the correlation of a synthesized set of wave spectra (calculated analytically) with a measured spectrum for a specific source position. To get to an matching profile (correlation vector) we first have to calculate the correlation coefficient of the $\mathbf{A}_{nm}$-vector and the $\mathbf{A}_{nm}$-matrix [5]. This synthesis matrix ($\mathbf{A}_{nm}$-matrix) holds all angle dependent source positions, so the correlation coefficient has to be calculated for all angle pairs $\{\varphi_l, \vartheta_l\}_{l=1\ldots L}$. The result of each correlation will be stated as in (1.17), where $\mathbf{A}_{nm,s}$ is a (NxM)-Matrix (N source positions and M sampling points) and denotes the synthesized spectrum. $\mathbf{A}_{nm}$ is a vector of length M and denotes the measured spectra. The highest correlation value of the $\mathbf{A}_{nmMATCH}$-vector for a successful source localization includes the sought angle pair (seen in Figure 1.1).

$$\mathbf{A}_{nmMATCH} = \frac{\mathbf{A}_{nm,s}^H\mathbf{A}_{nm}}{\sqrt{\mathbf{A}_{nm,s}^H\mathbf{A}_{nm,s}\mathbf{A}_{nm}^H\mathbf{A}_{nm}}} \tag{1.17}$$

1.2.4 Examining the ASL Algorithm

Since an analytically feasible acoustic source localization algorithm for arbitrary source direction is found, the next step is to examine the feasibility for different scenarios. The first empirical evaluation of the algorithm is made with a measurement setup fulfilling the free field condition. Impulse responses are obtained using a exponential sweep measurement technique and

[1] The configuration of the capsules on the sphere of the Eigenmike yield such a small condition number.

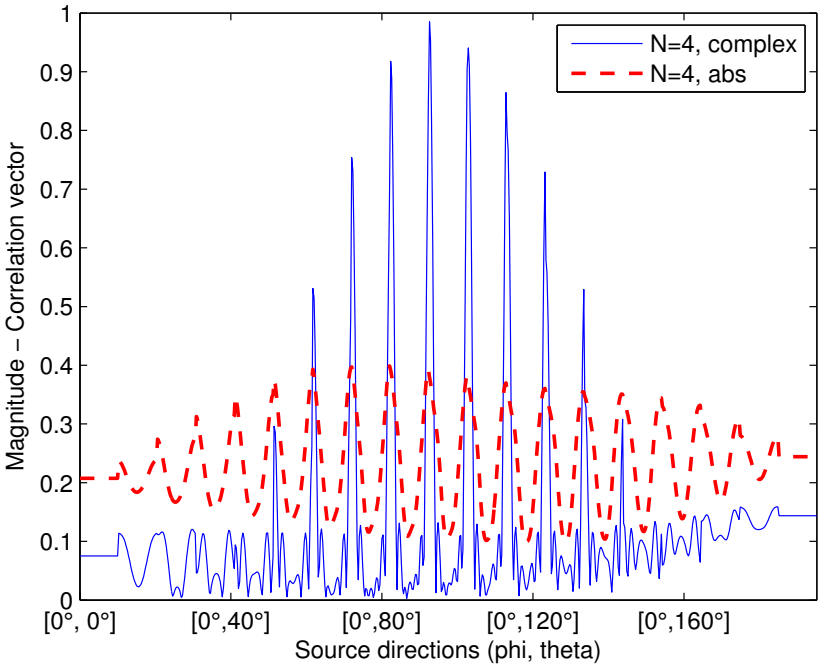

Figure 1.1: Matching profile $\mathbf{A}_{nmMATCH}$ for 4kHz over all source positions - red: absolute vector, blue: complex vector, sought direction at phi=0°,theta=90°, the x-axis shows all 612 source positions starting from the zenith

adequate windowing with a window with 20 samples fade in and 80 samples fade out to prevent spatial influences.
For further investigations the behavior of the algorithm by reducing the synthesis matrix - reducing the sets grid of possible source locations - to a minimum is examined. After minimizing the synthesis matrix, which holds the possible source locations, an vector interpolation approach is made. In theory, the three strongest components of the matching profile should hold the most valuable directional information. Therefore, these components are vectiorially summed up and weighted to gain an improvement for small synthesis sets.
Another task is examining the robustness of the algorithm for different window lengths, which is directly connected to a change of the D/R - ratio (ratio of the direct sound field to reverberant sound field) and similar to changing the distance to the acoustic sound source.
Depending on the radial distance of a sound source to the microphone array, a radial filter is necessary to focus the localization algorithm onto the source. These dependencies are also subject of interest. Investigations on the performance are also in relation to the structural characteristics of the microphone array and therefore the limits of the aperture have to be taken into account.

Chapter 2

Measurement

2.1 Measurement Setup

Both measurements were carried out at the Institute of Electronic Music and Acoustics (IEM) in the "IEM-CUBE" in Graz. The CUBE is an 10.3 x 12 x 4.8 m large room (reverberation time $RT_{60} \approx 0.7$s, $r_H = 1.9$m). For the first setup the task was to minimize the room influences and gain free field conditions. Therefore, a radius at $r_q = 0.7$m was chosen and a loudspeaker with nearly spherical radiation, one acoustic center and linear frequency response was used (see Figure 2.1, Figure 2.2 shows the measurement chain). The different source positions of the loudspeaker were precisely set up with the optitrack tracking system[2]. The microphone array was placed on a turntable and controlled via ethernet to prevent changes in orientation.
The first measurement was taken in 10° steps in azimuth and elevation. The microphone was positioned at a height of 1.4m. The measurement was reduced to 17 positions in the elevation because of the difficulty to position the loudspeaker underneath the array and the fact that the microphone array shows identical properties at top and bottom, resulting in a set of 612 source positions. The second measurement was taken at the same site but at a distance of 1.5m ($r_H = 1.9$m). The loudspeaker setup this time was a fixed system of 16 loudspeaker on a metallic ring, where no tracking of the source position was necessary. This set holds 480 source positions and was taken in 12° steps in azimuth and 11.25° steps in elevation.

[2]website: http://www.naturalpoint.com/optitrack/

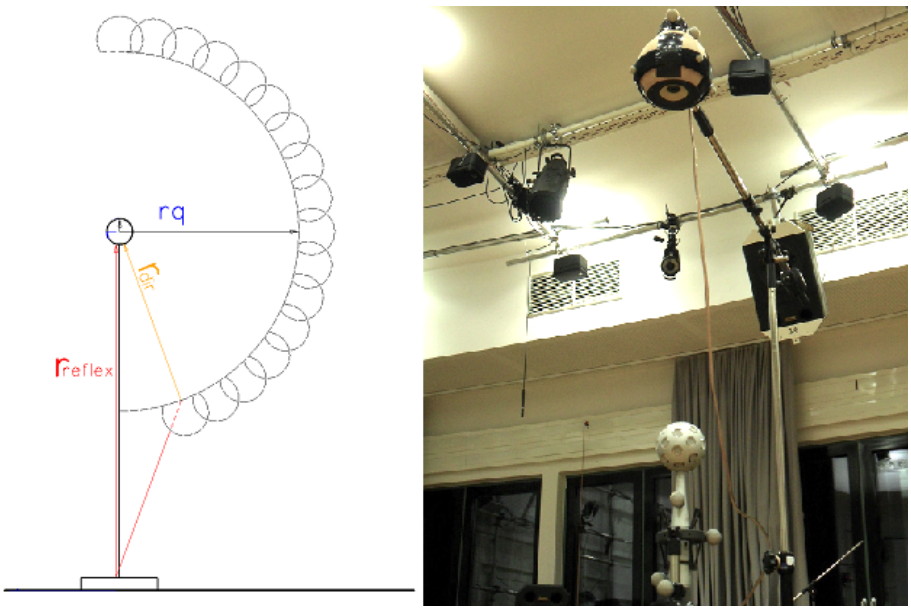

Figure 2.1: Measurement setup, left: measurement sketch, right: image of tracking and positioning the loudspeaker.

2.2 Limitation due to Measurement Arrangement and Array Structure

On the one hand, we experience an upper frequency limit, namely the aliasing frequency (can be seen as the spatial nyquist frequency), which is depending

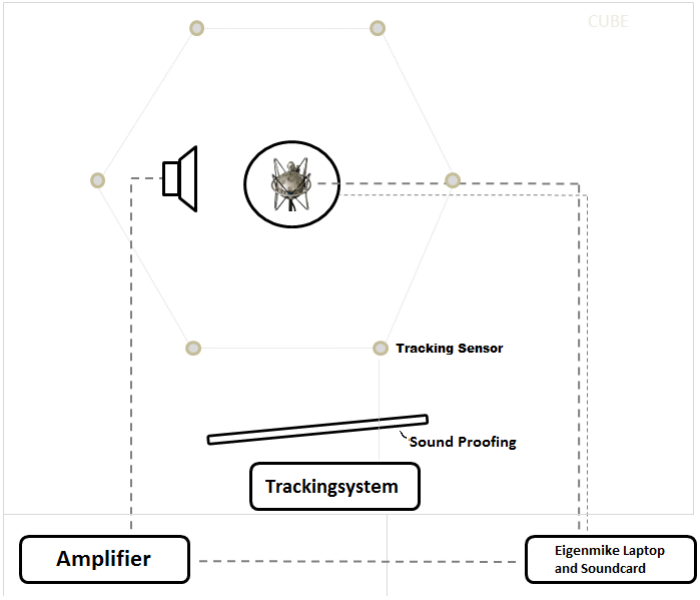

Figure 2.2: Measurement chain, arrangement of equipment, except the tracking system the noisy equipment was located outside, to minimize the noise influence of the tracking system fans a mobile soundproofing was used.

on the array spacing. The Eigenmike consists of 32 microphone capsules and is therefore an oversampled 4th order holographic array ($samplingpoints \equiv (N+1)^2$). The aliasing frequency can be expressed as follows:

$$f_{alias} = \frac{Nc}{r_{mic}\,2\pi} = 5.2\text{kHz}. \tag{2.1}$$

The lower frequency limitation is for practical reasons of the measurement setup. If we consider a loudspeaker at a radial distance of 0.7m, the lower frequency bound can be expressed as (free field condition):

$$\frac{2\pi f}{c} r \geqslant 2 \tag{2.2}$$

$$f_{lower} \geqslant 155.97\text{Hz} \tag{2.3}$$

The loudspeaker's frequency response shows the -3dB mark for low frequencies at 150Hz.

Chapter 3

Evaluation and Analysis

In Figure 3.1 we can see the overall performance of the proposed ASL algorithm, of course for the ideal case, where the free field condition is fulfilled. In this ideal case the algorithm still works beyond the aliasing frequency, because only little dynamic for a correct localization is necessary - no interferences are predominant. The same phenomenon occurs for very low frequencies. The matching vector has a low dynamic, but still enough to separate the sought direction. In case of low dynamics we included the phase term which should yield an improvement in the source localization. Due to the fact that the for low frequencies the amplitudes of adjacent capsules are a minimum an improvement is desired. The empirical evaluation shows that including the phase term also leads to an improvement of the algorithm for the rest of the interested frequency range (cf. Figure 1.1 and Figure 3.4). For the investigation of harsher conditions a second measurement set is evaluated. The radial distance to the source positions is more than double the distance for the first arrangement. This should enable the comparison of the ideal case with a case at a source radius with higher room influences and give an overview of the performance of the algorithm (see Sec.3.4).

3.1 Reducing the Synthesis Matrix

The idea of reducing the synthesis matrix is to generate a lookup-table for source localization with minimum computational costs. The microphone array is a 4th order system with 32 microphone capsules (to be precise, an overdetermined 4th order system, $(N+1)^2 < 32$) and it is obvious that dividing the sphere by the main lobe of a 4th order beam should give us the minimal sampling points necessary to still take the advantages of the HOA system. The approach leads us to a minimal synthesis set. If the synthesis matrix is thinned to this 25 points, the exact localization of the acoustic

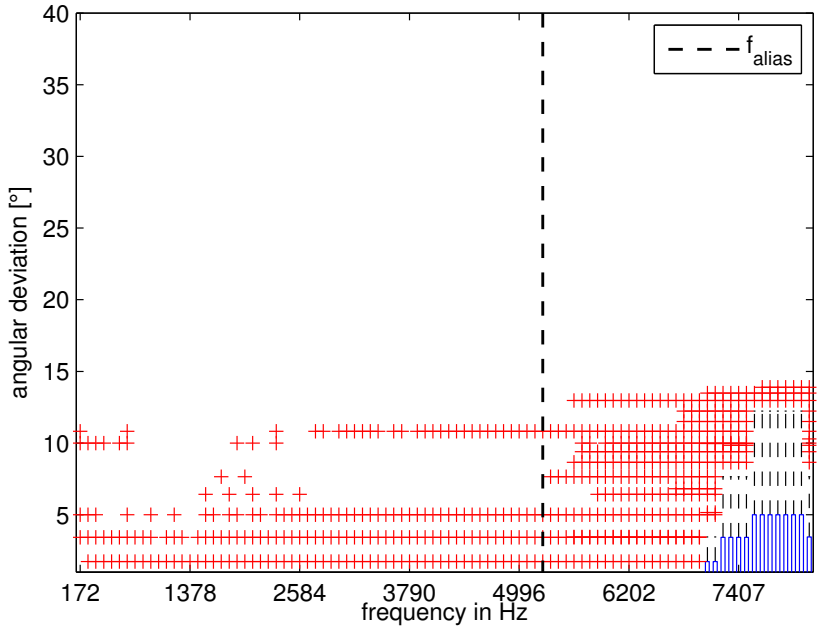

Figure 3.1: Boxplot of the angular deviation for all source positions over frequency for 612pt synthesis matrix.

source is not possible anymore. The actual position will be automatically mapped to the nearest neighbor in the lookup-table. With this knowledge and by knowing the 4th order beam angle at -3dB [6](cf.(3.1)), we can assume that the median in Figure 3.2 almost lies at the half of the beam width. Because the arrangement of the microphones does not correspond to a platonic solid, a deviation of the mean value away from half of the beamwidth can be expected. The deviation might be approximated via the ratio of the order to the number of the actual microphone capsules ($\mu_{approx} = 14.6°$).

$$\gamma_{beam} = \frac{187°}{N+1} = 37.4°. \tag{3.1}$$

If the synthesis matrix is thinned out to 49 points ($\gamma_{beam} = 26.7°$) the median deviation is reduced to 11° (cf. Figure 3.3). For the ideal case the algorithm is feasible beyond the alias frequency as long as the magnitudes of the spatial aliasing beams are smaller than the one in the sought direction. We can state that at least the reduction of the synthesis grid is feasible and leads also to a reduction of computational costs. Nevertheless, the intrinsic error increases with reduction of the lookup-table.

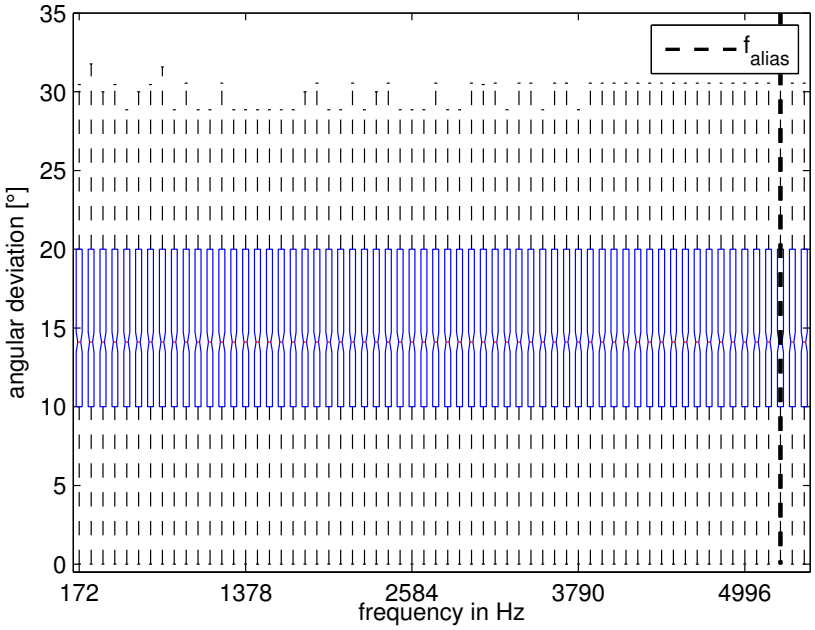

Figure 3.2: Boxplot of the angular deviation for all source positions over frequency for 25pt synthesis matrix.

3.2 Dependence of Spherical Harmonics Order and Phase Information

As we can see in Figure 1.1, including the phase information (N=4, complex) of the recorded signals yields a much higher dynamic. The phase information lies in the time difference of arrival at the capsules. The impulse responses are always calculated relatively to each other. In the case of the matching profile for 4 kHz we see a dynamic four times higher than without the complex matching approach.
If we have a look at the spherical harmonics order, we can see a reduction of ambiguity for higher order calculations in the matching profile which can not only be seen as an enhancement of dynamics, but also as an improvement for robust vector interpolation (cf. Figure 3.4). As already stated an increase of accuracy for low frequencies is also given. The higher the order, the smaller the side lobes of the spherical beam, which means also less ambiguities.

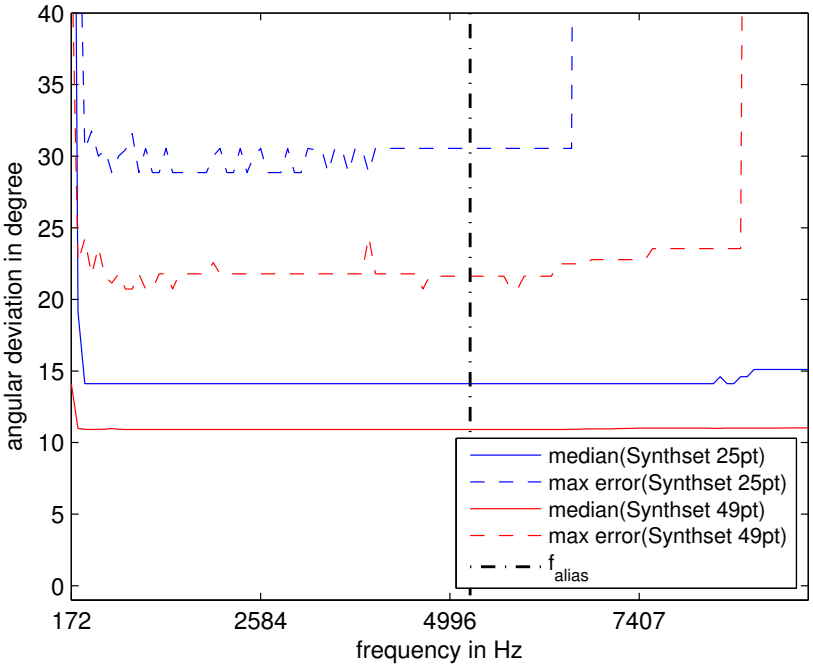

Figure 3.3: Angular deviation and maximum deviation for all source positions over frequency for 25pt synthesis matrix and 49pt.

3.3 Vector Interpolation Approach

In order to make the algorithm more robust concerning ASL, an approach of calculating the location via the vectorially summation of the strongest three components of the correlation vector (matching profile) is proposed. Two approaches were made: the first one calculates the sought source direction via the normalized summation, whereas by the second one the value of the component in the correlation vector is used as a weight. The position of the value in the correlation vector holds the information about the azimuth and elevation angle. Figure 3.5 depicts the reduction of the angle deviation for a reduced synthesis set of 25 points. A vector interpolation for high density synthesis grids seems due to the minimal deviation to the nearest neighbor not necessary at the moment. For realtime applications the interpolation has to be further evaluated.

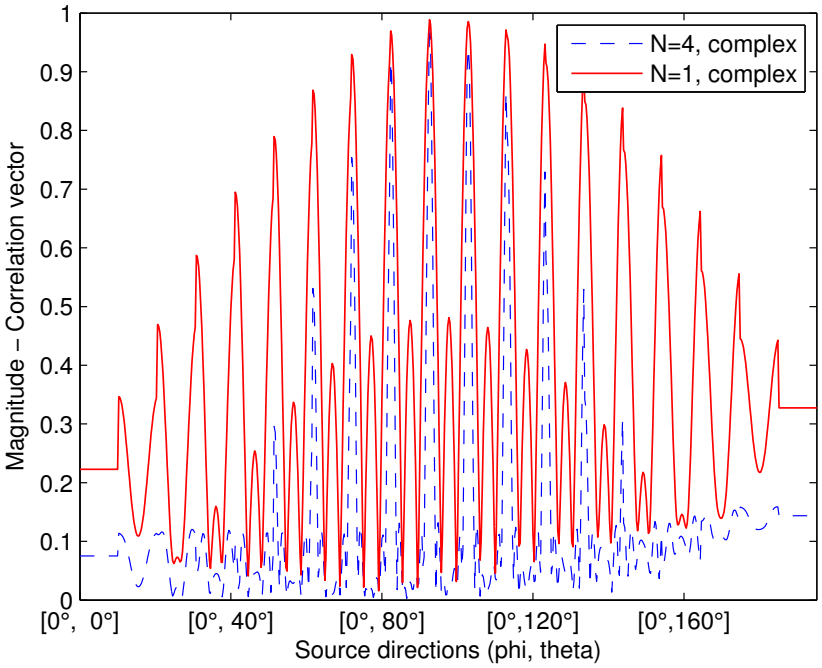

Figure 3.4: Matching profile $A_{nmMATCH}$ for 4kHz over all source positions - red: N=1, blue dashed: N=4, sought direction at phi=0°,theta=90°.

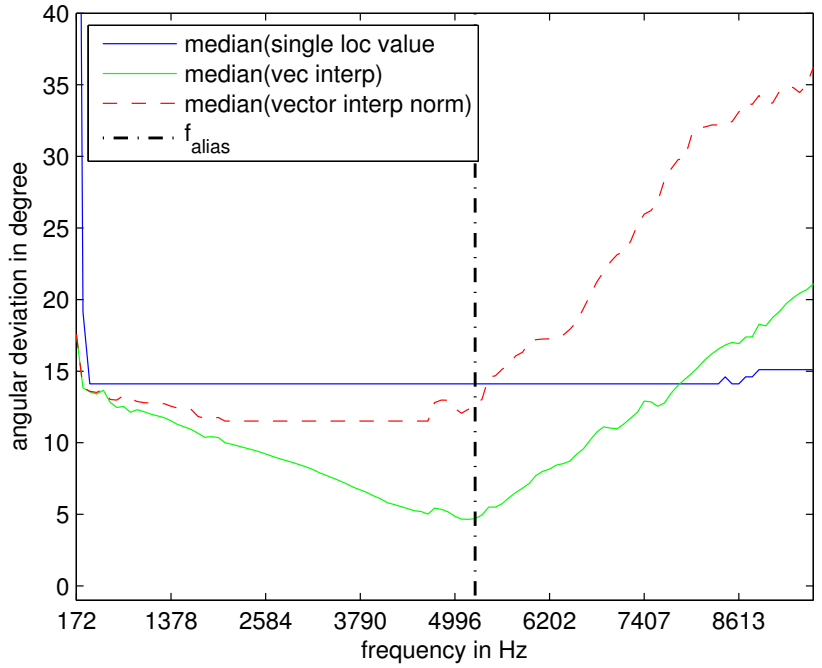

Figure 3.5: Median angular deviation for all source positions over frequency for 25 point synthesis set in comparison of two vector interpolation approaches. The normalized approach is only a vectorially summation of the three highest correlation values, whereas the second interpolation approach uses corresponding correlation dependent weighting.

3.4 D/R-Ratio - Direct-to-Reverberant Sound

To have a glance at a real performance situation and the impact on the accuracy of the algorithm, different window lengths were considered. The ratio of the direct sound field to the diffuse sound field was chosen as an adequate parameter (cf.(3.2)) to evaluate every source position based on a the predominant sound field and not only in relation to the source distance. The window length of the direct sound was chosen with 128 samples, based on an evaluation of the impulse responses. Figure 3.6 depicts that the longer the overall window length is chosen, the higher the deviation error gets and the D/R-ratio decreases. It can be seen as similar to changing the distance toward the source position. If we increase the distance between the microphone array and the acoustical source, the influences of disturbance increase, therefore the D/R-ratio decreases.

$$D/R = \frac{10log(\sum s_{dir}^2)}{10log(\sum s_{rev}^2)}. \tag{3.2}$$

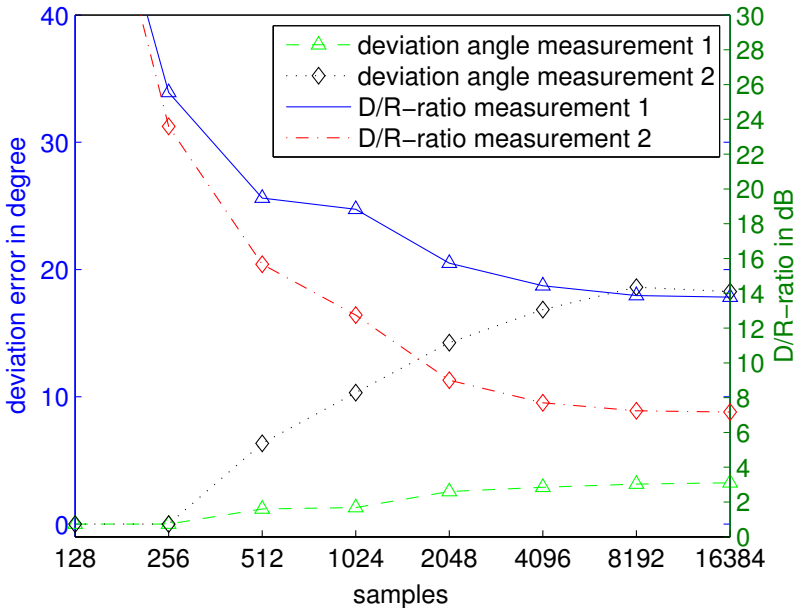

Figure 3.6: Direct sound to reverberant sound ratio versus the window length. For the first measurement set a radial distance of $r_q = 0.7$m toward the acoustic sources was chosen. The course of the D/R ratio for the first set shows only a 16dB range due to the measurement arrangement. Because examining the algorithm in a free field situation the room influences are a minimum. For the second measurement set less damping material is used and the distance of the radial arranged sources is almost the critical distance, $r_H = 1.9$m. Higher window lengths are in direct relation with a higher amount of room influences which is still visible in the plot.

The evaluation was made for both measurement sets (measurement set 1, $r_q = 0.7$m and measurement set 2, $r_q = 1.5$m). As we can see in Figure 3.6 the increase of the distance yields a larger angle deviation for higher window lengths.
If instead of an analytical lookup-table a measurement set is considered and for evaluation purposes windowed with a short window length, the almost same accuracy can be achieved. The advantage of this approach is that a possible mismatch of the microphone capsules are taken into account. However, a drawback would be that the windowing is directly connected to the resolution in the digital domain. If short windowed impulse responses after the SHT are used as a lookup-table, the correlation with measured data can only be made for existing frequencies in the spectrum due to the FFT-resolution. If the lookup-table and the measured data have the same window lengths no further calculations are necessary. But if the window length of the source signal is larger than the one of the lookup-table, averaging over the frequency bin of interest and its neighbors is crucial. The analytical solution is independent of the FFT-resolution and can be calculated for any frequency of interest.

3.5 Dependence of the Radial Filter

In Section 1.2.1 a method to calculate a synthesized spectrum was treated. Equation (1.12) consists of two multiplicands, whereas the second one states the spherical harmonics at specific source position. The first fractional term states the radialfilter, which focuses the spectrum at a specific source distance. The higher the order of the spherical harmonics, the higher the amplification of the associated frequency component gets (cf. Figure 3.7). Especially if a noise is added in the signal, the high amplification also enlarges the unwanted signal components. If we consider different focusing (also wrong focusing) for one and the same source position, it is possible to examine the impact on the source localization. The error angle is in direct relation with focusing on a source position via the radial filter (cf. Figure 3.8). Problems occur if the algorithm focuses on a source distance closer than the actual radius. If further away focused, no increase in the angular deviation is observed. As a limitation on focusing on a source the amplification of the radial filter has to be considered. It is not feasible to amplify a signal more than 60dB due to the dynamic range of the microphone array. Therefore, a maximum distance in regarding of the radial filter limits the focus.

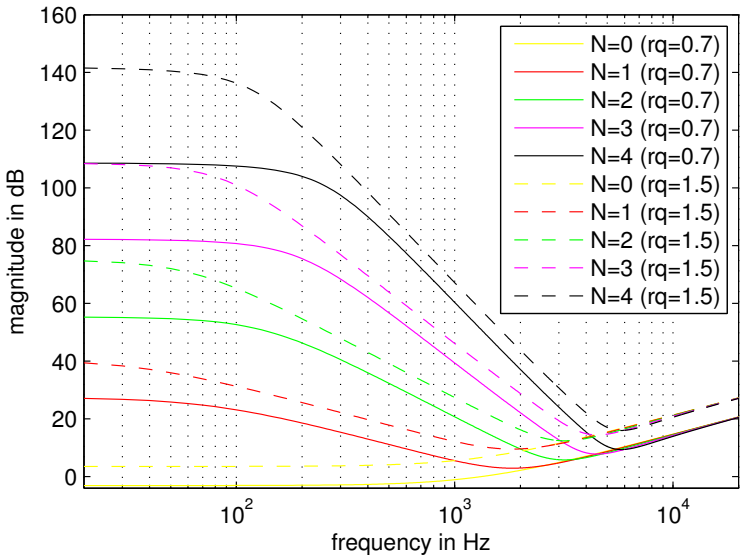

Figure 3.7: Radial filter for $r_q = 0.7$m and $r_q = 1.5$m, the radial filter show extremely high amplification for higher orders in the lower frequency range.

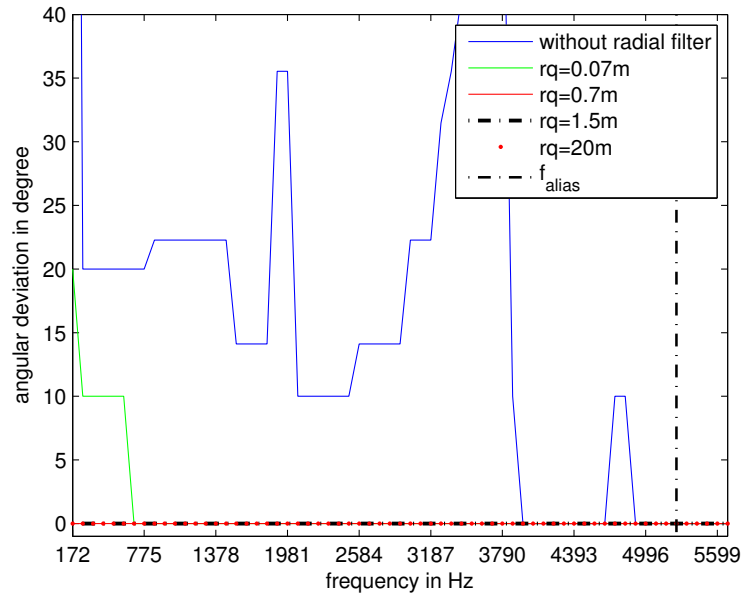

Figure 3.8: Angular deviation for one source position in relation to the focusing of the radial filter, $r_q = 0.7$m. If no radial filter is used, the source localization is not feasible. For focusing on sources nearer than the actual source distance a higher deviation can be observed. Focusing on source positions farther away than the actual position, the source localization is stable.

Chapter 4

Summary and Conclusion

First of all, the evaluation showed the feasibility of the algorithm for free field conditions. For this specific conditions the algorithm yields in the whole frequency range of interest (up to 5.2kHz) excellent results. Especially including the phase term in the source localization enhances the dynamic range extremely. The approach of using a higher order ambisonics system yields much more information about the source and direction than a conventional approach.
Second of all, it is shown that it is possible to reduce the synthesis grid in order to cut back computational costs. Nevertheless, an increase of the intrinsic deviation error is the result. If the synthesized matrix is cropped to a minimum, a vector interpolation yields a very decent enhancement, depending on the frequency, of up to 10 degree (cf. Section 3.3, Figure 3.5). Of course the best result in respect of the deviation error angle yields the source localization algorithm if the synthesis matrix is maximized.
Furthermore, it is shown that changing the window size and also changing the radial distance of the acoustic source is directly connected to the performance of the localization. The more influences of the room is omitted, the better the algorithm works. Especially room reflections with directional information impair the results.
Another treated subject is the impact of focusing on a specific radial distance and the result of wrong focusing. If the source is closer focused than actual located, the deviation error increases for low frequencies noticeable. On the other hand for focusing on positions further away as the actual source position neglectable errors occur. However, it is not feasible to focus on infinity, because this would yield high amplification for all frequency terms.

4.1 Outlook

There are several tasks that can be considered for further evaluation of the algorithm. Examining the feasibility of the source localization for different distances and positions with an audio signal, such as speech and music, would be very interesting. The low frequency range shows lower dynamics in the matching vector and therefore a closer look on real environment behavior should be made. Furthermore, the dependence of the phase information for higher frequencies and the phase term in respect to the correlation of audio signals for real-life conditions should be evaluated.
Investigating the angular deviation and dependence of focusing with the radial filter farther away as the actual acoustic source is necessary. Especially the relationship of the radial filter and the boost for low frequencies could hold necessary informations for practical purposes.

Bibliography

[1] Tervo, S.: Direction estimation based on sound intensity vectors. 17th European Signal Processing Conference, Scotland (2009), 700-704

[2] Freiberger, K. and Sontacchi, A.: Similarity-based sound source localization with a coincident microphone array. Proc. of the 14th Int. Conference on Digital Audio Effects (DAFx-11), France (2011), 185-190

[3] Earl G. Williams. Fourier Acoustics: Sound Radiation and Near- field Acoustical Holography. Academic Press, 1999

[4] Plessas, P.: Rigid sphere microphone arrays for spatial recording and holography. Master's thesis, Institute of Electronic Music and Acoustics University of Music and Performing Arts Graz, Austria (2009)

[5] Hohl, F.: Kugelmikrofonarray zur Abstrahlungsvermessung von Musikinstrumenten. Master's thesis, Institute of Electronic Music and AcousticsUniversity of Music and Performing Arts Graz, Austria (2009)

[6] Zotter, F.: Analysis and Synthesis of Sound-Radiation with Spherical Arrays. PhD thesis, Institute of Electronic Music and Acoustics University of Music and Performing Arts, Austria (2009)