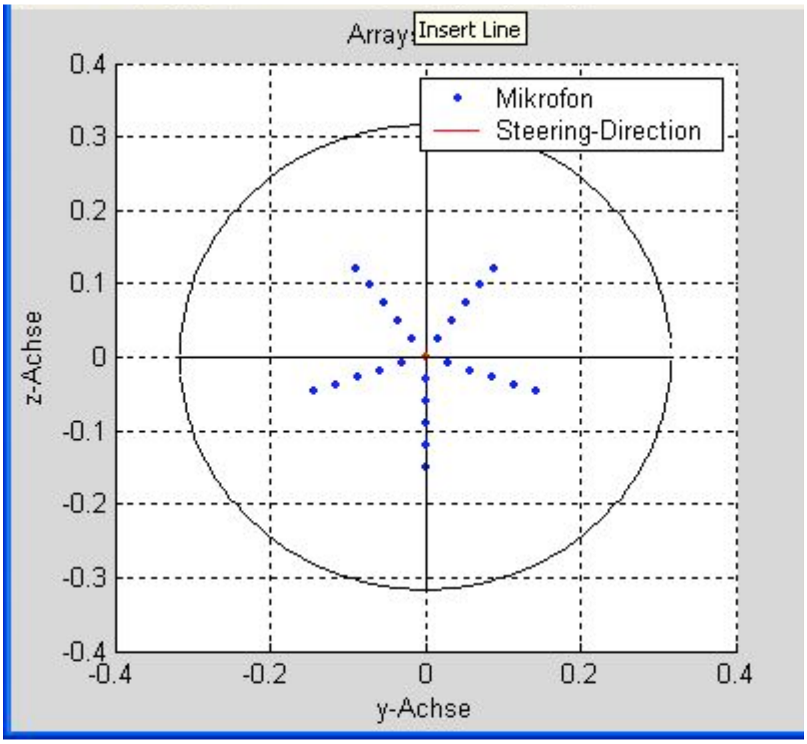
Akustische Dialogsegmentierung mit kleinen Mikrofonarrays

Toningenieur-Projekt
Lutz Pape

Durchgeführt am
Institut für Elektronische Musik und Akustik (IEM)
Graz, 2005



Betreuer: Dipl-Ing. Markus Noisternig

Aufgabenstellung:

Es soll ein kleines Mikrofonarray entworfen werden, mit Hilfe dessen es möglich ist über Delay & Sum Beamforming eine segmentierte (spurengetrennte) Aufzeichnung eines Dialoges zu erreichen. Die Geometrie des Arrays soll so beschaffen sein, dass über Steering die Positionen der Sprecher verfolgt werden können. Der zu übertragende Frequenzbereich beschränkt sich auf den der Sprache (Bandbreite ca. 4kHz).

Die Umsetzung des Projektes erfolgt über die Entwicklung eines Array-Simulationsprogrammes in Matlab. Mit diesem Programm soll zuerst eine geeignete Geometrie für die Mikrofonanordnung gefunden werden. Danach wird das Array mit den besten Simulationsergebnissen gebaut und mit Hilfe der Software Pure Data[1)] die geeignete Beamformer-Struktur erstellt. Die Ergebnisse der Simulationen sollen dann durch Vermessung der Richtcharakteristik des entwickelten Beamformers bei verschiedenen Steering-Directions überprüft werden.

1) Pure Data ist eine Open Source Programmierumgebung zur Echtzeit-Signalverarbeitung von Audiosignalen.

Inhalt:

I. Grundlagen des Arrayprocessing

Unter Array Processing versteht man das Senden oder Empfangen von Wellen, mit Hilfe einer Anordnung von mehreren Sensoren.
Es gibt heute schon viele Anwendungen für Array Processing (Sonar, Radar, Seismologie, Astrologie, Radio, Tomografie).
Sensor Arrays können als gesamplete Aperturen (repräsentativer Schnitt) betrachtet werden. Die akustische Anwendung zur Steuerung der Richtcharakteristik von Schallwandlern ist ein relativ neuer Zweig des Gebietes. Die Theorie ist aber für alle Anwendungen gleich und basiert auf der Wellenausbreitung.

II. Wellenausbreitung im Raum

Schallwellen breiten sich longitudinal aus. Die Moleküle des Mediums bewegen sich in Ausbreitungsrichtung der Schallwelle hin und her und verursachen so regionalen Über- und Unterdruck. Man kann über die Newtonschen Bewegungsgleichungen die Wellengleichung für ein ideales Medium herleiten [A. McCowan – Microphone Arrays a Tutorial]:

$$\nabla^2 x(t,\vec{r}) - \frac{\delta^2}{c^2 \cdot \delta t^2} x(t,\vec{r}) = 0 \qquad \vec{r} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

x(t,**r**) liefert den Druck einer Schallquelle zu einem bestimmten Zeitpunkt und an einem bestimmten Ort.
Die Lösung der Gleichung ergibt für die Kugelwelle:

$$x(t,\vec{r}) = -\frac{A}{4\pi r} e^{j(\omega t - kr)}$$

$$\text{mit: } k = \frac{2\pi}{\lambda}$$

und für die ebene Welle:

$$x(t,\vec{r}) = A e^{j(\omega t - \vec{k}\vec{r})}$$

$$\text{mit: } \vec{k} = \frac{2\pi}{\lambda}\begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{bmatrix}$$

Mit der Substitution $u = t – r/c$ bei der Kugelwelle und $u = t - \beta\mathbf{r}$ ($\beta = k/\omega$) bei der ebenen Welle erkennt man, dass sich ein Signal entweder durch zeitliches Sampling an einem bestimmten Ort oder durch räumliches Sampling zu einem bestimmten Zeitpunkt vollkommen rekonstruieren lässt. Es muss allerdings die Fouriertransformation existieren, da man das Signal als aus lauter Einzelwellen zusammengesetzt betrachtet.

$$\text{Kugelwelle}: x(t,\vec{r}) = -\frac{A}{4\pi r}e^{j\omega u}$$

$$\text{ebene Welle}: x(t,\vec{r}) = Ae^{j\omega u}$$

Das räumliche Sampling bildet die Basis für jede Art von Array Processing.

III. Kontinuierliche Aperturen

Allgemein wird eine Apertur als ein Bereich betrachtet, aus dem entweder Wellen austreten oder in den Wellen eintreten.
Eine Apertur, in die Wellen eintreten wird gilt als passive Apertur. In der Akustik bezeichnet man eine passive Apertur als einen Bereich/ ein Gerät, in dem akustische Signale in elektrische überführt werden (z.B. Mikrofon).

1. Aperturfunktion

In einem infinitesimal kleinen Volumen dV (passive Apertur) an einem Ort **r** tritt zu einer Zeit t das Signal x(t,**r**) ein. Wenn man dV an der Stelle **r** als linearen Filter mit der Impulsantwort a(t,**r**) betrachtet, kann das empfangene Signal x_R(t,**r**) als Faltung von x(t,**r**) mit a(t,**r**) betrachtet werden [A. McCowan – Microphone Arrays A Tutorial].

$$x_R(t,\vec{r}) = \int_{-\infty}^{+\infty} x(\tau,\vec{r})a(t-\tau,\vec{r})d\tau$$

Oder im Frequenzbereich:

$$X_R(f,\vec{r}) = X(f,\vec{r})A(f,\vec{r})$$

Der Term A(f,**r**) wird als die Aperturfunktion oder Empfindlichkeitsfunktion bezeichnet und definiert die Systemantwort als eine Funktion der räumlichen Position auf der Apertur.

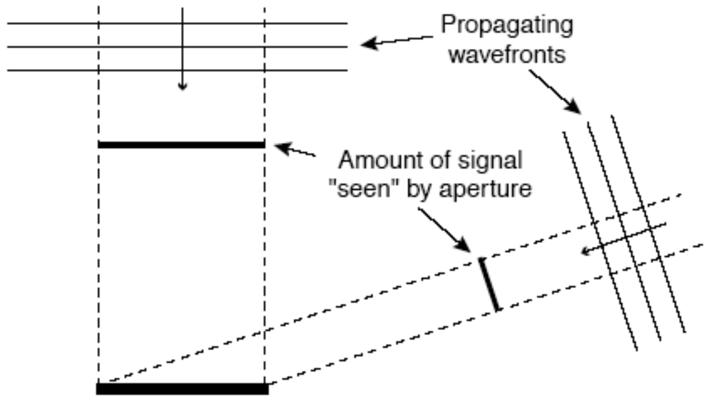
Hinweis:
Für die weiteren Betrachtungen wird nun von einer ebenen Wellenausbreitung ausgegangen, da das die erforderlichen Herleitungen extrem vereinfacht und auch in der Praxis Arrays nur selten unter Berücksichtigung der Kugelwellenausbreitung realisiert werden. Die meisten Anwendungen sind nur für Fernfeldbedingungen relevant. Zusätzlich ergeben sich meist auch für das Nahfeld nur geringfügig abweichende oder tolerierbare Ergebnisse.

2. **Directivity Pattern**

Die Systemantwort einer passiven Apertur ist direktiv und somit wird abhängig von der Einfallsrichtung des Signals nur ein gewisser Anteil des Signals „gesehen".

Die Systemantwort der Apertur als Funktion der Frequenz wird als Directivity Pattern oder Beam Pattern bezeichnet.

Die Aperturfunktion und das Directivity Pattern sind über eine Fouriertransformation verknüpft [A. McCowan – Microphone Arrays A Tutorial].

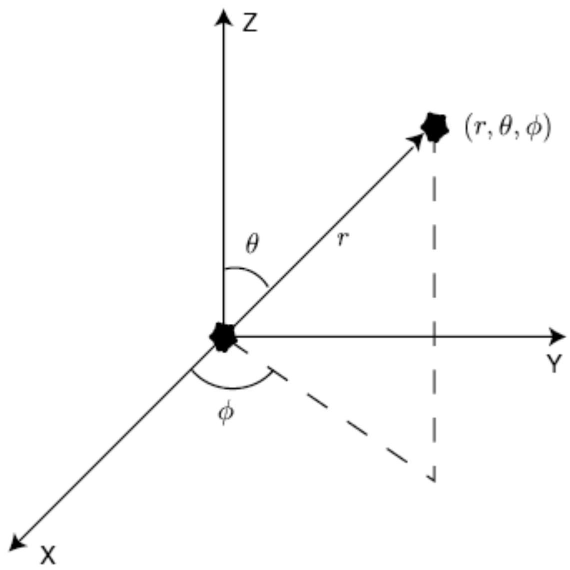


Das Fernfeld Directivity Pattern $D_R(f,\alpha)$ mit der Aperturfunktion $A_R(f,\mathbf{r})$ erechnet sich aus (F_r ist die räumliche Fouriertransformation):

$$D_R(f,\vec{\alpha}) = F_r\{A_R(f,\vec{r})\} = \int_{-\infty}^{+\infty} A_R(f,\vec{r}) e^{j2\pi\vec{\alpha}\vec{r}} d\vec{r}$$

$$\vec{r} = \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix}, \quad \vec{\alpha} = f\frac{k}{\omega} = \frac{1}{\lambda}\begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \end{bmatrix}$$

$$\lambda = \frac{c}{f}$$



r ist ein Punkt entlang der Apertur, **α** ist die Einfallsrichtung des Schalls mit dem Azimutwinkel ϕ und dem Elevationswinkel θ.
Die Frequenzabhänigkeit liegt in der Wellenlänge λ.

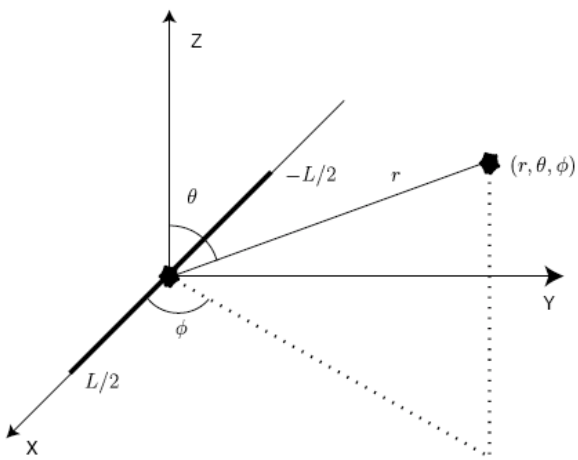
3. Lineare Apertur

Um die Eigenschaften eines Beam Patterns zu verdeutlichen wird im Folgenden eine lineare Apertur der Länge L entlang der x - Achse angenommen [A. McCowan – Microphone Arrays A Tutorial].

$$\vec{r} = \begin{bmatrix} x_a \\ 0 \\ 0 \end{bmatrix} \Rightarrow D_R(f,\alpha_x) = \int_{-L/2}^{L/2} A_R(f,x_a) e^{j2\pi\alpha_x x_a} dx_a$$

$$\rightarrow \alpha_x = \frac{\sin\theta\cos\phi}{\lambda}$$

$$\Rightarrow D_R(f,\theta,\phi) = \int_{-L/2}^{L/2} A_R(f,x_a) e^{j\frac{2\pi}{\lambda}\sin\theta\cos\phi x_a} dx_a$$



Gilt für ebene Wellen aus dem Fernfeld!
Bei linearen Aperturen wird eine Fernfeldbeziehung angenommen, wennn gilt:

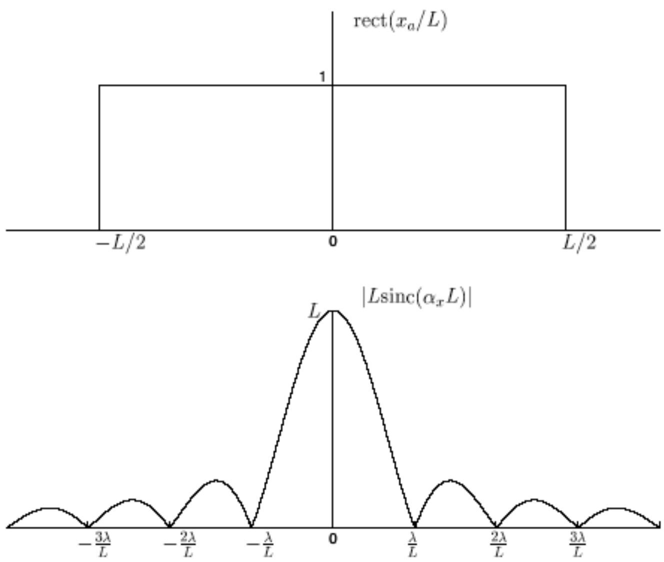
$$|r| > \frac{2L^2}{\lambda}$$

Für die weiteren Betrachtungen wird nun eine lineare Apertur, mit einheitlicher, frequenzunabhängiger Aperturfunktion angenommen:

$A_R(x_a)=\mathrm{rect}(x_a/L)$

=> $D_R(f,\alpha_x)=L\mathrm{sinc}(\alpha_x L)$

normalisiert mit 1/L und auf Winkelform gebracht:

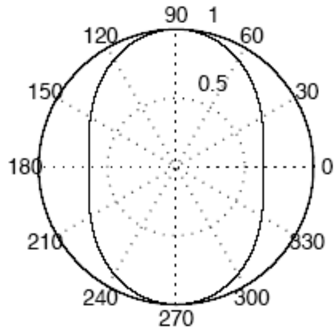


$$D_N(f,\theta,\phi) = \mathrm{sinc}\left(\frac{L}{\lambda}\sin\theta\cos\phi\right)$$

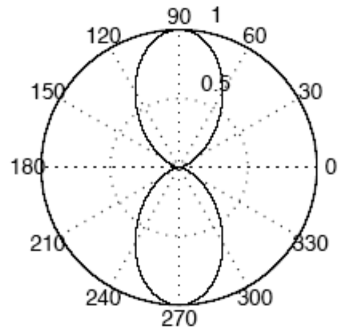
Aus diesem Directivity Pattern lässt sich nun das Polardiagramm (horizontales Beam Pattern) einer Apertur gewinnen, indem man nur den Azimutwinkel ϕ betrachtet:

$$D_N(f, \frac{\pi}{2}, \phi) = \mathrm{sinc}(\frac{L}{\lambda}\cos\phi)$$

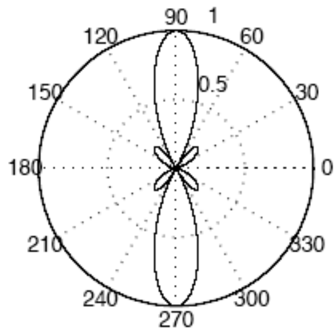
Man kann aus der Formel erkennen, dass die Nullstellenbreite (Beam Width) der sinc-Funktion von der Wellenlänge und damit von der Frequenz abhängt. Je höher die Frequenz, desto enger wird der Beam und desto mehr Nebenkeulen (Nullstellen) werden sichtbar.



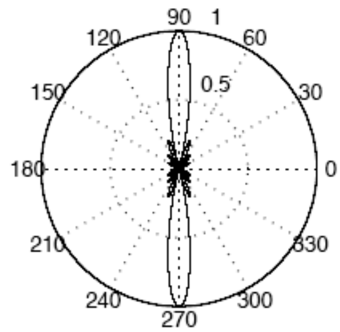
(a) $L/\lambda = 0.5$



(b) $L/\lambda = 1$



(c) $L/\lambda = 2$



(d) $L/\lambda = 4$

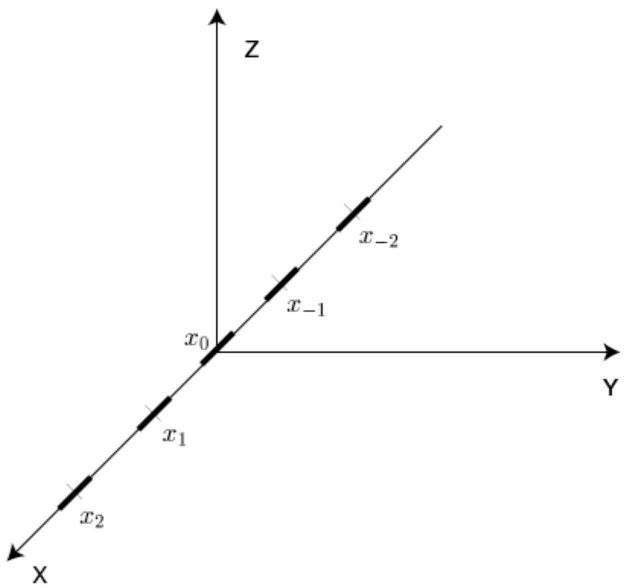
IV. Diskrete Sensorarrays

Ein Sensor Array kann als gesamplete Version einer kontinuierlichen Apertur betrachtet werden. Die Apertur wird hier nur durch eine endliche Anzahl von diskreten Punkten repräsentiert. Jeder Punkt kann selber als kontinuierliche Apertur angesehen werden. Die gesamte Antwort des Arrays ergibt sich aus der Superposition der einzelnen Antworten der Sensorpunkte und das daraus resultierende Ergebnis ist eine Aproximation an die äquivalente kontinuierliche Apertur, die gesampled wurde.

1. Lineares Array

Es wird ein lineares Array mit ungerader Sensoranzahl angenommen. Jeder Sensor n hat seine eigene komplexe Frequenzantwort $e_n(f,x)$, die zusätzlich mit einer Fensterfunktion $w_n(f)$ gewichtet wird (die Fensterfunktion tritt zu einem späteren Zeitpunk in genauere Betrachtung). Durch Superposition wird nun die komplexe Antwort des Arrays ermittelt [A. McCowan – Microphone Arrays A Tutorial]:

$$A(f,x_a) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e_n(f,x_a - x_n)$$



Zusätzlich wird $e_n(f,x)$ für alle Sensoren gleich vorausgesetzt und es kann vereinfacht werden.

Directivity Pattern für N gleiche Sensoren:

$$A(f,x_a) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)\delta(x_a - x_n) \Rightarrow D(f,\alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j2\pi\alpha_x x_n}$$

Wenn alle Sensoren den gleichen Abstand d haben ergibt sich für das horizontale Beam Pattern:

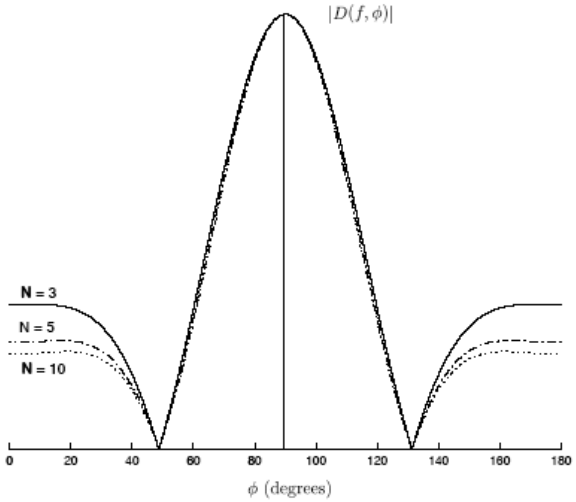
$$D(f,\alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j2\pi\alpha_x nd}$$

$$\Rightarrow D(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j\frac{2\pi}{\lambda}nd\cos\phi} = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f)e^{j\frac{2\pi f}{c}nd\cos\phi}$$

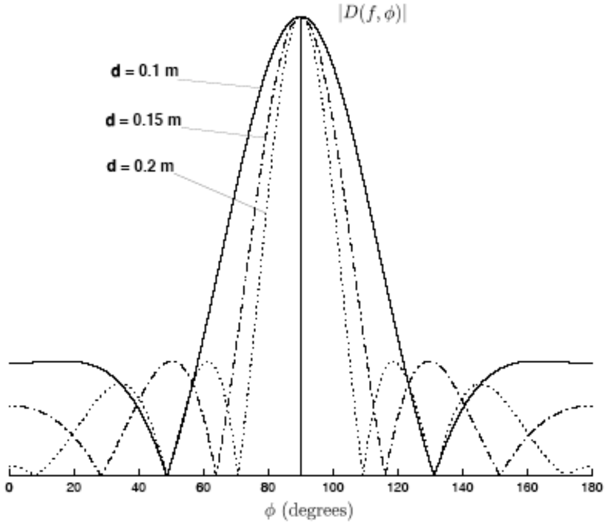
Man kann erkennen, dass das Beam Pattern von 3 Parametern abhängt:

- Die Anzahl der Sensoren N
- Der Abstand d zwischen den Sensoren
- Die Frequenz f

Die Länge der äquivalenten Apertur ist: L=Nd und die physikalische Länge des Arrays (N-1)d. Die folgenden Grafiken zeigen die jeweilige Änderung des Patterns bei der Variation der verschiedenen Parameter:



Directivity pattern for varying number of sensors (f=1 kHz,L=0.5 m)



Directivity pattern for varying effective array length (f=1 kHz, N=5)

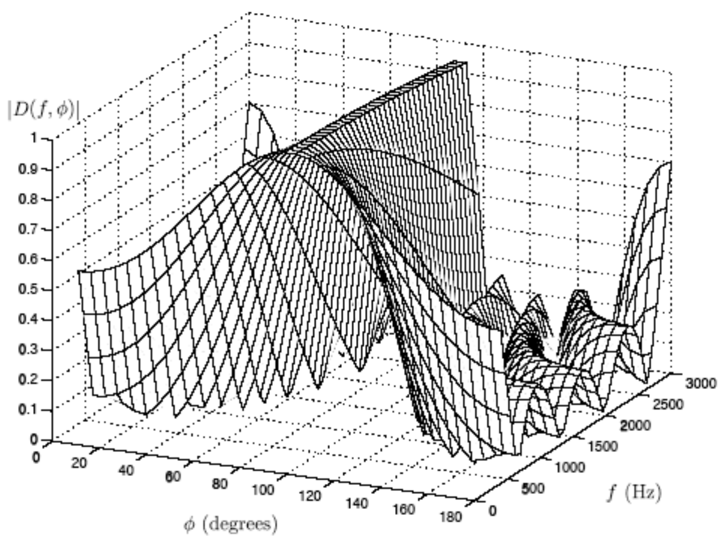
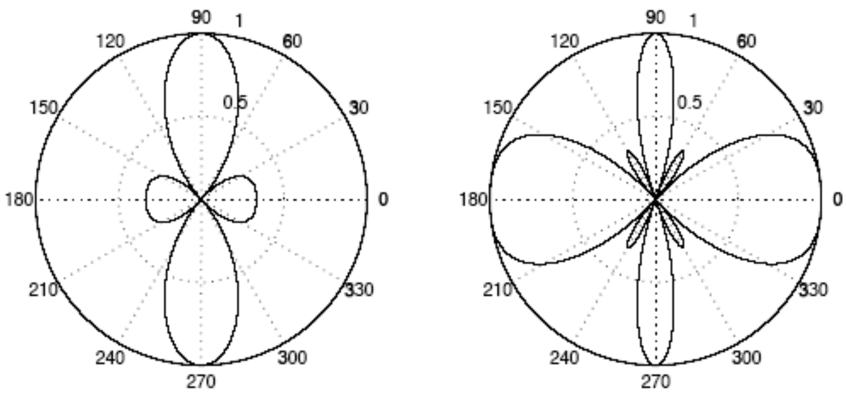


Figure 9: Directivity pattern for $400Hz \leq f \leq 3000Hz$ (N=5, d=.1 m)

2. räumliches Aliasing

Analog zum zeitlichen Aliasing, bei dem die Nyquist Frequenz die maximal abtastbare Frequenz bestimmt, gibt es beim räumlichen Sampling einen minimalen Abstand d zwischen den Sensoren, der hier das an der maximalen Frequenz f_{max} korrekt abgebildete Beam Pattern bestimmt. f_{xs} ist die räumliche Samplingfrequenz in Samples/m. Um nun kein räumliches Aliasing zu erhalten, muss der Abstand d zwischen den Sensoren der unten beschriebenen Bedingung genügen [A. McCowan – Microphone Arrays A Tutorial].

$$f_{x_s} = \frac{1}{d} \geq 2 f_{\max} \;\text{ mit }\; f_{x_s} = \frac{\sin\theta\cos\phi}{\lambda} \Rightarrow f_{x_{\max}} = \frac{1}{\lambda_{\min}} \Rightarrow d < \frac{\lambda_{\min}}{2}$$



(a) $d = \frac{\lambda_{\min}}{2}$ (b) $d = \lambda_{\min}$

3. Array Gain und Directivity Faktor

Der Array Gain beschreibt das Signal/ Rausch Verhältnis eines Arrays zwischen einem Referenzsensor und dem Ausgang des Arrays. G_d ist die aufgenommene Leistung des Nutzsignals aus der Hauptrichtung. G_n ist die mittlere Leistung aller Rauschquellen auf allen Sensoren [A. McCowan – Microphone Arrays A Tutorial].

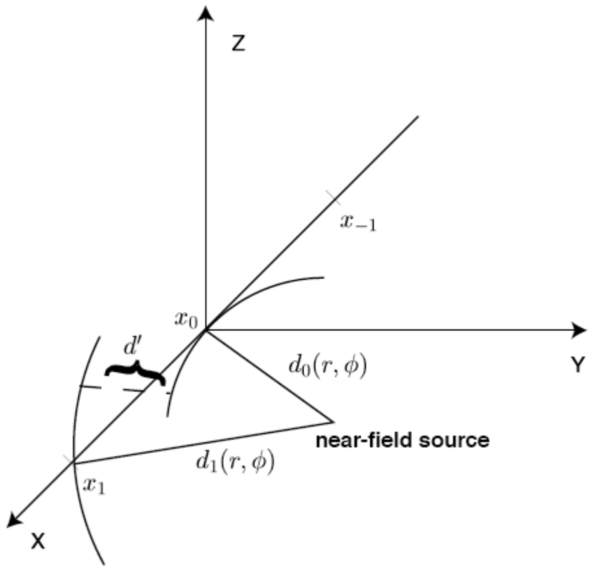
$$G_a = \frac{G_d}{G_n}$$

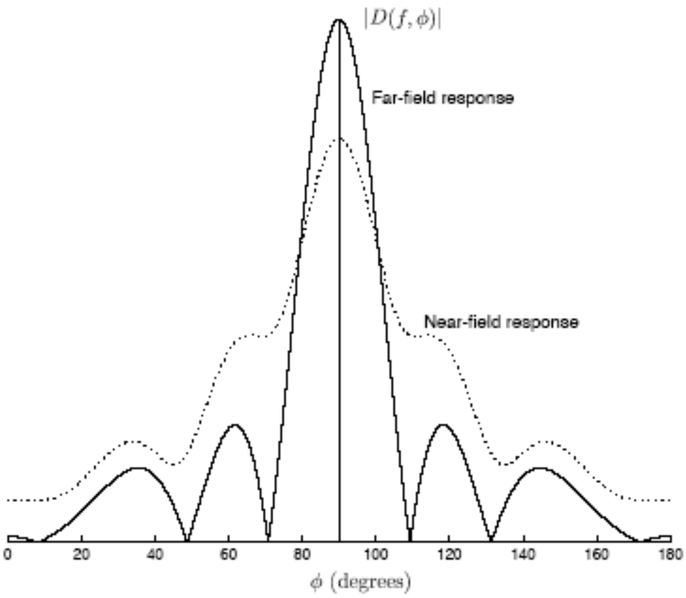
Mit einem diffusen Rauschfeld ergibt sich der Directivity Faktor: θ_0 und ϕ_0 geben die Richtung der Quelle an.

$$G_a(f,\theta_0,\phi_0) = \frac{\left|D(f,\theta_0,\phi_0)\right|^2}{\frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\left|D(f,\theta,\phi)\right|^2 \sin\theta d\theta d\phi}$$

V. Verhalten im Nahfeld

Es wird anhand zweier Diagramme kurz gezeigt, dass die oben hergeleiteten Ergebnisse nicht für das Nahfeld gelten können. Auf eine nähere Herleitung der Beziehungen wird jedoch verzichtet, da sich die Ergebnisse ohne Berücksichtigung des Nahfeldes nicht dramatisch verschlechtern [A. McCowan – Microphone Arrays A Tutorial].





Aufgrund der Krümmung der Kugelwelle verzerrt sich das Directivity Pattern, da die Phasenbeziehungen zwischen den Sensoren nicht mehr direkt sind. Es muss für eine richtige Beziehung der Abstand d’ mit berücksichtigt werden.

VI. Beamforming

1. Einführung

Erinnerung an das horizontale Fernfeld Directivity Pattern eines linearen Arrays:

$$D(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j\frac{2\pi f}{c} nd \cos\phi}$$

Bis jetzt wurde angenommen, dass die Fensterfunktion $w_n(f)$ alle Sensoren gleich gewichtet und zwar mit $w_n(f)=1/N$.
Die Fensterfunktion kann aber auch aus komplexen Gewichten bestehen [A. McCowan – Microphone Arrays A Tutorial].

$$w_n(f) = a_n(f) e^{j\varphi_n(f)}$$

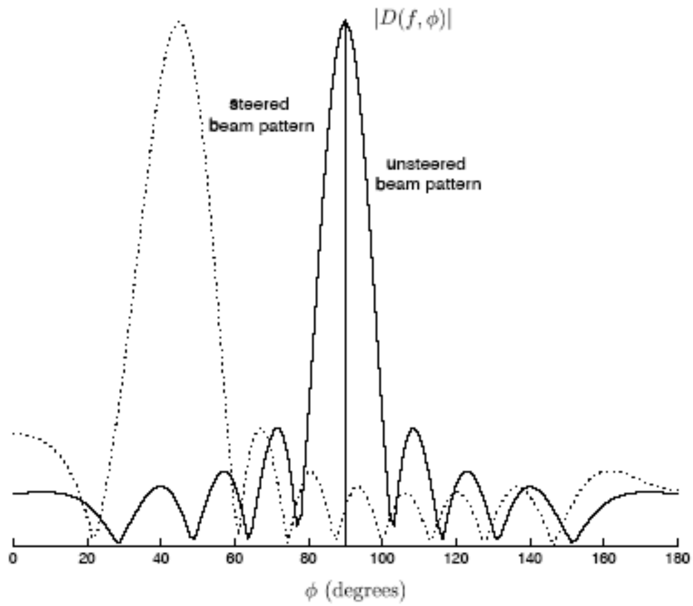
$a_n(f)$ (Amplitude) und $\varphi_n(f)$ (Phase) sind reell

Mit $w_n(f)$ ist es nun möglich die Amplitude und die Phase jedes Sensors unterschiedlich zu gewichten. Mit der Variation der Amplitudengewichte ist es möglich die Form des Beam Patterns zu verändern. Durch unterschiedliche Verzögerung der Phasen kann die Richtung der Hauptkeule variiert werden (Steering).
Das Directivity Pattern lässt sich anders anschreiben:

$$D(f,\phi) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} e^{j(2\pi\alpha_x nd + \varphi_n(f))} \text{ mit } a_n(f) = 1 \text{ und } \varphi_n(f) = -2\pi\alpha'_x nd$$

$$\alpha'_x = \frac{\sin\theta'\cos\phi'}{\lambda} \Rightarrow D'(f,\alpha_x) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} e^{j\frac{2\pi}{\lambda} nd(\alpha_x - \alpha'_x)} = D(f,\alpha_x - \alpha'_x)$$

Steering für ϕ´= 45°, f = 1kHz, N = 10, d = 0,15m



Die Betrachtung des Phasendelays im Zeitbereich zeigt, dass ein negativer Phase-Shift im Frequenzbereich eine Zeitverzögerung im Zeitbereich bedeutet (hier nur für ϕ´).
Es ist also möglich, ein Steering nur durch unterschiedliche Verzögerung der einzelnen Sensorsignale (bezogen auf einen Referenzsensor) zu erzielen. Das Delay ist äquivalent zu Laufzeit der ebenen Welle vom Referenzsensor zum n-ten Sensor.

$$\tau_n = \frac{\varphi_n}{2\pi f} = \frac{2\pi f n d \cos\phi'}{2\pi f c} = \frac{n d \cos\phi'}{c}$$

Diese Verzögerung und anschließende Summation der Sensorsignale bezeichnet man als

Delay & Sum Beamforming.

2. Delay & Sum Beamformer

Beim Delay & Sum Beamformer werden zuerst alle Sensorsignale so verzögert, dass sich die Hauptkeule in Richtung der Signalquelle ausrichtet. Danach werden alle Signale der Mikrofone addiert. Es entsteht ein Mono-Signal, dass die Information aus der Hauptrichtung betont und Störsignale aus den anderen Richtungen durch destruktive Interferenzen unterdrückt.

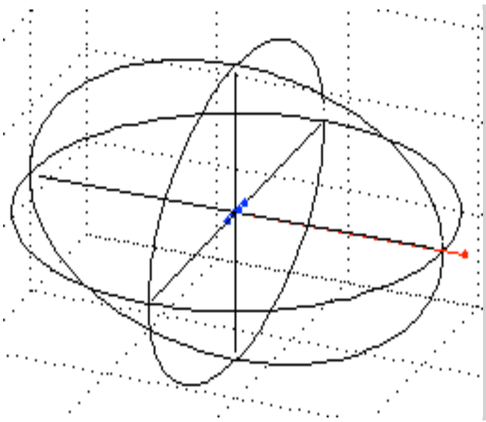
$$D(f,\phi) = \frac{1}{N}\sum_{n=1}^{N} e^{j\frac{2\pi f(n-1)d(\cos\phi - \cos\phi')}{c}} \qquad \tau_n = \frac{(n-1)d\cos\phi'}{c}$$

$$\Rightarrow Y(f) = \frac{1}{N}\sum_{n=1}^{N} X_n(f) e^{j\frac{-2\pi f}{c}(n-1)d\cos\phi'} \Rightarrow y(t) = \frac{1}{N}\sum_{n=1}^{N} x_n(t-\tau_n)$$

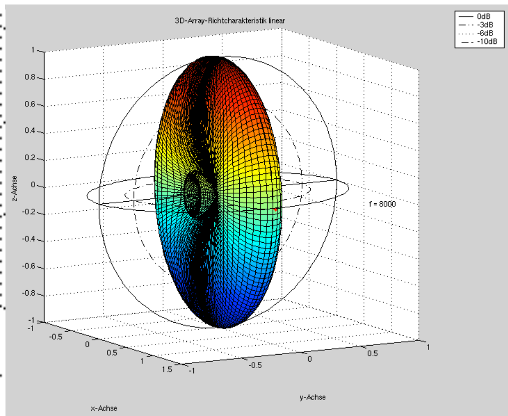
einige Matlab Simulationen:

Beispiel 1: lineares Array mit 4 Mikrofonen (d = 2 cm)

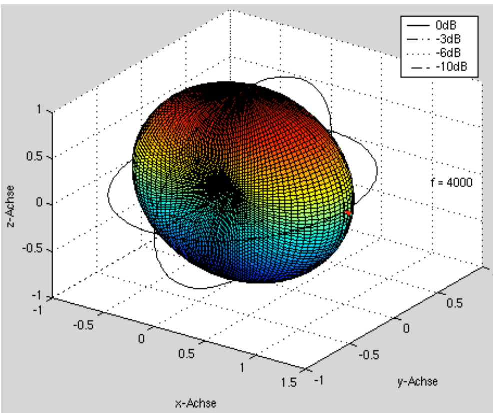
Geometrie



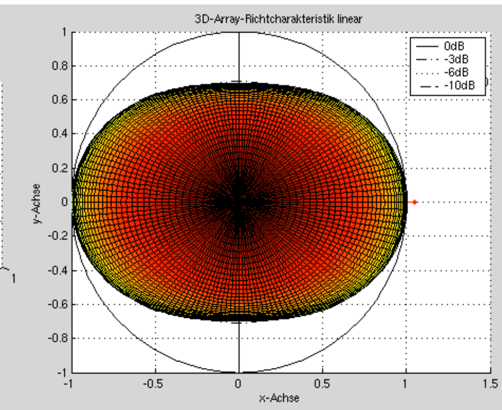
f = 8kHz



f = 4kHz

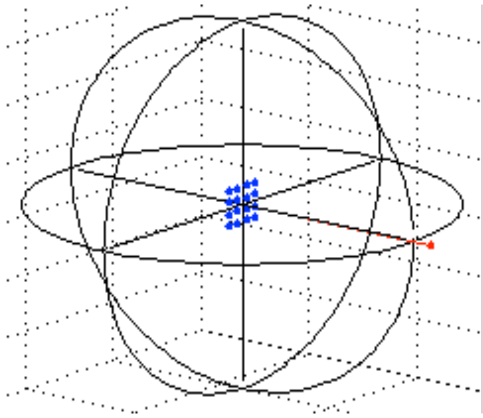


f = 2kHz

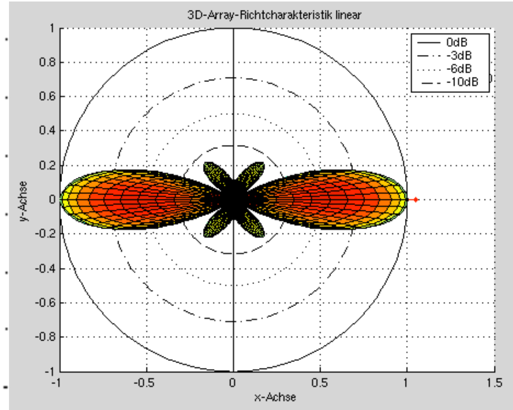


Beispiel 2: lineares, 2D-Array mit 16 Sensoren (d = 2cm)

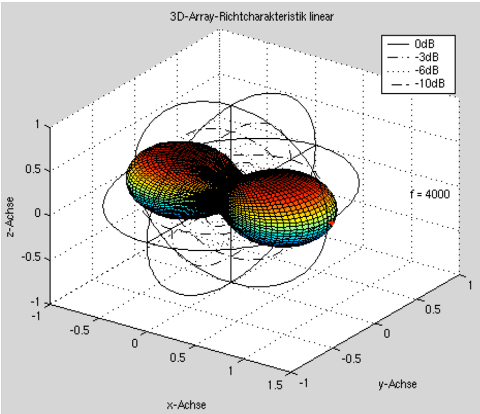
Geometrie:



f = 8kHz

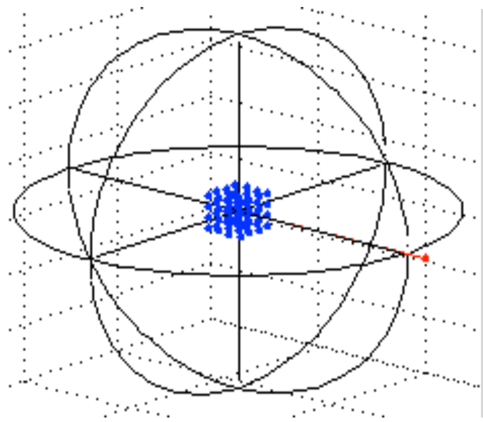


f = 2kHz

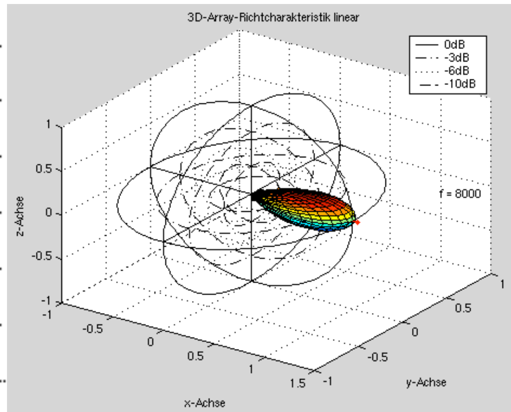


Beispiel 3: lineares 3D-Array mit 64 Mikrofonen (d = 2cm)

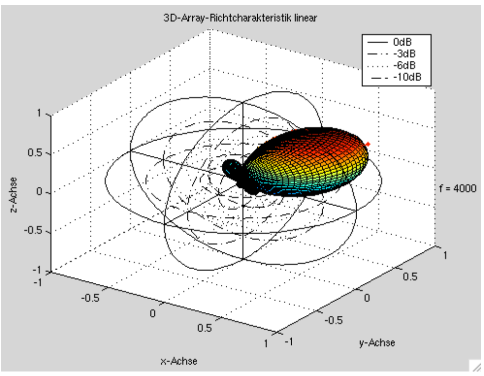
Geometrie:



f=8kHz

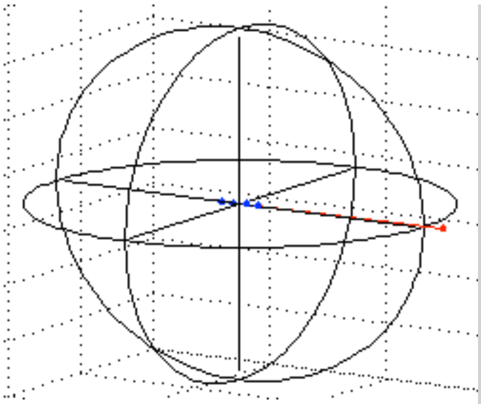


f = 4kHz mit 30° horizontalem Steering

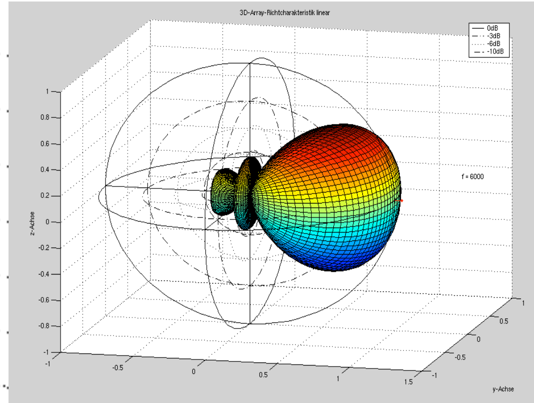


Beispiel 4: lineares Endfire-Array mit 4 Mikrofonen (d = 2cm)
Bei einem Endfire-Array ist die Bezugsrichtung in der Achse der Mikrofone.

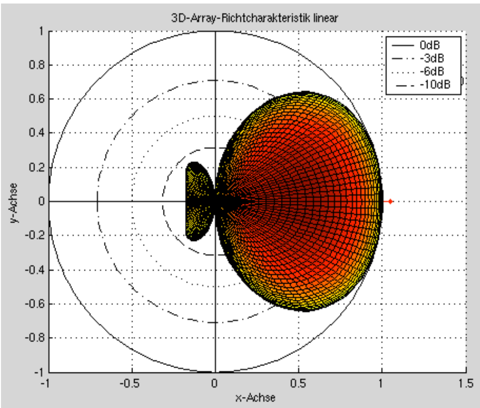
Geometrie:



f = 6kHz



f = 4kHz



VII. Der Arrayentwurf

1. Die Simulationstools in Matlab

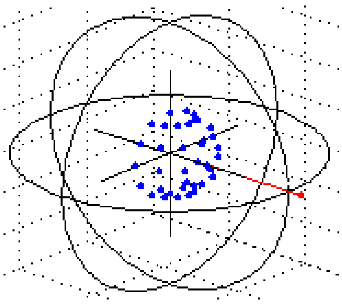
Zum Entwurf der Arraygeometrie wurde ein Simulationstool programmiert (Beam3D.m), mit dem es möglich ist, das richtungsbezogene Übertragungsverhalten eines Delay & Sum Beamformers mit beliebiger 3-dimensionaler Sensorstruktur zu simulieren. Als Simulationsergebnis erhält man einen
3-dimensionalen plot (Polardiagramm) der Richtcharakteristik des Beamformers. Es wird von Sensoren mit Kugelcharakteristik ausgegangen. Durch Modifikation dieses Patches entstanden weitere Tools (s. Dateianhänge), die es ermöglichen filmisch den Verlauf der Richtcharakteristik bei Variation verschiedener Parameter (Frequenz, Steering) darzustellen. Ein anderes Tool ermöglicht es direkt eine Tabelle mit Delays für verschiede Steering-Directions auszugeben. Die genaue Funktionsweise der Patches ist allerdings dem Quellcode und seiner Dokumentation zu entnehmen (s. Quellcodedateien auf Dokumentations-CD).

2. Entwurfspezifikationen

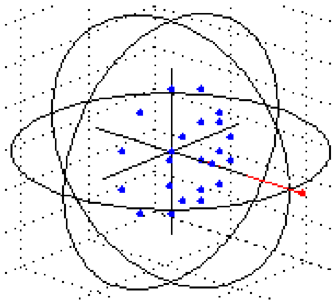
Meine Vorstellung war es, einen Beamformer zu entwickeln der sich besonders für die Sprachübertragung eignet und gleichzeitig mit möglichst wenig Mikrofonen einen möglichst breiten Steeringbereich bei geringer Arrayabmessung konstant abdeckt. Als Vorgabe habe ich mir einen Bereich von ±45° für Azimut- und Elevationswinkel gemacht. Die obere Grenzfrequenz sollte mindestens bei 6kHz liegen, was nach dem räumlichen Samplingtheorem einen maximalen Mikrofonabstand von 3cm bei einer Schallgeschwindigkeit von 340m/s bestimmt. Die untere Grenzfrequenz habe ich mir nicht festgelegt, da sie aufgrund der beschränkten Mikrofonanzahl variabel bleiben sollte, um in der geometrischen Anordnung der Mikrofone mehr Freiheiten zu haben. Die Richtwirkung sollte sich nur in eine Richtung ausprägen. Somit kamen entweder nur 3-dimensionale Arrays oder 2-dimensionale mit vorgerichteten Sensoren in Frage. Die Entscheidung fiel wegen des geringeren Materialaufwandes und der schlechteren Simulationsergebnisse mit 3D-Arrays auf letzteres. Auch wurden wegen der geforderten Eigenschaft des 2-dimensionalen Steerings nur rotationssymmetrische Arrays betrachtet, da diese bessere Eigenschaften dafür aufweisen sollten.

3. Entwurfs- und Simulationsergebnisse

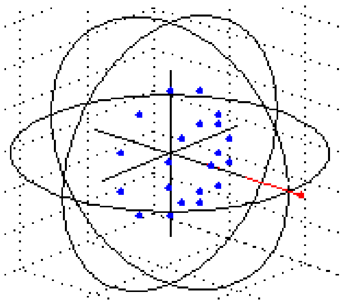
Es wurden folgende Anordnungen simuliert:



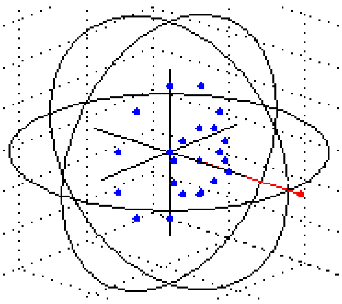
Halbkugel mit 37 Mikrofonen
R=10cm



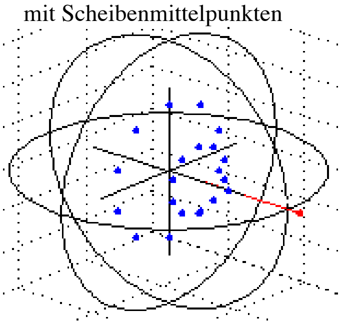
Kugelkappe mit 23 Mikrofonen
R=16cm
mit Scheibenmittelpunkten



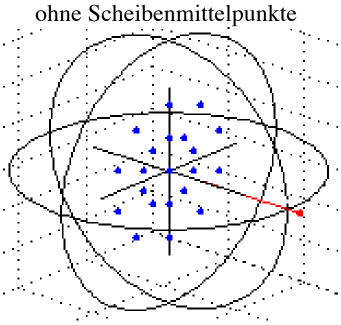
Kugelkappe mit 21 Mikrofonen
R=16cm
ohne Scheibenmittelpunkte



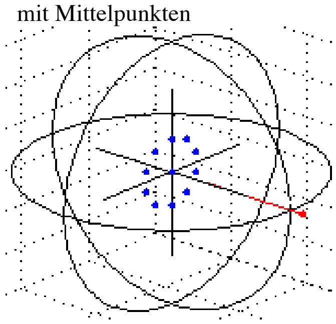
Kegel mit 23 Mikrofonen
R=h=16cm r=8cm
mit Mittelpunkten



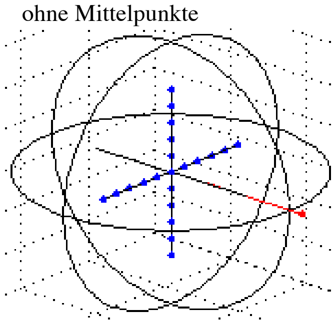
Kegel mit 21 Mikrofonen
R=h=16cm r=8cm
ohne Mittelpunkte



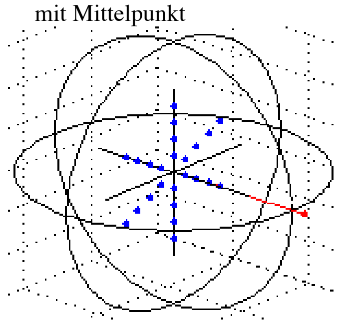
Scheibe mit 21 Mikrofonen
R=16cm r=8cm
mit Mittelpunkt



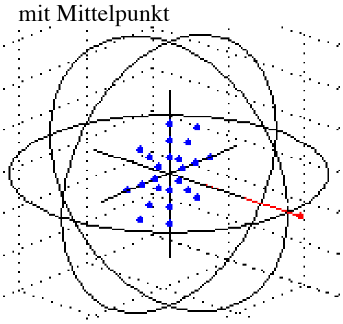
Scheibe mit 11 Mikrofonen
R=8cm
mit Mittelpunkt



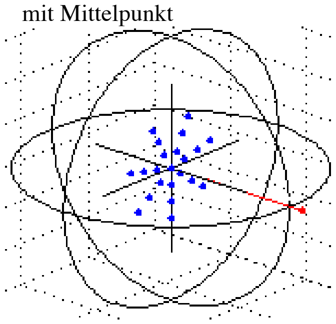
Kreuz mit 21 Mikrofonen
d=4cm =>l=40cm
mit Mittelpunkt



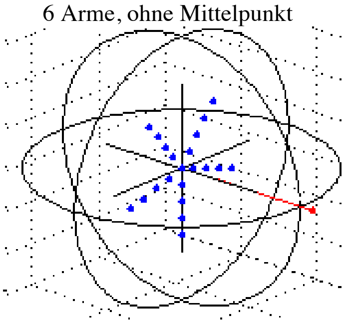
Stern mit 24 Mikrofonen
d=4cm => R=16cm
6 Arme, ohne Mittelpunkt



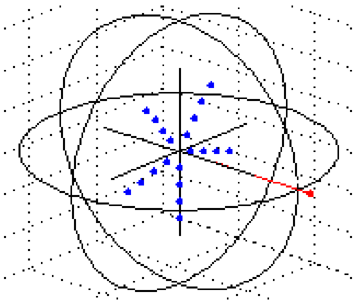
Stern mit 24 Mikrofonen
d=4cm => R=12cm
8 Arme ohne Mittelpunkt



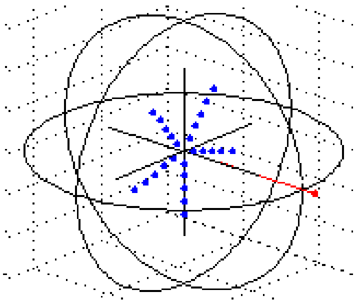
Stern mit 22 Mikrofonen
d=4cm =>R=12cm
7 Arme mit Mittelpunkt



Stern mit 21 Mikrofonen
d=4cm => R=16cm
5 Arme mit Mittelpunkt



Stern mit 20 Mikrofonen
d=4cm => R=16cm
5 Arme ohne Mittelpunkt



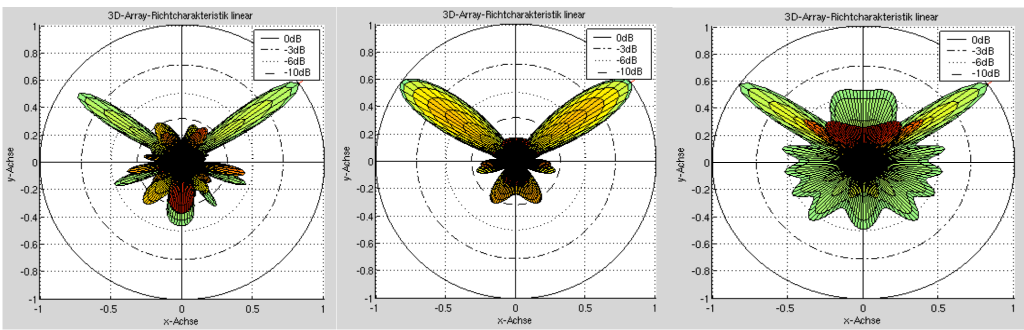
Stern mit 25 Mikrofonen
d=3cm => R=15cm
5 Arme ohne Mittelpunkt

Ergebnisse:

Die Ermittlung des am besten geeigneten Arrays erfolgte rein empirisch. Alle Geometrien sind so ausgelegt, dass die 0°-Richtung frontal einfällt (gekennzeichnet durch die rote Linie in der x-Achse). Auffallend war gleich zu Anfang, dass bei mehrdimensionalen Anordnungen der kleinste Abstand zwischen den Mikrofonen nicht die obere Grenzfrequenz bestimmt, sondern der kleinste Abstand, der in der Projektionsfläche der Einfallsrichtung herrscht. Die obere Grenzfrequenz liegt bei mehreren getesteten Arrays über 6kHz, obwohl diese einen Mikrofonabstand von >=3cm haben.

Es hat sich außerdem in den Simulationen gezeigt, dass 3-dimensionale Anordnungen zu viele Mikrofone beanspruchen und eher schlechte Ergebnisse als 2- dimensionale Anordnungen mit gleicher Mikrofonanzahl beim Steering zeigen, da sich viele Nebenkeulen ausbilden. Eine 2-dimensionale Ausdehnung lieferte in jedem Fall konstantere Ergebnisse für den vorgegebenen Steering-Bereich (Ausnahme: Kreuz).

Beispiel: Kugelkappe mit 21 Mikrofonen f=5kHz, Azimut=35°, Elevation=0° (links)
7-armiger Stern mit 22 Mikrofonen gleiche Parameter (Mitte)
Kreuz mit 21 Mikrofonen gleiche Parameter (rechts)



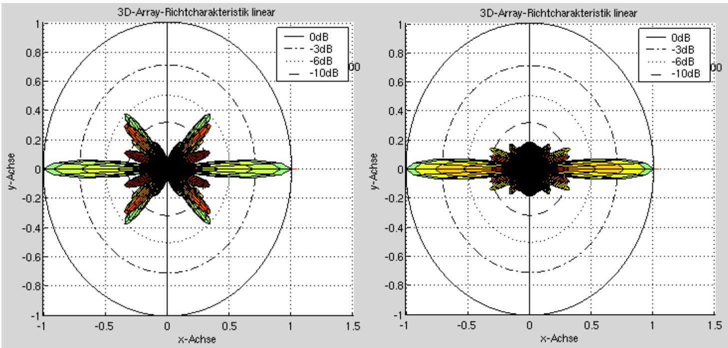
Deshalb wurden die weiteren Beobachtungen auf 2-dimensionale Arrays beschränkt.

Es wurden in den weiteren Simulationen prinzipiell 2 Ansätze für 2-dimensionale Arraygeometrien verfolgt:

- ringförmige Mikrofonanordnungen auf einer Scheibe
- sternförmige Mikrofonanordnungen auf einer Scheibe

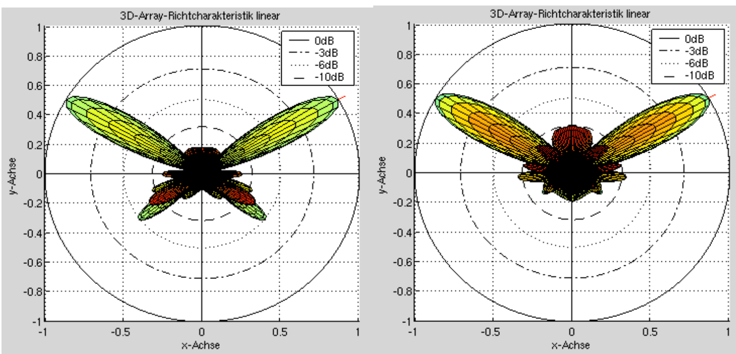
Die Simulationsergebnisse der sternförmigen Anordnung liefern deutlich höhere Grenzfrequenzen. Während z. B. eine Ringanordnung bei 6kHz schon starkes Aliasing zeigt, ist bei der Sternanordnung mit dem gleichen Radius und der gleichen Anzahl von Mikrofonen bis etwa 8kHz kein Aliasing erkennbar.

Beispiel: Ring-Scheibe mit 21 Mikrofonen, R=16cm, f=6kHz, Steering=0° (links)
5-armige Stern-Scheibe mit 21 Mikrofonen, gleiche Parameter (rechts)



Auch ist das Steering-Verhalten einer Sternanordnung merklich besser als das einer Ringanordnung.

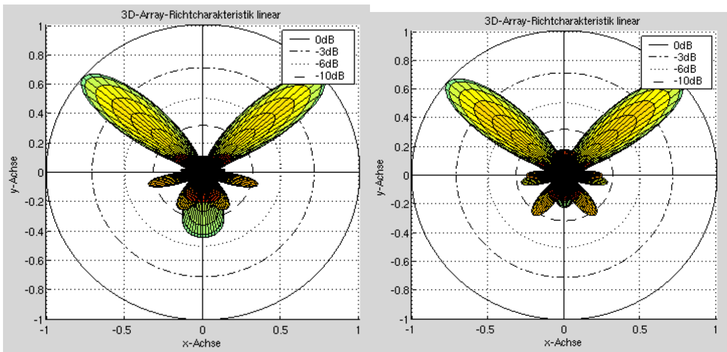
Beispiel: Ring-Scheibe mit 21 Mikrofonen, R=16cm, f=4kHz, Steering=30° (links)
5-armige Stern-Scheibe mit 21 Mikrofonen, gleiche Parameter (rechts)



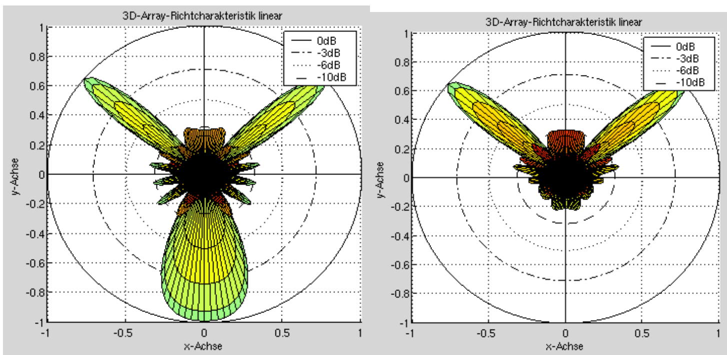
Bei der Ringanordnung sind schon bei 30° starke Sidelopes erkennbar.

Somit fallen alle weiteren Betrachtungen nur noch auf die sternförmigen Anordnungen.
Hier hat sich gezeigt, dass Aufbauten mit einer geraden Anzahl von Armen schlechtere Simulationsergebnisse beim Steering zeigen als solche mit einer ungeraden Anzahl.

Beispiel 1: 8-armiger Stern mit 24 Mikrofonen, R=12cm, f=5kHz, Steering=40° (links)
7-armiger Stern mit 22 Mikrofonen, gleiche Parameter (rechts)



Beispiel 2: 6-armiger Stern mit 24 Mikrofonen, R=12cm, f=5kHz, Steering=40° (links)
5-armiger Stern mit 25 Mikrofonen, R=15cm, gleiche Parameter (rechts)



Die besseren Ergebnisse im Vergleich zu Ringanordnungen und Sternanordnungen mit gerader Arm-Anzahl rühren vermutlich daher, dass sich bei einer Anordnung mit ungerader Arm-Anzahl keine Mikrofone direkt gegenüberliegen und somit auch nicht in der Projektionsebene. Das sorgt für ein effektiveres Spacing, bei dem die Mifrofonabstände in der Projektionsebene zur Mitte hin sehr klein werden und sich keine Mikrofon überlagert (embedded spacing). Die Geometrie sollte also eher eine ungerade Anzahl von Armen aufweisen.
Die Wahl der Geometrie fiel nach weiteren Tests auf die sternförmige Anordnung mit 5 Armen und 25 Mikrofonen. Dieser Aufbau hat die geforderten Spezifikationen am besten erfüllt. Um die Keulenbildung nach hinten zu verhindern, werden Mikrofone mit Nierencharakteristik verwendet.

Für detaillierte Simulationen des gewählten Arrays und einiger anderer Geometrien verweise ich auf die in der Dokumentations-CD beigelegten Simulationsfilme.

4. Portierung des Beamformers nach PD

Mittels des Matlab-Patches „getdelays.m“ wurde eine Tabelle erstellt, die alle erforderlichen Verzögerungszeiten der Mikrofone für alle geforderten Steering-Directions enthält. Der Steering-Bereich wurde hierbei mit einer Genauigkeit von 5° aufgelöst.
In PD ist nun ein Patch erstellt worden, der die Mikrofonsignale der gewählten Arraygeometrie so miteinander verrechnet, dass ein Delay & Sum Beamformer entsteht, mit dem es möglich ist bis zu einer oberen Grenzfrequenz von etwa 8,5kHz ±45° in Azimut- und Elevationsrichtung auszuschwenken. Dies geschieht, indem für eine eingestellte Steering-Direction die dazugehörigen Delays aus der Tabelle ausgelesen und die Mikrofonsignale damit entsprechend verzögert und summiert werden. Für die Programmiertechnischen Details bitte ich den beigelegten Patch direkt einzusehen.
Um nun eine Dialogsegmentierung erreichen zu können müssen auf das gleiche Array nun zwei dieser Beamformer geschaltet werden. Es ist dann möglich 2 Sprecher getrennt voneinander aufzunehmen und zu verfolgen. Würde man nun noch einen Spracherkennungs- und Tracking-Algorithmus implementieren, währe dieser Vorgang automatisiert möglich. Die Implementierung jener Tracking- und Erkennungsalgorithmen würde jedoch den Rahmen der Projektarbeit übersteigen und sollen somit als Ausblick vermerkt bleiben.

VIII. Bau und Vermessung des Arrays

1. Aufbau des Arrays

Das Mikrofonarray sollte eigentlich aus 25 Nierenmikrofonen AKG Q300T bestehen. Es hat sich aber herausgestellt, dass diese Mikrofone eine Tonaderspeisung haben und eine Richtcharakteristik, die 45° zur Montageebene gerichtet ist. Zusätzlich neigt sich die Richtcharakteristik schon sehr der Superniere entgegen. Aufgrund des zu hohen Konstruktionsaufwandes und der zu engen Richtcharakteristik wurde entschieden, das Array mit Kugelmikrofonen AKG Q400 zu realisieren. Der Nachteil hierbei ist die in der Simulation auch zu sehende Keulenbildung in die hintere Halbkugel. Da die einzelnen Mikrofone aber auf eine Plexiglasscheibe montiert werden, ist eine

Abschattung des von hinten einfallenden Schalls für hohe Frequenzen zu erwarten und somit eine reduzierte hintere Keule. Für tiefe Frequenzen wird eine Kugelcharakteristik erwartet.
Jedoch ist die Konstruktion mit diesen Mikrofonen für Test- und Vermessungszwecke vollkommen ausreichend.

Aufbau:
25 Phantomgespeiste Mikrofone AKG Q400 – montiert auf einer Plexiglasscheibe in der ermittelten Geometrie. Dabei wurde immer das akustische Zentrum der Mikrofonkapsel genau auf den Samplingpunkt gesetzt. Die Kabel wurden durch Löcher nach hinten abgeführt, durch einen Schrumpfschlauch zusammengefasst und am Ende mit XLR-Steckern versehen.

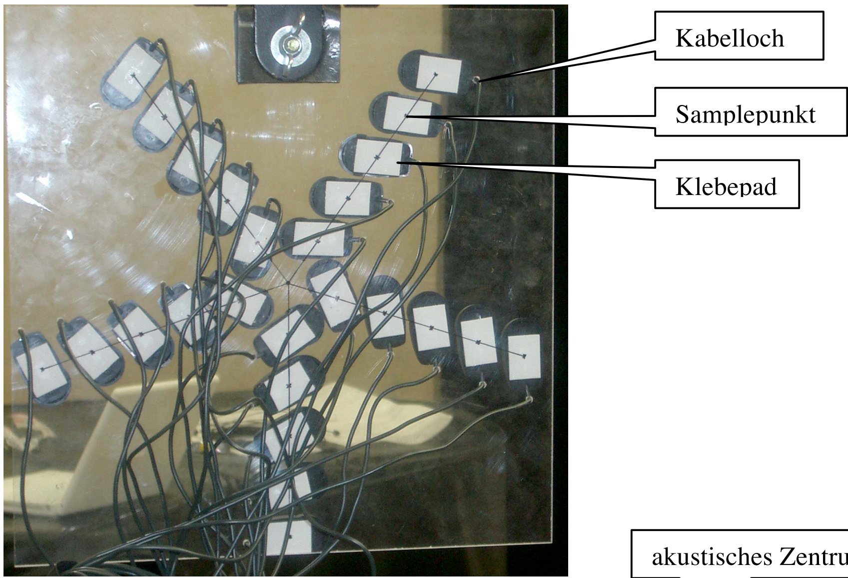


Abbildung: Array von hinten

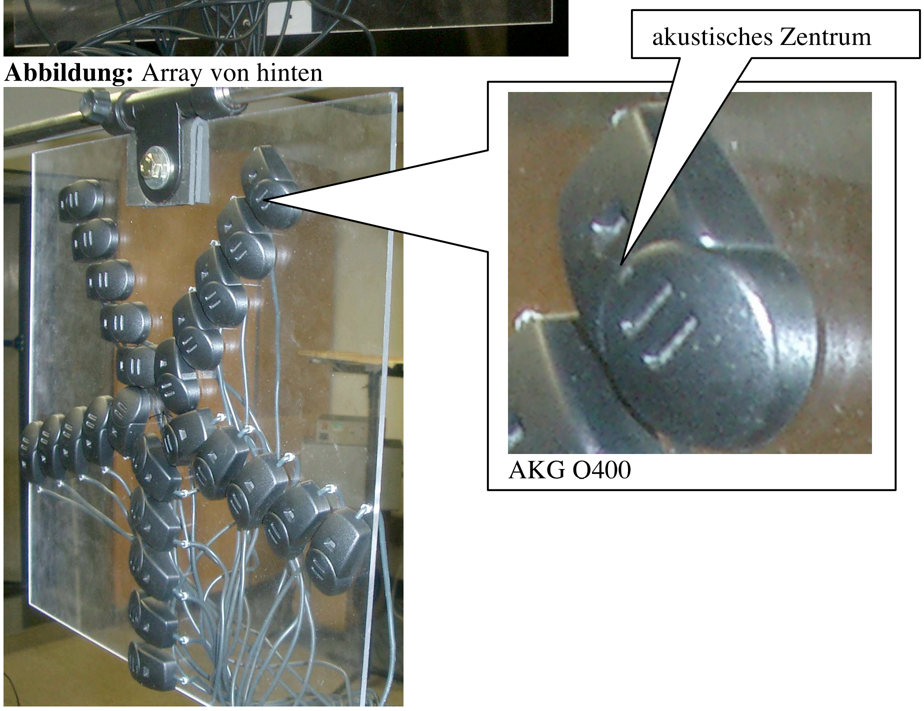


Abbildung: Array von vorne

Damit das Array richtig arbeitet, muss es genau wie in den Bildern gezeigt aufgehängt werden, da sonst die Steeringwinkel im PD-Patch nicht mehr stimmen. Die Mikrofone sind wie folgt nummeriert:

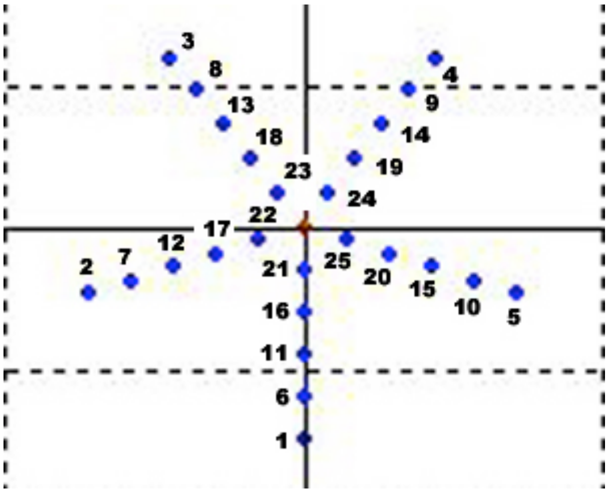


Abbildung: Nummerierung der Mikrofone

2. Messmethode und Messaufbau

Messmethode:

Die Richtcharakteristik des Arrays wurde mittels der Messmethode von Angelo Farina ermittelt.

Durch einen Sinus-Sweep, der über einen Lautsprecher abgespielt wird, kann man die Impulsantwort des Beamformers erhalten indem man eine Entfaltung der Systemantwort mit dem gesendeten Signal vornimmt.

$$y(t) = x(t) * h(t) \xrightarrow{FFT} Y(f) = X(f) \cdot H(f)$$

$$\Rightarrow h(t) = IFFT\left[\frac{Y(f)}{X(f)}\right]$$

Misst man nun die Impulsantworten rund um das Array herum, so erhält man aus den Spektren die Informationen für die Richtdiagramme über den gesamten Frequenzbereich. Eine genauere Beschreibung der Messmethode kann dem AES Paper 5093 (D-4) – Simultaneous Measurement of Impulse Response and Distorsion with a Swept-Sine Technique von Angelo Farina entnommen werden.

Messaufbau:

Gemessen wurde im reflexionsarmen Aufnahmeraum des Tonstudios am Institut für Breitbandkommunikation der technischen Universität Graz.

Die Messungen wurden mit Hilfe eines von DI. Franz Zotter erstellten PD-Patches durchgeführt. Dieser Patch vollzieht die Entfaltung nach Angelo Farina und kann am Institut für elektronische Musik und Akustik (IEM) der Universität für Musik und darstellende Kunst in Graz eingesehen werden.

Grundsätzlich wurde ,um ein besseres S/NR zu erhalten, eine Impulsantwort über 5 gemittelte Messungen ermittelt und dann als .wav File abgespeichert. Das Array war auf einem Drehteller montiert, der ebenfalls über einen PD-Patch angesteuert, das Array in 5°-Schritten sich um die eigene Achse drehen lies. Der Sweep wurde von einem tropfenförmigen Lautsprecher (besonders kugelförmige Schallabstrahlung) ausgestrahlt. Der Abstand zwischen Array und Lautsprecher betrug 1,4m. Array und Lautsprecher waren in einer Höhe von 1,25m montiert (2,5m Raumhöhe). Bei dieser Aufstellung waren wegen der sehr kurzen Impulsantworten des Arrays keine Störungen durch Raum-Reflexionen zu erwarten und die Fernfeldbedingung erfüllt. Die Mikrofonsignale gelangten über 4 Presonus Digimax

AD-Umsetzer durch ADAT Lightpipes @ 24 bit, 44,1kHz in den Messrechner. Um eine korrekte Funktionalität des Beamformers zu erreichen, wurden vor den Messungen alle Mikrofonkanäle mittels ausgespieltem weißen Rauschen auf gleiche Übertragungspegel gebracht.

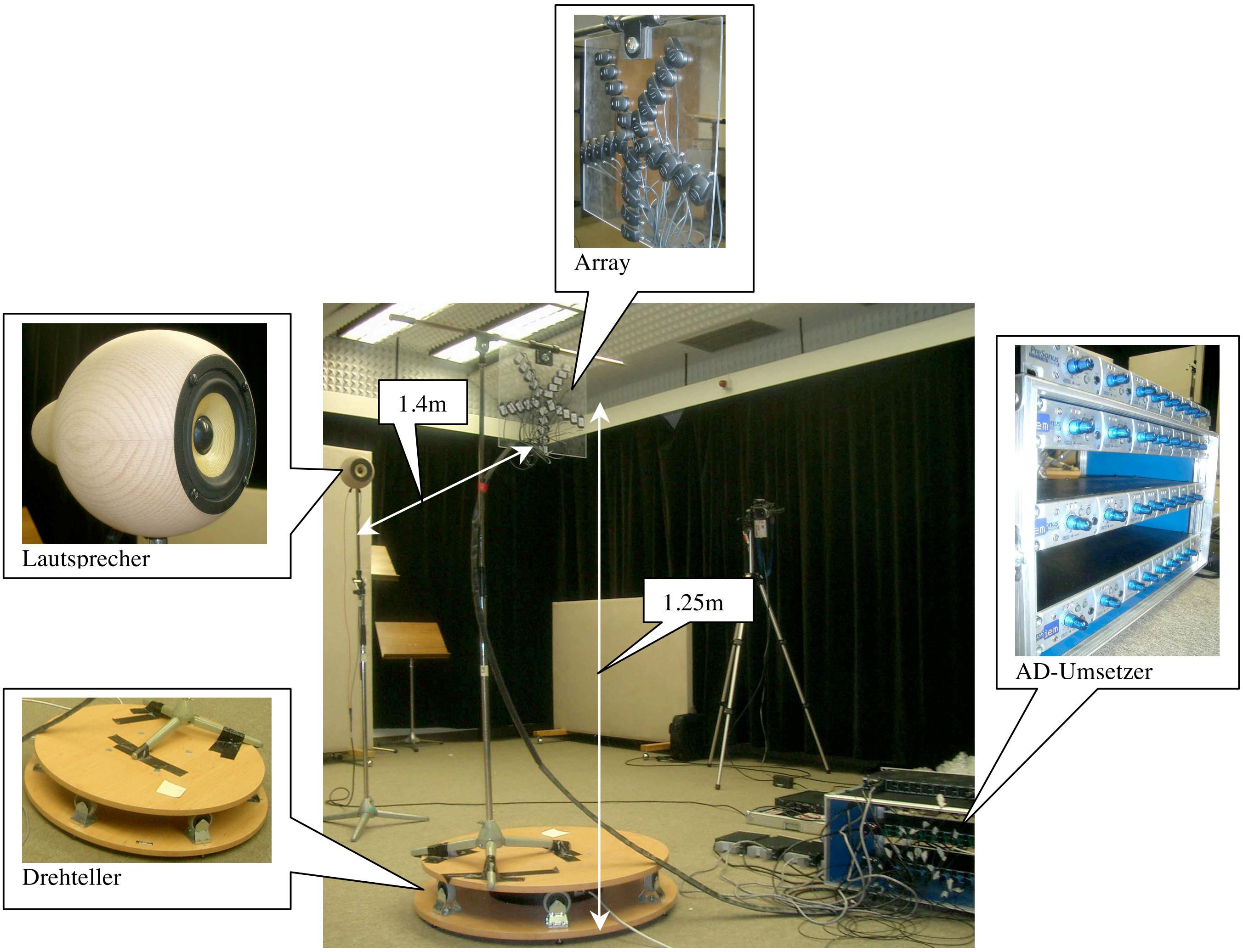


Abbildung: Messaufbau

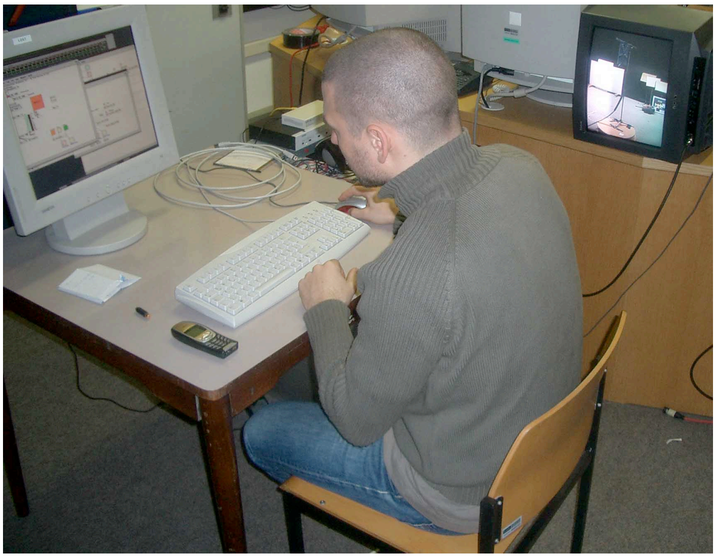


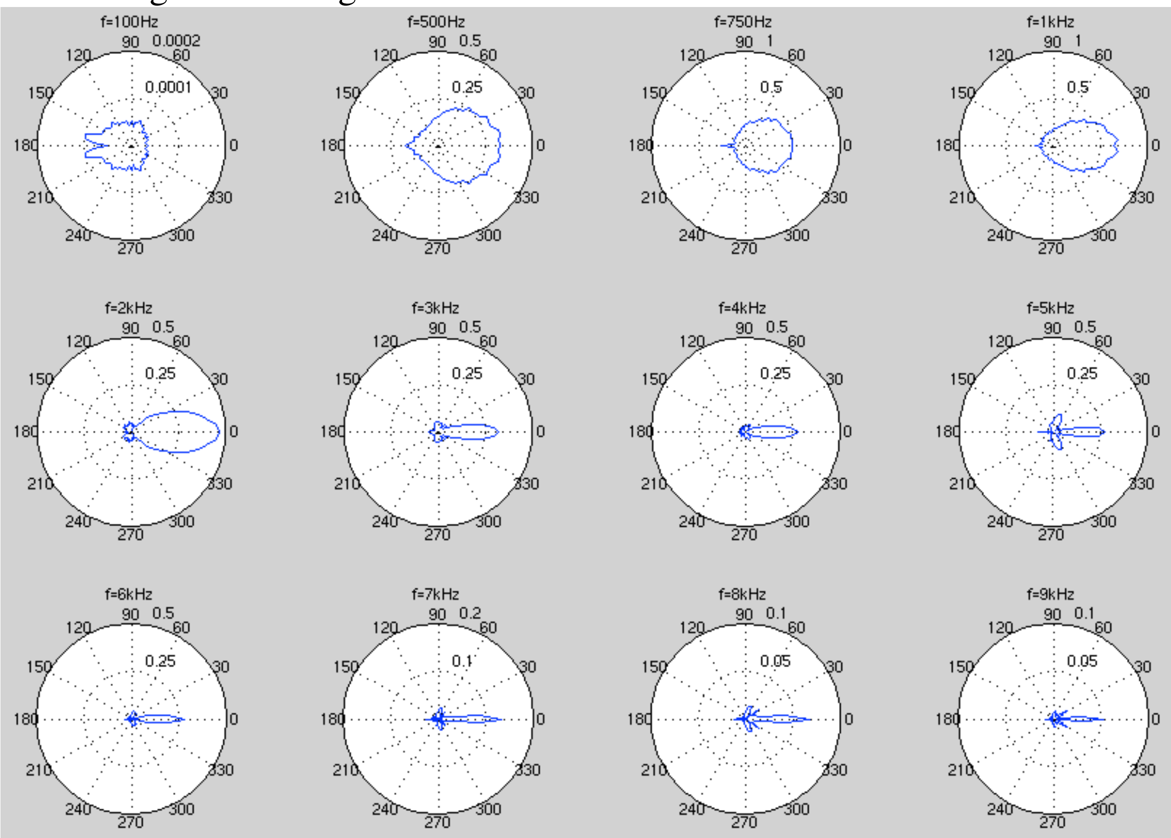
Abbildung: Messplatz

IX. Auswertung der Messergebnisse

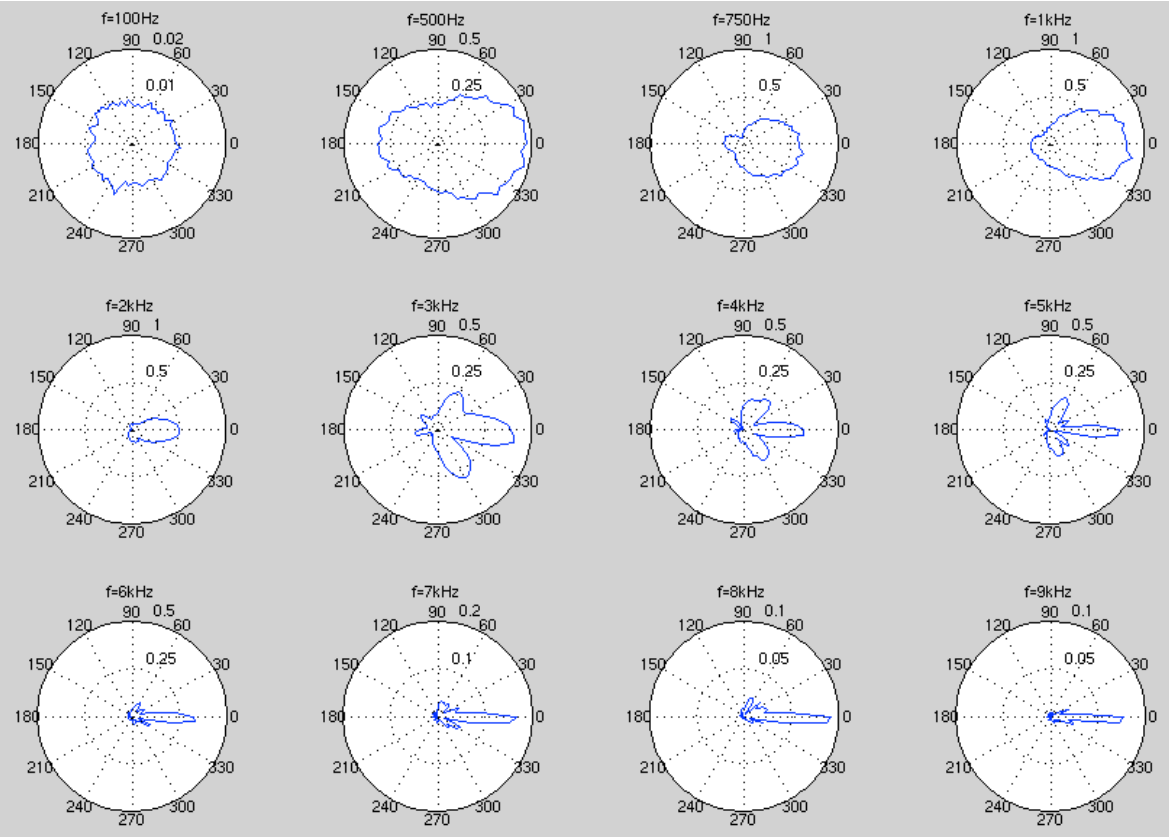
Die Richtcharakteristik des Arrays wurde für 0° Steering in Azimut und Elevation vermessen, für -30°, +45°, +15° Azimut-Steering in horizontaler Richtung und für +30° Elevations-Steering in vertikaler Richtung. Jedes Datenset besteht aus 72 Messpunkten im Abstand von jeweils 5°. Die gemessenen Impulsantworten wurden für jedes Set in MatLab importiert und in ihre Betragsspektren überführt. Man erhält die Richtdiagramme über den gesamten Frequenzbereich, indem man für jedes Diagramm die entsprechenden Frequenzpunkte in den Spektren abtastet und zusammen mit der Winkelinformation in einem Polardiagramm darstellt.

Ergebnisse:

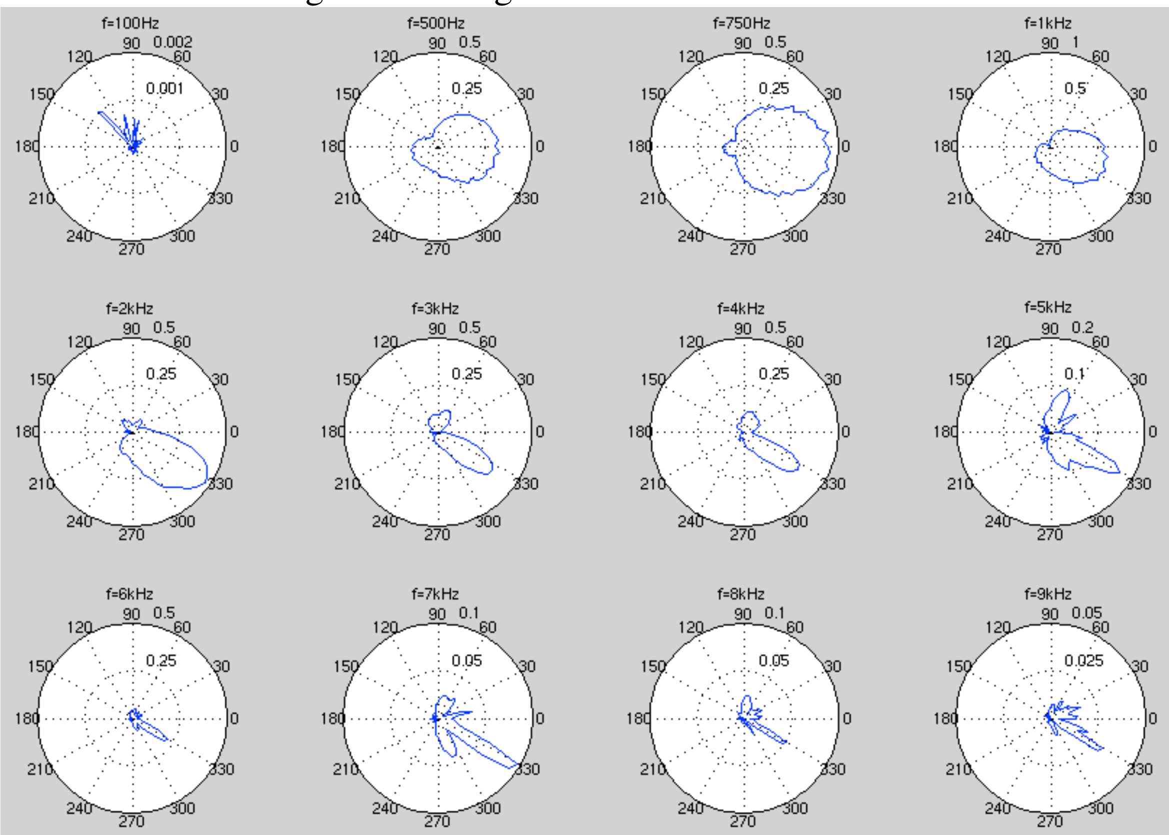
0° Steering horizontal gemessen:



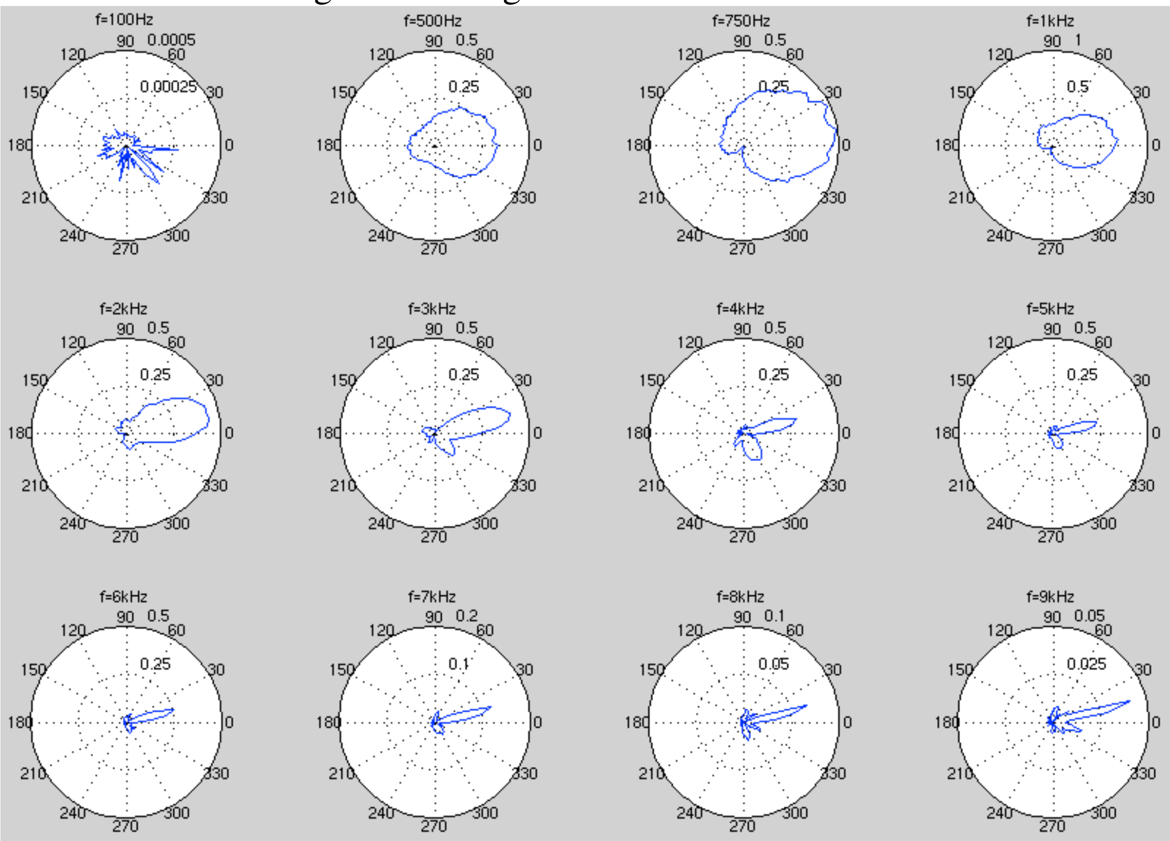
0° Steering vertikal gemessen:



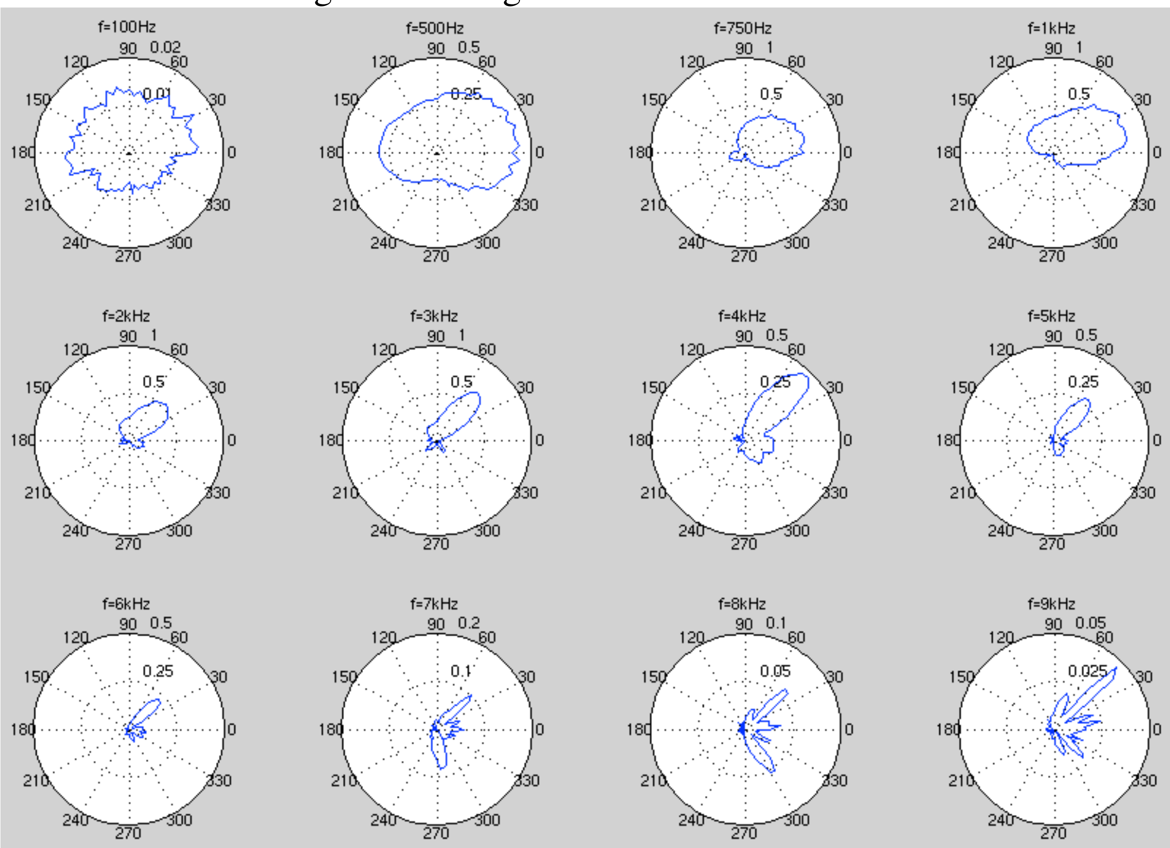
-30° Azimut-Steering horizontal gemessen:



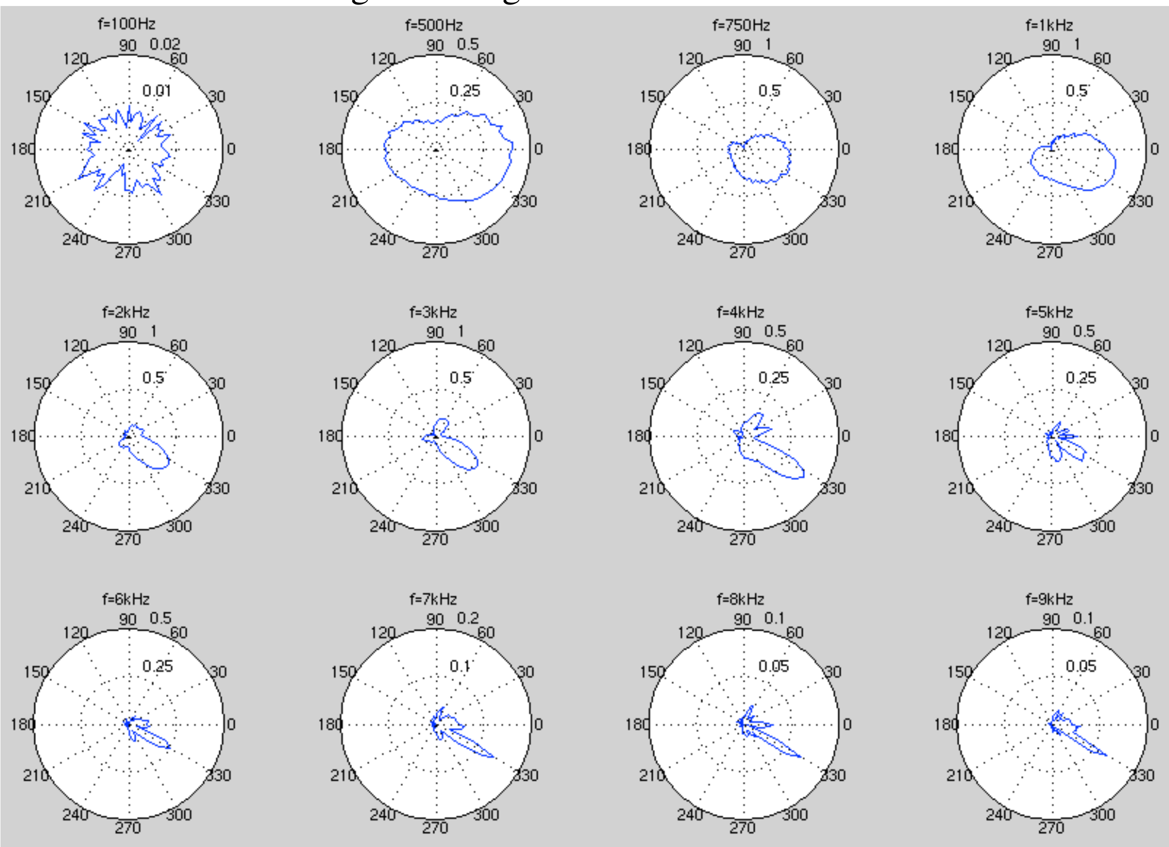
+15° Azimut-Steering horizontal gemessen:



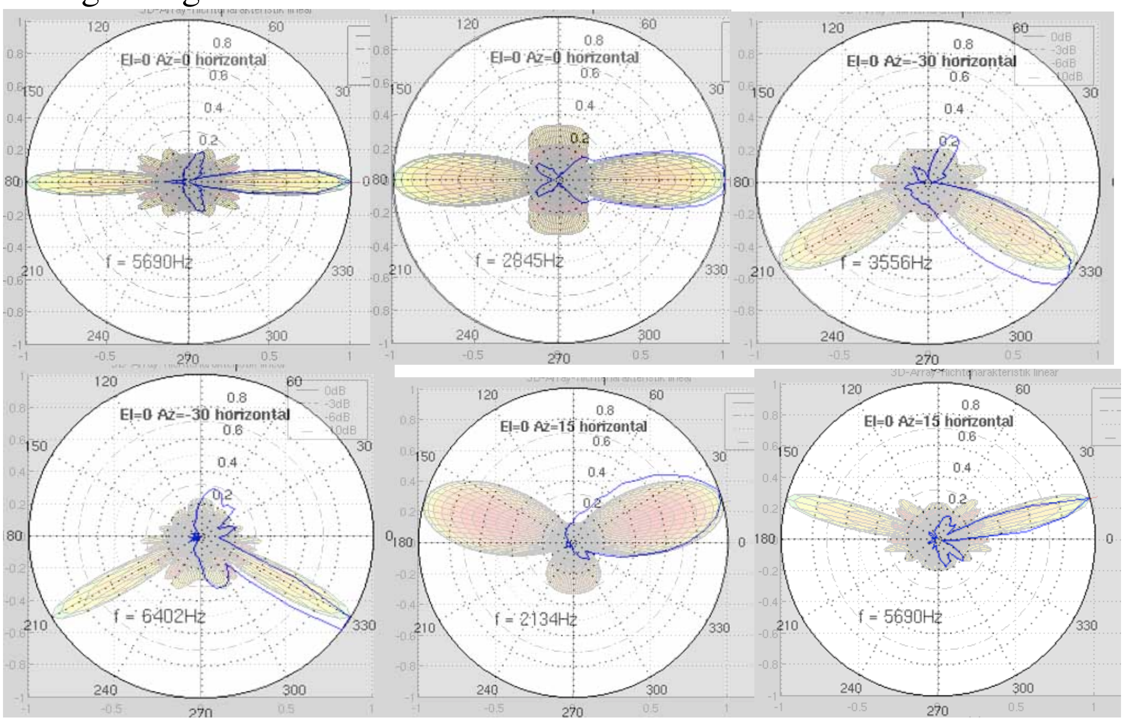
+45° Azimut-Steering horizontal gemessen:

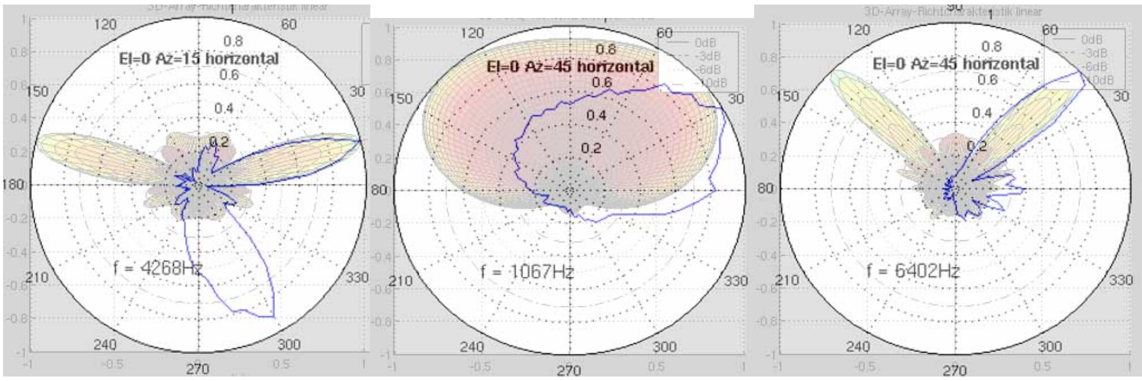


-30° Elevations-Steering vertikal gemessen:



- einige Vergleiche:



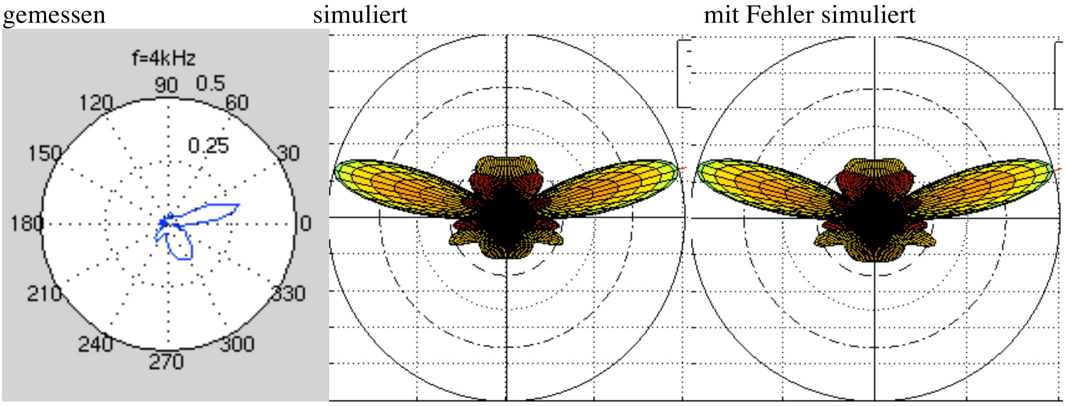


Die Diagramme zeigen nur einen Ausschnitt der Ergebnisse. Die gesamten Messergebnisse liegen als Filme auf der Projektdokumentations-CD bei. Ich möchte bemerken, dass die in den Richtplots aufscheinenden Verstärkungsfaktoren nicht zu beachten sind, da sie zum einen durch den Frequenzgang der einzelnen Mikrofone und durch Anti-Aliasing Filter im Beamformer und zum anderen durch den Messpatch einem nicht deterministischen Gewichtungsfaktor unterliegen. Die Grundformen der Richtcharakteristiken und die Nebenkeulendämpfungen können jedoch als verwertbar angesehen werden. Ziel dieser Messungen war es doch nur, die Simulationsergebnisse qualitativ zu bestätigen und die Funktionalität des Beamformers zu zeigen. Prinzipiell lässt sich sagen, dass die Messungen auf der einen Seite überraschend gute Ergebnisse bieten und auf der anderen Seite auch unerwartete Abweichungen von den Simulationen zeigen. Das grundsätzliche Verhalten der Simulation hat sich in den Messungen bestätigt. Die Keulenbreite stimmt mit der simulierten überein. Sehr auffällig ist, dass sich trotz der Kugelmikrofone keine Keule in die hintere Halbkugel ausbildet. Es scheint, als ob die Plexiglasscheibe eine ausreichende Abschirmung gegen von hinten einfallenden Schall bietet. Zusätzlich lassen sich aber im Vergleich zu den Simulationen stärkere Nebenkeulen erkennen. Besonders auf den Filmen zeigt sich immer in einem kleinen Frequenzbereich eine abhängig von der Steering-Direction mehr oder weniger starke Nebenkeulenbildung. Auch variiert dieser Frequenzbereich mit der Steering-Direction. Da das Simulationsprogramm aber als verifiziert gilt, gibt es für mich nur zwei mögliche Erklärungen für dieses Phänomen. Entweder entstehen diese Nebenkeulen durch Reflexionseffekte auf der Plexiglasplatte oder es kommt zu einer konstruktiven Interferenz, weil das im Beamformer verwendete Delay für die Mikrofonsignale nicht die benötigte Genauigkeit aufweisen kann. Dies wirkt sich dann genau so wie ein Positionierungsfehler aus.

Ich halte die Wahrscheinlichkeit des zu ungenauen Delays für größer, da sich in den Simulationen der Scheiben- und Sterngeometrien eine prinzipielle Tendenz zur Ausbildung von Nebenkeulen dieser Art gezeigt hat. So kann ein dem Positionierungsfehler gleichzusetzender Fehler zu einem Ausbruch dieser Nebenkeulen führen.
Bei einer Samplingfrequenz von 44,1kHz ergibt sich ein minimal darstellbares Delay von 1 Sample = 0,022676ms. Die kleinsten im Beamformer vorkommenden Delays sind jedoch geringer. Zusätzlich müssen die von Matlab ausgerechneten Delays aufgrund des in PD verwendeten Delay-Externals immer auf ganze Samples gerundet werden. Es liegt also nahe, dass dies der maßgebliche Grund für die Abweichungen von der Simulation ist. Leider haben sich aber nach der Einbeziehung eben dieses Fehlers in der Matlab Simulation keine besonderen Veränderungen der Richtcharakteristik gezeigt. Hierzu wurden zuerst die berechneten Delays auf ganze Samples gerundet und danach wieder in ms zurückgerechnet, sodass theoretisch der von PD gemachte Fehler in die Simulation mit einfließt.

Beispiel: 4kHz, 15° Azimut-Steering

gemessen simuliert mit Fehler simuliert



Der maximale Rundungsfehler der hier in diese Simulation mit eingeht beläuft sich auf 71,5% (sind bei allen Mikrofonen unterschiedlich) und zeigt keine ersichtlichen Auswirkungen. Dieser maximale Delayfehler würde bei diesem Array einem Positionierungsfehler von ± 2,145cm entsprechen (3cm Mikrofonabstand x 71,5%).
Es bleibt also nur zu vermuten, dass die im realen Array auftretenden Nebenkeulenbildungen durch Reflexions- und Streuungseffekte an der Plexiglasscheibe entstehen.

Schematische Rechnung:

Delays aus Simulation/0,022676ms = Sampleteile
Runden der Sampleteile = ganze Samples
Ganze Samples x 0,022676ms = mit Rundungsfehler behaftete Delays
|Delaydifferenz| = |Delays aus Simulation – mit Rundungsfehler behaftete Delays|
Fehler = (|Delaydifferenz|/|Delays aus Simulation|) x 100%

X. Ausblick

Durch die Hohe Frequenzabhängigkeit ist es bei einem reinen Delay & Sum Beamformer unmöglich einen konstanten Frequenzgang zu erhalten. Das bedeutet zum Einen, dass das Ausgangssignal entzerrt werden muss und zu Anderen, dass aus unterschiedlichen Richtungen unterschiedliche Frequenzen verschieden stark übertragen werden. Um dieses Problem in den Griff zu bekommen muss man den Delay & Sum Ansatz erweitern. Zum Beispiel könnte man durch nested Arrays die Keulenbreite anpassen oder überhaupt auf einen anderen Ansatz ausweichen, wie etwa den Minimum Varianz Beamformer. Auch braucht man für einen halbwegs breitbandigen Beamformer eine sehr große Anzahl von Kanälen. Dadurch wird der Hardwareaufwand (Mikrofone, Wandler, Soundkarten oder DSP´s?), die damit verbundenen Kosten und die benötigte Rechenleistung sehr hoch. Zusätzlich müssen die geometrischen Abmessungen des Arrays für eine genaue Abbildung der tiefen Frequenzen sehr groß sein. Das macht eine unauffällige oder überhaupt eine Anwendung sehr schwierig.
Bezüglich meines Aufbaus könnte es vielleicht zu besseren Messergebnissen kommen, wenn man die Plexiglasscheibe mit einem absorbierenden Material versehen würde. Wenn die Nebenkeulen durch Reflexionen und Streuungen an der Oberfläche zustande kämen, so könnte man dies so sicher reduzieren, da die Nebenkeulen immer in einem höheren Frequenzbereich auftreten. Es ist aber prinzipiell nicht auszuschließen, dass der Fehler an einer anderen Stelle liegt, weil auch die Messkette aufgrund ihrer experimentellen Natur bei den Messungen nicht vollkommen zuverlässig gearbeitet hat.

XI. Quellen

- Diplomarbeit Stefan Warum – Abschnitt Beamforming, erstellt und einzusehen am Institut für elektronische Musik und Akustik der Universität für Musik und darstellende Kunst in Graz.

- VDI Fortschrittbericht Nr. 429, Reihe 10, Dipl.-Ing. Carsten Sydow – selbstausrichtende Mikrofonarraysysteme, VDI Verlag, ISBN: 3-18-342910-1

- Paper von Ian A. McCowan – Microphone Arrays, A Tutorial, http://www.idiap.ch/~mccowan/arrays/tutorial.pdf

- Michael Brandstein, Darren Ward – Microphone Arrays, Techniques and Applications, Springer Verlag, ISBN: 3540419535

- Angelo Farina - AES Paper 5093 (D-4) – Simultaneous Measurement of Impulse Response and Distorsion with a Swept-Sine Technique