

Synthetisierte Aufnahmecharakteristika auf Basis spektraler Subtraktion koinzidenter Mikrofone

Bachelorarbeit aus Aufnahmetechnik 1, SE

Nils Meyer-Kahlen

Betreuung: Dr. Alois Sontacchi

Graz, 13. Juli 2016



institut für elektronische musik und akustik



Kurzfassung

Ausgehend von einem planaren, koinzidenten Mikrofonarray wird in dieser Arbeit ein Algorithmus zur Synthetisierung von steuerbaren Aufnahmecharakteristika untersucht. Mithilfe des Verfahrens lassen sich stark gebündelte Richtcharakteristika mit einem 3dB-Aufnahmewinkel von ca. 25.1° erreichen, die sich in der x-y-Ebene beliebig ausrichten lassen. Um diese Richtwirkungen zu erzielen, wird spektrale Subtraktion verwendet. Zwischenschritte sind dabei die Erzeugung von FIR-Filter in der Eingangs- und Ausgangsstufe, die Bildung eines B-Formats und dessen Weiterverarbeitung zu virtuellen Mikrofonsignalen. Die Evaluierung der erzeugten Richtcharakteristika erfolgt durch über sphärische harmonische Basisfunktionen interpolierte 3D-Darstellungen sowie SNR-Messungen. Zusätzlich wird ein Quellortungsalgorithmus vorgestellt. Eine mögliche Anwendung des Verfahrens liegt zum Beispiel bei Telefonkonferenzen, wo die Aufnahmecharakteristik an die Richtung eines aktiven Sprechers angepasst werden kann, um Störgeräusche aus anderen Richtungen zu minimieren.

Abstract

Starting from a planar, coincident microphone array an algorithm for synthesizing controllable recording characteristics is examined. Using the tested method directional characteristics with a 3 dB recording angle of approximately 25.1° can be created. These can be steered towards any angle in the horizontal plane. To achieve these directed patterns spectral subtraction is used. Apart from this, important processing steps are the generation of FIR-filters for the algorithms input and output stage, the formation of a B-format and its further processing, aimed at forming three virtual cardioid microphones. The evaluation of the results is done by presenting 3D-directivity plots with data, which has been interpolated by the use of spherical harmonic base-functions. Also SNR-measurements are shown. Furthermore an algorithm for source localization using the created recording characteristics is presented. A possible application for the system lies in teleconferencing, where it could be used to detect an active speaker and adapt to its location, thus eliminating noise from other directions.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 5 |
| 2 | Funktionsweise des Verfahrens | 7 |
| 2.1 | Aufbau des Arrays | 7 |
| 2.2 | Der Algorithmus | 8 |
| 2.2.1 | Berechnung des B-Formats | 10 |
| 2.2.2 | Berechnung dreier virtueller Nieren | 12 |
| 2.2.3 | Bildung von Summen und Differenzsignal | 13 |
| 2.2.4 | Spektrale Subtraktion | 14 |
| 3 | Messung der Impulsantworten des Arrays | 18 |
| 4 | Filterdesign | 20 |
| 4.1 | Mikrofonverzerrung | 20 |
| 4.2 | Ausgangsverzerrung | 23 |
| 5 | Auswertung | 25 |
| 5.1 | Darstellung der Richtcharakteristika | 25 |
| 5.2 | Erzeugte Aufnahmecharakteristika | 30 |
| 5.3 | Andere Evaluierungsmethoden | 34 |
| 6 | Quelllokalisierung | 37 |
| 7 | Schlussfolgerung und Ausblick | 41 |
| A | Anhang | 43 |
| A.1 | Herleitung der Normierungsfaktoren für die Berechnung des B-Formats | 43 |
| A.2 | Echtzeitimplementation in Pure Data | 45 |
| A.2.1 | Implementation | 45 |
| A.2.2 | Bedienung | 46 |

1 Einleitung

In dieser Arbeit wird ein Algorithmus zur Synthetisierung von Aufnahmecharakteristika implementiert und getestet.

In der Literatur sind viele Verfahren bekannt, um variable Richtcharakteristika zu erzeugen. Diese Verfahren werden als Beamforming bezeichnet. Zu den gängigen Methoden des konventionellen Beamformings zählen zum Beispiel Delay-and-Sum-, bzw. Filter-and-Sum-Beamforming und Differentielle Mikrofonarrays. Zu den genannten Verfahren sind Mikrofonarrays unterschiedlicher Geometrien notwendig. Das modale Beamforming ist ein Beamforming-Ansatz, bei dem eine Zerlegung in sphärische harmonische Basisfunktionen variable Beampatterns ermöglicht. Diese Methode wird in [Zau12] behandelt, wozu dasselbe Array verwendet wurde, wie in der vorliegenden Arbeit. Demnach konnten die richtungsabhängigen Impulsantwortmessungen verwendet werden, die im Rahmen dieser Arbeit gemessen wurden. Außerdem wurde auf eine Funktion zur Darstellung von über sphärische Harmonische interpolierte 3D-Richtcharakteristika zurückgegriffen.

Das generelle Ziel aller Verfahren ist eine räumliche Filterung des Schallfeldes. Bestimmte Richtungen sollen verstärkt, andere unterdrückt werden. Dies ermöglicht auch der hier vorgestellte Algorithmus. Nur ist er in der Form der erzeugten Charakteristika verhältnismäßig unflexibel. Es stellt sich allerdings die Frage, ob es in Anwendungssituationen, wie zum Beispiel einer Telefonkonferenz, notwendig ist, beliebig komplizierte Beampatterns zu erzeugen. In den meisten Situationen besteht die wichtige Aufgabe darin, eine Richtung stark hervorzuheben und die anderen möglichst stark zu bedämpfen. Dabei sollen Signalverzerrungen jeglicher Art vermieden werden und die Richtwirkung über alle Frequenzen stabil sein. Dies leistet der vorgestellte Algorithmus, wie in der folgenden Arbeit gezeigt wird.

Das Verfahren wird in [Rei11b] im Zusammenhang mit einem Array aus drei koinzident angeordneten Druckgradientenempfängern beschrieben. Die Firma AKG hat zu Forschungszwecken bereits mehrere derartige Prototypen hergestellt. Für diese Arbeit wurde ein Array mit sechs Mikrofonen verwendet.

Der Algorithmus beruht darauf, zunächst aus den Mikrofonsignalen ein flaches B-Format zu berechnen. Das B-Format stellt vier Kanäle dar, die Schallsignale aus unterschiedlichen Richtungen tragen. Der W-Kanal stellt den omnidirektionalen Schalldruckanteil dar. Die Aufnahme des X-, Y- und Z-Kanals ist schallschnelleabhängig und somit gerichtet. Die Kanäle werden durch drei orthogonale Achsen dargestellt, die in x-, y, und z-Richtung zeigen. Für den vorgestellten Algorithmus ist zunächst keine Kontrolle in z-Richtung vor-

gesehen. Daher wird das so genannte „flache“ B-Format verwendet, welches nur aus W-, X- und Y-Kanal besteht.

Durch Überlagerung und Gewichtung der B-Format-Signale werden drei Signale mit den Richtwirkungen dreier Nieren erzeugt, die jeweils um 120° zueinander geneigt sind. Schließlich werden diese unter der Verwendung von spektraler Subtraktion weiterverarbeitet. Es entsteht eine stark gebündelte Richtcharakteristik, die in der x-y-Ebene zu jedem beliebigen Winkel gedreht werden kann.

Der Aufbau der Arbeit gestaltet sich folgendermassen. Nachdem in Kapitel 2 der Aufbau des Arrays geschildert wird, findet eine detaillierte Beschreibung des Algorithmus statt. Daraufhin wird in Kapitel 3 auf die Messung der Impulsantworten eingegangen. Im nachfolgenden Kapitel 4 wird beschrieben, wie Filter erzeugt wurden, welche die Aufgabe haben, den Frequenzgang der Mikrofone aneinander anzupassen und den Frequenzgang des Ausgangs zu linearisieren. Anschließend erfolgt die Auswertung der mit dem Algorithmus erzielten Richtcharakteristika in Kapitel 5. Außerdem wird in Kapitel 6 ein Quellortungsalgorithmus präsentiert, der die Richtung einer Schallquelle erkennen kann. In Kapitel 7 befindet sich die Schlussfolgerung sowie ein Ausblick auf zukünftige Erweiterungsmöglichkeiten. Zusätzlich ist ein Anhang vorhanden, in dem sich Herleitungen sowie eine Beschreibung und Bedienunghinweise zur Echtzeitimplementation in Pure Data befinden.

2 Funktionsweise des Verfahrens

2.1 Aufbau des Arrays

Für alle folgenden Implementierungen und Messungen wurde das planare Array mit sechs Mikrofonen verwendet, welches in Abb. 1 gezeigt ist. Das Array besitzt einen zentral angeordneten Druckempfänger (Kugelmikrofon) und fünf Druckgradientenempfänger mit Nierencharakteristik. Diese sind gleichmäßig auf einem Kreis mit dem Radius $r = 2,1 \text{ cm}$ angeordnet. Somit liegen zwischen den 0° -Achsen der Mikrofone jeweils 72° . Für das Array wurden Grenzflächenmikrofone verwendet, bei denen die Schalleintrittsöffnungen, die den Schall auf die Vorder- bzw. Rückseite der Membran lenken, auf einer Seite liegen. Dadurch wird die flache Bauform des Arrays ermöglicht.

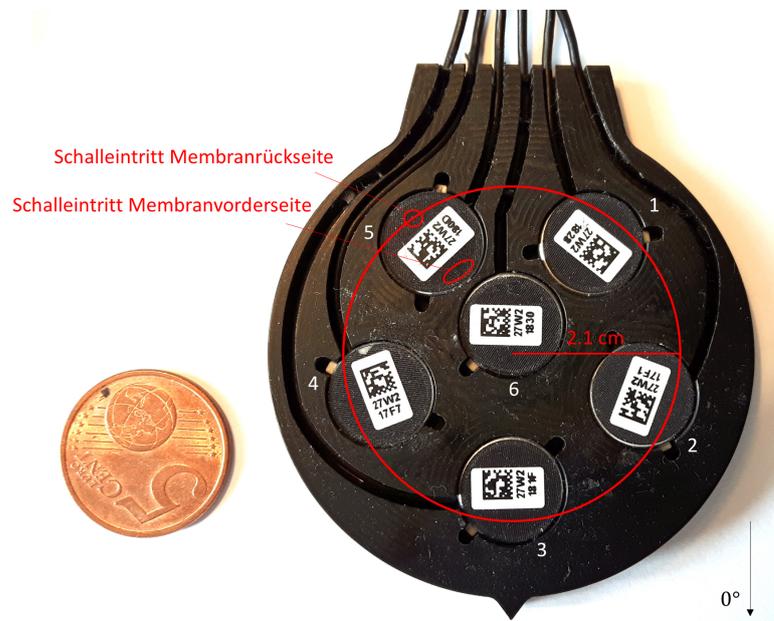


Abbildung 1: Das Array mit sechs Mikrofonen und Größenvergleich; Nummerierung der Mikrofone und Angabe der gewählten 0° -Richtung. Bei Mikrofon 5 sind die Schalleintrittsöffnungen in Richtung der Vorder- und Rückseite der Membran markiert. Die Schalleintrittsöffnungen zur Membranrückseite liegen auf einem Kreis dem Radius $r = 2.1 \text{ cm}$. Die Schalleintrittsöffnung zur Membranrückseite des Kugelmikrofons liegt im Zentrum des Kreises.

Die Grundvoraussetzung für den implementierten Algorithmus ist die Koinzidenz der Mikrofone. Im optimalen Fall sollten sich bei einem koinzidenten Mikrofonarray die akustischen Zentren¹ der Mikrofone am gleichen Ort befinden.

1. Das akustische Zentrum ist der Ort, von dem Kugelwellen auszugehen scheinen, wenn der Wandler als Quelle betrieben wird. [JFR04]

Da dies aufgrund der Konstruktion nicht unbedingt sichergestellt werden kann, lässt sich mit der Bedingung arbeiten, dass die akustischen Zentren in einer Kugel liegen, deren Radius nicht größer als die doppelte größte Membranabmessung ist [Rei11a]. Durch näher aneinander liegende akustische Zentren verbessern sich die Ergebnisse. Die Bündelung bleibt bis hin zu höheren Frequenzen erhalten. Neben dem vorgestellten Mikrofonarray eignet sich im Allgemeinen jede Mikrofonanordnung, aus dessen Ausgangssignalen sich ein flaches B-Format (W-, X- und Y-Kanal) berechnen lässt, um den beschriebenen Algorithmus zu verwenden. Dass die vorliegende Anordnung gegenüber anderen Möglichkeiten, speziell dem Soundfield-Mikrofon (wie zuerst in [CG77] vorgestellt), gewisse Vorteile aufweist, wird in [Rei11b] beschrieben.

2.2 Der Algorithmus

Das folgende Kapitel behandelt die Grundelemente des getesteten Algorithmus.

Die **Berechnung des flachen B-Formats (2.2.1)** aus den Eingangssignalen ermöglicht eine beliebige Rotation der Richtcharakteristik und wird für die **Berechnung dreier virtueller Nieren (2.2.2)** benötigt. Diese drei Signale entsprechen dreier um 120° geneigter Nieren und stellen die Basis für die **Bildung von Summen- und Differenzsignal (2.2.3)** dar. Summen und Differenzsignal werden dann durch den entscheidenden Schritt der **Spektralen Subtraktion (2.2.4)** voneinander betragsmäßig subtrahiert, was eine stark gebündelte Charakteristik erzeugt. Alle diese Schritte sind im Blockschaltbild in Abb. 2 erkennbar. Die FIR-Filter (Finite-Impulse-Response-Filter) werden separat in Kapitel 4 behandelt.

Obwohl die ersten Verarbeitungsschritte bis hin zur spektralen Subtraktion theoretisch auch im Analogen geschehen könnten, wird hier angenommen, dass die Mikrofonsignale bereits digitalisiert wurden.

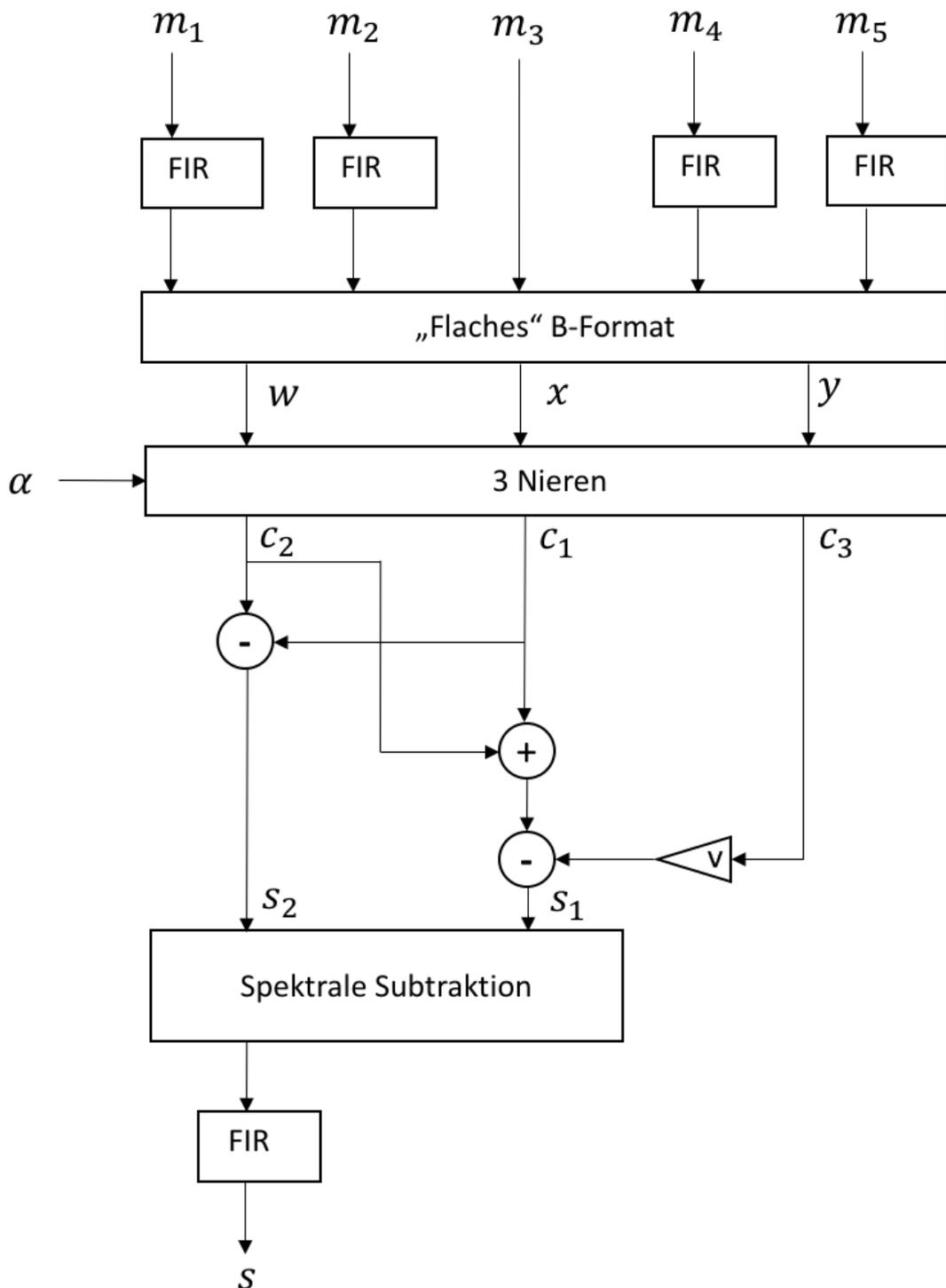


Abbildung 2: Blockschaltbild des Algorithmus, bestehend aus FIR-Filtern zur Eingangs-entzerrung, der Berechnung des flachen B-Formats, der Bildung dreier virtueller Nieren, der Bildung von Summen- und Differenzsignal, spektraler Subtraktion und einem FIR-Filter zur Ausgangsentzerrung.

2.2.1 Berechnung des B-Formats

In [Rei11a] wurde ein Algorithmus vorgestellt, der es ermöglicht, eine stark gerichtete Charakteristik aus den Signalen dreier um jeweils 120° zueinander geneigten Nierenmikrofone zu synthetisieren. Die Patentansprüche beziehen sich hierbei auf eine Anordnung von zwei bis drei derartig verwendeten Mikrofonen. [Rei11b] bezieht sich auf eine Anordnung aus mindestens drei Mikrofonen. In diesem Patent wurde ein Schritt am Anfang der Signalverarbeitung hinzugefügt, der die Flexibilität deutlich erhöht. Aus den tatsächlichen Mikrofonensignalen wird ein so genanntes flaches B-Format berechnet und aus diesem dann wieder die Signale dreier um 120° geneigter Nieren. Der Vorteil in der Bildung des flachen B-Formats liegt darin, dass die Hauptrichtung der erzeugten Charakteristik durch Variation eines einzigen Parameters α beliebig rotiert werden kann. Die möglichen Richtungen, in welche die Hauptkeule des gesamten synthetisierten Signals zeigen kann, ist somit nicht mehr wie in [Rei11a] auf sechs limitiert, sondern kann beliebig auf einem Kreis gewählt werden.

Die Berechnung des flachen B-Formats funktioniert wie folgt. Die Eingangssignale der einzelnen Mikrofone werden in Abhängigkeit ihrer geometrischen Lage und Ausrichtung addiert bzw. subtrahiert, um räumlich unabhängige Signale zu erhalten. Darauf folgt die Normierung der Richtcharakteristika. Die fünf Druckgradientenempfänger besitzen jeweils näherungsweise eine Nierencharakteristik. Eine Nierencharakteristik setzt sich aus einem omnidirektionalen Anteil und einem Anteil mit Achtercharakteristik zusammen. Eine solche Überlagerung lautet

$$\Gamma(\varphi) = \frac{a + b \cdot \cos(\varphi)}{a + b}, \quad (1)$$

wobei $\Gamma(\varphi)$ Richtungsfaktor genannt wird.

Die Konstanten a und b bestimmen die Form der Richtcharakteristik. Für normale Nieren, wie sie hier angenommen werden, mit einem 3 dB-Winkelbereich von $\pm 65.1^\circ$ und einer theoretischen Rückwärtsdämpfung von $-\infty$ dB gilt $a = b = 1$. Zu beachten ist, dass alle idealen Formeln des Richtungsgrades tatsächlich nur in einem sehr kleinen Frequenzbereich eine gute Annäherung an das tatsächliche Mikrofonverhalten liefern. Der Frequenzbereich liegt üblicherweise im Bereich um 1 kHz.

Durch Überlagerung der Mikrofon-signale $\{m_1[n], \dots, m_5[n]\}$ wird nun das flache B-Format berechnet. Die Signale $w[n]$, $x[n]$ und $y[n]$ stellen den W-, X- und Y-Kanal dar. Das erste Signal, das des W-Kanals, stellt einen rein omnidirektionalen Anteil dar und entsteht aus der Überlagerung der Signale aller Nierenmikrofone. Im vorliegenden Fall könnte dazu auch das mittig angeordnete Kugelmikrofon verwendet werden. Um zu zeigen, dass der Algorithmus in der vorliegenden Implementierung aber nicht davon abhängt, ein zusätzliches Kugelmikrofon zur Verfügung zu haben, wurde mit der minimalen Anzahl von Mikrofonen gearbeitet.

$$w[n] = \frac{m_1[n] + m_2[n] + m_3[n] + m_4[n] + m_5[n]}{2.5} \quad (2)$$

Der X-Kanal entspricht einem Mikrofon mit Achtercharakteristik, bei welchem die Richtung positiver Phase nach 0° zeigt. Es wurde festgelegt, dass 0° der Richtung des 3. Mikro-fons entspricht (siehe Abb. 1). Zur Berechnung des X-Kanals sind bei dieser Anordnung drei Mikrofon-signale notwendig.

$$x[n] = \frac{(m_3[n] - m_1[n]) + (m_3[n] - m_5[n])}{1 - \cos(\frac{4\pi}{5})} \quad (3)$$

Zur Berechnung des Y-Kanals ist die Differenzbildung zweier Mikrofon-signale notwendig. Der Y-Kanal steht orthogonal zum X-Kanal und stellt ebenfalls eine Acht dar.

$$y[n] = \frac{(m_2[n] - m_4[n])}{\sin(\frac{2\pi}{5})} \quad (4)$$

Die Normierungsfaktoren $c_w = 2.5$, $c_x = 1 - \cos(\frac{4\pi}{5})$ und $c_y = \sin(\frac{2\pi}{5})$ werden berechnet, indem die winkelabhängigen Funktionen des Richtungsgrades der Mikrofone in die Vorschriften zur Berechnung von W-, X- und Y-Kanal eingesetzt werden und ein Koeffizientenvergleich mit der gewünschten Charakteristik durchgeführt wird. Für eine genaue Herleitung wird auf den Anhang A.1 verwiesen. Abb. 3 zeigt das normierte flache B-Format.

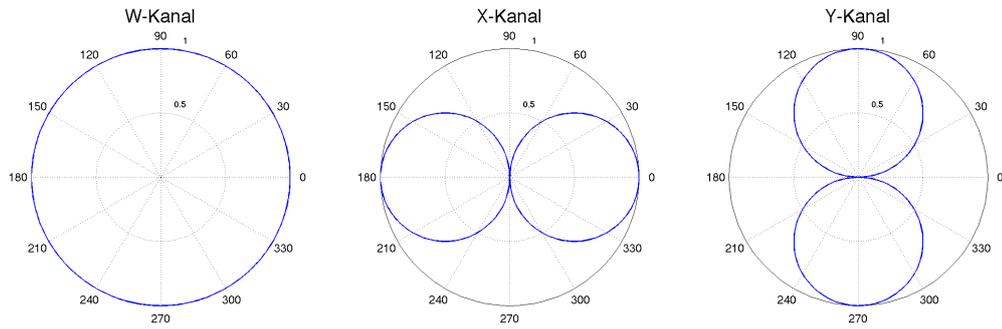


Abbildung 3: Flaches B-Format

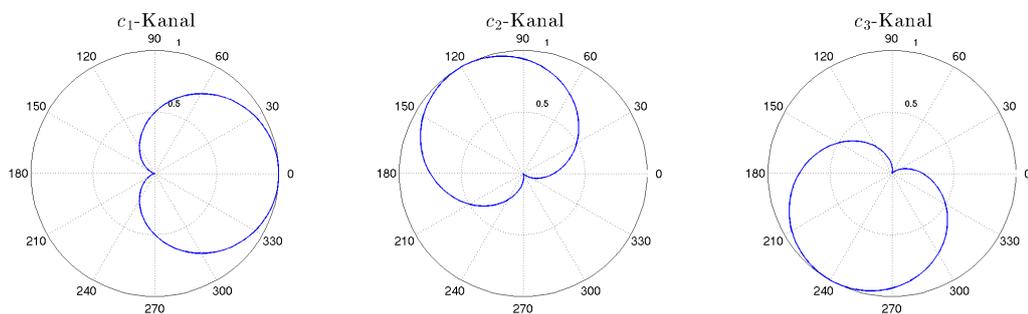
2.2.2 Berechnung dreier virtueller Nieren

Ausgangspunkt der weiteren Signalverarbeitung sind die Signale $\{c_1[n], c_2[n], c_3[n]\}$, die dreier um jeweils 120° geneigten Nieren entsprechen und aus den Signalen des B-Formats erzeugt werden können. $\bar{\alpha}$ entspricht der Hauptrichtung des Signals der ersten virtuellen Niere c_1 .

$$c_1[n] = \frac{w[n] + x[n]\cos(\bar{\alpha}) + y[n]\sin(\bar{\alpha})}{2} \quad (5)$$

$$c_2[n] = \frac{w[n] + x[n]\cos(\bar{\alpha} + \frac{2\pi}{3}) + y[n]\sin(\bar{\alpha} + \frac{2\pi}{3})}{2} \quad (6)$$

$$c_3 = \frac{w[n] + x[n]\cos(\bar{\alpha} + \frac{4\pi}{3}) + y[n]\sin(\bar{\alpha} + \frac{4\pi}{3})}{2} \quad (7)$$

Abbildung 4: Abgeleitete virtuelle Mikrofoncharakteristika zur Hauptrichtung $\bar{\alpha} = 0^\circ$

2.2.3 Bildung von Summen und Differenzsignal

Bei der Bildung des Summensignals wird abgesehen von der Hauptrichtung ein weiterer Parameter v eingeführt, der die Form der erzeugten Charakteristik beeinflusst. Es wird das Summensignal

$$s_1[n] = c_2[n] + c_1[n] - v \cdot c_3[n] \quad (8)$$

und das Differenzsignal

$$s_2[n] = c_2[n] - c_1[n] \quad (9)$$

gebildet.

Um den Einfluss des Parameters v zu untersuchen, ist in Abb. 5 der Plot der Richtcharakteristik von $s_1[n]$ mit fünf unterschiedlichen Werten von v dargestellt. Es ist erkennbar, dass für $v = 0.5$ wieder eine typische Niere in Richtung 60° entsteht. Bei kleinerem v ist die Rückwärtsdämpfung geringer. Bei Werten über $v = 0.5$ geht die Richtcharakteristik von $s_1[n]$ in Richtung einer Superniere. Die Rückwärtsdämpfung wird geringer, im Gegensatz dazu wird die Hauptkeule schmaler. Für sehr große Werte ist es möglich, eine in die entgegengesetzte Richtung zeigende Niere zu erzeugen. Für den in [Rei11a] vorgestellten Algorithmus ist dies von Bedeutung, da so die Hauptrichtung um 180° drehbar ist. Für den hier vorgestellten Algorithmus aus [Rei11b] ist dies irrelevant, da die Hauptrichtung ohnehin beliebig rotiert werden kann.

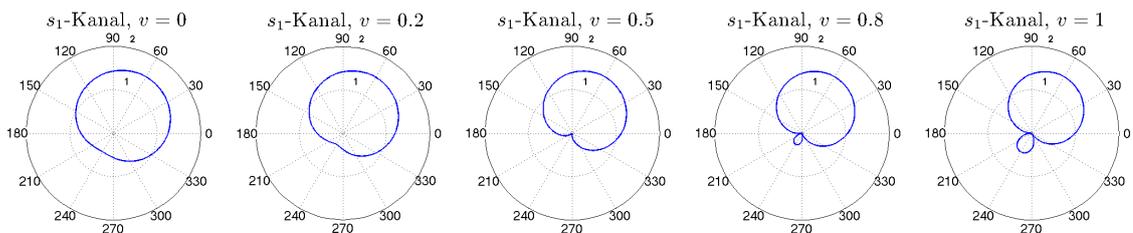


Abbildung 5: Varianten der synthetisierten Richtcharakteristik mit $v = [0, \dots, 1]$; $\bar{\alpha} = 0^\circ$

Das Differenzsignal $s_2[n]$ besitzt die Richtcharakteristik einer Acht. Sie steht orthogonal zur Hauptrichtung des Summensignals. Abb. 6 zeigt die Richtcharakteristik des Differenzsignals $s_2[n]$.

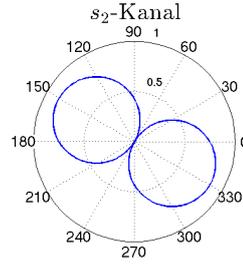


Abbildung 6: Richtcharakteristik des erzeugten Differenzsignals $s_2[n]$; $\bar{\alpha} = 0^\circ$

Durch die relative Lage der Richtcharakteristika der beiden erzeugten Signale $s_1[n]$ und $s_2[n]$ ist es notwendig, von einer gewählten Hauptrichtung α jeweils 60° abzuziehen, um $\bar{\alpha}$ zu erhalten, das als Argument für die Erzeugung der virtuellen Mikrofone dient (siehe Gleichung 5-7).

$$\bar{\alpha} = \alpha - 60^\circ \quad (10)$$

2.2.4 Spektrale Subtraktion

Der entscheidende Schritt zur verstärkten Bündelung ist die spektrale Subtraktion. In diesem Verarbeitungsschritt werden die gebildeten Signale $s_1[n]$ und $s_2[n]$ einem spektralen Subtraktionsalgorithmus unterworfen.

Hierzu werden die Signale zunächst in Blöcke aufgeteilt und mit einem Hammingfenster $w[n]$ gewichtet. Die Blocklänge beträgt $T = 1024$ Samples und die benachbarten Blöcke überlappen sich um jeweils die Hälfte. Aus einem kontinuierlichen Eingangssignal werden $k = 0, \dots, \infty$ Blöcke erzeugt. Innerhalb eines Blocks liegt n im Bereich $n = 0, \dots, (T - 1)$. Der k -te Block des Eingangssignals ergibt sich zu

$$s_{1,k}[n] = s_1\left[n + \frac{kT}{2}\right] \cdot w[n].$$

Diese Blöcke werden nun mittels der Diskreten Fourier Transformation (DFT) in den Frequenzbereich transformiert. Durch die Wahl der Blocklänge als Zweierpotenz kann ein effizienter „Fast Fourier Transform“-Algorithmus verwendet werden.

$$S_{1,k}(f) = DFT\{s_{1,k}\}$$

$$S_{2,k}(f) = DFT\{s_{2,k}\} \quad (11)$$

Es entstehen somit die Spektren $S_{1,k}(f)$ und $S_{2,k}(f)$.

In jedem Block wird das Betragsspektrum $|S_{2,k}(f)|$ vom Betragsspektrum $|S_{1,k}(f)|$ subtrahiert. Unter bestimmten Voraussetzungen können bei spektraler Subtraktion hörbare Artefakte entstehen. Dies ist vor allem dann der Fall, wenn bei einem konkreten Frequenzwert f_i gilt, dass $|S_{2,k}(f_i)| > |S_{1,k}(f_i)|$ ist. Ein negativer Amplitudenwert entsteht. Da dies physikalisch unmöglich ist, wird der Wert im einfachsten Fall auf 0 gesetzt. Es ist denkbar, dass sich zwischen mehreren auf 0 gesetzten Werten Bins befinden, bei denen $|S_{2,k}(f_i)| < |S_{1,k}(f_i)|$ ist. Diese treten dann nach der Rücktransformation als tonale Komponenten hervor, da im entsprechenden Block im Zeitbereich ein Sinuston bei der Frequenz f_i entsteht. Eine Methode, um diese Artefakte zu minimieren, ist die Einführung eines Parameters β , der die Amplitude eines Schwellenwertes bestimmt. Amplitudenwerte, bei denen $(|S_{1,k}(f_i)| - |S_{2,k}(f_i)|) < \beta \cdot |S_{1,k}(f_i)|$ ist, werden nicht auf 0 gesetzt, sondern auf den mit β gewichteten Wert des Spektrums $S_{1,k}(f_i)$. Die Sinustöne, die durch herausstechende Werte entstehen, werden dadurch maskiert. Diese Methode ist in ähnlicher Form bei der Reduktion von Störuschen bekannt [BSM79].

$$S_d(f) = \begin{cases} |S_{1,k}(f)| - |S_{2,k}(f)|, & (|S_{1,k}(f)| - |S_{2,k}(f)|) > \beta \cdot |S_{1,k}(f)| \\ \beta \cdot |S_{1,k}(f)|, & (|S_{1,k}(f)| - |S_{2,k}(f)|) < \beta \cdot |S_{1,k}(f)| \end{cases} \quad (12)$$

Im Idealfall sind überhaupt keine störenden Artefakte mehr hörbar. In der Praxis zeigt sich, dass sehr gute Ergebnisse mit $\beta = 0.005$ erzielt werden.

Vor der Rücktransformation in den Zeitbereich wird dem Differenzspektrum die Phase von Signal $s_{1,k}$ aufgeprägt. Sie entspricht der Phase des Signals aus der Richtung, die hervorgehoben werden soll. Somit wird das Ausgangssignal als die inverse DFT des Differenzspektrums mit der aufgeprägten Phase des Signals aus der gewünschten Richtung berechnet.

$$s_k[n] = IDFT\{S_{d,k} \cdot e^{j \cdot \arg(S_{1,k}(f))}\} \quad (13)$$

Nach der Rücktransformation werden die Blöcke mit dem Overlap-and-Add-Verfahren zu einem kontinuierlichen Ausgangssignal $s[m]$ zusammengefügt.

$$s[m] = s_k[n] + s_{k-1}[n + \frac{T}{2}] \quad (14)$$

Betrachtet man die Richtcharakteristik, so sollte der Schritt der spektralen Subtraktion dazu führen, dass der Absolutbetrag des Signals $s_{2,k}$ von $s_{1,k}$ subtrahiert wird. Würde dies im Zeitbereich geschehen, so ergäbe sich bei der einen Seite der Acht, welche die Empfindlichkeit von $s_2[n]$ darstellt, durch die negative Phase eine Addition. Durch die Subtraktion der Amplitudenspektren im Frequenzbereich und die anschließende Aufprä-

gung der Phase von $s_{1,k}$ gelingt es aber, den Betrag von $s_{2,k}$ zu subtrahieren.

In Abb. 7 wird die synthetisierte Charakteristik für $v = 0, 0.2, 0.4$ für den idealen Fall, in dem zum Beispiel eine Achtercharakteristik bei 90° eine Dämpfung von $-\infty$ dB aufweist, gezeigt.

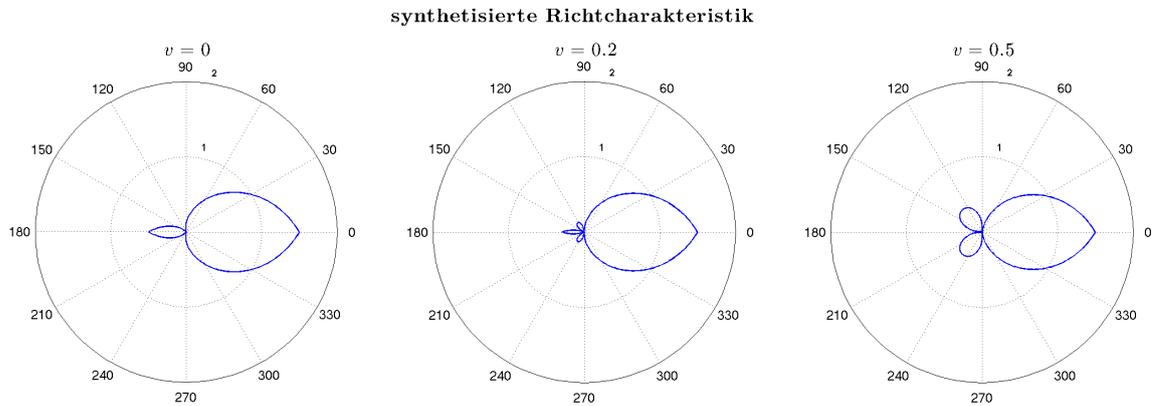


Abbildung 7: Synthetisierte Richtcharakteristik für $v = [0, 0.2, 0.5]$, $\alpha = 0^\circ$

Richtungsmaß. Bis hierher wurde zur Beschreibung der Richtcharakteristik die Größe Richtungsgrad $\Gamma(\theta)$ angegeben. Um die erstellten Richtmuster im Folgenden mit den real gemessenen und solchen, die mit anderen Verfahren erzeugt wurden, vergleichen zu können, wird von nun an die häufig verwendete logarithmische Größe Richtungsmaß gezeigt [ZZ93]. Das Richtungsmaß ergibt sich aus dem Richtungsgrad als

$$D(\theta) = 20 \cdot \log(\Gamma(\theta)). \quad (15)$$

Hierbei muss eine Grenze festgelegt werden, bis zu welcher der logarithmische Wert dargestellt wird. In einer idealen Darstellung ist es möglich, dass sehr kleine Werte bis hin zu $\Gamma = 0$ auftreten. Logarithmisch würde dies $D = -\infty$ bedeuten und kann nicht mehr aufgetragen werden. Wie in den späteren Diagrammen mit gemessenen Werten wird ein in der Praxis ausreichender Bereich von 30 dB dargestellt.

Richtungsmaß der synthetisierten Richtcharakteristik für $v = 0.5$

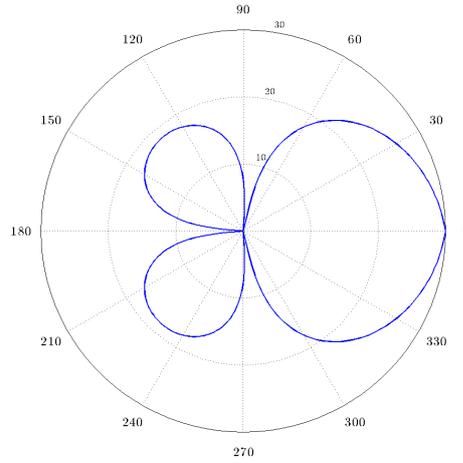


Abbildung 8: Richtungsmaß der idealen synthetisierten Richtcharakteristik mit $v = 0.5$; $\alpha = 0^\circ$

3 Messung der Impulsantworten des Arrays

Die Impulsantwortmessungen des Arrays lagen bereits vor. Sie sind im Zusammenhang mit der Diplomarbeit [Zau12] gemessen worden. Im folgenden Abschnitt sollen die wichtigen Eckdaten dieser Messung aufgezeigt werden. Details werden in der genannten Arbeit beschrieben.

Zur Bestimmung der Richtcharakteristika der Mikrofone werden Impulsantwortmessungen aus möglichst vielen unterschiedlichen Anregungsrichtungen vorgenommen.

Im Bezug auf die Winkelangaben werden die Bezeichnungen hier mit dem sphärischen Koordinatensystem, wie es zum Beispiel bei [Wil99] definiert wurde, konsistent gehalten. Der Azimutwinkel φ ist definiert als Winkel zur x-Achse in der x-y-Ebene. Der Zenitwinkel ϑ ist als Winkel zur z-Achse definiert.

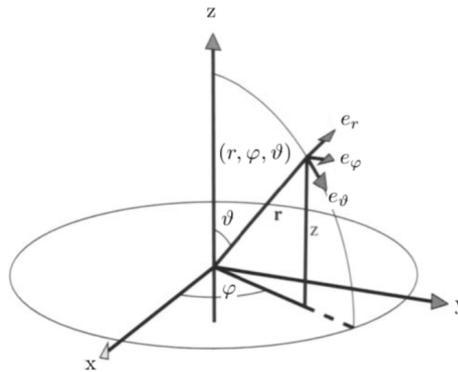


Abbildung 9: sphärisches Koordinatensystem aus [Wil99]

Die Transformation von kartesischen Koordinaten auf sphärische Koordinaten erfolgt über

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = r \begin{pmatrix} \cos(\varphi) \sin(\vartheta) \\ \sin(\varphi) \sin(\vartheta) \\ \cos(\vartheta) \end{pmatrix} = r \boldsymbol{\theta}. \quad (16)$$

Für die Messung wurden acht Lautsprecher auf einem Viertelkreis entlang des Zenits aufgebaut. Es ergibt sich für den Zenitwinkel zwischen $\vartheta = 0, \dots, 90^\circ$ eine Winkelauflösung von $\Delta\vartheta = 11.5^\circ$. Nun wurden die Impulsantworten der Mikrofone jeweils für diese acht Positionen bestimmt und das Array anschließend um $\Delta\varphi = 10^\circ$ weitergedreht. Es wurde somit unter 36 Azimutwinkeln gemessen. Durch Messung an diesen 36 Azimut- und 8 Zenitwinkeln ergeben sich für jedes Mikrofon 288 Messpositionen auf einer Hemisphäre.

An jedem dieser Messpunkte wurde mittels der Methode des exponentiellen Sinussweeps die Impulsantwort der sechs Mikrofone bestimmt. Das Vorgehen entsprach hierbei der in [Far00] vorgestellten Methode.

Die Impulsantwort ergibt sich demnach als Inverse Diskrete Fourier Transformation (IDFT) der durch das Anregungsspektrum dividierten DFT der gemessenen Systemantwort der $i = 0, \dots, (L = 288)$ Lautsprecher zu den $j = 1, \dots, 6$ Mikrofonen.¹

$$h_{ij}[n] = IDFT \left\{ \frac{DFT(s_{ij}[n])}{DFT(s_{sweep}[n])} \right\} \quad (17)$$

Da die aus den Impulsantworten bestimmbaren Frequenzgänge der Mikrofone mit dem Frequenzgang der Messausrüstung überlagert sind, wurde zunächst eine Testmessung mit einem Messmikrofon anstelle des Arrays durchgeführt. Mithilfe des damit gemessenen Referenzfrequenzgangs wurden Filter erzeugt, mit denen die gemessenen Mikrofon-signale gefiltert wurden, um den Einfluss der Messausrüstung zu minimieren.

Da die Messung nicht in einer reflexionsarmen Messumgebung durchgeführt wurde, erfolgte zur Unterdrückung der Reflexionen, die durch den Raum oder die Messausrüstung verursacht werden, eine Fensterung der Impulsantworten. Dazu wurden die Impulsantworten auf 256 Samples begrenzt. Bei einer Abtastrate von 44,1 kHz entspricht dies einer Zeit von ca. 6 ms. Bei einer Schallgeschwindigkeit von ca. 340 $\frac{m}{s}$ ergibt sich dadurch ein Ausbreitungsweg von ca. 2 m. Das bedeutet, dass Reflexionen über Wege, die um mehr als 2 m länger sind als der Direktschallweg, unterdrückt werden.

Allerdings wird durch die Fensterung die untere Grenzfrequenz des Verfahrens beeinflusst. Um ein verlässliches Ergebnis sicherzustellen, muss mindestens eine Periode einer Schwingung in die Impulsantwort passen, um die Schwingung darstellen zu können. Bei 6 ms entspricht dies einer unteren Grenzfrequenz von ungefähr $f_u \approx 170 Hz$.

Die Impulsantworten wurden außerdem noch zeitlich aneinander angepasst, da sich durch die Abweichung der Mikrofone vom Halbkugelmittelpunkt eine Phasendifferenz ergibt.

Somit liegen nun für alle sechs Mikrofone des Arrays Impulsantwortmessungen für 288 Schalleinfallrichtungen vor. Diese werden zur Analyse der synthetisierten Richtcharakteristika verwendet.

1. Das sechste, zentral angeordnete Mikrofon wird für diese Arbeit nicht verwendet.

4 Filterdesign

4.1 Mikrofontzerrung

Nachdem die Messergebnisse vorliegen, können die FIR-Filter (Finite-Impulse-Response-Filter) der Eingangsstufe des Algorithmus, wie in Abbildung 2 (oben) gezeigt, konstruiert werden. Bei den vier Filtern, die sich direkt nach den Mikrofoneingängen befinden, geht es darum, die produktionsbedingten Abweichungen unter den Mikrofonen auszugleichen. Hierbei wird das Mikrofon 3, welches in die gewählte 0° -Richtung zeigt, als Referenz verwendet. Abb. 10 zeigt die Frequenzgänge der Mikrofone 1-5. In Abb. 11 ist das Verhältnis der Mikrofone 1, 2, 4 und 5 zum Bezugsmikrofon 3 zu sehen. Es werden die Messungen mit einem Zenitwinkel von $\vartheta = 90^\circ$ und dem Azimutwinkel, welcher der 0° -Achse des jeweiligen Mikrofons am nächsten ist, verwendet.

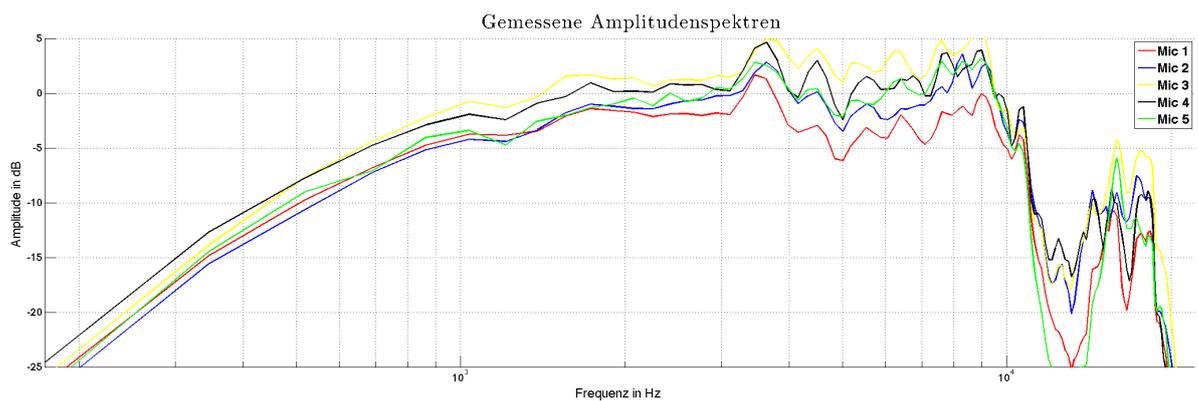


Abbildung 10: Amplitudenspektren der Mikrofone

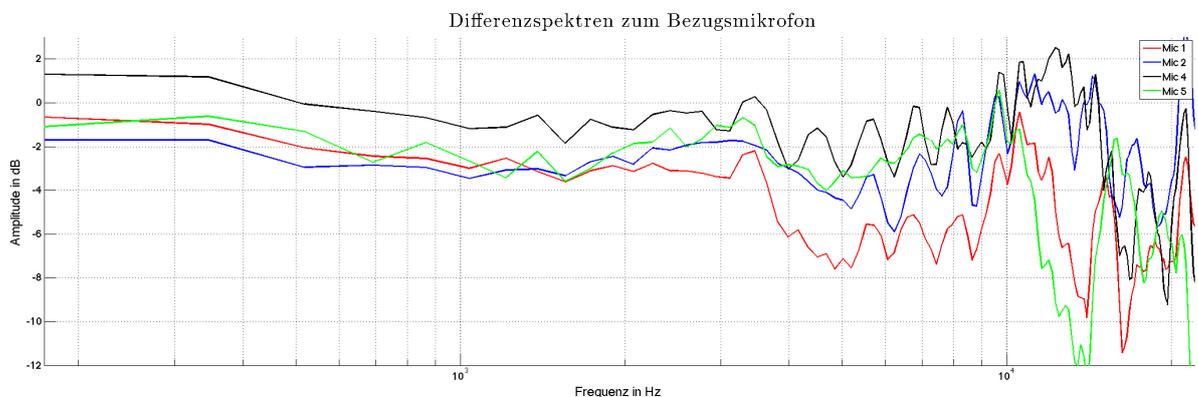


Abbildung 11: Differenzspektren zwischen Bezugsmikrofon (Mic 3) und den anderen Mikrofonen des Arrays (Mic 1, 2, 4 und 5)

Mit dem Ziel die Amplitudengänge zwischen Bezugsmikrofon (Mic 3) und den anderen Mikrofonen (Mic 1, 2, 4 und 5) anzupassen, wird nun für jedes der vier anderen Mikrofone ein FIR-Filter entworfen.

Um die Phasenverschiebungen möglichst gering zu halten, wurden diese Filter mittels Cepstrum minimalphasig ausgeführt.

Damit man ein minimalphasiges Filter erhält, wurde dazu zunächst die DFT der Impulsantwort vom Referenzmikrofon (Mikrofon 3) berechnet.

$$M_3(f) = DFT\{h_3[n]\} \quad (18)$$

Außerdem wird die DFT der jeweiligen Impulsantworten der anderen vier Mikrofone berechnet. Da die folgenden Verarbeitungsschritte für alle Mikrofone gleichermaßen ausgeführt werden, wird allgemein der Index $j = [1, 2, 4, 5]$ verwendet.

$$M_j(f) = DFT\{h_j[n]\} \quad (19)$$

Es wird der Betrag des Quotienten der Fouriertransformierten des Bezugsmikrofons $M_3(f)$ zum jeweiligen Mikrofon $M_j(f)$ gebildet. Stellt man sich vor, dass bei einer konkreten Frequenz f_i die Amplitude des auszugleichenden Mikrofons j größer ist als die des Bezugsmikrofons, so wird $\hat{M}_j(f)$ kleiner und das fertige Filter wird bei dieser Frequenz dämpfen.

$$\hat{M}_j(f) = \left| \frac{M_3(f)}{M_j(f)} \right| \quad (20)$$

Wie in Abb. 11 erkennbar, sind die Abweichungen zwischen den Mikrofonen zum Referenzmikrofon vor allem im hohen Frequenzbereich groß. Der Ausgleich großer Abweichungen bei hohen Frequenzen ist allerdings nicht sinnvoll, da eine Auswirkung auf die Dynamik zu erwarten ist. Da ein Großteil der Energie bei typischen Quellsignalen unter 10 kHz liegt und bei sehr hohen Frequenzen die Qualität der Richtcharakteristika abnimmt, wurde der Frequenzgang der Filter ab dieser Grenze auf 1 gesetzt.

Ab 10 kHz wird also $\hat{M}_j(f) = 1$ festgelegt. Dadurch, dass der Wert des vorigen Frequenzbins allerdings stark von 1 abweichen kann, könnte es hier zu einem sehr un stetigen Frequenzgang kommen. Es wird der Verlauf der fünf Werte vor dem ersten Frequenzbin, der auf 1 gesetzt wird, mittels linearer Interpolation geglättet.

In Abb. 12 sind die interpolierten Amplitudenverläufe der Ausgleichsfilter der Mikrofone zu sehen. Auf allen Amplitudendiagrammen ist das einseitige Spektrum gezeigt. Das

gesamte Spektrum enthält zusätzlich noch eine an der halben Samplingrate gespiegelte Version. Alle beschriebenen Eingriffe müssen auch im gespiegelten Spektrum vorgenommen werden. Hierzu wird das einseitige Spektrum bearbeitet, dann gespiegelt und kopiert.

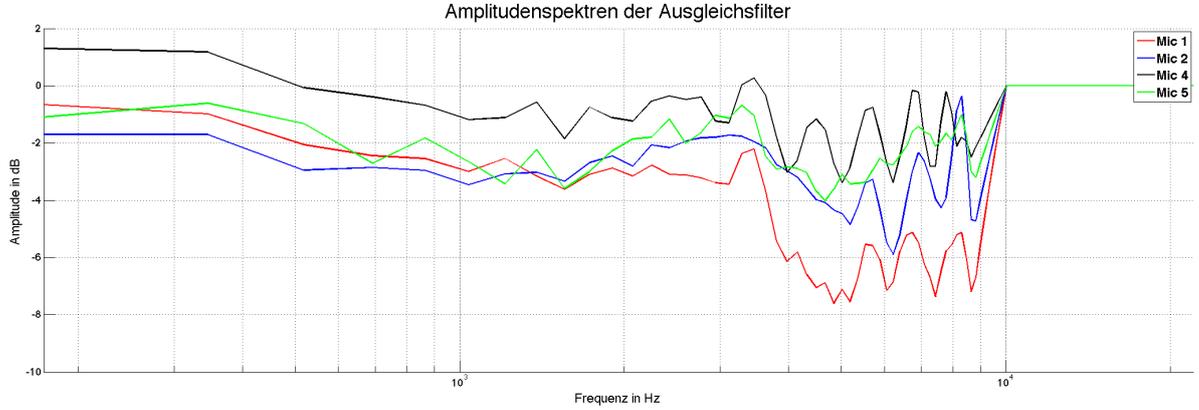


Abbildung 12: Amplitudengang der Ausgleichsfilter; Ab 10 kHz wurde der Amplitudengang auf 0 dB gesetzt. Von 9 kHz an ist der durch lineare Interpolation geglättete Verlauf erkennbar.

Nachdem nun diese Vorgaben im Frequenzbereich realisiert wurden, kann mittels Cepstrum das minimalphasige System mit diesem Amplitudenverlauf gefunden werden. Hierzu wird der Logarithmus von $\hat{M}_j(f)$ gebildet und das Ergebnis in den Zeitbereich zurücktransformiert.

$$\hat{m}_j[n] = IDFT\{\log(\hat{M}_j(f))\} \quad (21)$$

Um das minimalphasige System zu finden, wird nun eine Fensterung von $\hat{m}_j[n]$ mit $win[n]$ vorgenommen. Das Fenster sorgt dafür, dass nur die erste Hälfte von $\hat{m}_j[n]$ erhalten bleibt, wobei die äußeren Werte halbiert werden. Nach der Fensterung wird $\hat{m}_j[n]$ erneut in den Frequenzbereich transformiert.

$$\hat{M}_{j,MP}(f) = DFT\{\hat{m}_j[n] \cdot win[n]\} \quad (22)$$

Anschließend werden diese Werte entlogarithmiert und anschließend in den Zeitbereich zurücktransformiert. Das Ergebnis ist die Impulsantwort eines minimalphasigen FIR-Filters mit dem vorgegeben Amplitudenverlauf.

$$\hat{m}[n]_{j,MP} = IDFT\{10^{\hat{M}_{j,MP}(f)}\} \quad (23)$$

Da die Impulsantworten schnell abklingen, kann die Länge auf 64 Samples verringert werden, um in weiterer Folge unnötigen Rechenaufwand einzusparen.

Abb. 13 zeigt den Einfluss der Filterung am Beispiel des zweiten Mikrofons. In rot ist der Amplitudenverlauf vor der Filterung erkennbar. Angestrebt wird der in grün gezeigte Verlauf des Bezugsmikrofons (Mic 3). In blau ist der Verlauf von Mikrofon 2, nachdem die Impulsantwort dieses Mikrofons mit der soeben bestimmten Impulsantwort des Filters gefaltet wurde ($h_2[n] * \hat{m}_{2,MP}[n]$), ersichtlich. Man erkennt, dass bei Frequenzen kleiner als 10 kHz die blaue Kurve (gefiltertes Mikrofon 2) auf der grünen (angestrebter Verlauf) liegt. Ab 10 kHz liegt die blaue Kurve auf der roten (ungefiltertes Mikrofon 2). Es wurde also im tiefen Frequenzbereich der Amplitudengang von Mikrofon 2 an jenen von Mikrofon 3 angepasst und im hohen Frequenzbereich keine Änderung bewirkt. Das gewünschte Verhalten ist damit erreicht.

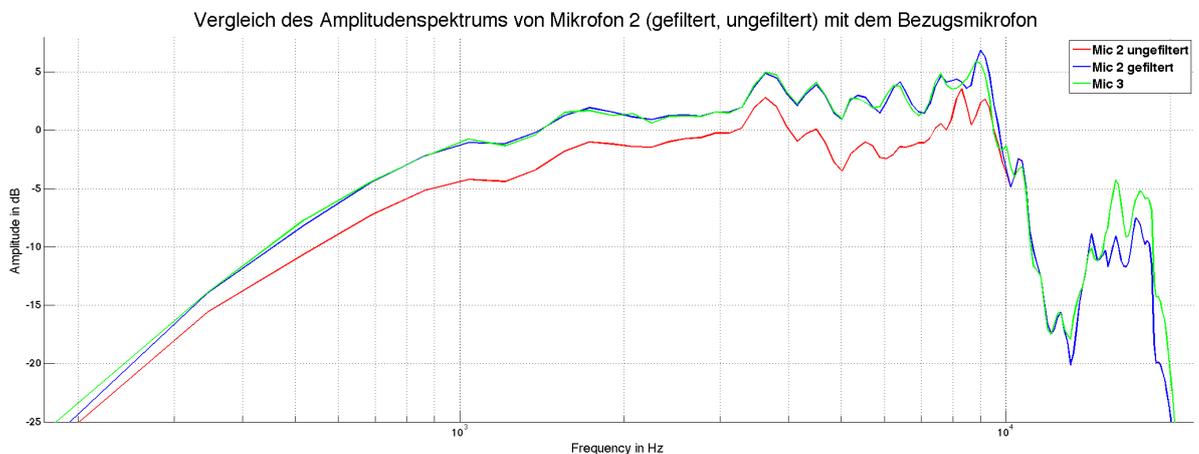


Abbildung 13: Mikrofontzerrung durch die Ausgleichsfilter in der Eingangsstufe des Algorithmus am Beispiel von Mikrofon 2

4.2 Ausgangsentzerrung

Das Anpassen der Mikrofonfrequenzgänge untereinander stellt die korrekte Funktion des Algorithmus sicher. Im Bezug auf den Frequenzgang des gesamten, synthetisierten Ausgangssignals ist damit allerdings noch keine Optimierung erreicht. Bei Betrachtung dieses gesamten Frequenzgangs stellt man fest, dass eine deutliche Absenkung der tiefen Frequenzen stattfindet (Abb. 14). Dies lässt sich bereits erahnen, wenn man den Frequenzgang der Mikrofone betrachtet. Um eine gute Filterfunktion für das synthetisierte Ausgangssignal zu finden, wären mehrere grundsätzliche Ansätze denkbar.

Man könnte für eine Hauptrichtung die Frequenzgänge aller Richtungen mitteln und den mittleren invertierten Frequenzgang als Entzerrung verwenden, ähnlich wie bei diffusfeldentzerrten Mikrofonen. Alternativ könnte man einfach den invertierten Frequenzgang

der Hauptrichtung verwenden und mit diesem entzerren. Dies entspräche einer Freifeld-entzerrung.

Da die Dämpfungen der anderen Richtungen im Verhältnis zur Hauptrichtung sehr groß sind, wird hier die zweite Variante gewählt. Die Impulsantwort des Filters wird auf die gleiche Art und Weise wie bei den Filtern der Eingangsstufe bestimmt. Zusätzlich wird im tiefen Frequenzbereich nachjustiert, um eine möglichst glatte Kurve zu erhalten.

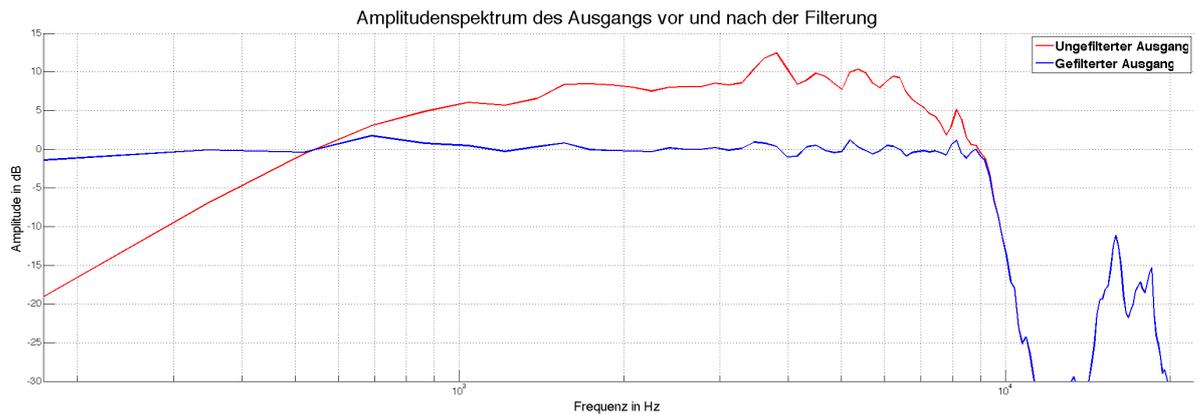


Abbildung 14: Amplitudenspektrum des gesamten Systems vor und nach der Filterung mit dem Ausgangsfilter am Ende des Algorithmus. Das Filter dient der Verbesserung des Frequenzgangs des gesamten Systems.

5 Auswertung

5.1 Darstellung der Richtcharakteristika

Es besteht hiermit die Möglichkeit Signale mit starker Bündelung zu erzeugen. Es ist notwendig, eine geeignete Darstellung zu finden, damit die aus den Impulsantworten berechnete Richtmuster auch visualisiert und verglichen werden können. Hierzu konnte auf die Matlab Funktion *"balloon_holo"* zurückgegriffen werden, die sich dazu eignet, aus Impulsantworten dreidimensionale Grafiken der Richtwirkung zu erzeugen. Die Funktion ist ebenfalls im Zusammenhang mit [Zau12] entstanden.

Um eine optimale grafische Aufbereitung zu bieten, verwendet die Funktion Interpolation über die Zerlegung in Kugelflächenfunktionen (sphärische Harmonische). So kann die Dichte der Datenpunkte erhöht werden und es entsteht eine Grafik, die für das Auge besser geeignet ist.

Theorie SHT. Das zugrunde liegende Prinzip der verbesserten Darstellung ist die Discrete Spherical Harmonics Transform (DSHT). Das genaue Vorgehen und die entsprechenden Hintergründe werden in [Zau12] aufgezeigt und hier in gekürzter Form wiedergegeben.

Zunächst werden die Gewichte der Kugelflächenfunktionen bestimmt. Die Kugelflächenfunktionen stellen die Lösungen des Winkelanteils der Helmholtzgleichung² dar, nachdem eine Variablenseparation, die Winkelanteil und radialen Anteil trennt, durchgeführt wurde.

Die reellwertigen Kugelflächenfunktionen sind definiert als

$$Y_n^m(\theta) = N_n^{|m|} P_n^{|m|} \Phi_m(\varphi) = N_n^{|m|} P_n^{|m|}(\cos(\vartheta)) \cdot \begin{cases} \sin(m\varphi) & \text{for } m < 0 \\ \cos(m\varphi) & \text{for } m > 0 \end{cases} \quad (24)$$

Hierbei wird n als die Ordnung und m als der Grad bezeichnet. $P_n^{|m|}$ sind Legendre Polynome mit dem Argument $\cos(\vartheta)$ und stammen von der Lösung des Zenitanteils der Helmholtzgleichung. $\Phi_m(\varphi)$ stammen von der Lösung des Azimutanteils.

2. Die Helmholtzgleichung lautet $\Delta p(r, \omega) + k^2 p(r, \omega) = 0$ und beschreibt die Wellenausbreitung des Schalldrucks im Frequenzbereich (Δ ist der Laplace-Operator).

$N_n^{|m|}$ sind Normalisierungsfaktoren und definiert als

$$N_n^{|m|} = \sqrt{\frac{(2n+1)(2-\delta_m)(n-m)!}{4\pi(n+m)!}} \quad (25)$$

Die Kugelflächenfunktionen sind ein orthonormaler Satz an Funktionen. Das bedeutet, dass sie die Orthonormalitätsbedingung erfüllen.

$$\int_{S^2} Y_{n'}^{m'}(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} = \delta_{nn'} \delta_{mm'} \quad (26)$$

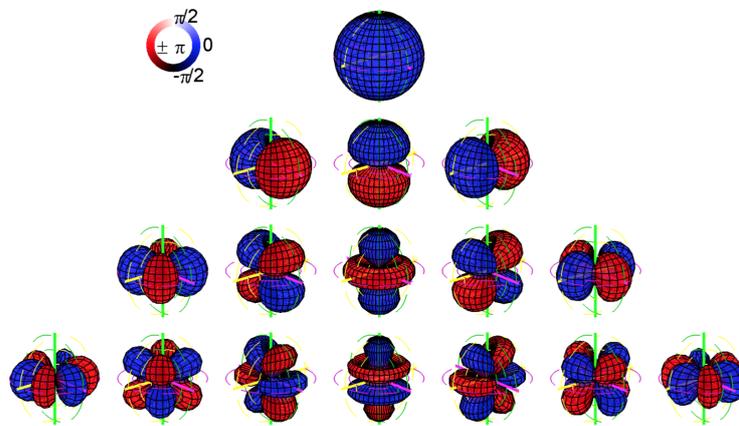


Abbildung 15: Kugelflächenfunktionen bis zur Ordnung 3 [Pom08]

Aus diesen Funktionen kann jede beliebige quadratisch integrierbare Funktion, die auf einer Kugel definiert ist, zusammengesetzt werden.

Hierzu muss man in Analogie zur Fouriertransformation die Abbildung der beliebigen Funktion $f(\boldsymbol{\theta})$ auf die Basisfunktionen berechnen. Das Integral des Produkts der Funktion $f(\boldsymbol{\theta})$ und den Basisfunktionen Ordnung 0 bis Ordnung ∞ über alle Winkel $\boldsymbol{\theta}$ wird Spherical Harmonics Transform (SHT) genannt.

$$SHT\{f(\boldsymbol{\theta})\} = \int_{S^2} f(\boldsymbol{\theta}) Y_n^m(\boldsymbol{\theta}) d\boldsymbol{\theta} = \gamma_n^m \quad (27)$$

Die Inverse SHT ist die unendliche Summe über die mit γ_n^m gewichteten sphärischen harmonischen Funktionen.

$$ISHT\{\gamma_n^m\} = \sum_{n=0}^{\infty} \sum_{m=-n}^n Y_n^m(\boldsymbol{\theta}) \gamma_n^m = f(\boldsymbol{\theta}) \quad (28)$$

Im vorliegenden Fall der Messpunkte ist die winkelabhängige Charakteristik allerdings nicht für alle Winkel bekannt, sondern liegt an den diskreten Lautsprecherpositionen vor. Ebenfalls in Analogie zur Fouriertransformation wird für den Fall diskreter Datenpunkte die DSHT definiert. Wo bei der SHT eine unendliche Summe von Basisfunktionen gebildet wird, muss hier eine Bandbegrenzung erfolgen. Es werden nur die Funktionen bis zur Ordnung N verwendet.

Bestimmung der Koeffizienten der sphärischen Harmonischen. Um auf die Gewichte der sphärischen Harmonischen und damit auf eine Summe aus Funktionen zu kommen, die dann an beliebigen Stellen ausgewertet werden kann, müssen zunächst einige Vektoren definiert werden.

Der Vektor $\mathbf{h}_j(\omega)$ enthält die Impulsantworten der Mikrofone $j = 1, \dots, 5$ zur jeweiligen Richtung $\boldsymbol{\theta}_i$ der Lautsprecher $i = 1, \dots, L$. Die Einträge des Vektors \mathbf{p} entsprechen dem räumlich erzeugten Schalldruck des jeweiligen Lautsprechers.

$$\mathbf{h}_j(\omega) = \begin{pmatrix} h_j(\boldsymbol{\theta}_1, \omega) \\ h_j(\boldsymbol{\theta}_2, \omega) \\ \vdots \\ h_j(\boldsymbol{\theta}_L, \omega) \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_L \end{pmatrix} \quad (29)$$

Hierbei ist $\mathbf{p} = [\delta_{i'v}]$, sodass mit

$$\mathbf{h}_j(\boldsymbol{\theta}_i) = \mathbf{h}_j^T \mathbf{p} \quad (30)$$

die Impulsantwort des Lautsprechers i zum Mikrofon j ausgewählt werden kann, wobei $\delta_{i'i'}$ eine Kronecker-Delta-Funktion am Lautsprecher i ist.

Ziel der Interpolation soll die Bestimmung einer Formel sein, mit der nicht nur die Messdaten der diskreten Winkel $\boldsymbol{\theta}_i$ ausgewählt werden, sondern zusätzliche Stützstellen an beliebigen Winkeln $\boldsymbol{\theta}$ berechnet werden können.

Dazu wird zunächst die Verteilung der Lautsprecher durch eine Summe aus Deltafunktionen dargestellt.

$$p(\boldsymbol{\theta}) = \sum_{i=1}^L \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_i) p_i \quad (31)$$

Die Koeffizienten der Modenstärke für die Quelle ergeben sich laut [Pom08] als

$$\phi_n^m = \sum_{i=1}^L Y_n^m(\boldsymbol{\theta}_i) p_i. \quad (32)$$

Dadurch, dass nur Moden bis zur Ordnung N verwendet werden, kann dies in eine Vektorschreibweise überführt werden.

$$\boldsymbol{\phi}_N = \mathbf{D}_N \mathbf{p} \quad (33)$$

wobei in \mathbf{D}_N die sphärischen Harmonischen enthalten sind.

$$\mathbf{D}_N = \begin{bmatrix} \mathbf{y}_N^T(\boldsymbol{\theta}_1) \\ \mathbf{y}_N^T(\boldsymbol{\theta}_2) \\ \vdots \\ \mathbf{y}_N^T(\boldsymbol{\theta}_L) \end{bmatrix} \quad (34)$$

Der erzeugte Schalldruck zu einem beliebigen Winkel ergibt sich nun aus der Pseudoinversen der Matrix mit den sphärischen Harmonischen \mathbf{D}_N^+ multipliziert mit der zuvor bestimmten Stärke der Mode.

$$\mathbf{p} = \mathbf{D}_N^+ \boldsymbol{\phi}_N \quad (35)$$

Durch Einsetzen der von $\boldsymbol{\phi}_N = \mathbf{y}_N(\boldsymbol{\theta})$ können nun Schalldruckwerte berechnet werden, sodass die interpolierte Impulsantwort für jeden beliebigen Winkel bestimmt werden kann.

$$h_j(\boldsymbol{\theta}) = \mathbf{h}_j^T \mathbf{D}_N^+ \mathbf{y}_N(\boldsymbol{\theta}) \quad (36)$$

Reduktion der Ordnung. Es können durch die Abtastung an endlichen Punkten nur sphärische Harmonische bis zu einer bestimmten Ordnung ausgewertet werden. Da in diesem Fall nur die geraden Harmonischen angeregt werden können, ergibt sich die theoretische maximale Ordnung nach der Formel

$$L \geq \frac{(N + 1)(N + 2)}{2}. \quad (37)$$

Bei $L = 288$ erhält man somit eine maximale Ordnung von $N = 22$. Die tatsächlich verwendbare maximale Ordnung hängt allerdings stark von der Anordnung der Abtastpunkte ab (vergl. [Pom08]). Für das hier verwendete Schema resultiert eine maximale Ordnung von $N = 14$. Durch weitere Reduktion der Ordnung lassen sich die Richtungsdiagramme zusätzlich glätten.

Abb. 16 zeigt eine derartig interpolierte Darstellung des Mikrofons 3 bei 1 kHz, erstellt mit *baloon_holo*. Zusätzlich wird rechts ein Schnitt entlang der x-y-Ebene der dreidimensionalen Darstellung angezeigt.

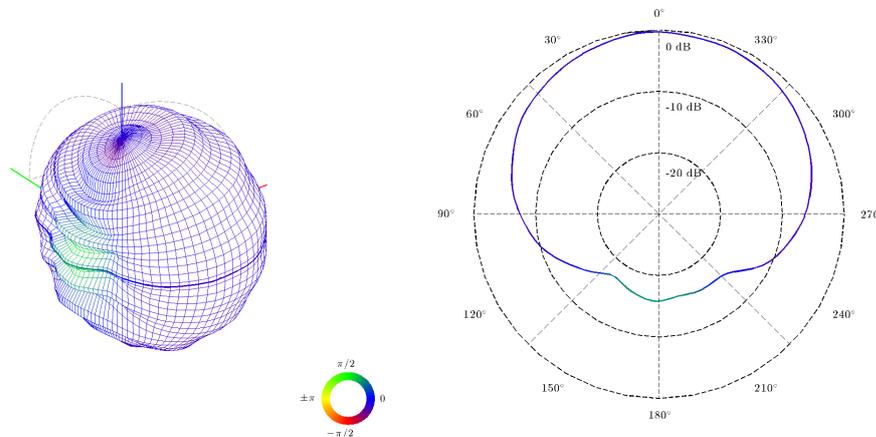


Abbildung 16: (links) Interpolierte 3D-Richtcharakteristik des unbearbeiteten Mikrofons 3 bei 1 kHz; (rechts) Schnitt durch die 3D-Darstellung entlang der x-y-Ebene; Die Farbe gibt die Phasenlage an.

5.2 Erzeugte Aufnahmecharakteristika

Mithilfe der Impulsantwortmessungen und der Funktion zur Darstellung der interpolierten Daten können nun Plots der synthetisierten Aufnahmecharakteristika erzeugt werden. Dazu wurden die Impulsantwortmessungen der Mikrofone jeder Richtung dem Algorithmus zugeführt und die resultierenden Impulsantworten an die Funktion zur Darstellung übergeben. Die Ergebnisse sind in Abb. 17 dargestellt.

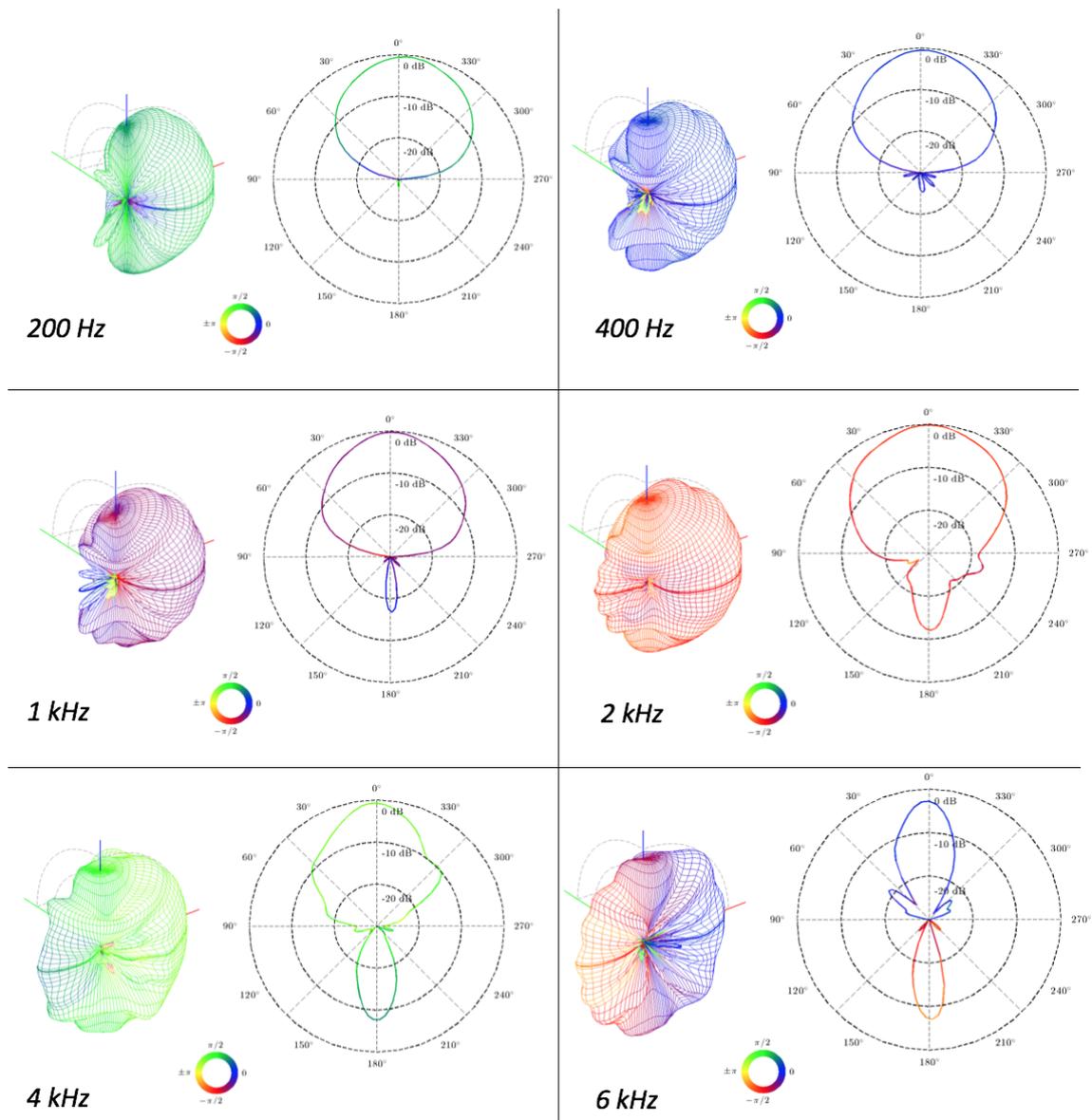


Abbildung 17: Mögliche synthetisierte Aufnahmecharakteristika für $\alpha = 0^\circ$, $v = 0.4$ bei unterschiedlichen Frequenzen $f = [200\text{Hz}, \dots, 6\text{kHz}]$

Bei allen Frequenzen zeigt sich, dass stark gerichtete Charakteristika entstanden sind. Die Dämpfung in alle anderen Richtungen, abgesehen von der Hauptrichtung, ist sehr groß. Schall aus den Richtungen $\pm 90^\circ$ wird fast vollständig unterdrückt. Besonders bei tiefen Frequenzen ist auch die Rückwärtsdämpfung groß. Zu höheren Frequenzen sinkt die Rückwärtsdämpfung, liegt allerdings bei 2 kHz noch über 10 dB. Bei 4 kHz liegt sie leicht unter 10 dB. Zu 6 kHz sinkt sie zwar weiter, aber es ist eine noch stärkere Bündelung nach vorne zu erkennen. Die Richtcharakteristika sind also frequenzabhängig.

Vergleicht man die gemessenen Charakteristika nun mit den idealen Modellen aus Kapitel 2, so fällt auf, dass Breite und Form der Hauptkeule über einen breiten Frequenzbereich sehr gut übereinstimmen. In den Plots für $v = 0.4$ sind die erwarteten Nebenkeulen allerdings weniger deutlich erkennbar. Dies gilt zumindest für den Schnitt entlang der x-y-Ebene bei tiefen Frequenzen. Da die Form und Ausprägung der Nebenkeulen von Faktor v abhängen, werden in Abb. 18 die Schnitte durch die x-y-Ebene für unterschiedliche Werte von v gezeigt. Die Ausprägungen der Nebenkeulen entsprechen nicht der Erwartung. Auch bei höheren Werten für v ist nur eine einzige Keule in 180° -Richtung erkennbar. Der Grund hierfür liegt in der endlichen Winkelauflösung sowie in den Abweichungen der Mikrofoncharakteristika von den idealen Annahmen. Bei einigen Frequenzen sind Asymmetrien zu erkennen. Dennoch wird deutlich, dass die Dämpfung nach hinten bei $v = 0.5$ am größten ist.

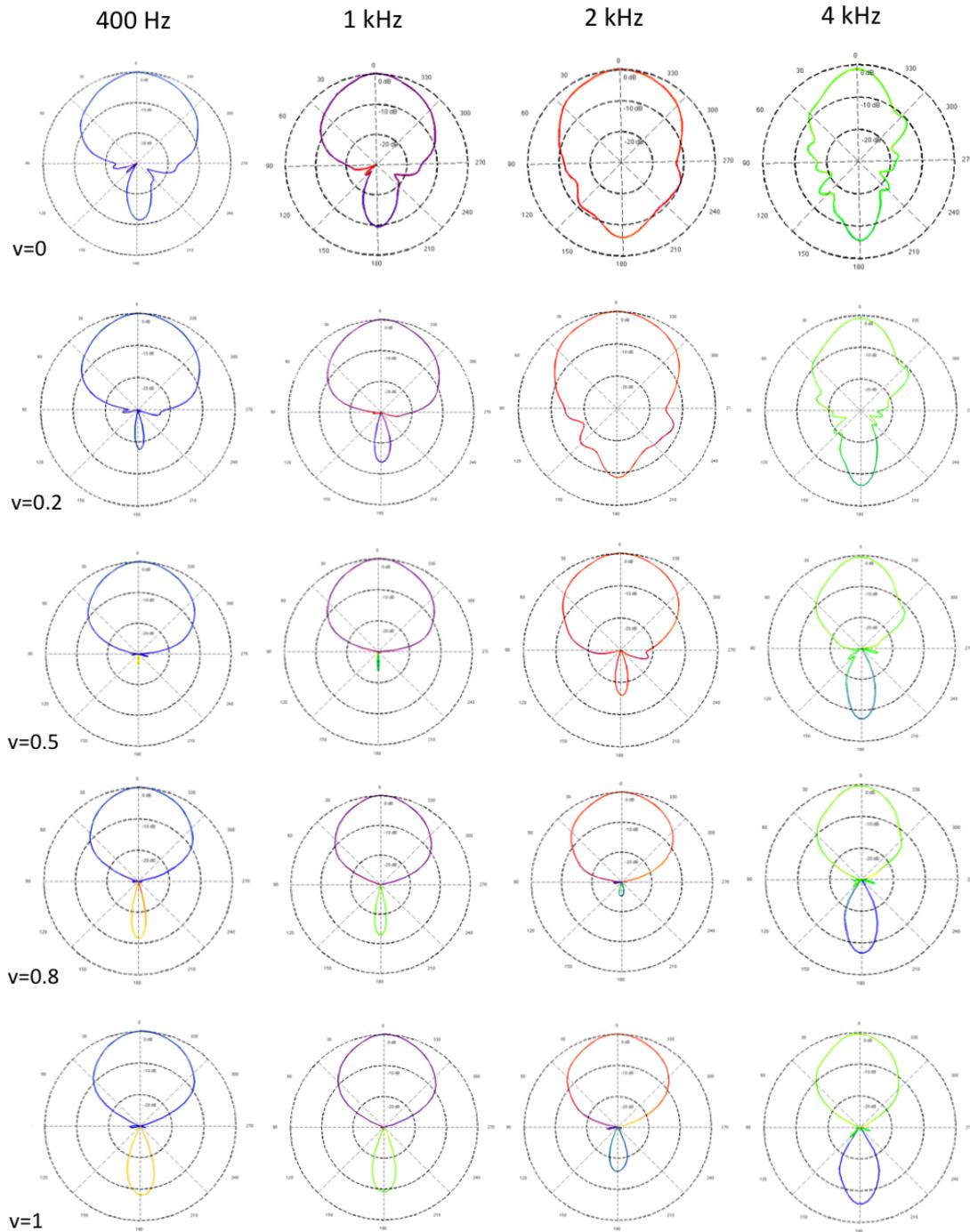


Abbildung 18: Vergleich der Richtcharakteristik für unterschiedliche Werte von v bei Frequenzen $f = [400\text{Hz}, \dots, 4\text{kHz}]$, $\alpha = 0^\circ$; Die 0° -Grad Richtung zeigt nach oben. Die eingezeichneten Kreise stellen jeweils -10 dB Abfall dar.

Vergleich zu einer mit modalem Beamforming erzeugten Charakteristik. In Abb. 19 ist die erzeugte Charakteristik im Vergleich zu einem Hypernieren-Beampattern gezeigt, das mit modalem Beamforming entstanden ist. Das mit modalem Beamforming erzeugte Pattern zeigt in z-Richtung, was aus der Abbildung nicht ersichtlich wird. Dies ist mit der vorliegenden Implementierung des untersuchten Algorithmus nicht möglich. Was die Richtwirkung betrifft, sind die Ergebnisse jedoch vergleichbar. Es zeigt sich, dass das mit spektraler Subtraktion erzeugte Pattern über die Frequenzen 400 Hz - 4 kHz stabiler ist. Zudem ist die Unterdrückung der 90°-Richtung deutlich stärker.

Ferner ist zu erkennen, dass die Phase beim mit modalem Beamforming erzeugten Pattern über den Winkel nicht konstant bleibt. Bei dem mit spektraler Subtraktion erzeugten Pattern ändert sich die Phase über den Winkel nicht, dahingegen ändert sie sich über die Frequenz. Dieser Verlauf ist stetig und könnte, falls störend, mit einem Allpassfilter ausgeglichen werden.

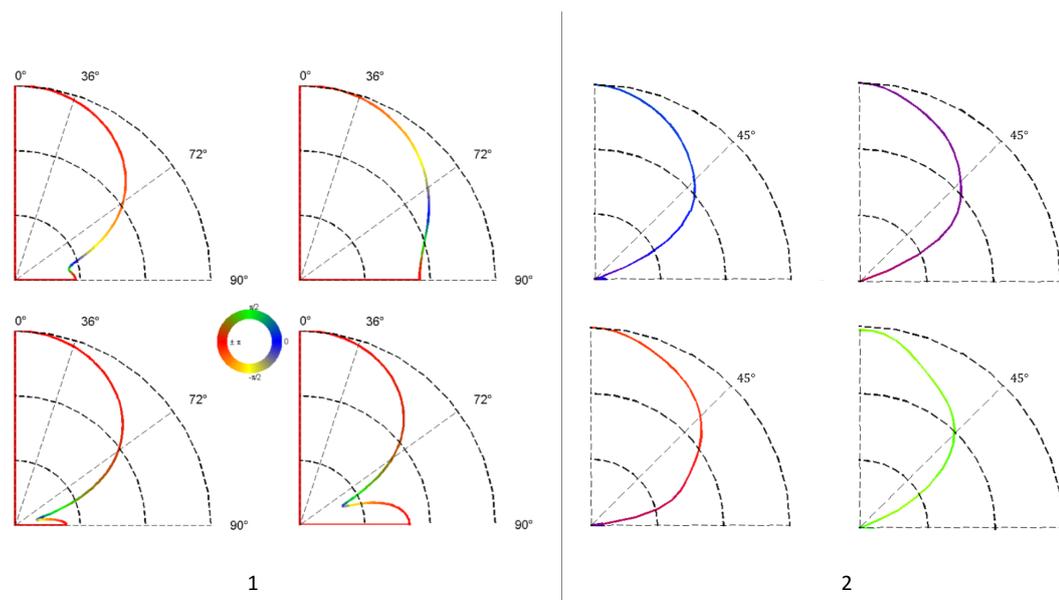


Abbildung 19: Vergleich der erzeugten Charakteristika von (1) mit modalem Beamforming erzeugt [Zau12] zu (2) mit spektralem Beamforming erzeugt, jeweils von links oben nach rechts unten bei 400 Hz, 1 kHz, 2 kHz und 4 kHz; $\alpha = 0^\circ$

5.3 Andere Evaluierungsmethoden

Außer der Darstellung dieser dreidimensionalen Richtcharakteristikdiagramme sind noch andere Möglichkeiten der Auswertung denkbar.

Zum Beispiel kann jede beliebige der 288 Lautsprecher-Impulsantworten mit einem Signal gefaltet werden. Dies entspricht dann einem Abspielen eines Signals über die jeweilige Lautsprecherposition.

Als Signal kann Musik oder Sprache verwendet werden, um einen akustischen Test der Funktion durchzuführen. Wenn der Winkel α von der virtuellen Quellposition wegbewegt wird, hört man den starken Abfall der Lautstärke. Ein akustischer Test ist vor allem deshalb sinnvoll, da Störungen (z.B. Verzerrungen, Artefakte) mit dem Ohr leicht erkannt werden können. Durch Faltung der Impulsantworten mit Testsignalen wurden auch die vorgefertigten Beispiele im Pure Data-patch erzeugt (siehe A.2).

Außerdem kann der Pegelabfall des Schalls aus einer beliebigen Richtung im Verhältnis zur Hauptrichtung, der aus dem Richtdiagramm abgelesen werden kann, am laufenden Signal verifiziert werden. Hierzu werden zwei unterschiedliche Rauschsignale durch Faltung mit den Impulsantworten h_j mit $j = 1, \dots, 5$ aus einer jeweils anderen Richtung im Raum platziert.

Das erste Rauschsignal $f_s[n]$ wird mit den Impulsantworten aus einer Richtung θ_I gefaltet, die der Lautsprecherposition $i = I$ entspricht. Dieses Signal betrachten wir als das Nutzsignal. Das zweite Rauschsignal $f_n[n]$ wird mit den Impulsantworten aus einer anderen Richtung θ_{II} entsprechend der Lautsprecherposition $i = II$ gefaltet. Wir betrachten dieses als Störsignal. Die Signale werden addiert und dem Algorithmus zugeführt.

$$f_{s+n}[n] = h_{j,I}[n] * f_s[n] + h_{j,II}[n] * f_n[n] \quad (38)$$

Anschließend wird die Kreuzkorrelation des Ausgangs des gesamten Algorithmus $f_o[n]$ mit den beiden Eingangssignalen gebildet. Für beispielsweise das Ausgangssignal $f_o[n]$ und das Nutzsignal $f_s[n]$ ergibt sich die diskrete Kreuzkorrelation zu

$$R_{f_o, f_s}[n] = \sum_{m=-\infty}^{\infty} f_o[m] f_s[m+n]. \quad (39)$$

Da die beiden Rauschsignale f_s und f_n zueinander und untereinander dekorreliert sind, ergibt sich in den Kreuzkorrelationsfunktionen ein scharfer Peak. Dieser entsteht an der Stelle, wo der Ausgang und der Eingang zeitlich genau übereinander liegen. Die Höhe

des Peaks gibt Auskunft darüber, wie sich das Verhältnis der Amplitude des Signals am Ausgang zur Amplitude des Signals am Eingang verhält.

Der 20-fache Logarithmus des Verhältnisses der Maxima der Kreuzkorrelationen lässt sich dann als Signal-Stör-Verhältnis (SNR) definieren.

$$SNR = 20 \log \left(\frac{|R_{f_o, f_s}|_{max}}{|R_{f_o, f_n}|_{max}} \right) \quad (40)$$

Für einen Test wird für beide Signale Oktavrauschen im 1 kHz Frequenzband verwendet. Das Nutzsinal wird in der horizontalen Ebene ($\vartheta = 90^\circ$) auf die Azimutposition $\varphi = 0^\circ$ projiziert. Das Störsignal bewegt sich, ebenfalls in der Horizontalebene, in Schritten von $\Delta\varphi = 10^\circ$ um das Array. Als Hauptrichtung wird $\alpha = 0^\circ$, also die Richtung des Nutzsinals, gewählt. Die Genauigkeit erhöht sich mit der Länge der Signale, deswegen wird eine Signallänge von mehreren Sekunden gewählt.

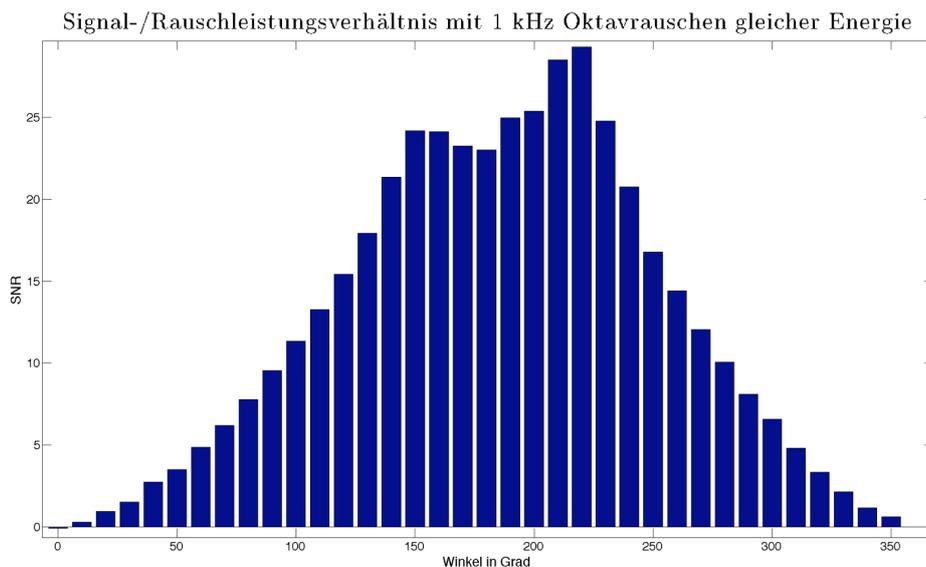


Abbildung 20: Signal-Stör-Verhältnis zweier 1 kHz Oktavrauschsignale; Eines statisch bei $\varphi = 0^\circ$ das andere dynamisch; Es vollführt eine Rotation um das Array mit Schritten von $\Delta\varphi = 10^\circ$; $v = 0.5$

Bei der Berechnung des SNR mit zwei Oktavrauschsignalen (Abb. 20) lässt sich das Ergebnis aus den Richtdiagrammen verifizieren. Hierbei ist eine leichte Asymmetrie erkennbar. Vergleicht man dieses Ergebnis mit dem aus Abb. 18 (Spalte 2, Zeile 3), so fällt auf, dass diese Asymmetrie hier nicht erkennbar ist. Es bildet sich allerdings auf der rechten, unteren Seite eine Ausbeulung, wenn die Frequenz steigt. Bei 2 kHz wird diese

dann besonders deutlich. Weil bei der Messung mit Oktavrauschen gearbeitet wurde, ist es plausibel, dass diese Asymmetrie auch hier erkennbar ist, da das Oktavrauschen eine gewisse Bandbreite aufweist. Ein Oktavfilter ist ein Filter mit relativer Bandbreite. Es gilt für den Zusammenhang von Mittenfrequenz f_m sowie oberer und unterer Grenzfrequenz f_o, f_u [Moe12].

$$f_o = 2f_u \quad f_m = \sqrt{f_u f_o} \rightarrow f_m = \sqrt{\frac{f_o^2}{2}} \rightarrow f_o = \sqrt{2} f_m = \sqrt{2} \cdot 1000 \text{ Hz} \approx 1414 \text{ Hz} \quad (41)$$

Es zeigt sich somit der Vorteil dieser Messmethode. Sie vermittelt einen guten Eindruck der Dämpfung für Signale mit unterschiedlichen Bandbreiten. Im Besonderen können auch Signale verwendet werden, die ähnlich solcher Signale aus tatsächlichen Anwendungsfeldern sind.

6 Quelllokalisierung

Durch die Fähigkeit der Erzeugung stark gerichteter Aufnahmecharakteristika ist es möglich, einen Quellortungsalgorithmus zu verwenden, um die Position einer Schallquelle zu bestimmen.

Dies könnte zum Beispiel in einer Anwendungssituation vorstellbar sein, in welcher sich das Array in der Mitte befindet und sich die Hauptrichtung der synthetisierten Charakteristik je nach Position eines wechselnden Sprechers anpassen soll. Ein typisches Beispiel wäre eine Telefonkonferenzsituation.

Um zu demonstrieren, dass eine Ortung möglich ist, wurde hier ein einfacher Algorithmus zur Lokalisation implementiert. Hierzu werden für Blocks von einigen Millisekunden (in diesem Fall $\approx 46 \text{ ms}$) Signale unterschiedlicher Hauptrichtungen berechnet und die Signalenergie verglichen.

Maximale Winkelauflösung. Um zu evaluieren, welche Winkelauflösung sinnvoll ist, wird die Breite der Hauptkeule des synthetisierten Signals betrachtet. Als Grenze wurden -3 dB Abfall zur Hauptrichtung gewählt. Es wurde bestimmt, dass dieser Abfall bei einer Winkeldifferenz von ca. 25.1° gegeben ist (siehe Abb. 21).

Richtungsmaß der synthetisierten Richtcharakteristik für $v = 0.5$

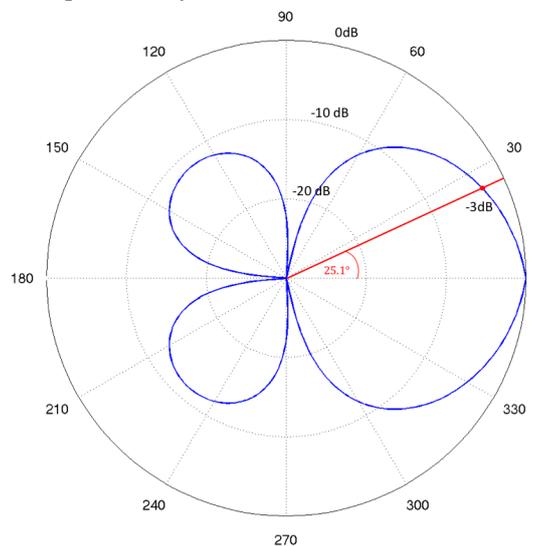


Abbildung 21: Winkel, unter dem die Hauptkeule um 3 dB abfällt

Mit diesem Ergebnis wird die Winkelauflösung mit 24° festgelegt, da sich der 360° -Bereich so gleichmäßig an 15 Punkten abtasten lässt.

Erster Test. In einem ersten Versuch zeigt sich, dass die Quellortung prinzipiell funktioniert. Ein eine Sekunde langes 1 kHz Oktavrauschen wurde der Reihe nach jeweils mit den Impulsantworten der 36 Lautsprecher auf der x-y-Ebene gefaltet. Es ergibt sich somit ein „räumlicher Sweep“, bei dem das Rauschsignal einmal in 10° -Schritten um das Array wandert. Mit der soeben bestimmten Winkelauflösung von $\Delta\alpha = 24^\circ$ wird die Lokalisation durchgeführt. In einem zeitlichen Block von $T = 2048$ Samples werden für die 15 Winkel das synthetisierte Ausgangssignal berechnet und die Signalenergie bestimmt. Diese wird hier definiert als

$$E = \frac{1}{T} \sum_{n=0}^T s^2[n]. \quad (42)$$

Der absolute Wert ist hierbei irrelevant. In Abb. 22 ist die Energie als Farbe ausgedrückt. Hellere Bereiche bedeuten eine größere Energie. Auf der y-Achse ist der Winkel aufgetragen. Der räumliche Sweep ist deutlich nachvollziehbar.

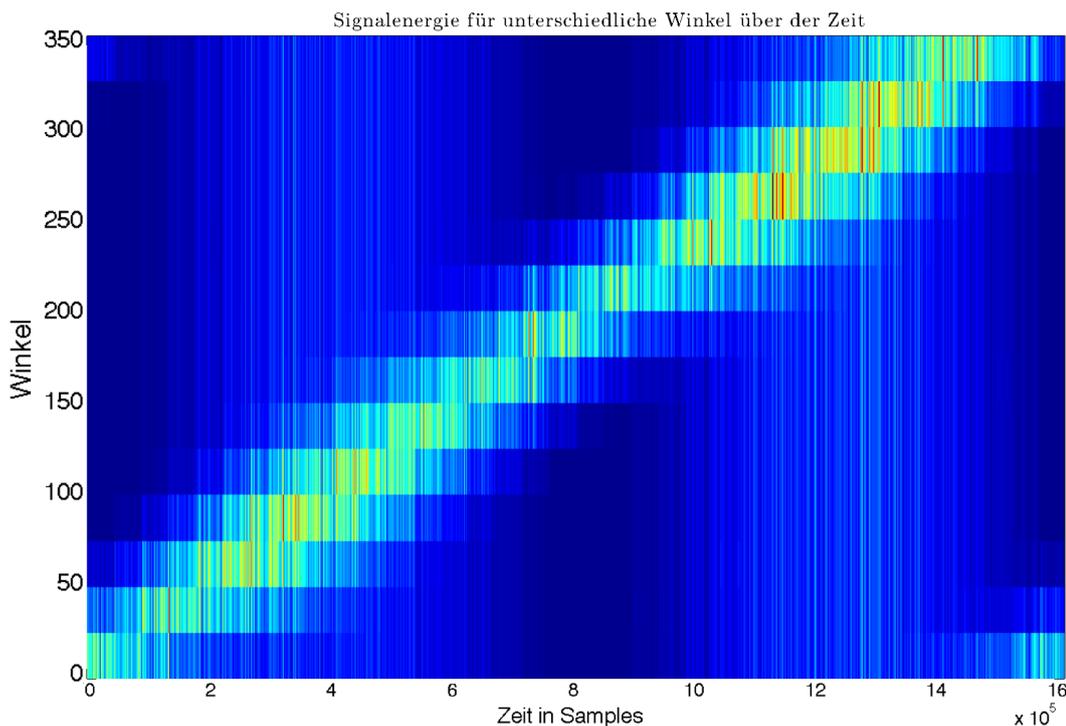


Abbildung 22: Bestimmte Signalenergie E für unterschiedliche Winkel über der Zeit mit 1 kHz Oktavrauschen an 15 Quellpositionen, analysiert mit einer Winkelauflösung von $\Delta\alpha = 24^\circ$

Im folgenden Schritt wird jeweils in jedem Block der Winkel ausgewählt, bei dem die größte Energie messbar war. Dieser ist in Abb. 23 unter der Signalamplitude für ein synthetisiertes Signal mit dem Winkel $\alpha = 0^\circ$ über die Zeit aufgetragen.

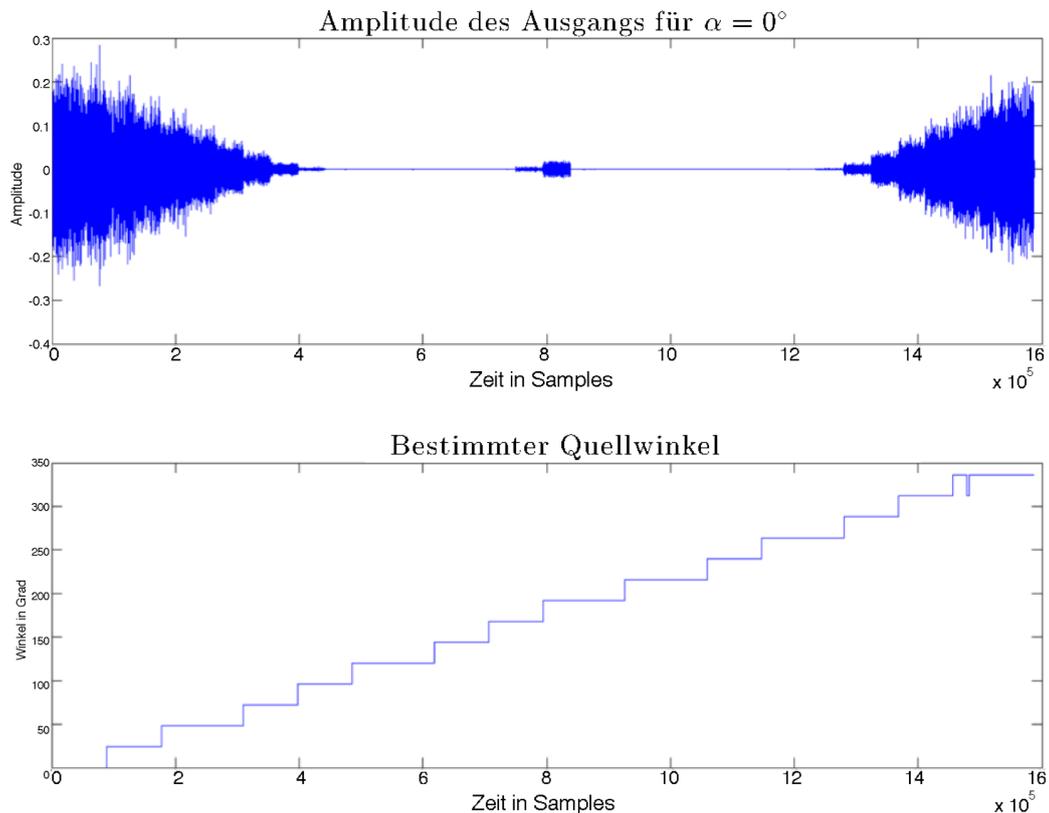


Abbildung 23: Quellortungstest mit 1 kHz Oktavrauschen, welches sich in $\Delta\varphi = 10^\circ$ -Schritten um das Array bewegt; (oben) Signalamplitude für den Winkel $\alpha = 0^\circ$; (unten) durch Auswahl des Winkels, unter dem die größte Energie bestimmt wurde, geschätzter Quellwinkel

Der Winkel ist hier wie in Abb. 22 sehr gut nachvollziehbar. Durch die Unterschiede in der Schrittgröße der Quellbewegung und der Winkelauflösung entsteht keine gleichmäßige Treppenfunktion.

Praxisnaher Test. Für einen zweiten Test wurde eine anwendungsnahe Situation gewählt. Das Array liegt hierzu auf einem Schreibtisch und ein männlicher Sprecher bespricht es zunächst aus der Richtung $\varphi \approx -45^\circ$ und dann aus der Richtung $\varphi \approx 45^\circ$. Die Entfernung beträgt ca. 1 m. Es sind Stadtgeräusche (eine mittelgroße Straße an einem normalen Arbeitstag, Bauarbeiten, Vögel) vorhanden. Das Signal wurde im breiten Frequenzbereich 100 Hz - 5 kHz bandpassgefiltert. Es wurden Zeitfenster von 4410 Samples $\hat{=}$ 100 ms gewählt.

Trotz des relativ kleinen Pegelabstandes von Signal und Störgeräuschen lässt sich die Quellposition erraten (siehe Abb. 24).

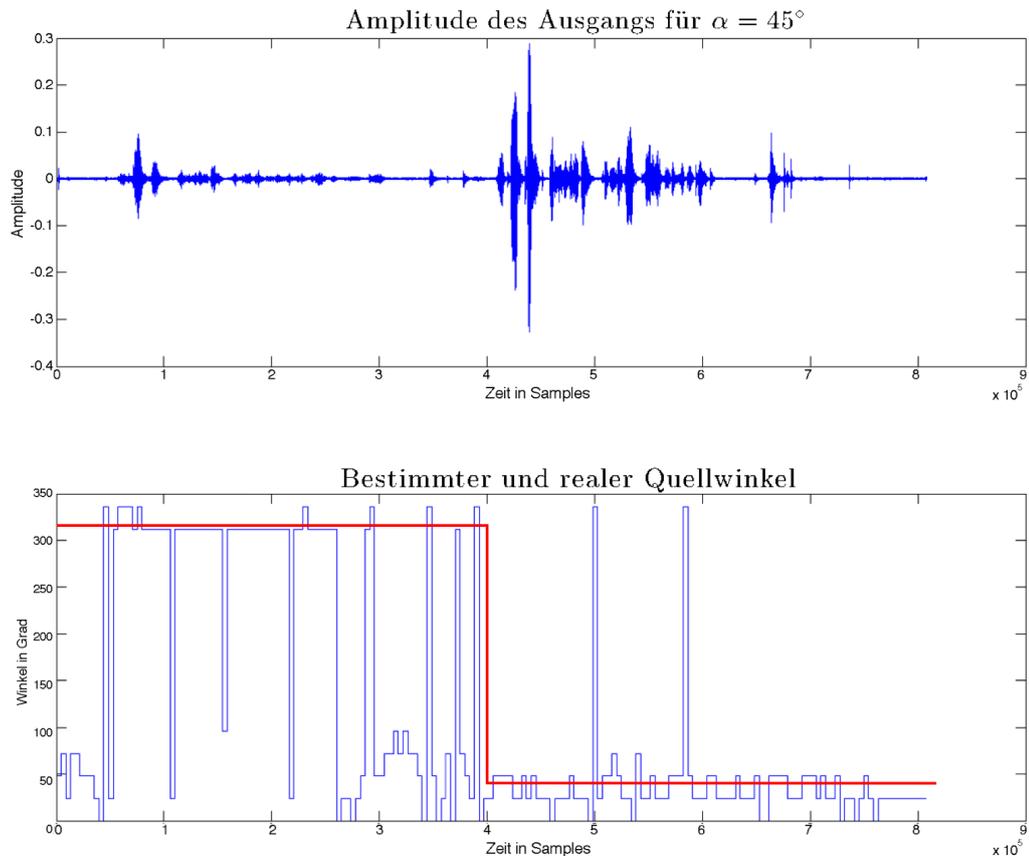


Abbildung 24: Quellortungstest mit Sprache, welche zuerst unter dem Winkel $\varphi \approx -45^\circ$ und dann unter $\varphi \approx +45^\circ$ auftritt; (oben) Signalamplitude für den Winkel $\alpha = 45^\circ$; (unten) durch Auswahl des Winkels, unter dem die größte Energie bestimmt wurde, geschätzter Quellwinkel in blau, tatsächlicher Quellwinkel in rot

Ausreißer sind in beide Richtungen zu erkennen. Diese treten vor allem in den kurzen Sprachpausen auf. Abhilfe verschaffen unter Umständen größere Zeitfenster. Allerdings verschlechtert sich dadurch die zeitliche Auflösung. Einen stärkeren Einfluss hat es außerdem eine Energieschwelle einzuführen, unterhalb dieser die bestimmte Richtung in einem Zeitfenster nicht verwendet, sondern die vorige beibehalten wird. Abb. 25 zeigt das Ergebnis. Es werden nur Winkelabschätzungen übernommen, wenn die Maximalenergie für einen Winkel in einem Zeitfenster einen bestimmten Grenzwert überschreitet. In einer Situation wie dieser ist es unproblematisch, diese Grenze durch Ausprobieren zu bestimmen. In einer Echtzeitsituation wäre das nicht so einfach möglich. Hierzu könnte zum Beispiel, wenn es um Sprachübertragung geht, eine Sprechpausenerkennung (Voice Activity Detector, VAD) verwendet werden, bei der die Energieschwelle dynamisch angepasst wird. Hierzu sind diverse Algorithmen bekannt [PSJ⁺02].

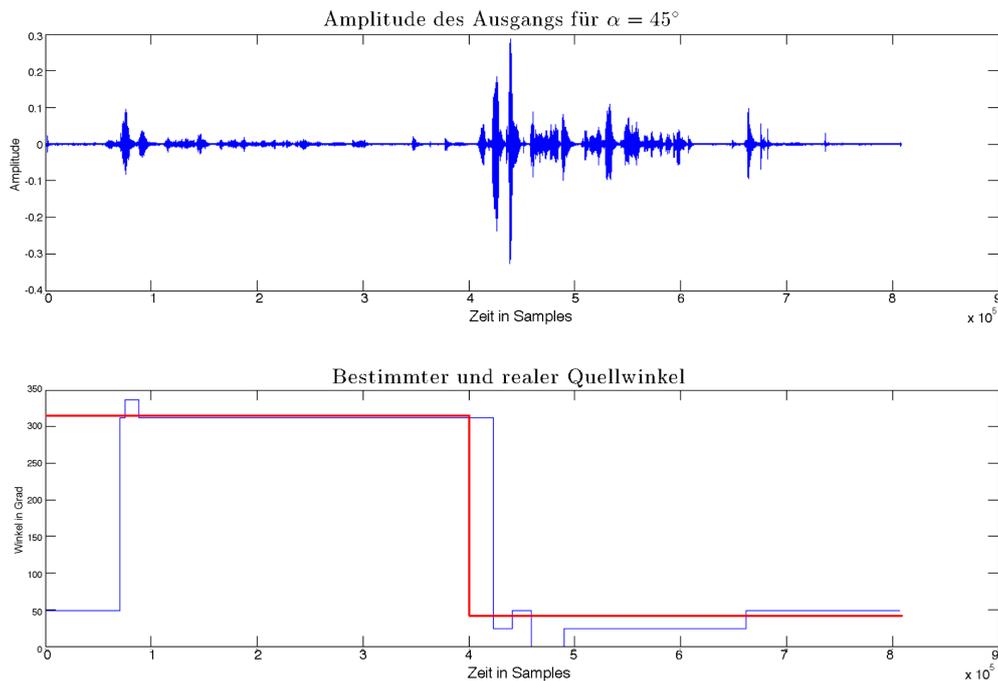


Abbildung 25: Praxistest mit einem Sprecher bei $\varphi \approx \pm 45^\circ$; (oben) Amplitude des Ausgangs mit der Hauptrichtung $\alpha = 45^\circ$; (unten) tatsächlicher Quellwinkel in rot und bestimmter Quellwinkel in blau; Im Vergleich zu Abb. 24 wurde zusätzlich eine Energieschwelle eingeführt, die in einem Zeitfenster überschritten werden muss, damit der bestimmte Winkel übernommen wird.

7 Schlussfolgerung und Ausblick

Abschließend lässt sich festhalten, dass sich der Algorithmus gut dazu eignet, gebündelte Aufnahmecharakteristika in beliebige Richtungen auf der x-y-Ebene zu erzeugen und somit Störgeräusche, die aus anderen Richtungen kommen, auszublenden. Sowohl die Richtdiagramme als auch die Messung des Signal-/Störverhältnisses belegen dies. Die Phase über alle Winkel ist hierbei pro Frequenz konstant. Die Flexibilität in der Form der erreichten Charakteristik ist jedoch gering. Der Rechenaufwand ist niedrig, sodass eine Echtzeimplementation problemlos möglich ist. Auch Quellortung lässt sich erreichen. Durch seine kompakten Abmessungen ließe sich das Array in Kombination mit dem Algorithmus an vielen Stellen verwenden.

Ausblick. Um den Algorithmus zu erweitern, kann noch untersucht werden, wie am besten eine Kontrolle der z-Richtung erreicht werden kann. Hierzu muss der Z-Kanal des B-Formats ebenfalls berechnet werden. Um die Ergebnisse zu evaluieren, müsste die Funktion zur Darstellung erweitert werden.

Bezüglich einer konkreten Anwendung wäre es zum Beispiel möglich den Algorithmus in ein System zu integrieren, welches sich für die Übertragung von wechselnden Sprechern in beispielweise einer Telefonkonferenz eignet. Es wäre möglich, die durch die Quellortung bestimmte Richtung automatisch dazu zu verwenden und die Aufnahmecharakteristik dorthin auszurichten. Ein wechselnder Sprecher könnte so verstärkt und Störgeräusche aus anderen Richtungen minimiert werden.

A Anhang

A.1 Herleitung der Normierungsfaktoren für die Berechnung des B-Formats

Zur Berechnung der Normierungsfaktoren werden die Funktionen des Richtungsgrades der Nierenmikrofone eingesetzt. Wichtig hierzu sind die Symmetrien der Winkelfunktionen,

$$\cos(-\varphi) = \cos(\varphi), \quad \sin(-\varphi) = -\sin(\varphi),$$

sowie das Additionstheorem für Winkelfunktionen [Teu96]

$$\cos(\varphi + \gamma) = \cos(\varphi)\cos(\gamma) - \sin(\varphi)\sin(\gamma).$$

Der Normierungsfaktor für den X-Kanal c_x ergibt sich nun durch Einsetzen der Funktionen des Richtungsgrades der Mikrofone 1, 3 und 5,

$$\Gamma_{m_1}(\varphi) = \frac{1 + \cos(\varphi)}{2}, \quad \Gamma_{m_3}(\varphi) = \frac{1 + \cos(\varphi + \frac{4\pi}{5})}{2}, \quad \Gamma_{m_5}(\varphi) = \frac{1 + \cos(\varphi - \frac{4\pi}{5})}{2},$$

in die in Kapitel 2.2.1 angegebene Formel.

$$\begin{aligned} \Gamma_X(\varphi) &= \frac{(\Gamma_{m_1}(\varphi) - \Gamma_{m_3}(\varphi)) + (\Gamma_{m_3}(\varphi) - \Gamma_{m_5}(\varphi))}{c_x} = \\ &= \frac{\cos(\varphi) - \cos(\varphi + \frac{4\pi}{5})}{2c_x} + \frac{\cos(\varphi) - \cos(\varphi - \frac{4\pi}{5})}{2c_x} = \\ &= \frac{1}{2c_x} \cdot [\cos(\varphi) - \cos(\varphi)\cos(\frac{4\pi}{5}) - \sin(\varphi)\sin(-\frac{4\pi}{5}) + \cos(\varphi) - \cos(\varphi)\cos(-\frac{4\pi}{5}) - \sin(\varphi)\sin(\frac{4\pi}{5})] = \\ &= \frac{1}{2c_x} \cdot [\cos(\varphi) - \cos(\varphi)\cos(\frac{4\pi}{5}) + \cos(\varphi) - \cos(\varphi)\cos(\frac{4\pi}{5})] = \\ &= \frac{1}{2c_x} \cdot \cos(\varphi) \cdot (1 + 1 - \cos(\frac{4\pi}{5}) - \cos(\frac{4\pi}{5})) \stackrel{!}{=} \cos(\varphi) \end{aligned}$$

Durch Koeffizientenvergleich ergibt sich

$$\frac{1}{2c_x} \cdot (2 - 2\cos(\frac{4\pi}{5})) = 1 \rightarrow c_x = 1 - \cos(\frac{4\pi}{5}) \approx 1.8090.$$

Für den Y-Kanal wird gleichermaßen vorgegangen. Die Funktionen des Richtungsgrades der Mikrofone 2 und 4,

$$\Gamma_{m_2}(\varphi) = \frac{1 + \cos(\varphi + \frac{2\pi}{5})}{2}, \quad \Gamma_{m_4}(\varphi) = \frac{1 + \cos(\varphi - \frac{2\pi}{5})}{2},$$

werden in die Formel aus Kapitel 2.2.1 eingesetzt.

$$\begin{aligned} \Gamma_Y(\varphi) &= \frac{(\Gamma_{m_2}(\varphi) - \Gamma_{m_4}(\varphi))}{c_y} = \frac{1}{2c_y} \cdot [1 + \cos(\varphi + \frac{2\pi}{5}) - \cos(\varphi - \frac{2\pi}{5})] \\ &= \frac{1}{2c_y} \cdot [\cos(\varphi)\cos(\frac{2\pi}{5}) - \sin(\varphi)\sin(\frac{2\pi}{5}) - \cos(\varphi)\cos(-\frac{2\pi}{5}) + \sin(\varphi)\sin(-\frac{2\pi}{5})] \\ &= \frac{1}{2c_y} \cdot [-\sin(\varphi)\sin(\frac{2\pi}{5}) - \sin(\varphi)\sin(\frac{2\pi}{5})] \stackrel{!}{=} \cos(\varphi + \frac{\pi}{2}) = -\sin(\varphi) \\ &= \frac{1}{2c_y} \cdot [\sin(\frac{2\pi}{5}) + \sin(\frac{2\pi}{5})] = 1 \rightarrow c_y = \sin(\frac{2\pi}{5}) \approx 0.9511 \end{aligned}$$

Die Herleitung wird vor allem für den Normierungsfaktor c_w deutlich übersichtlicher, wenn der Zusammenhang

$$\begin{aligned} \cos(\varphi+\gamma)+\cos(\varphi-\gamma) &= \cos(\varphi)\cos(\gamma)-\sin(\varphi)\sin(\gamma)+\cos(\varphi)\cos(\gamma)+\sin(\varphi)\sin(\gamma) = \\ &= 2\cos(\varphi)\cos(\gamma) \end{aligned}$$

verwendet wird. Für den W-Kanal lassen sich damit die symmetrisch zur 0° -Achse liegenden Mikrofone sofort zusammenfassen und es ergibt sich

$$\begin{aligned} \Gamma_w(\varphi) &= \frac{\Gamma_{m_1}(\varphi) + \Gamma_{m_2}(\varphi) + \Gamma_{m_3}(\varphi) + \Gamma_{m_4}(\varphi) + \Gamma_{m_5}(\varphi)}{c_w} = \\ &= \frac{5}{2c_w} + \frac{1}{2c_w} (\cos(\varphi) + 2\cos(\varphi)\cos(\frac{2\pi}{5}) + 2\cos(\varphi)\cos(\frac{4\pi}{5})) = \\ &= \frac{5}{2c_w} + \frac{1}{2c_w} \cos(\varphi) (1 + 2\cos(\frac{2\pi}{5}) + 2\cos(\frac{4\pi}{5})) = \\ &= \frac{5}{2c_w} + \frac{1}{2c_w} \cos(\varphi) \cdot 0 = \frac{5}{2c_w} \stackrel{!}{=} 1 \rightarrow c_w = \frac{5}{2} = 2.5. \end{aligned}$$

A.2 Echtzeitimplementation in Pure Data

Um den behandelten Algorithmus auch in der Praxis verwenden zu können, wurde eine Implementation in der grafischen Programmiersprache Pure Data (Pd) realisiert.

Die Verwendung ist einfach und das Programm eignet sich sowohl zur Echtzeitverarbeitung zusammen mit dem getesteten Mikrofonarray als auch zur Demonstration der Funktion mit im Vorhinein aufgenommenen Signalen.

A.2.1 Implementation

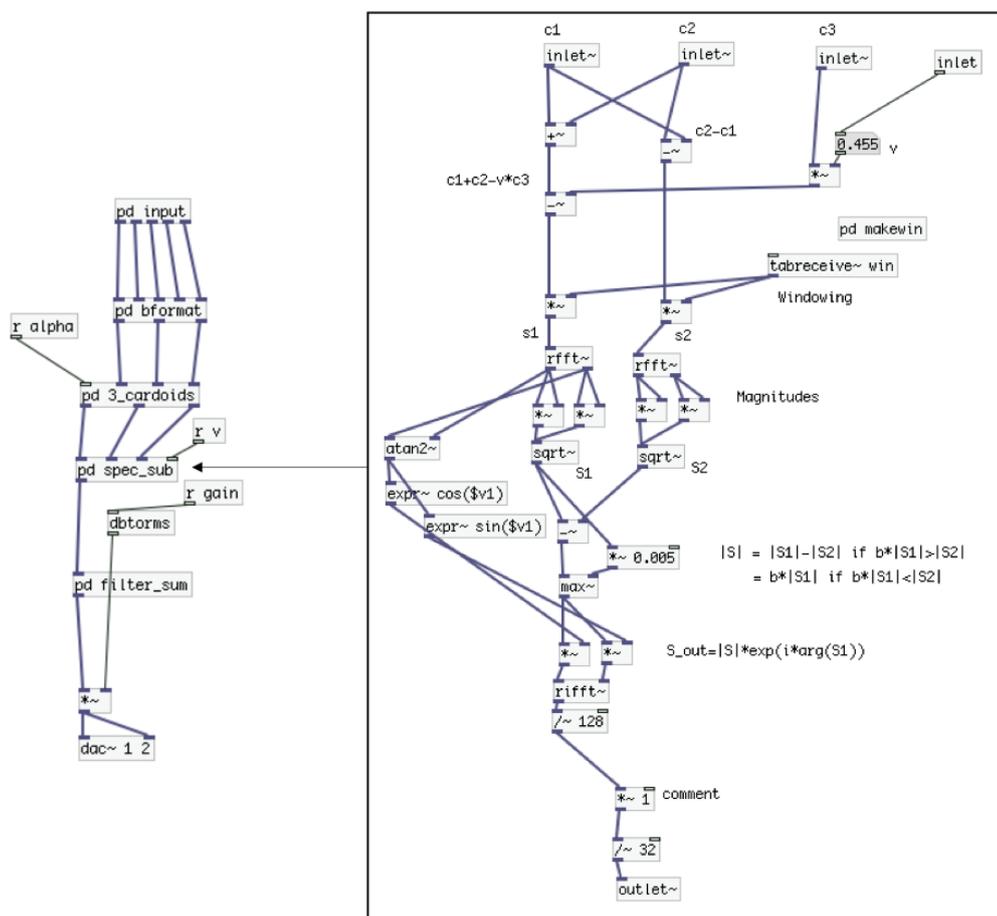


Abbildung 26: Implementation in Pd: subpatch *guts* mit subpatch *spec_sub*

In Abb. 26 ist links das subpatch *guts* zu sehen. Während das mainpatch die Bedienoberfläche enthält, ist hier der eigentliche Algorithmus aufgebaut. Einer der großen Vorteile einer grafischen Programmiersprache wie Pd ist es, dass man sich beim Programmieren am Blockschaltbild orientieren kann.

Dadurch, dass der Aufbau von *guts* dem Blockschaltbild in Abb. 2 sehr ähnlich sieht, ist auch die Funktion leicht nachvollziehbar. Das subpatch *spec_sub* enthält die spektrale Subtraktion und ist rechts abgebildet.

Kompatibilität. Die vollständige Funktion ist unter Pd extended gegeben. Es muss die Library *flatspace iemlib* vorhanden sein, da zur Filterung der Eingänge und des Ausgangs das FIR~Objekt zur Verwendung kommt. Ist die Library nicht vorhanden, kann der Algorithmus durch Deaktivierung der Filterung (siehe "Bedienung") aber trotzdem verwendet werden. Das atan2~Objekt aus der *cyclone flatspace ggee*-Library muss jedoch in jedem Fall vorhanden sein. Es ist notwendig, um die Phase von s_1 zu berechnen.

A.2.2 Bedienung

In Abb. 27 ist die Bedienoberfläche zu sehen. Im Folgenden werden kurz die Optionen umrissen.

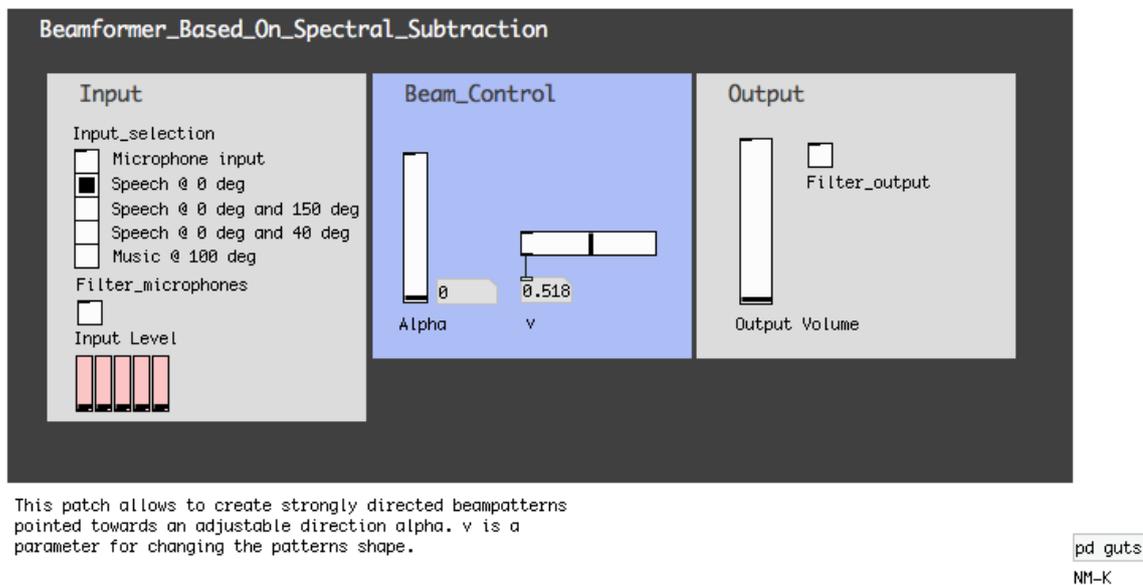


Abbildung 27: Bedienoberfläche der Pd-Implementation

Input. In der ersten Sektion befindet sich die Auswahl, welcher Eingang verwendet wird. Ist *Microphone input* selektiert, so werden die fünf Nierensignale des angeschlossenen Arrays als Eingang verwendet. In diesem Fall sollten diese vom

Mikrofon 1 ($\varphi = +144^\circ$) bis zu Mikrofon 5 ($\varphi = -144^\circ$) der Reihe nach an den ersten fünf Audioeingängen angeschlossen sein. Um die richtigen Eingänge zu wählen, muss man unter *guts\input* die entsprechenden dac-Kanäle einstellen. Die 5 Slider unter *Input Level* zeigen die Eingangspegel der Mikrofone an.

Ist die Eingangsoption *speech @ 0 deg* gewählt, so ist kein angeschlossenes Array notwendig. Es wird als Sprachsignal eine weibliche Sprecherin durch Faltung mit der Impulsantwort des ersten Lautsprechers ($\varphi = 0^\circ, \vartheta = 90^\circ$) in den virtuellen Raum projiziert (siehe Kapitel 5.3). Für die Option *Music @ 100 deg* wird in gleicher Art und Weise vorgegangen, nur ist das Musiksignal unter ($\varphi = 100^\circ, \vartheta = 90^\circ$) zu finden. Für das Demonstrationsbeispiel *Speech @ 0 deg and 150 deg* ist jeweils eine Sprecherin unter $\varphi = 0^\circ$ und $\varphi = 150^\circ$ zu finden. Durch Variation der Hauptrichtung α kann dann eine räumliche Filterung vorgenommen werden. Bei *Speech @ 0 deg and 40 deg* ist die gleiche Situation vorhanden, nur dass die Sprecherinnen deutlich näher aneinander liegen. Die Option *Filter_microphones* ermöglicht ein Ein- und Ausschalten der Ausgleichsfilter für die Mikrofone untereinander (vgl. Kap. 4.1).

Beam Control. In dieser Sektion erfolgt die Kontrolle über den Algorithmus. Mit dem vertikalen Regler lässt sich der Winkel der Hauptrichtung α einstellen und mit dem horizontalen Slider v der Parameter zur Form der synthetisierten Charakteristik (siehe Kap. 2.2).

Output. In der letzten Sektion befindet sich der Regler *Output Volume*, mit dem die gesamte Lautstärke des synthetisierten Signals eingestellt werden kann. Mit *Filter_Output* besteht die Möglichkeit, die Entzerrung des Ausgangssignals ein- oder auszuschalten. Die Ausgangskanäle werden in *guts* unten eingestellt.

Literatur

- [BSM79] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP 79*, vol. 4, 1979.
- [CG77] P. Craven and M. Gerzon, “Coincident microphone simulation covering three dimensional space and yielding various directional outputs,” US Patent 4,042,779, 1977.
- [Far00] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *Audio Engineering Society Convention 108*, 2000.
- [JFR04] F. Jacobsen, S. B. Figueroa, and K. Rasmussen, “A note on the concept of acoustic center,” *J. Acoust. Soc. Am.*, vol. 115, no. 2, pp. 1468–1473, 2004.
- [Moe12] M. Moeser, *Technische Akustik, 9. Auflage*. VDI, 2012.
- [Pom08] H. Pomberger, “Angular and radial directivity control for spherical loudspeaker arrays,” Diplomarbeit, Institut für Elektronische Musik und Akustik, Kunstuni Graz, Technical University Graz, Graz, A, 2008.
- [PSJ+02] R. V. Prasad, A. Sangwan, H. Jamadagni, C. M.C, R. Sah, and V. Gaurav, “Comparison of voice activity detection algorithms for voip,” in *Proceedings of the Seventh International Symposium on Computers and Communications (ISCC 020)*, vol. 2, 2002, pp. 1530–1546.
- [Rei11a] F. Reining, “Method for synthesizing a microphone signal,” Europäisches Patent EP 2 208 362 B1, 2011.
- [Rei11b] ———, “Microphone arrangement comprising 3 pressure gradient transducers,” Europäisches Patent EP 2 208 360 B1, 2011.
- [Teu96] B. Teubner, *Taschenbuch der Mathematik*. Teubner Verlagsgesellschaft, 1996.
- [Wi199] E. G. Williams, *Fourier Acoustics*. Academic Press, 1999.
- [Zau12] M. Zaunschirm, “Modal beamforming using planar circular microphone arrays,” Diplomarbeit, Institut für Elektronische Musik und Akustik, Kunstuni Graz, Technical University Graz, Graz, A, 2012.
- [ZZ93] M. Zollner and E. Zwicker, *Elektroakustik, 3. Auflage*. Springer-Lehrbuch, 1993.