

Richtungshörmodell mit Elevation und Korrektur der Mehrdeutigkeiten

Diplomarbeit

von **Sebastian Preis**

Graz, 7. Juni 2011

Beurteiler: Robert Höldrich, O. UnivProf. Dr. techn. DI. Mag.

Betreuer: Franz Zotter, Dr. DI

Matthias Frank, DI



institut für elektronische musik und akustik



Abstract

Ziel der folgenden Arbeit ist es, ein Lokalisationshörmodell zu entwickeln und in einer Simulation wie auch in einer Echtzeitanwendung umzusetzen. Die Grundlage des Modells bildet das Lokalisationsmodell von Breebaart, welches in dieser Arbeit genauer untersucht und erweitert wird. Gängige Lokalisationsmodelle detektieren nur Richtungen in der Horizontalebene, wobei stets eine Vorne-/Hintenverwechslung der Schallquellen auftreten kann. Der Lokalisationsbereich des Modells, welches in dieser Arbeit entwickelt wird, wird um die Elevation erweitert. Eine weitere Aufgabe ist es die auf den Elevationsbereich erweiterten Mehrdeutigkeiten (cone-of-confusion) zu minimieren. Dafür wird eine Lokalisationsszene mit mehreren Stellungen der Kopfdrehung um seine Achse analysiert. Mit der Annahme, dass die Lokalisationsszene statisch ist und sich der Kopf in dieser bewegt, können die Analyseergebnisse mit ihren unterschiedlichen Mehrdeutigkeiten so zusammengefasst werden, dass ein eindeutigeres Ergebnis entsteht. Abschließend soll das Modell dazu befähigen, einen Rückschluss auf die wahrgenommene Breite einer Schallquelle zuzulassen.

The main goal of this diploma thesis is to investigate an auditory localization model aiming at a real-time implementation. The model is based on the model of Breebaart which is reviewed in this thesis, considering potential improvement. Present models are able to detect sources regarding their azimuth angle in the horizontal plane with front-/back confusion. Within the presented work, an extension of the models are sought to provide localization of the elevation angle of sound sources. Moreover, the remaining ambiguities (cone-of-confusion) can be corrected by additional means. For that purpose, the localisation scene is analyzed using several orientations of the head with regard to its rotation axis. Under the assumption of a static localization scene in which the head moves, the results of the various ambiguous analyses can be joined to obtain a unique result. Concluding, this model provides an estimation of the apparent source width.

Inhaltsverzeichnis

1	Einführung in die Lokalisation	4
1.1	Grundlegende Begriffe	5
1.2	Das Ohr	6
1.2.1	Das Außenohr	7
1.2.2	Das Mittelohr	8
1.2.3	Das Innenohr	10
1.3	Eigenschaften des Gehörs	12
1.3.1	Phase Locking	12
1.3.2	Komprimierende Eigenschaft des Gehörs	12
1.3.3	Frequenzgruppenbildung des Gehörs	13
1.4	Lokalisation	15
1.4.1	Lokalisation in der Horizontalebene	15
1.5	Dreidimensionale Lokalisation	17
1.6	Modelle für eine binaurale Lokalisation	19
1.7	Wahrgenommene Schallquellengröße und Lokalisationsunschärfe	21
2	Binaurales Lokalisationsmodell	23
2.1	Auditive Peripherie	24
2.1.1	Übertragungspfad Außen- und Mittelohr	24
2.1.2	Übertragungspfad Innenohr	27
2.1.3	Neuronaler Übertragungspfad	28
2.2	Binaurale Lokalisationsstufe	31
2.2.1	Das EI-Muster	33
2.2.2	EI-Mustervergleich	40
2.2.3	Verringerung der Richtungsmehrdeutigkeiten durch eine simulierte Kopfdrehung	44
2.3	Berechnung der Quellenbreite und der Lokalisationsunschärfe	48
2.4	Echtzeitumsetzung des Lokalisationsmodells mit Pd	53

3	Evaluierung des Lokalisationsmodells	55
3.1	Ergebnisse unter simulierten Bedingungen	55
3.1.1	Eine Quelle mit weißem Rauschen	55
3.1.2	Eine Quelle mit Schmalbandrauschen	58
3.1.3	Eine Sinusquelle	60
3.1.4	Eine Quelle mit menschlicher Stimme	62
3.1.5	Zwei korrelierte Quellen	64
3.1.6	Zwei nichtkorrelierte Quellen	65
3.2	Ergebnisse in einem echten Raum	68
4	Zusammenfassung	73

Kapitel 1

Einführung in die Lokalisation

Motiviert wurde diese Arbeit durch die Idee ein computerbasiertes Tool zu entwickeln, welches es ermöglicht die genaue Richtung einer Schallquelle zu lokalisieren und deren Quellenbreite zu bestimmen. Dieses Tool wäre eine Brücke zwischen realen oder durch Lautsprecher wiedergegeben Schallquellen und der menschlichen Wahrnehmung, und soll die Antwort dieser auf unterschiedlichste Lokalisationsszenen abbilden. Auf diese Art könnten wahrnehmungsbezogene menschliche Reaktionen ermittelt werden, die damit natürlich auch Rückschlüsse auf die Funktionalität und Vergleichbarkeit von unterschiedlichen Wiedergabesystemen zulassen. Bisher musste jede Fragestellung an die Hörwahrnehmung durch Hörversuche mit Versuchspersonen beantwortet werden. Jeder Versuch mit Versuchspersonen bedeutet viel Aufwand und Zeit. Durch ein referenziertes computerbasiertes Tool könnten zumindest Vorabuntersuchungen ohne Versuchspersonen durchgeführt werden. Auch könnte dabei auf den Versuchsaufbau verzichtet werden, da man die menschliche Wahrnehmung durch eine geeignete Simulation abbilden könnte. In dieser Diplomarbeit wird daher ein Lokalisationshörmodell entwickelt, welches möglichst exakt mit der menschlichen Hörwahrnehmung übereinstimmt. Deshalb werden die physiologischen Eigenschaften des Ohrs und der Nervenleitung möglichst genau umgesetzt und die Abbildung der Lokalisationsrichtung, Quellenbreite und Lokalisationsunschärfe mit den Antworten von bekannten Hörversuchen verifiziert.

Im ersten Kapitel werden die Grundlagen des menschlichen Gehörs und der Lokalisation beschrieben. Das zweite Kapitel widmet sich der Umsetzung des Lokalisationshörmodells und der Ermittlung der Quellenbreite sowie der Lokalisationsunschärfe. Im dritten Kapitel wird die Funktionalität des Modells evaluiert.

1.1 Grundlegende Begriffe

Um die physiologischen und psychoakustischen Eigenschaften des Gehörs beschreiben zu können, müssen die damit zusammenhängenden und grundlegenden Begriffe erklärt werden.



Abbildung 1.1: Horizontalebene (links), Medianebene (rechts) [Maj10]

Die interaurale Achse befindet sich zwischen den oberen Kanten der beiden Ohrkanäle. Die in Abb. 1 grün eingezeichnete Horizontalebene wird von der interauralen Achse und der Augenunterkante aufgespannt. Die Medianebene (Abb. 2 grün) ist orthogonal zur Horizontalebene und zur interauralen Achse [Bla74]. Sie teilt den Kopf in zwei Hälften.

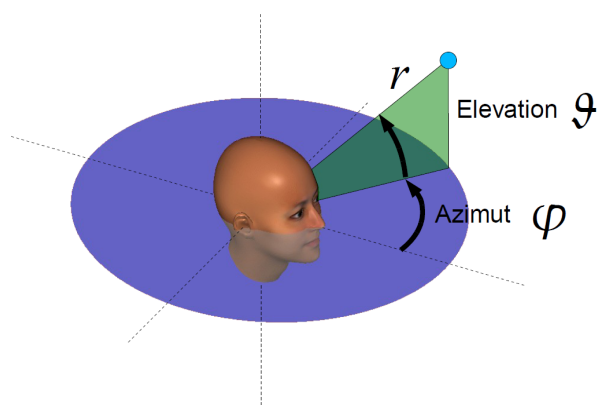


Abbildung 1.2: Position einer Schallquelle [Maj10]

Die Position einer Schallquelle wird in Polarkoordinaten angegeben. Hierbei ist φ der im mathematisch positiven Sinn definierte Azimutwinkel, ϑ der Elevationswinkel und r (Radius) die Entfernung der Quelle (siehe Abb. 1.2).

1.2 Das Ohr

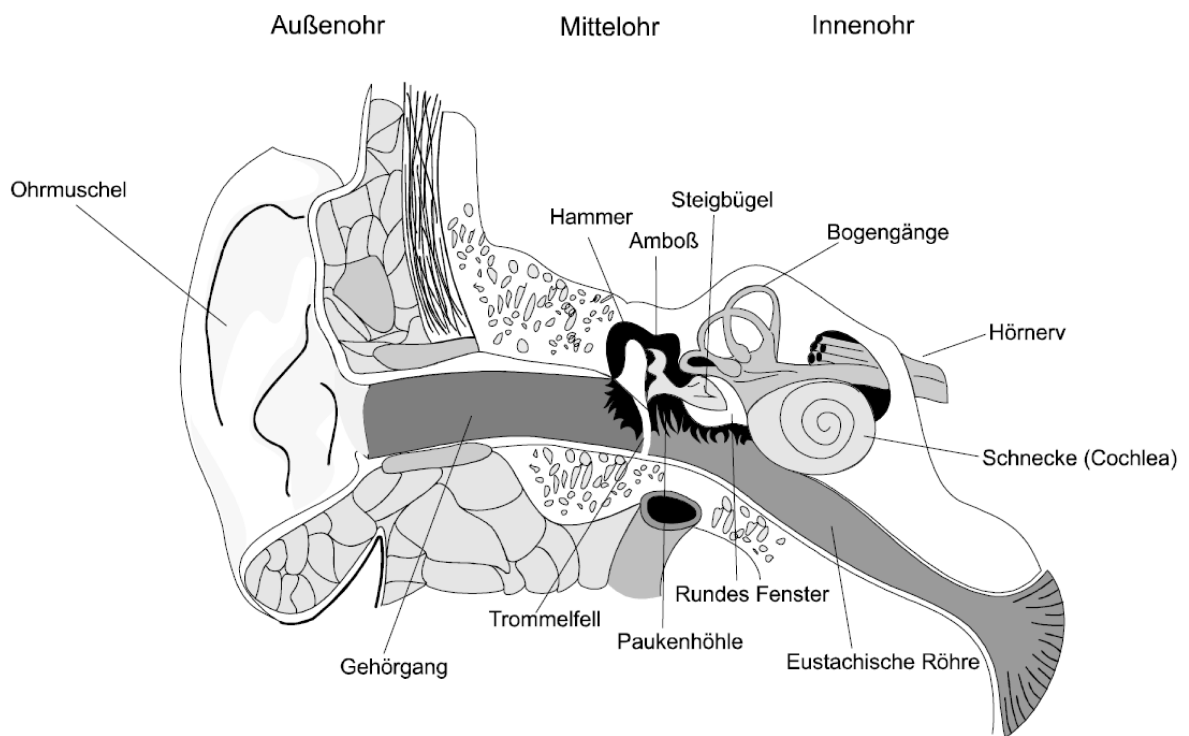


Abbildung 1.3: Physiologie des Ohrs [Lab06]

1.2.1 Das Außenohr

Das Außenohr besteht aus der Ohrmuschel (Pinna) und dem Gehörgang. Der Gehörgang ist leicht verformt, wodurch das Trommelfell geschützt wird. Die Pinna wirkt wie ein akustischer Trichter und bündelt hohe Frequenzen. Der Gehörgang verhält sich wie ein $\lambda/4$ Resonator (siehe Abb. 1.4). Damit befinden sich die Resonanzfrequenzen bei $\lambda/4$, $3\lambda/4$, $5\lambda/4$ usw., also bei allen ungeradzahigen Vielfachen der ersten Resonanzfrequenz [Pfl00].

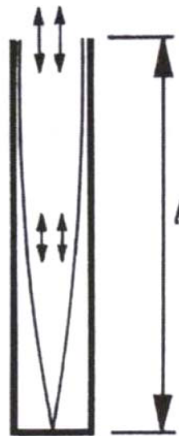


Abbildung 1.4: Resonanzfrequenz bei $l=\lambda/4$ eines $\lambda/4$ Resonators

Die Länge l des Gehörgangs beträgt ca. 0.02 m. Daraus folgt die erste Resonanzfrequenz

$$\frac{\lambda}{4} = 0.02m \quad ,$$

$$f = \frac{c}{\lambda} = \frac{340}{0.08} = 4.2kHz \quad . \quad (1.2.1)$$

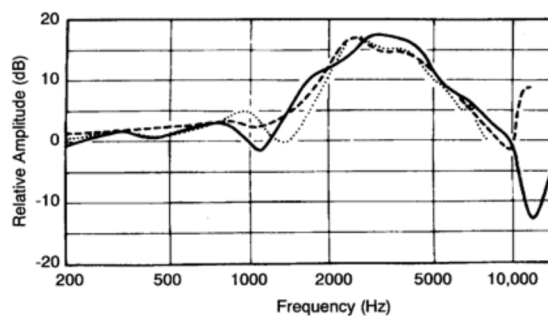


Abbildung 1.5: Übertragungsfunktion des Außenohrs bei frontal platzierter Schallquelle [Gel97]

Da das Trommelfell schräg im Gehörgang liegt, werden stehende Wellen bei hohen Frequenzen vermieden.

Ist die Wellenlänge des einfallenden Schalls klein (ab 4 kHz) gegenüber der Pinnaabmessung, kommt es zu einer richtungsabhängigen Filterung durch Überlagerung des Direktschalls mit den von der Pinna verursachten Reflexionen. Dies ist wichtig für die Lokalisation in der Medianebene.

1.2.2 Das Mittelohr

Das Mittelohr setzt sich aus den Gehörknöchelchen Hammer, Amboss und Steigbügel (siehe Abb. 1.3) zusammen und dient als Hebelübersetzung zwischen dem Trommelfell und dem ovalen Fenster (Innenohr) [Pff00]. Zwischen der Luft und der sich in der Schnecke befindenden Lymphflüssigkeit besteht auf Grund der unterschiedlichen Impedanzen eine Fehlanpassung. Würden Schallwellen direkt auf das ovale Fenster treffen, würden sie reflektiert werden.

Das Mittelohr erreicht eine Anpassung durch folgende 3 Eigenschaften:

- eine große Trommelfellfläche wird auf eine kleine Fläche (ovales Fenster) übersetzt: 22:1
- Krümmung des Trommelfells (Prinzip der gekrümmten Membran): 1:2
- Hebelverhältnis (langer Hammergriff und kurzer Ambossfortsatz): 1:1.2

Dadurch ergibt sich eine ca. 50-fache Kraftverstärkung pro Flächeneinheit [Lab06].

Das Mittelohr verhält sich ähnlich einem Feder-Masse-System das bei 1 kHz eine optimale Anpassung zwischen Luft und Lymphflüssigkeit besitzt. Ab Frequenzen größer 1 kHz wird die Übertragung durch Massenträgheit bedämpft [Pfl00], wodurch das Mittelohr eine Tiefpasscharakteristik aufweist (siehe Abb.1.6).

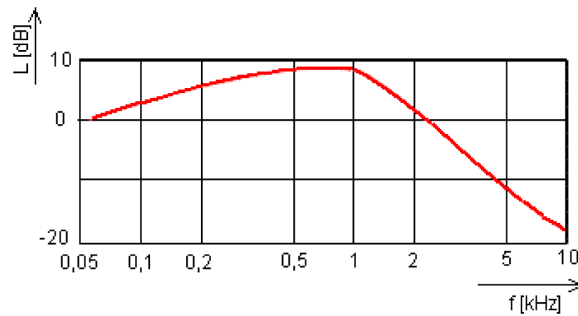


Abbildung 1.6: Übertragungsfunktion des Mittelohrs mit Tiefpasscharakteristik [Wup]

Eine weitere Aufgabe des Mittelohrs ist der Druckausgleich zwischen dem Druck außerhalb und innerhalb des Trommelfells. Das Mittelohr ist durch die Eustachische Röhre mit dem Nasen-Rachenraum verbunden. Durch Schlucken und Gähnen wird ein Druckausgleich erzeugt.

Die Übertragungsfunktion des Aussen- und Mittelohrs entspricht der inversen 100 Phon-Kurve aus ISO 226 (siehe Abb. 1.7).

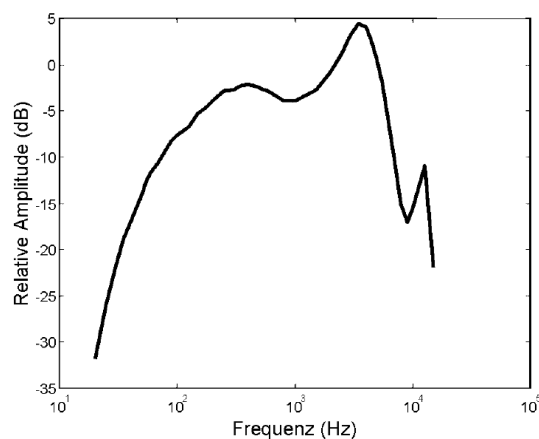


Abbildung 1.7: Übertragungsfunktion des Außenohrs und Mittelohrs [Lab06]

1.2.3 Das Innenohr

Das Innenohr besteht aus dem Gleichgewichtsorgan und der Schnecke (Cochlea). Die Cochlea hat $2\frac{1}{2}$ Windungen und ist im ausgerollten Zustand 35 mm lang [Lab06].

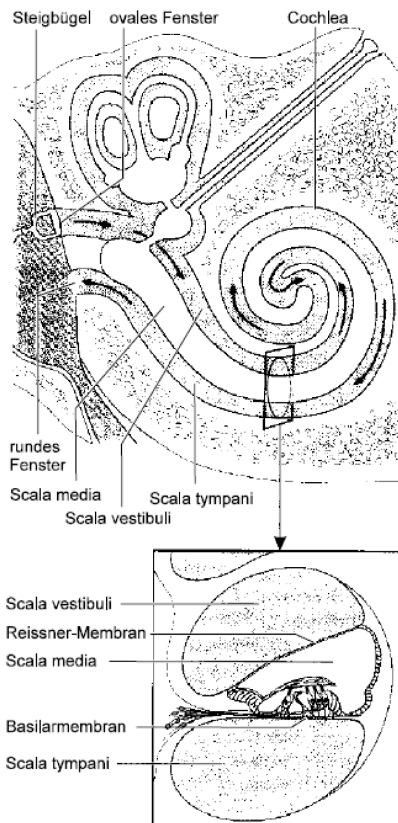


Abbildung 1.8: Schematischer Aufbau des Innenohrs: Längsschnitt (oben) Querschnitt (unten) [KKD08]

In der Cochlea verlaufen 3 parallele Kanäle: die Scala Tympani, Scala Media und die Scala Vestibuli. Die Basilarmembran liegt zwischen der Scala Tympani und der Scala Media. Die Scala Tympani und die Scala Vestibuli sind mit Perilymphe und die Scala Media mit Endolymphe gefüllt, wodurch sich eine Spannungsdifferenz von 40 mV ergibt, die wiederum an der Basilarmembran anliegt [Lab06].

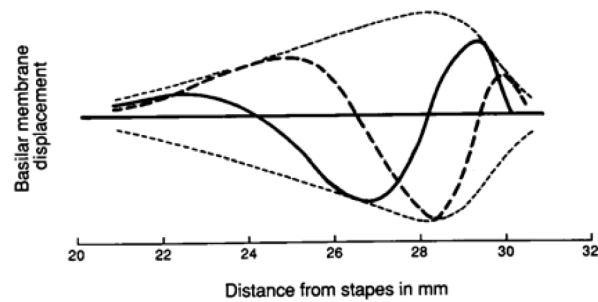


Abbildung 1.9: Ort des frequenzabhängigen Maximums einer Wanderwelle [Lab06]

Schallwellen, die nach der Übersetzung vom Mittelohr über das ovale Fenster in die Scala Vestibuli eingepreßt werden, erzeugen Druckunterschiede in der Lympflüssigkeit, die sich von der Scala Vestibuli in die Scala Tympani über eine Öffnung (Helikotrema) fortpflanzen und dort, durch das am Ende der Scala Tympani liegende runde Fenster, ausgeglichen werden (siehe Abb. 1.8). Durch die Druckunterschiede entsteht an der Basilarmembran eine Wanderwelle (siehe Abb. 1.9). Der Ort auf der Membran an dem sich das Maximum der Wanderwelle befindet ist frequenzabhängig. Dies lässt sich durch die breiter werdende Fläche wie durch eine Abnahme der Steifigkeit der Basilarmembran erklären. So befindet sich das Maximum hoher Frequenzen am Anfang der Basilarmembran und verlagert sich bei Abnahme der Frequenz zum Ende der Membran.

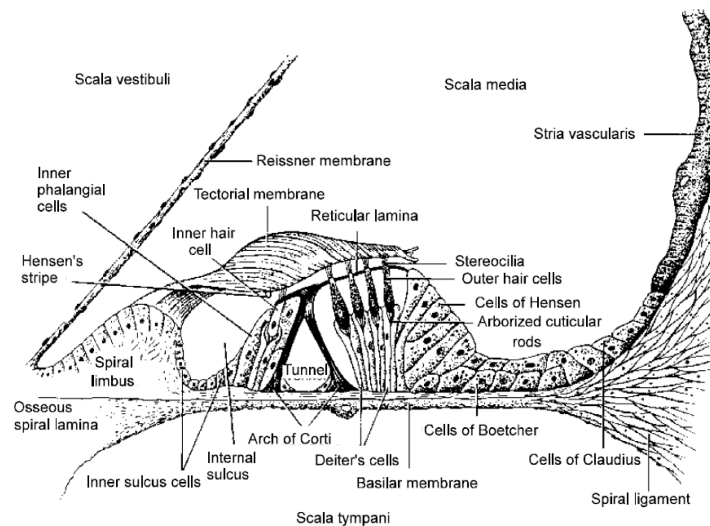


Abbildung 1.10: Stark vergrößerte Darstellung eines Cochleaerschnittes [GGF89]

1.3 Eigenschaften des Gehörs

1.3.1 Phase Locking

Die sich auf der Basilarmembran befindenden Haarzellen, die über die Haare (Stereozilien) mit der Tektorialmembran verbunden sind (siehe Abb. 1.10) werden durch Scherkräfte zwischen Basilarmembran und Tektorialmembran ausgelenkt und depolarisieren. Dadurch wird ein Aktionspotential ausgelöst (Spikes). Die Depolarisation kann nur durch die Auslenkung in einer Richtung ausgelöst werden. Dies bedeutet, dass bei einer Schwingung eine in Spikes kodierte Halbwellenrichtung stattfindet. Die Neuronen feuern synchron zur Phase des halbwellenrichteten Signals, dies nennt sich Phase Locking (siehe Abb. 1.11). Die maximale Feuerrate liegt bei 1000 Spikes/s (absolute Refraktionszeit: 1ms). Dadurch nimmt die die Präzision des Phase Lockings ab 1 kHz kontinuierlich ab, da die Refraktionszeit zu lange dauert. Dies bedeutet, dass die Phaseninformation mit steigender Frequenz verloren geht, da einzelne Phasen vermehrt ausgelassen werden.

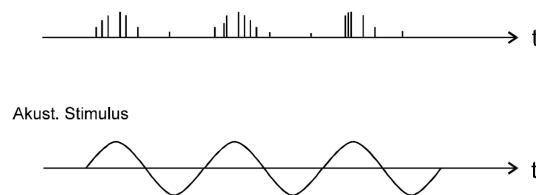


Abbildung 1.11: Phase Locking der Neuronen: Spikes synchron zur Phase der oberen Halbwellen (oben) akustischer Stimulus (unten) [Lab06]

1.3.2 Komprimierende Eigenschaft des Gehörs

Es gibt 3 unterschiedliche Neuronen die sich durch ihre Schwellen unterscheiden. Das ist der minimale Unterschied bei dem Veränderungen der neuronalen Antwort (Spikes/s) registriert werden. Neuronen mit niedrigen Schwellen kodieren eine kleinere Dynamik und lösen den Dynamikumfang genauer auf (20 dB kleiner Dynamikumfang bis 60 dB großer Dynamikumfang). Die 3 Neuronentypen unterscheiden sich zusätzlich durch ihre Spontanfeuerrate, das ist die Feuerrate, die ohne Eingangsstimulus erfolgt. Neuronen mit niedriger Schwelle haben eine höhere Spontanfeuerrate.

Da Neuronen mit höheren Schwellen einen größeren Dynamikbereich auflösen, ist die Eingangs-Ausgangsfunktion der Basilarmembran, wie in Abb. 1.12 ersichtlich nichtlinear (Steigung nimmt ab). Das Signal wird komprimiert.

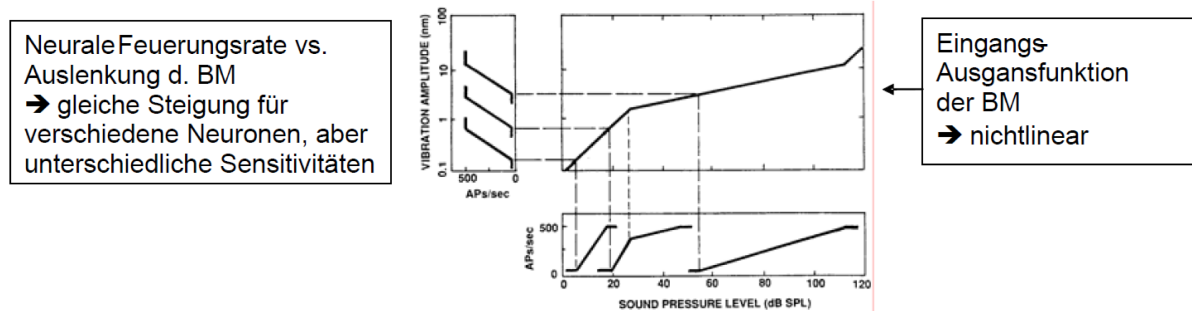


Abbildung 1.12: Komprimierte Ein- Ausgangsfunktion aufgrund unterschiedlicher Auflösung und unterschiedlichen Schwellen der Neuronen [Moo95]

1.3.3 Frequenzgruppenbildung des Gehörs

Der Mensch fasst bei der Bildung der Lautheitsempfindung einen bestimmten Frequenzbereich zusammen, den man Frequenzgruppenbreite nennt. Innerhalb dieser Frequenzgruppe wird die vorhandene Energie aufsummiert. Man kann die Frequenzgruppenbreite durch die Bestimmung der Mithörschwelle eines Testsignals innerhalb eines Schmalbandrauschens (siehe Abb. 1.13) mit variabler Bandbreite veranschaulichen [Lab06].

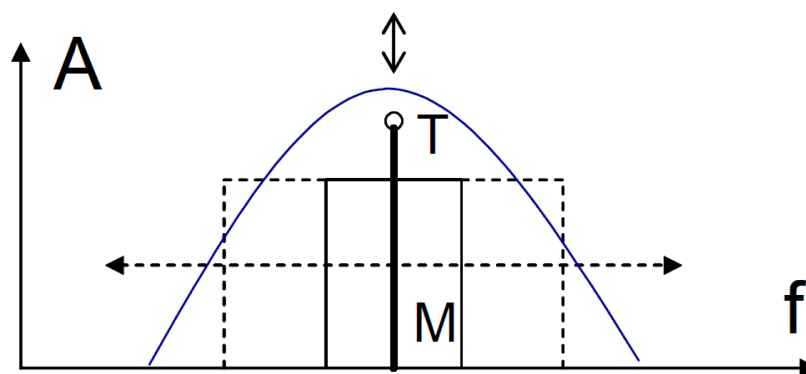


Abbildung 1.13: Bestimmung der Mithörschwelle: Testsignal mit variablen Pegel bei Mittenfrequenz eines Schmalbandrauschens mit variabler Bandbreite [Lab06]

Vergrößert man die Bandbreite des Schmalbandrauschens, so nimmt die Mithörschwelle zu, bis das Schmalbandrauschen die kritische Bandbreite überschritten hat und bleibt anschließend konstant. Die Frequenzgruppenbreite ist mit einem gedachten auditiven Filter vergleichbar. Schallanteile die außerhalb des Filters liegen, haben keinen Einfluss auf die Bildung der Mithörschwelle [Pfl00].

Die Frequenzgruppenbreite ist frequenzabhängig und nimmt mit steigender Frequenz zu, was sich mit dem frequenzabhängigen Ort des Maximums der Wanderwelle auf der Basilmembran erklären lässt. Tiefe Frequenzen beeinflussen die auditiven Filter im höheren Frequenzbereich, da sie sich durch den Abschnitt der Basilmembran fortpflanzen, der für die Detektion der hohen Frequenzen zuständig ist. Deshalb fasst der Mensch im höheren Frequenzbereich eine größere Frequenzbandbreite für die Lautheitsempfindung zusammen.

Die Barkskala teilt den vom Menschen wahrgenommenen Frequenzbereich in einzelne auditive Filter (Bandpassfilter) ein. Die Umrechnung von Bark in die Mittenfrequenz erfolgt nach Traunmüller [Pfl00] [Zot04] [Tra90] mit

$$\frac{z}{[Bark]} = T f m_{Bark} \left\{ \frac{f}{[Hz]} \right\} \approx \frac{26,81 \frac{f}{1000}}{1,96 + \frac{f}{1000}} - 0,53 \quad , \quad (1.3.1)$$

$$\frac{f}{[kHz]} = T f m_{Bark}^{-1} \left\{ \frac{z}{[Bark]} \right\} \approx 1,96 \frac{z + 0,53}{26,28 - z} \quad . \quad (1.3.2)$$

1.4 Lokalisation

Um Schallquellen zu lokalisieren muss der Mensch aus den vorhandenen 2 Ohrsignalen die Parameter mit den Richtungsinformationen detektieren. Die Signale werden durch die eigenen Körperausmaße des Menschen beeinflusst und beinhalten deshalb körperabhängige Parameter. Da die 2 Ohren des Menschen an 2 gegenüberliegenden Punkten des Kopfes liegen, beeinflusst der Körper die Parameter Pegel (Abschattung), Laufzeit (unterschiedlicher Detektionszeitpunkt der 2 Ohren) und die spektrale Zusammensetzung des Signals (Körperform).

1.4.1 Lokalisation in der Horizontalebene

In der Horizontalebene wird eine Schallquelle durch interaurale Pegeldifferenzen (ILD) und interaurale Zeitdifferenzen (ITD) lokalisiert.

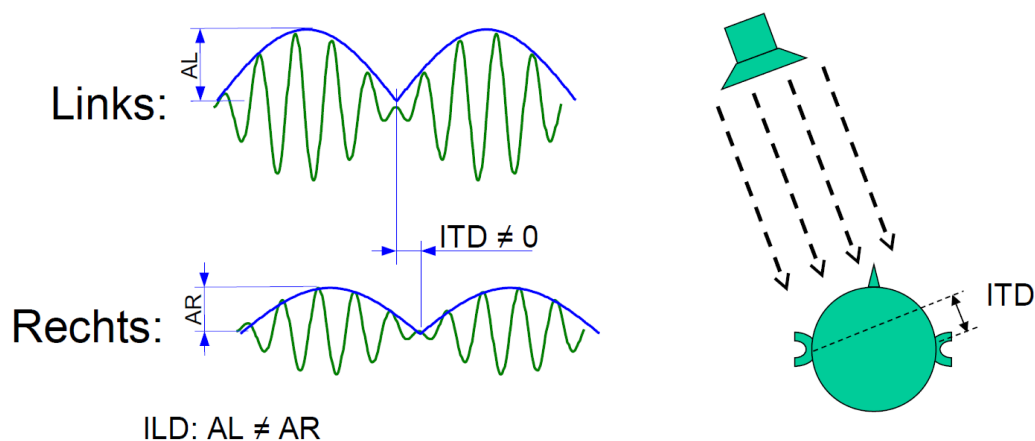


Abbildung 1.14: Laufzeit und Pegeldifferenz der beiden Ohrsignale bei einer Schallquelle [Maj10]

Die Pegeldifferenz basiert auf der Abschattung durch den Kopf. Das Maximum bzw. das Minimum liegt bei $\pm 90^\circ$ Azimutwinkel (siehe Abb. 1.15).

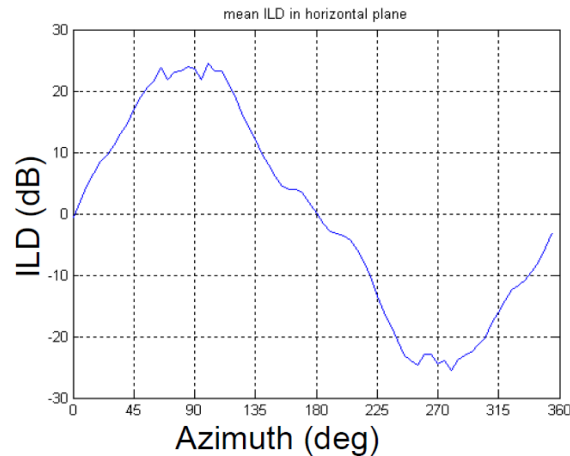


Abbildung 1.15: Durchschnittliche Pegeldifferenz abhängig vom Azimutwinkel gemessen an einem Kunstkopf [Maj10]

Es existiert eine Abschattung, wenn die Wellenlänge klein gegenüber dem Kopfdurchmesser ist, sonst wird der Schall um den Kopf gebeugt. Es gilt bei dem Kopfdurchmesser $d = 0,17 \text{ m}$:

$$\lambda [m] < 0,17m \quad ,$$

$$\frac{c[\frac{m}{s}]}{f[\frac{1}{s}]} < 0,17m \quad ,$$

$$f \left[\frac{1}{s} \right] > \frac{340}{0,17} \frac{1}{s} \quad ,$$

$$f [Hz] > 2000Hz \quad .$$

Da die Ausbreitungsgeschwindigkeit von Schall in Luft begrenzt ist ($c = 340 \text{ m/s}$) existiert für Schallquellen die nicht direkt vor bzw. hinter dem Kopf liegen eine winkelabhängige Laufzeitdifferenz, die wiederum bei $\pm 90^\circ$ ihr Maximum bzw. Minimum hat (siehe Abb. 1.16).

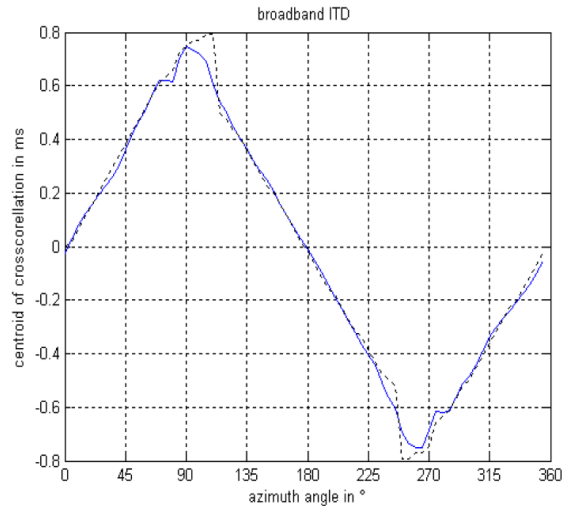


Abbildung 1.16: Durchschnittliche Laufzeitdifferenz abhängig vom Azimutwinkel gemessen an einem Kunstkopf [Maj10]

Ist die halbe Wellenlänge eines Signals kleiner dem Kopfdurchmesser (vgl. Nyquist Theorem) tritt Aliasing in der Laufzeitdifferenz auf. Dies bedeutet, dass $\lambda/2 > 0.17$ m ($f < 1000$ Hz) sein muss, damit die ITD eindeutig ist.

1.5 Dreidimensionale Lokalisation

Die vorher beschriebenen Gegebenheiten gelten nur für die vordere Horizontalebene. Aufgrund der symmetrischen Eigenschaften des Kopfes existiert eine Mehrdeutigkeit in der Pegel- und Laufzeitdifferenz, sobald man den Detektionsbereich auf die hintere Horizontalebene bzw. Medianebene erweitert. Diese Mehrdeutigkeiten liegen genau am Rand des „cone-of-confusion“ (siehe Abb.1.17).

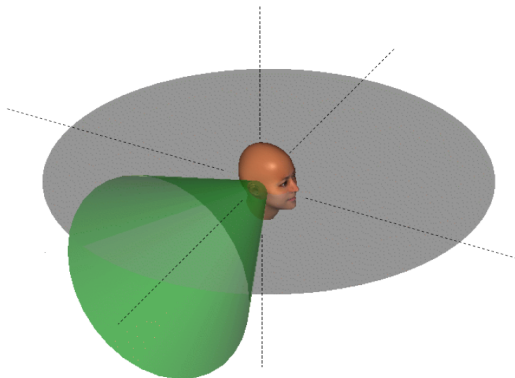


Abbildung 1.17: Durch die Symmetrieeigenschaften des Kopfes gibt es eine Mehrdeutigkeit der ILD und der ITD, die im cone-of-confusion ersichtlich ist [Maj10]

Um die Mehrdeutigkeit in der Richtungsinformation zu verringern, werden die spektralen Eigenschaften des Signals analysiert. Diese werden durch Abschattung bzw. Reflexion an Pinna, Torso und Kopf hervorgerufen. So wird es überhaupt erst möglich eine Schallquelle außerhalb der Horizontalebene zu lokalisieren. Zusätzlich wird die Mehrdeutigkeit durch Kopfbewegungen abgeschwächt. In [MRT98] wurde bewiesen, dass die Vorne-/Hinten Mehrdeutigkeit in der Horizontalebene durch eine Kopfdrehung eindeutig geringer wird (siehe Abb.1.18 und Abb.1.19). Es wird ein Versuch beschrieben, bei dem eine Versuchsperson die Mikrofonsignale eines Kunstkopfes hört. Über ein Head-Tracking-System kann die Kopfdrehung der Versuchsperson auf den Kunstkopf übertragen werden. Wird die Kopfdrehung übertragen, so wird eine Verwechslung fast vollständig verhindert.

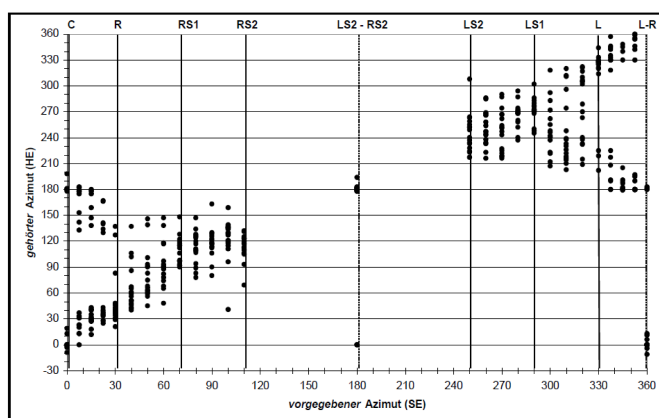


Abbildung 1.18: Vorne-/Hinten Mehrdeutigkeit bei starrem Kunstkopf [MRT98]

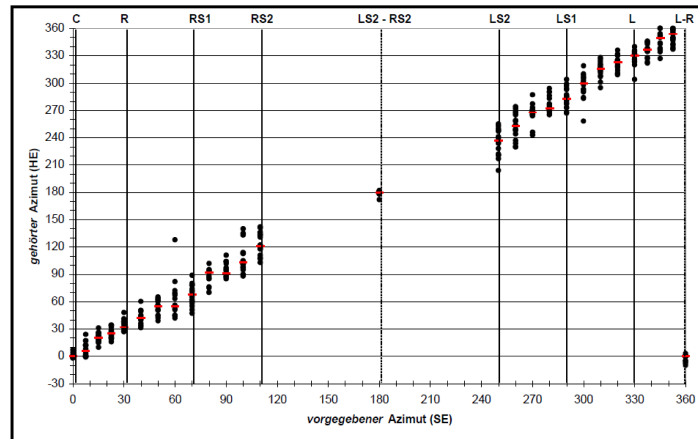


Abbildung 1.19: Verringerte Vorne-/Hinten Mehrdeutigkeit bei bewegtem Kunstkopf [MRT98]

1.6 Modelle für eine binaurale Lokalisation

Jeffress hat 1948 ein Lokalisationsmodell entwickelt, das bis heute verwendet wird und weiterentwickelt wurde [Jef48] [WB06]. Es basiert auf Neuronen die für die Detektion der Laufzeitunterschiede zuständig sind. Die Detektionseinheiten werden EE-Zellen genannt (excitation-excitation). Dies bedeutet, dass diese Zellen auf das linke und rechte Ohrsignal reagieren. Das maximale Aktionspotential einer EE-Zelle wird nur bei einer bevorzugten Laufzeitdifferenz erreicht. Die Neuronen vergleichen die Phasen der beiden Signale. Je größer die Übereinstimmung der Phasen, desto stärker ist die neuronale Antwort. In Abb.1.20 ist dies durch eine Multiplikation dargestellt. Damit ist dieses Modell einem Korrelationsverfahren sehr ähnlich.

In dem Modell werden die 2 Eingangssignale gegenläufig an den Verzögerungsstufen mit einem konstanten ΔT zeitverzögert (siehe Abb.1.20) und an den Detektionsstufen verglichen (multipliziert und aufsummiert). Dies bedeutet, dass 2 identische Signale ein erhöhtes Aktionspotential bei der mittleren Detektionseinheit hervorrufen.

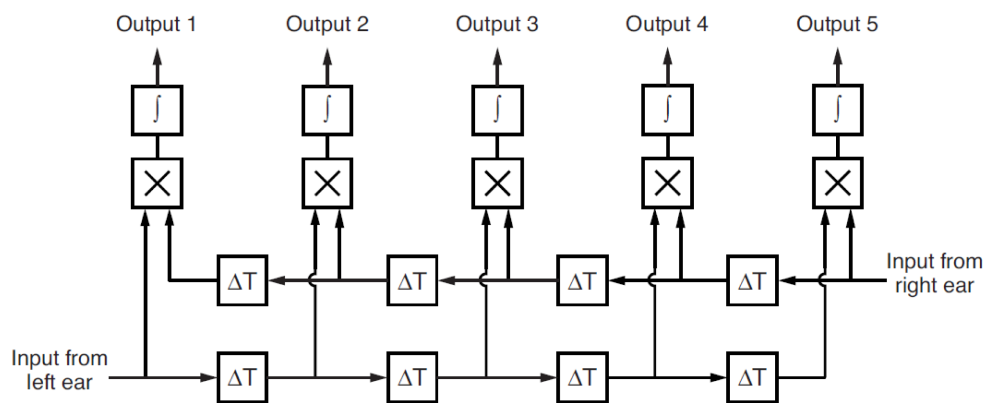


Abbildung 1.20: Schematische Darstellung des Lokalisationsmodells von Jeffress [WB06]

Breebaart hat das Modell von Jeffress weiterentwickelt (siehe Abb. 1.21). Die Detektion dieses Lokalisationsmodell ist zusätzlich zur Laufzeitdifferenz auch von der Pegeldifferenz abhängig.

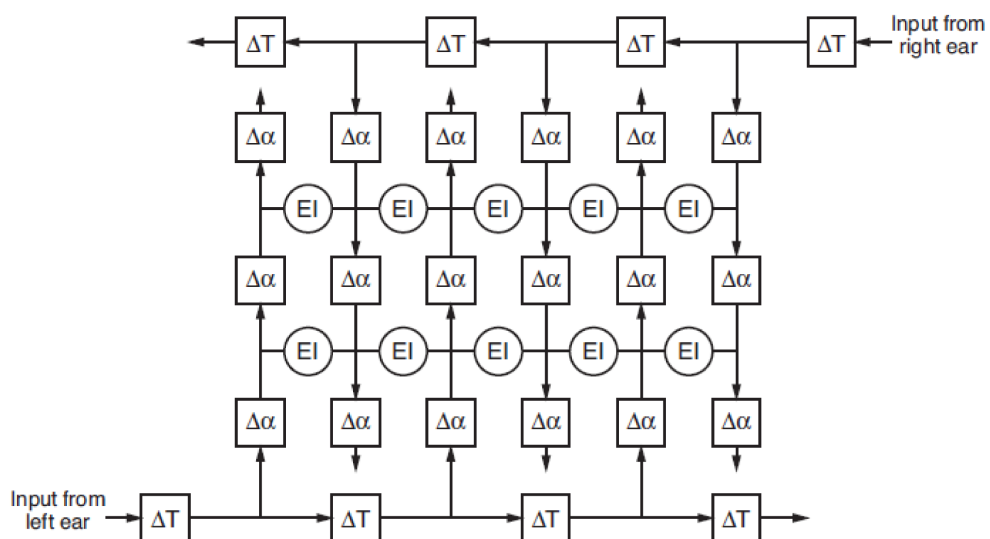


Abbildung 1.21: Schematische Darstellung des Lokalisationsmodells von Breebaart [WB06]

In der schematischen Darstellung des Lokalisationsmodells von Breebaart werden die beiden Ohrsignale gegenläufig zeitverzögert und bedämpft, wodurch eine Lokalisation der Laufzeitdifferenzen und Pegeldifferenzen ermöglicht wird. In den Detektionseinheiten werden die Ohrsignale voneinander subtrahiert. Der Ort des Minimums ist der Ort der maximalen Übereinstimmung und damit die detektierte Pegel- und Laufzeitdifferenz.

Die EI-Zellen (excitation-inhibition) haben ähnlich wie die EE-Zellen ein Maximum des Aktionspotentials bei einer bestimmten Laufzeitdifferenz. Zusätzlich ist dieses Maximum von der Pegeldifferenz abhängig. Die EI-Zellen hemmen eines der beiden Ohrsignale zu einem bestimmten Grad. Anders ausgedrückt wird das Signal bedämpft, was in der schematischen Darstellung durch den Dämpfungsfaktor $\Delta\alpha$ dargestellt wird.

1.7 Wahrgenommene Schallquellengröße und Lokalisationsunschärfe

In der menschlichen Wahrnehmung gibt es so etwas wie eine wahrgenommene Schallquellengröße. Sehr früh bemerkte man in der Raumakustik, dass seitliche Reflexionen den von einer Bühne kommenden Schall nicht nur lauter, sondern auch das Hörereignis größer erscheinen lässt.

Präsentiert man einer Versuchsperson einen Direktschall und eine einzelne Reflexion, wird die Größe der Schallquelle, abhängig von der Zeitverzögerung und dem Pegel der Reflexion, unterschiedlich groß wahrgenommen [SB95]. Die Schallquellengröße von einem Direktschall mit einer Reflexion wird immer größer wahrgenommen als ohne Reflexion. Ein ähnliches Verhalten der Wahrnehmung lässt sich bei einer durch einen Lautsprecher präsentierten Schallquelle (ein Lautsprecher) und einer durch 2 Lautsprecher (Stereofonie) präsentierten Phantomschallquelle erkennen. Beide Schallquellen werden der Versuchsperson am gleichen Ort frontal präsentiert. Die Phantomschallquelle wird als größer wahrgenommen.

Die IACC kann als eine Maß für eine wahrgenommene Quellenbreite verwendet werden [FMS11]. Durch die Berechnung der IACC von Matthias Frank wird gezeigt (zum Zeitpunkt dieser Arbeit noch nicht veröffentlicht), dass die Quellenbreite einer einzelnen realen Quelle von der Position der Quelle abhängt. Die Quellengröße ist bei ca. 60° und 120° Azimutwinkel am breitesten, da dort die IACC am kleinsten ist (siehe Abb. 1.22).

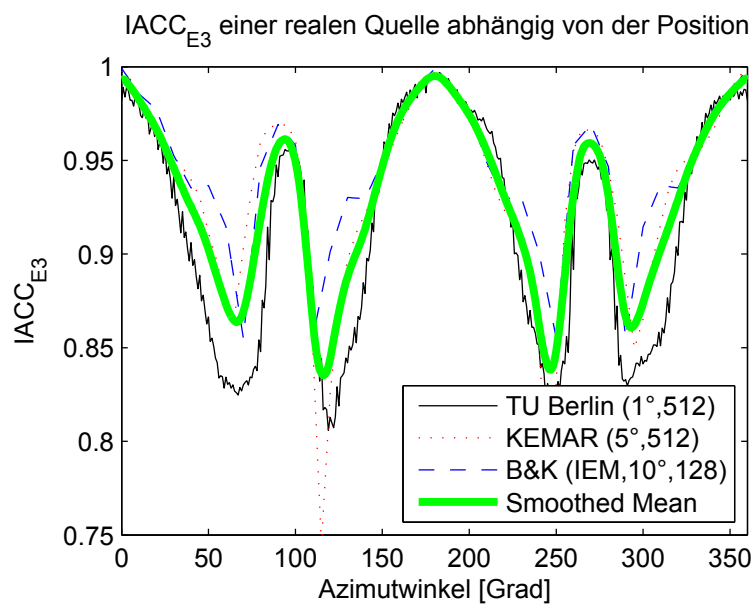


Abbildung 1.22: Berechnung der $IACC_{E3}$ abhängig von der Quellenposition von Matthias Frank

Als Lokalisationsunschärfe versteht man in [Bla74] die kleinste Änderung eines Merkmals eines Schallereignisses, das zu einer wahrgenommenen Ortsänderung führt. In den Hörversuchen in [Bla74] wird für die Lokalisationsunschärfe in der Horizontalebene folgendes Ergebnis erzielt: Die Lokalisationsunschärfe nimmt von 0° bis 90° zu und dann zu 180° wieder ab. Vergleicht man dieses Ergebnis mit obiger Grafik wird ersichtlich, dass kein zwingender Zusammenhang zwischen Lokalisationsunschärfe und Quellenbreite besteht (Matthias Frank).

Kapitel 2

Binaurales Lokalisationsmodell

In diesem Kapitel wird die Implementierung des binauralen Lokalisationsmodells beschrieben. Es gliedert sich in folgende Stufen (siehe Abb. 2.1): der auditiven Peripherie, der binauralen Lokalisationsstufe nach Breebaart und einer höheren Verarbeitungsstufe. Die höhere Verarbeitungsstufe ist ein Mustervergleich, bei dem das ähnlichste Muster die wahrscheinlichste Schallquellenrichtung angibt. Anschließend wird die Richtungsmehrdeutigkeit am cone-of-confusion durch eine simulierte Kopfdrehung herausgerechnet und die Wahrscheinlichkeiten der vom Modell lokalisierten Richtungen dargestellt, wodurch auch auf die wahrgenommene Breite einer Quelle geschlossen werden kann.

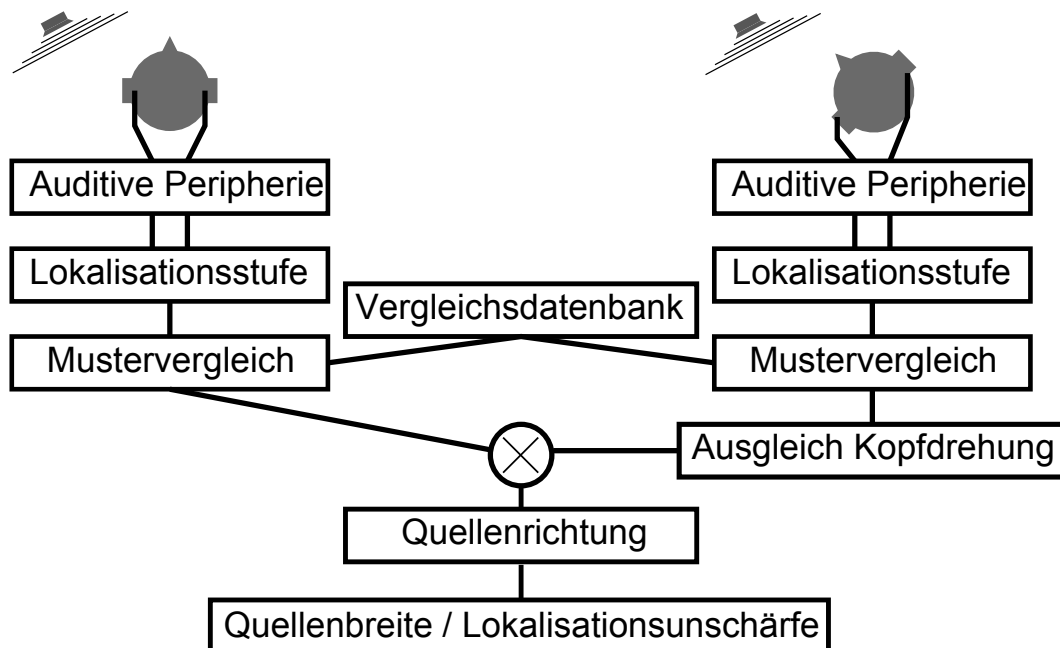


Abbildung 2.1: Skizze des binauralen Lokalisationsmodells

2.1 Auditive Peripherie

Die auditive Peripherie besteht aus dem Außen-, Mittel- und Innenohr sowie der neuronalen Verarbeitung. Ziel dieses Modells ist es der menschlichen Signalverarbeitung möglichst ähnlich zu sein, dafür werden die Ohrsignale durch auditive Signalverarbeitung so bearbeitet, dass sie an der binauralen Lokalisationsstufe den Signalen gleichen, welche am Hörnerv anliegen. Es wird hier jede Verarbeitungsstufe wie ein Übertragungspfad, mit einer definierten Übertragungsfunktion, behandelt.

2.1.1 Übertragungspfad Außen- und Mittelohr

Der Übertragungspfad des Außen- und Mittelohrs wird durch die von der Länge des Hörkanals abhängigen Resonanzfrequenz bei 4.2 kHz ($\lambda/4$ Resonator) und durch das Federmassesystem des Mittelohrs bestimmt (vgl. Abschnitt 1.2.1 und 1.2.2).

Die Außenohrübertragungsfunktion wird mittels eines Peakfilters [ZA02] implementiert. Der Peakfilter besteht aus einem Allpassfilter 2. Ordnung dessen Ausgangssignal von dem Eingangssignal abgezogen wird. Ein Allpassfilter 2. Ordnung hat ein glattes Frequenzspektrum und einen Phasenverlauf von 0° bis -360° . Durch die Subtraktion wird die Amplitude im resultierenden Frequenzgang für tiefe und hohe Frequenzen gegen 0 bedämpft und für Frequenzen um die Resonanzfrequenz verdoppelt. Danach wird das Signal gewichtet zu dem Eingangssignal addiert, wodurch die Verstärkung des Peaks kontrolliert werden kann (siehe Abb. 2.2).

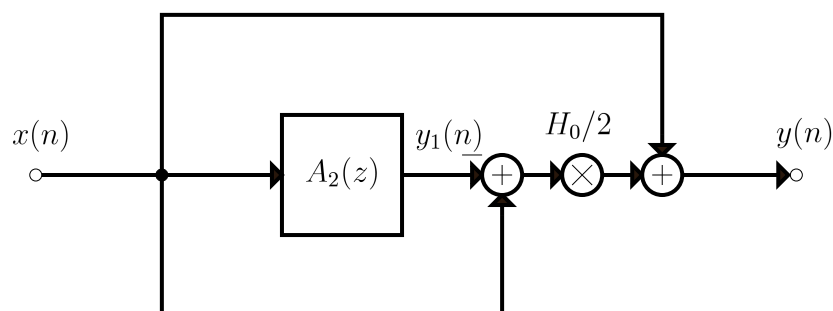


Abbildung 2.2: Signalflussplan eines Peakfilters

Für die Übertragungsfunktion ergibt sich im z -Bereich:

$$H_{Peak}(z) = 1 + \frac{H_0}{2} [1 - A_2(z)] \quad . \quad (2.1.1)$$

Umgeformt nach [Zot04] ergibt sich für die Übertragungsfunktion:

$$H_{Peak}(z) = \frac{\left[\frac{(1+r_{res}^2)(1-g)}{2} + g \right] - 2r_{res}\cos(\omega_{res})z^{-1} + \left[\frac{(1+r_{res}^2)(1+g)}{2} - g \right] z^{-2}}{1 - 2\cos(\omega_{res})z^{-1} + r_{res}z^{-2}} \quad . \quad (2.1.2)$$

Die Übertragungsfunktion des Mittelohrs, welches sich wie ein Feder-Masse-System verhält, wird mit einem Tiefpass und einem Hochpass modelliert. Hohe Frequenzen werden durch die Massenträgheit und tiefe Frequenzen durch die Steifigkeit des Trommelfells und der Gehörknöchelchen bedämpft. Beide Filter besitzen eine Biquadstruktur mit doppelter Pol- und Nullstelle

$$H_{Hp}(z) = \frac{(1 - r_{1,Hp}z^{-1})^2}{(1 - r_{2,Hp}z^{-1})^2} \quad , \quad (2.1.3)$$

$$H_{Tp}(z) = \frac{(1 + r_{1,Tp}z^{-1})^2}{(1 + r_{2,Tp}z^{-1})^2} \quad . \quad (2.1.4)$$

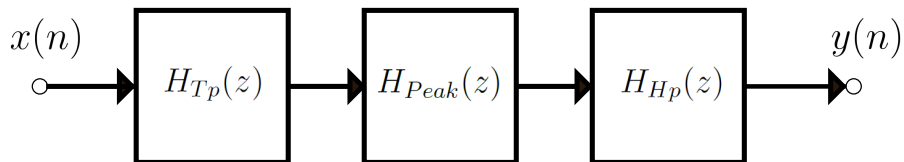


Abbildung 2.3: Signalflussplan für das Außen- und Mittelohr

Werden Hochpass-, Tiefpass- und Peakfilter in Serie geschaltet, resultiert die Gesamtübertragungsfunktion für das Außen- und Mittelohr (siehe Abb. 2.4).

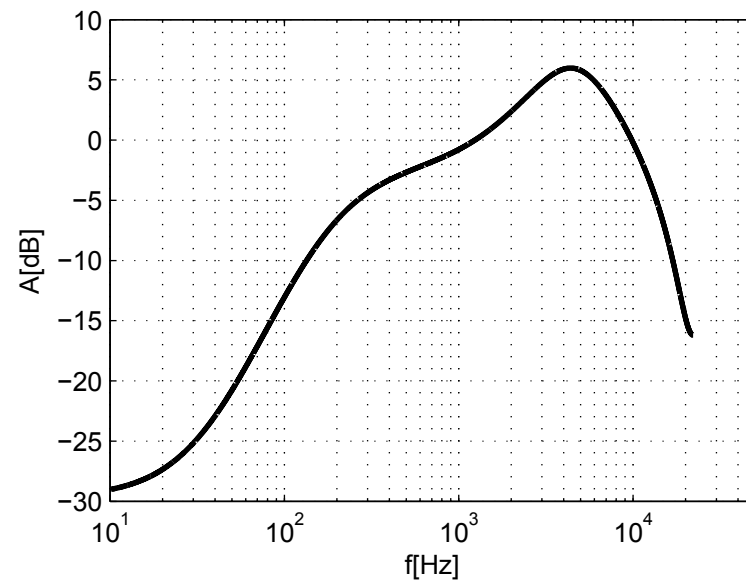


Abbildung 2.4: Übertragungsfunktion für Außen- und Mittelohr

2.1.2 Übertragungspfad Innenohr

Die auditive Signalverarbeitung im Innenohr wird durch seine frequenzgruppenbildende Eigenschaft bestimmt (siehe Abschnitt 1.3.3).

Für die Lautheitsempfindung fasst der Mensch die Energie innerhalb eines auditiven Filters zusammen, dessen Bandbreite mit steigender Frequenz zunimmt. Deshalb wird das Signal mit einer Barkbandfilterbank gefiltert. Bei dieser Filterbank wird das vom Mittelohr kommende Signal mit 22 auditiven Filtern in 22 separate Signale geteilt.

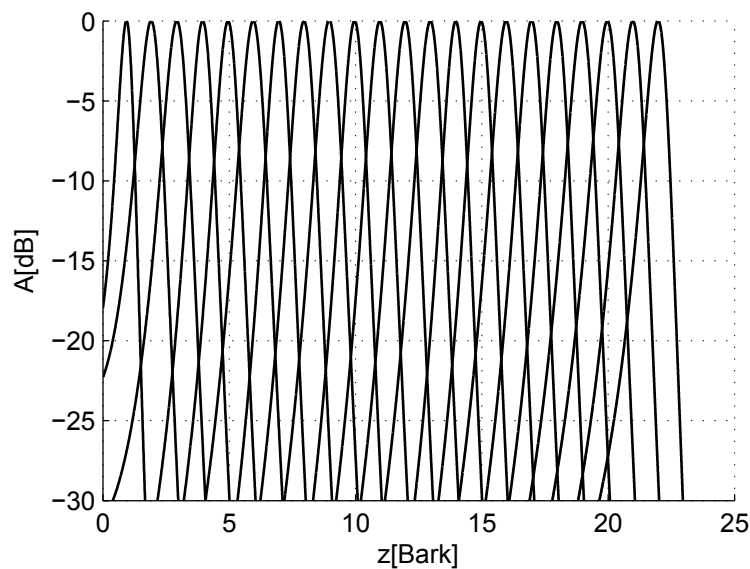


Abbildung 2.5: Übertragungsfunktionen aller 22 Barkbandfilter

Die auditiven Filter sind in der Barkskala dargestellt. In Abbildung 2.5 sieht man 22 Barkbandfilter die sich bei -8 dB überlappen. Die Bandbreite aller Filter ist identisch 1 Bark zwischen den beiden Flanken bei -8 dB. Da die Transformation von Hertz in Bark nichtlinear ist, ist die Bandbreite der auditiven Filter im Frequenzbereich nicht identisch und nimmt mit steigender Frequenz zu. Die Rücktransformation von Bark in Hertz ist in Gleichung 1.3.2 ersichtlich. Die Mittenfrequenzen der Filter liegen bei 118 Hz bis 10300 Hz.

Für die Implementierung eines auditiven Filters wird ein Allpolfilter 4. Ordnung mit einer Nullstelle unter der Mittenfrequenz und einem Nullstellenpaar oberhalb der Mittenfrequenz verwendet [Zot04] [LAH02]. Durch die Nullstelle und das Nullstellenpaar wird die Sperrdämpfung der unteren und oberen Filterflanke erhöht.

Die Übertragungsfunktion eines Barkbandfilters ist im z -Bereich gegeben mit:

$$H_{BARK}(z) = g_{BARK}(1 - r_0 z^{-1}) \frac{1 - 2r_z \cos(\omega_z z^{-1}) + r_z^2 z^{-2}}{[1 - 2r \cos(\omega) z^{-1} + r^2 z^{-2}]^4}, \quad (2.1.5)$$

$$g_{BARK} = \frac{[(1 - r) \sqrt{1 - 2r \cos(2\omega) + r^2}]^4}{\sqrt{(1 - \cos^2(\omega)) + \sin^2(\omega)}}. \quad (2.1.6)$$

2.1.3 Neuronaler Übertragungspfad

In der neuronalen Verarbeitungsstufe werden mechanische Schwingungen als elektrische Impulse kodiert. Ab hier wird eine Frequenz und ihr Pegel in Aktionspotentialen weitergeleitet. Die physiologischen Grundlagen können in Abschnitt 1.3.1 und 1.3.2 nachgelesen werden. In diesem Modell wird nicht die Umwandlung von Pegel in einzelne Impulse mit einer Aktionspotentialdichte durchgeführt, sondern eine dichtere Feuerrate wird weiterhin durch einen größeren Pegel der Schwingung dargestellt.

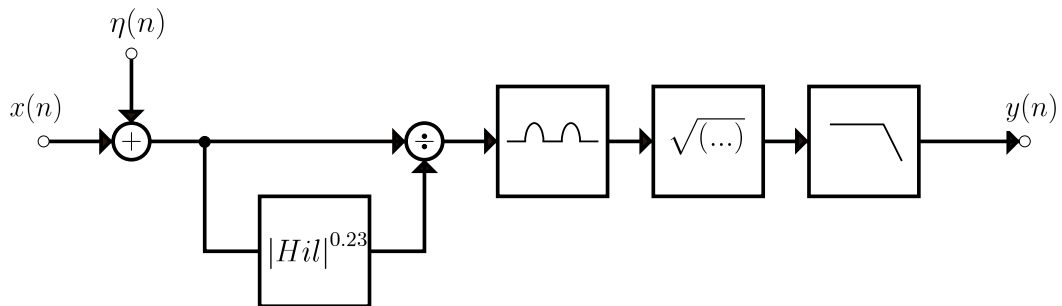


Abbildung 2.6: Signalflussplan des neuronalen Übertragungspfad

Der neuronale Übertragungspfad besteht aus einer Halbwellengleichrichtung, da die Depolarisation der Haarzellen nur in einer Richtung stattfindet, einer nichtlinearen Dynamikübertragung und einer Tiefpassfilterung, die bei höheren Frequenzen die Phaseninformation im Signal glättet. Zusätzlich wird dem Signal, aufgrund der Spontanfeurrate der Neuronen, Rauschen hinzugefügt. Die komprimierende Eigenschaft des Gehörs wird durch eine punktweise durchgeführten Division mit der komprimierten Einhüllenden erzeugt [BvdPT99]. Als Einhüllende wird der Betrag des hilbertransformierten Signals verwendet. Zusätzlich wird diese nichtlineare Eigenschaft verstärkt, indem die Wurzel gezogen wird. Es wird ein Tiefpass 4. Ordnung mit einer Grenzfrequenz bei 800 Hz verwendet.

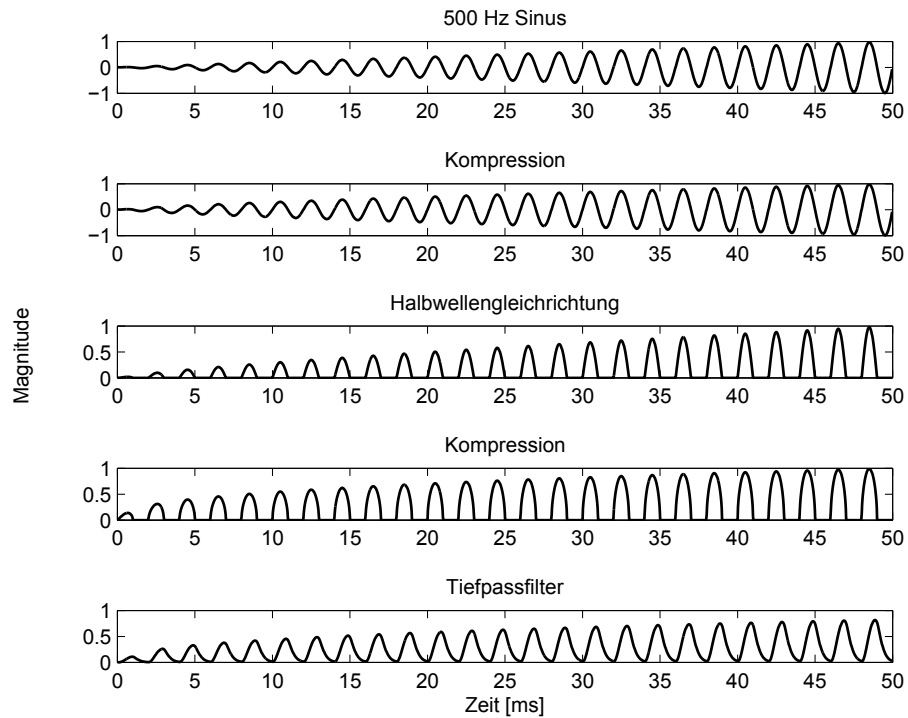


Abbildung 2.7: Neuronale Signalverarbeitungsschritte mit 500 Hz Sinus

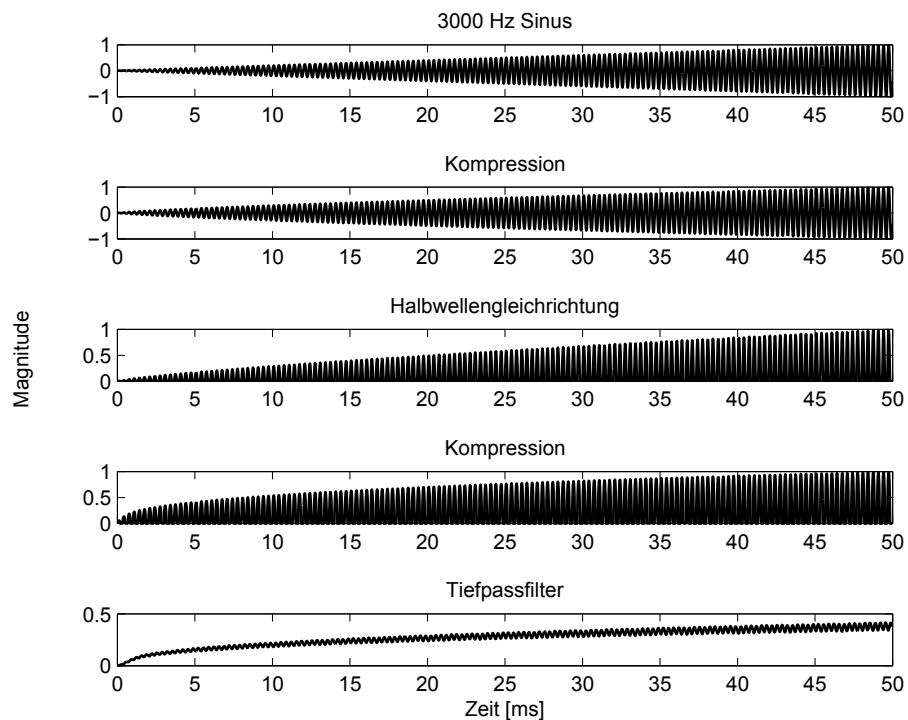


Abbildung 2.8: Neuronale Signalverarbeitungsschritte mit 3000 Hz Sinus

In den Abbildungen 2.7 und 2.8 sind die einzelnen Verarbeitungsschritte des neuronalen Übertragungspfades (siehe Abb. 2.6) für zwei unterschiedliche Eingangssignale dargestellt.

Die Eingangssignale sind einschwingende Sinustöne bei 500 Hz und 3000 Hz. Es ist deutlich erkennbar, dass bei höheren Frequenzen die Phaseninformation verloren geht und ein geglättetes Signal an die nächst höhere Ebene (Gehirn) weitergeleitet wird.

2.2 Binaurale Lokalisationsstufe

Als Grundlage der Lokalisationsstufe dient das Modell von Breebaart [BvdPK01], welches erweitert wurde.

Linkes und rechtes Ohrsignal werden gegenläufig mit $\Delta\tau$ zeitverzögert und mit $\Delta\alpha$ gedämpft. Die Signale werden anschließend an den EI-Knotenpunkten voneinander subtrahiert (siehe Abb. 2.9). Durch Betragsbildung und eine Rückkopplung entsteht ein gemittelt Minimum an demjenigen EI-Knotenpunkt, an dem sich die Signale am ähnlichsten sind. Fügt man alle EI-Knotenpunkte zu einer 2 dimensional Matrix (EI-Muster) zusammen, erkennt man an dem Minimum der Matrix den Laufzeit- und Pegelunterschied der 2 Ohrsignale. Dieses Modell basiert auf dem Prinzip der TDOA-Schätzung (time delay of arrival) mit der Minimum-Differenz-Methode. Das EI-Muster wird für alle 22 Signalpaare berechnet, welche aus der neuronalen Übertragungsstufe kommen und vorher durch die Barkbandfilterbank entstehen. Dadurch werden die Laufzeit- und Pegeldifferenzen in 22 Frequenzbändern berechnet, welche den relevanten Hörbereich des Menschen abdecken.

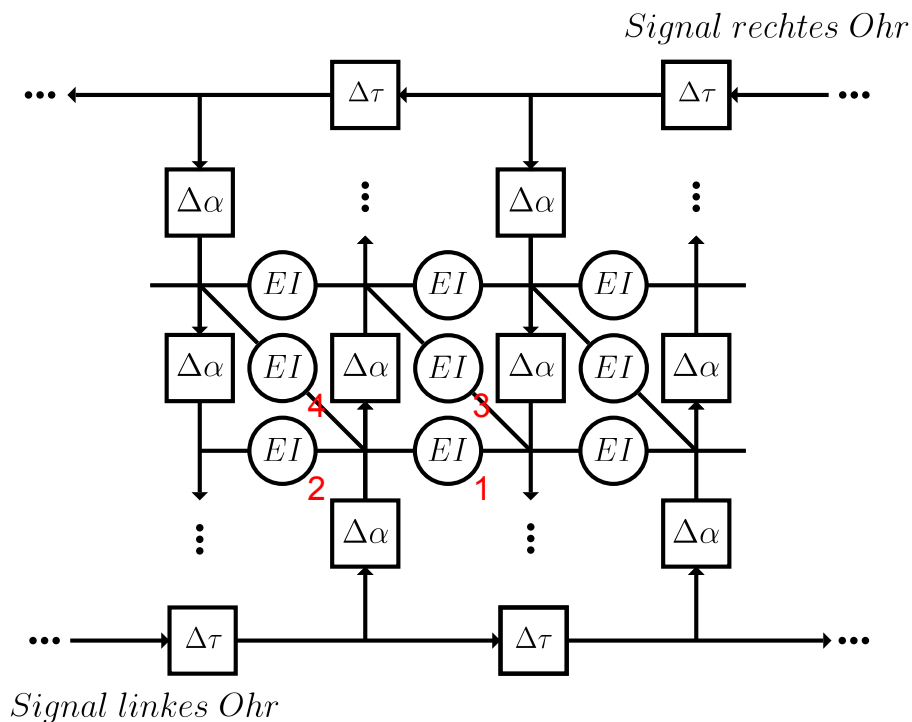


Abbildung 2.9: Signalflussplan der Lokalisationsstufe: es werden Differenzen der einzelnen Verzögerungs- und Dämpfungsstufen gebildet

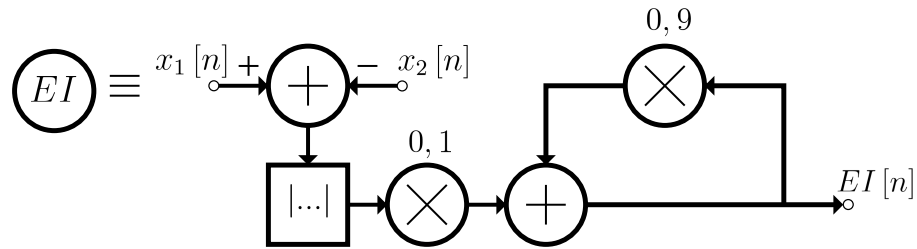


Abbildung 2.10: Es wird der Betrag der Differenz an jedem EI Knotenpunkt rekursiv gemittelt

Eine Erneuerung des Signalflusses ist in Abbildung 2.9 diagonal dargestellt. Es wird eine zusätzliche Signaldifferenz gebildet, um die Auflösung der Pegeldifferenzen zu erhöhen. Ohne diese Zwischenstufe würden bei einer geraden Anzahl der Dämpfungsschritte Signale, die am linken und rechten Ohr identisch sind, kein Minimum in der Mitte des EI-Musters ergeben, da sie durch die gleichgroßen Dämpfungsschritte niemals am gleichen Knotenpunkt den selben Pegel aufweisen können. Außerdem würde man nur bei jedem zweiten Dämpfungsschritt eine Differenz bilden können. Das bedeutet es werden weniger Multiplikation benötigt um die selbe Auflösung zu erhalten wie das ursprüngliche Modell von Breebaart.

2.2.1 Das EI-Muster

Bei genauerer Betrachtung ergeben sich 4 unterschiedliche Typen von EI-Knotenpunkten im EI-Muster (in Abb. 2.9 rot mit 1,2,3 und 4 gekennzeichnet):

- b ... Barkband
 L ... Anzahl Dämpfungsstufen
 M ... Anzahl Verzögerungsstufen

$$\begin{aligned}
 EI'_{1[b,l,m,n]} &= \frac{\left| \left[10^{\frac{\Delta\alpha}{20}}\right]^l x_{L[b,\Delta\tau m+n]} - \left[10^{\frac{\Delta\alpha}{20}}\right]^{L+1-l} x_{R[b,(M+1-m)\Delta\tau+n]} \right|}{\left[10^{\frac{\Delta\alpha}{20}}\right]^l + \left[10^{\frac{\Delta\alpha}{20}}\right]^{L+1-l}}, \\
 EI'_{2[b,l,m,n]} &= \frac{\left| \left[10^{\frac{\Delta\alpha}{20}}\right]^{L+1-l} x_{R[b,(M+2-m)\Delta\tau+n]} - \left[10^{\frac{\Delta\alpha}{20}}\right]^l x_{L[b,\Delta\tau m+n]} \right|}{\left[10^{\frac{\Delta\alpha}{20}}\right]^{L+1-l} + \left[10^{\frac{\Delta\alpha}{20}}\right]^l}, \\
 EI'_{3[b,l,m,n]} &= \frac{\left| \left[10^{\frac{\Delta\alpha}{20}}\right]^l x_{L[b,\Delta\tau m+n]} - \left[10^{\frac{\Delta\alpha}{20}}\right]^{L+2-l} x_{R[b,(M+1-m)\Delta\tau+n]} \right|}{\left[10^{\frac{\Delta\alpha}{20}}\right]^l + \left[10^{\frac{\Delta\alpha}{20}}\right]^{L+2-l}}, \\
 EI'_{4[b,l,m,n]} &= \frac{\left| \left[10^{\frac{\Delta\alpha}{20}}\right]^{L-l} x_{R[b,(M+2-m)\Delta\tau+n]} - \left[10^{\frac{\Delta\alpha}{20}}\right]^l x_{L[b,\Delta\tau m+n]} \right|}{\left[10^{\frac{\Delta\alpha}{20}}\right]^{L-l} + \left[10^{\frac{\Delta\alpha}{20}}\right]^l}.
 \end{aligned}$$

Die Dämpfung vom linken und rechten Signal wird an jedem Knotenpunkt so normiert, dass ihre Summe immer 0 dB ergibt. Insgesamt bleiben das linke und rechte Signal immer gleich laut (siehe Abb. 2.11).

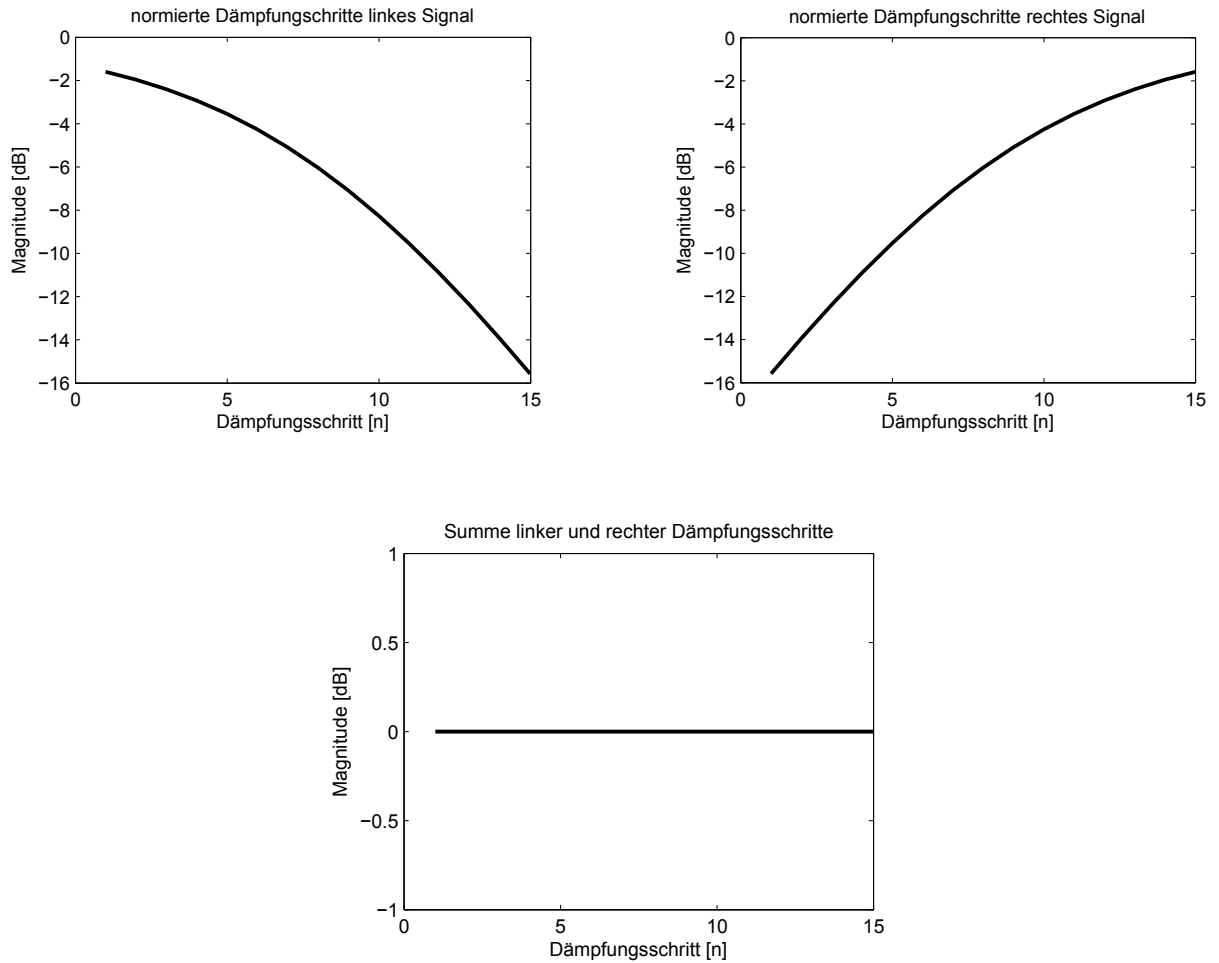


Abbildung 2.11: Verlauf der Dämpfungsschritte für linkes und rechtes Signal

Nach der Betragsbildung werden die Differenzen durch eine Rückkopplung zeitlich gemittelt

$$EI_{1,2,3,4[b,l,m,n]} = 0, 1EI'_{1,2,3,4[b,l,m,n]} + 0, 9EI'_{1,2,3,4[b,l,m,n-1]} \quad .$$

Das EI-Muster kann nun als 2 dimensionale LxM Matrix dargestellt werden

$$EI_{[n]} = \begin{bmatrix} EI_{1[b,1,1,n]} & EI_{2[b,1,2,n]} & EI_{1[b,1,3,n]} & \dots & EI_{1[b,1,M,n]} \\ EI_{3[b,2,1,n]} & EI_{4[b,2,2,n]} & EI_{3[b,2,3,n]} & \dots & EI_{3[b,2,M,n]} \\ EI_{1[b,3,1,n]} & EI_{2[b,3,2,n]} & EI_{1[b,3,3,n]} & \dots & EI_{1[b,3,M,n]} \\ EI_{3[b,4,1,n]} & EI_{4[b,4,2,n]} & EI_{3[b,4,3,n]} & \dots & EI_{3[b,4,M,n]} \\ \dots & \dots & \dots & \dots & \dots \\ EI_{1[b,L,1,n]} & EI_{2[b,L,2,n]} & EI_{1[b,L,3,n]} & \dots & EI_{1[b,L,M,n]} \end{bmatrix} .$$

In den nachfolgenden Grafiken wird das Verhalten der EI-Muster und damit die Funktionsweise des erweiterten Lokalisationsmodells von Breebaart ersichtlich. Als Eingangssignale werden Impulse verwendet. Der Impuls des linken Eingangs wird zeitverzögert oder gedämpft. Durch die Lage des Minimums innerhalb des EI-Musters werden Laufzeit- und Pegeldifferenzen detektiert. Für eine bessere Darstellung wird auf die auditive Peripherie und auf die Normierung der Dämpfungsschritte verzichtet. Durch die Erweiterung des Breebaart Modells kann jede Laufzeit- und Pegeldifferenz dargestellt werden. Die EI-Muster werden mit den Dämpfungsstufen $\Delta\alpha = 1\text{dB}$, den Verzögerungsstufen $\Delta\tau = 22,5\mu\text{s}$ (1 Sample bei 44,1 kHz), $L = 15$ und $M = 37$ berechnet (n = Sample).

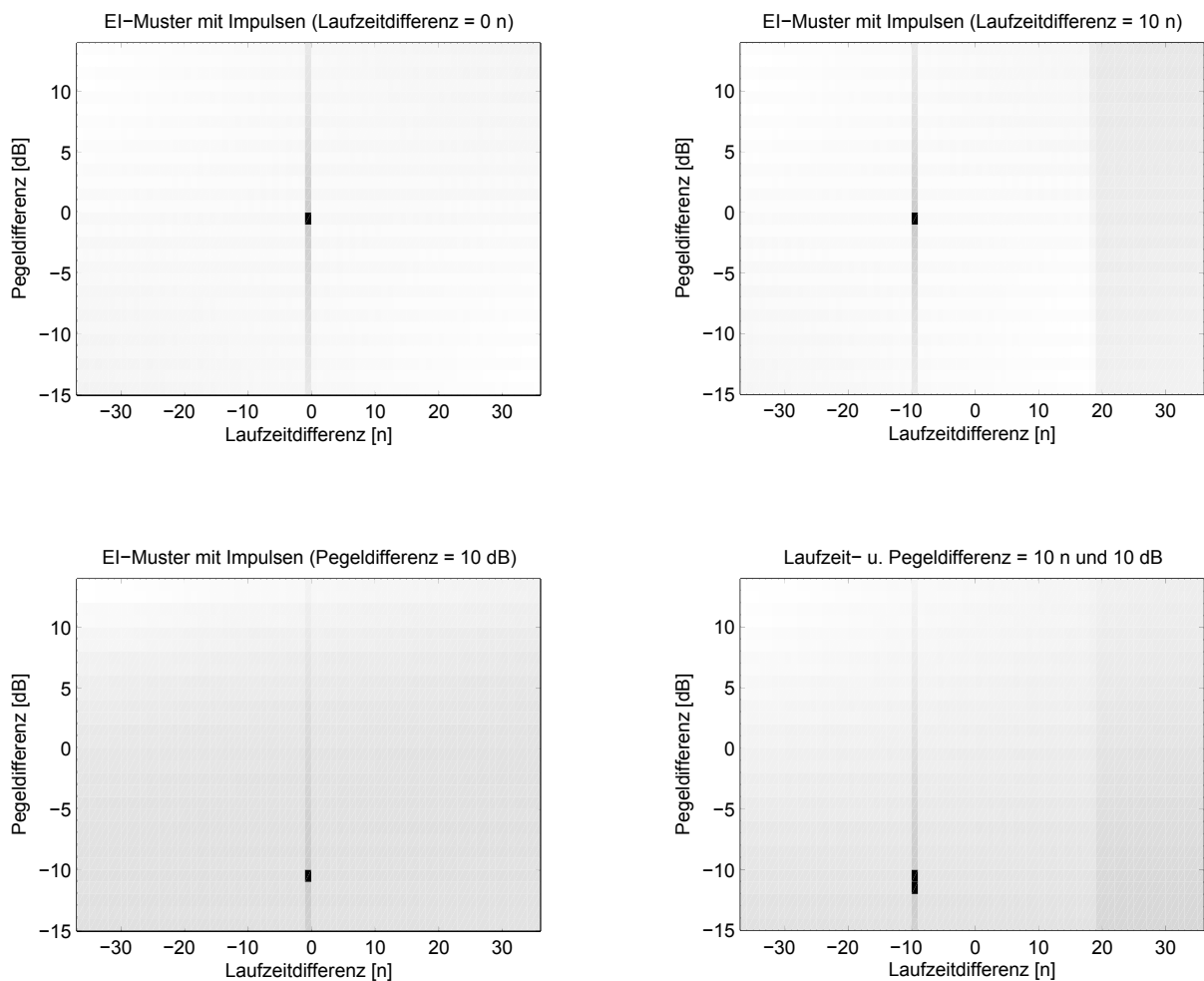


Abbildung 2.12: EI-Muster mit Laufzeit- und Zeitverzögerungen

Wird als Eingangssignal eine Sinusschwingung mit 100 Hz oder 10000 Hz ohne Zeit- bzw. Laufzeitdifferenz genommen, werden bei 10000 Hz mehrere Minima bzw. Laufzeitdifferenzen detektiert und es entsteht räumliches Aliasing (siehe Abb. 2.13). Das eindeutigste Minimum wird aber richtig in der Mitte des Musters detektiert. Schall pflanzt sich bei Raumtemperatur mit ca. 340 m/s fort. Für 17 cm (Kopfdurchmesser) benötigt er 0,5 ms. Das sind bei einer Samplingfrequenz von 44,1 kHz 22,5 Samples. In diesem Verzögerungsbereich tritt für Frequenzen < 1000 Hz kein räumliches Aliasing auf. Da ein Kunstkopf eine räumliche Ausdehnung besitzt, beugen sich Schallwellen um den Körper, was zu einer Vergrößerung der Laufzeitdifferenzen bei tiefen Frequenzen führt. Deshalb werden die Ohrsignale um jeweils 37 Samples zeitverzögert.

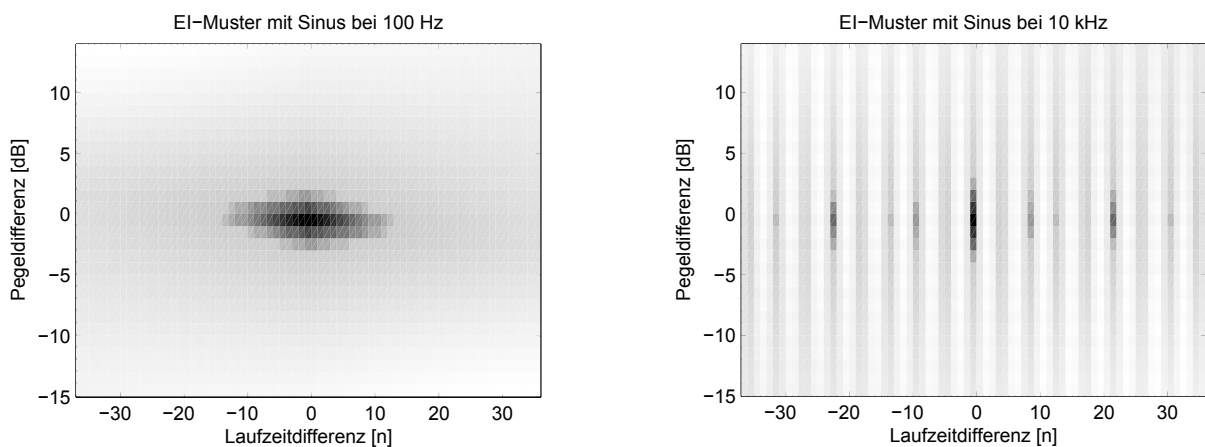


Abbildung 2.13: EI-Muster mit Sinustönen bei 100 Hz und 10000 Hz

In der Simulation mit Matlab [Mata] wird weißes Rauschen mit den HRIRs eines Kunstkopfes gefaltet, um die Richtungsinformation einer Schallquelle in das Eingangssignal einzuprägen. HRIRs (head-related impuls response) sind Impulsantworten, die je einen Übertragungspfad zwischen einem Punkt im Raum und dem Hörkanal angeben. Die Eingangssignale durchlaufen die gesamte auditive Peripherie bevor sie die Lokalisationsstufe erreichen.

Die resultierenden EI-Muster im 7. Barkband ($f_c = 765$ Hz) mit $\Delta\alpha = 1$ dB, $\Delta\tau = 22,5 \mu s$ (1 Sample), $L = 15$ und $M = 37$ sehen folgendermaßen aus:

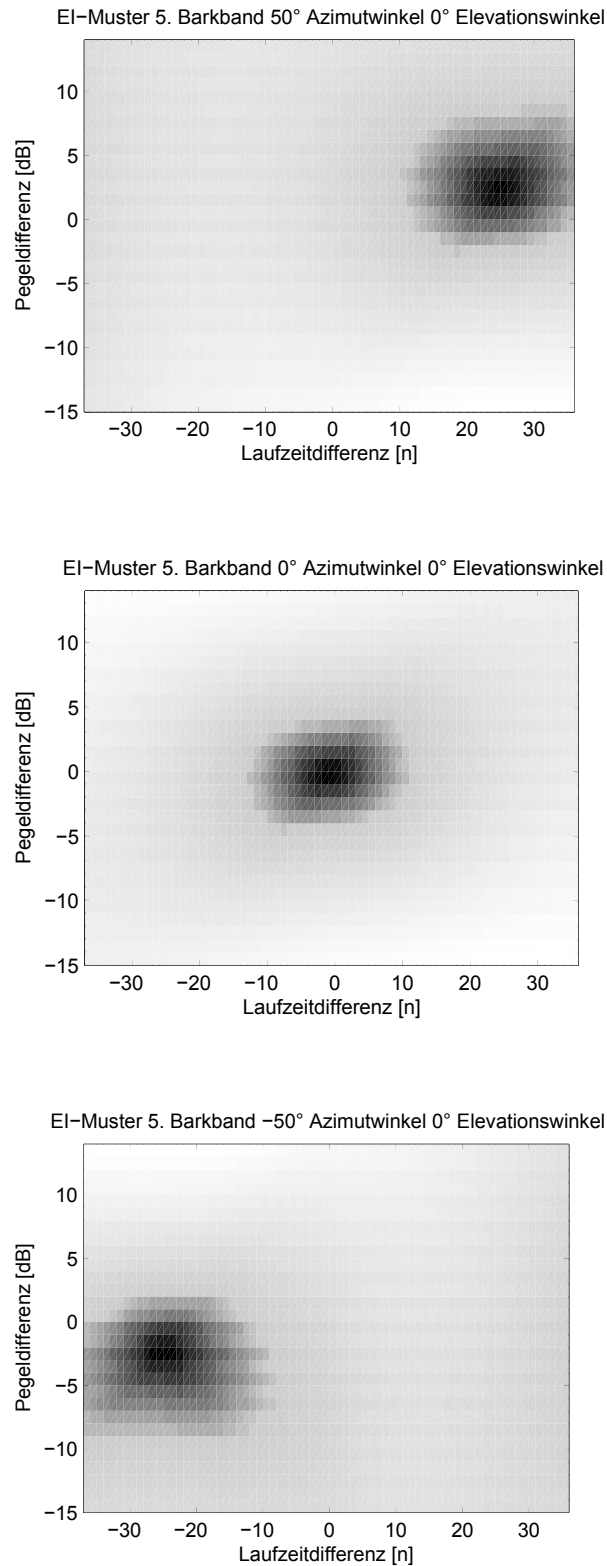


Abbildung 2.14: Die 3 Grafiken zeigen EI-Muster mit einer Schallquelle bei -50° Azimutwinkel und 0° Elevationswinkel oben; 0° Azimutwinkel 0° Elevationswinkel mitte, -50° Azimutwinkel 0° Elevationswinkel unten

In den 3 Abbildung in 2.14 sind eindeutig die Minima erkennbar, durch die Laufzeit- und Pegelunterschiede im EI-Muster kodiert sind.

Betrachtet man das EI-Muster in einem höheren Barkband, zum Beispiel im 15. Barkband, dann sind die Laufzeitunterschiede im EI-Muster, durch den zu großen Ohrenabstand, nicht eindeutig und räumliches Aliasing tritt auf (siehe Abb. 2.15).

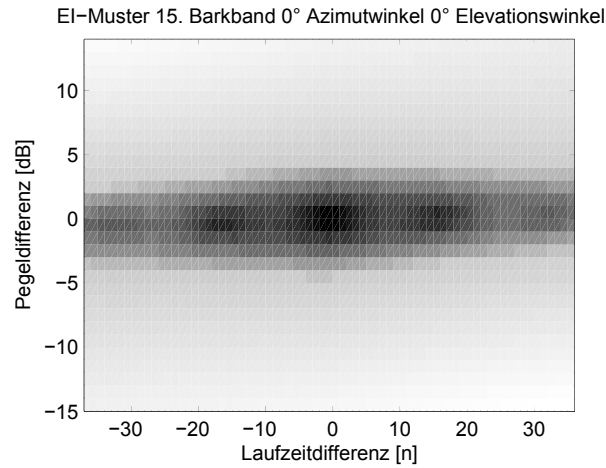


Abbildung 2.15: Räumliches Aliasing im EI-Muster mit Schallquelle bei 0° Azimutwinkel und 0° Elevationswinkel (15. Barkband)

Durch das räumliche Aliasing kommt es bei dem Mustervergleich zu keinen Verwechslungen, da sich die Minima unterschiedlicher Richtungen in der y-Achse durch die Pegeldifferenz verschieben (siehe Abb. 2.16). Das bedeutet, das System ist stabil gegenüber räumlichem Aliasing.

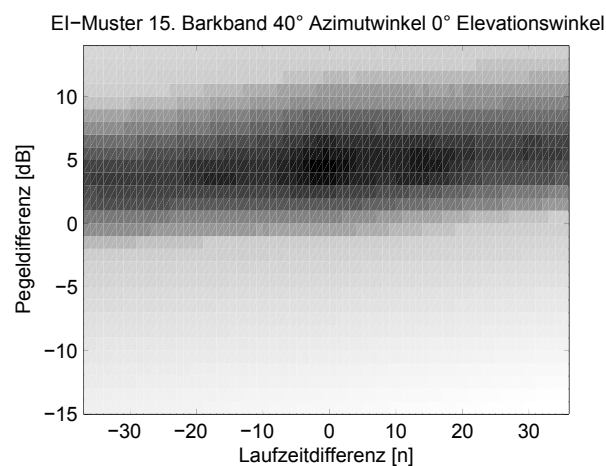


Abbildung 2.16: EI-Muster mit Schallquelle bei 40° Azimutwinkel und 0° Elevationswinkel (15. Barkband)

2.2.2 EI-Mustervergleich

Der EI-Mustervergleich kann als eine Art zweidimensionale Korrelation betrachtet werden, bei dem das von der Simulation berechnete EI-Muster mit Mustern aus einer Datenbank verglichen wird. Hierbei kann die zeitliche Verschiebung einer normalen Korrelation als eine Winkelverschiebung betrachtet werden. Das bedeutet, dass das zu erkennende EI-Muster mit einem bekannten EI-Mustern aus einer bestimmten Richtung verglichen wird. Der Wert der Ähnlichkeit wird in einer Matrix eingetragen und anschließend wird das EI-Muster aus der nächsten Richtung für den Vergleich herangezogen. Dies wird wiederholt, bis der Vergleich mit allen EI-Mustern aus der Datenbank (3D) durchgeführt wurde. Das Maximum der Matrix zeigt die wahrscheinlichste Richtung der Schallquelle.

Die Vergleichsdatenbank wird erstellt, indem die aus der Simulation resultierenden EI-Muster für alle Barkbänder in verschiedenen Richtungen berechnet und gespeichert werden. Für die Eingangssignale mit einer Richtungsinformation wird ein HRIR-Datensatz von einem Kunstkopf (BK 4128C) mit weißem Rauschen (5000 Samples) ohne Rauminformation gefaltet. Der HRIR-Datensatz hat eine Auflösung in 10° Schritten (Azimut = 0° bis 360° , Elevation = -50° bis 90°). Dadurch ergeben sich für jedes Barkband 15×36 EI-Muster für den Vergleich.

Bevor das von der Simulation neu berechnete EI-Muster mit den Mustern aus der Datenbank verglichen wird, werden die Minima der Muster hervorgehoben, damit der Vergleich eindeutiger wird und Verwechslungen verringert werden. Hierfür wird das Verhalten der Arkustangensfunktion genutzt, um das EI-Muster zu komprimieren. Die Muster werden so komprimiert, dass die Minima hervorgehoben werden indem der Wertebereich über den Minima komprimiert wird. Der Faktor β wird für jedes Barkband separat empirisch definiert, sodass sich für jedes Barkband die Verläufe der Minima im selben Wertebereich befinden.

Die Komprimierungsfunktion ist im eindimensionalen Bereich gegeben mit

$$y = \frac{2}{\pi} \operatorname{atan}(x\beta) \quad .$$

Für ein positives Eingangssignal befindet sich der Wertebereich des Arkustangens zwischen 0 und $\pi/2$. Durch die Normierung mit $\pi/2$ wird das Ergebnis in den Bereich 0 bis 1 geschoben (siehe Abb. 2.17). Durch den Faktor β kann die Grenze bestimmt werden, ab dem das EI-Muster in den Sättigungsbereich der Komprimierung gerät. Für das Hervorheben der Minima wurde eine Arkustangensfunktion gewählt, da sie einer Sigmoidfunktion [Matb] ähnlich ist, mit der Aktionsverläufe von Neuronen simuliert werden.

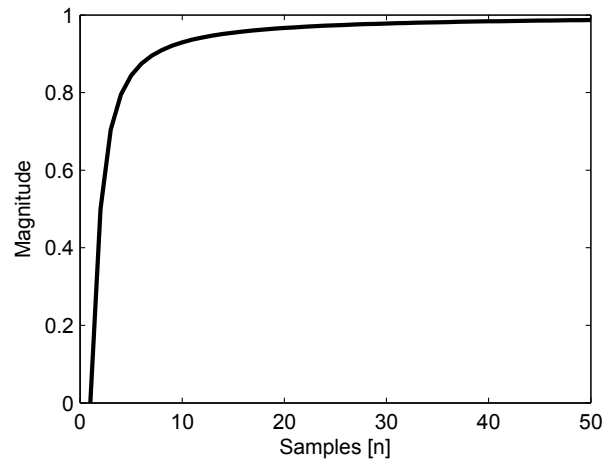
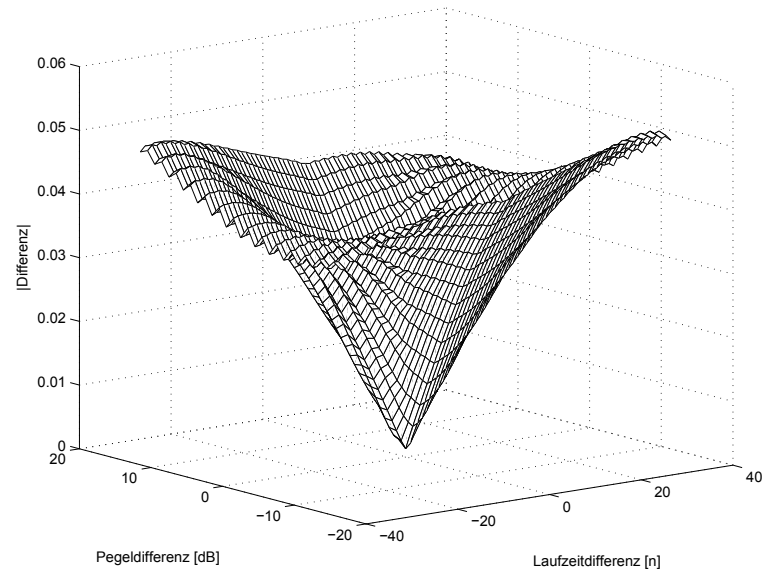
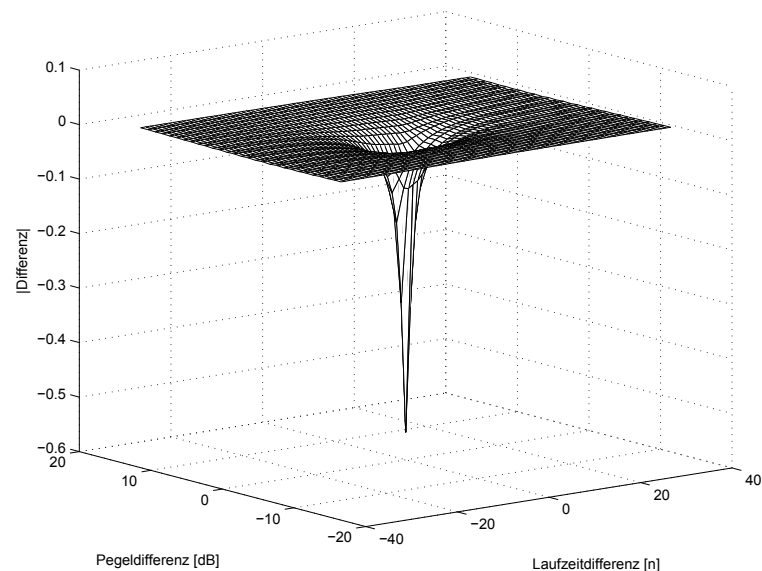


Abbildung 2.17: Normierte Arkustangensfunktion

Für den Vergleich wird das berechnete EI-Muster mit allen Mustern der Vergleichsdatenbank multipliziert. Vor der Multiplikation werden bei allen EI-Mustern der Gleichanteil abgezogen. Anschließend werden die Muster auf die Frobeniusnorm normiert. Zwei identische Muster (gleiche Laufzeit und Pegeldifferenz) besitzen an derselben Stelle im Muster ein negatives Minimum. Durch die Multiplikation wird das Minimum ein deutlich ausgeprägtes positives Maximum.

Abbildung 2.18: EI-Muster mit Schallquelle bei 0° Azimutwinkel und 0° ElevationswinkelAbbildung 2.19: Normiertes EI-Muster mit Schallquelle bei 0° Azimutwinkel und 0° Elevationswinkel nach Abzug des Gleichanteils und der Arkustangensfunktion

Das von der Simulation berechnete EI-Muster wird mit jedem EI-Muster der Vergleichsdatenbank in einem Barkband multipliziert. Die resultierenden Muster werden jeweils Spalten- und Zeilenweise aufsummiert und in eine 15×36 Matrix gefüllt, die einer Lokalisationsmatrix entspricht. Dies bedeutet es wird beim Mustervergleich der Korrelationskoeffizient berechnet. Sind zwei Muster vollständig identisch ist dieser 1. Allgemein ergibt sich der Korrelationskoeffizient mit:

$$Korr_{(x,y)} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} . \quad (2.2.1)$$

Dieser Vorgang wird für alle Barkbänder durchgeführt. Die resultierenden 22 Lokalisationsmatrizen werden addiert, wodurch die endgültige Lokalisationsmatrix entsteht (siehe Abb. 2.20).

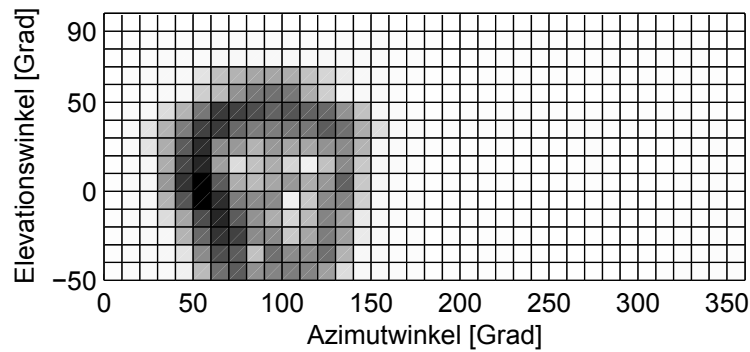


Abbildung 2.20: Lokalisationsmatrix mit Schallquelle bei 50° Azimutwinkel und 0° Elevationswinkel

Das Maximum der Lokalisationsmatrix ist bei 50° Azimutwinkel und 0° Elevationswinkel. Die Schallquelle wird somit richtig lokalisiert. Es ist auch der cone-of-confusion erkennbar, der durch die Mehrdeutigkeiten innerhalb der EI-Muster zu Stande kommt. In Abbildung 2.21 wird die Lokalisationsmatrix auf einer Kugel aufgetragen wodurch der cone-of-confusion besser dargestellt wird.

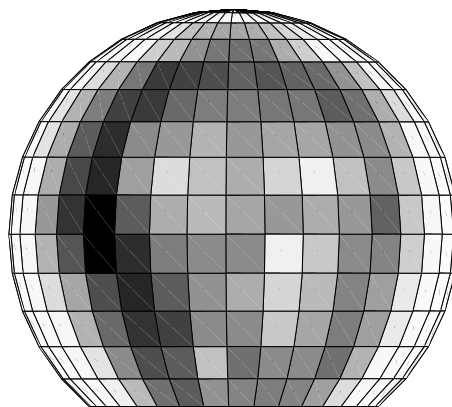


Abbildung 2.21: Lokalisationsmatrix mit einer Schallquelle bei 50° Azimutwinkel und 0° Elevationswinkel auf einer Kugel dargestellt

2.2.3 Verringerung der Richtungsmehrdeutigkeiten durch eine simulierte Kopfdrehung

Aufgrund der in Abschnitt 1.4 erwähnten Eigenschaften, kommt es zu einer Mehrdeutigkeit in der Lokalisation einer Richtung. Diese Mehrdeutigkeit ist auch in den EI-Mustern zu finden. So sind sich EI-Muster entlang des cone-of-confusion oft sehr ähnlich, wobei diese Ähnlichkeit durch die frequenzabhängigen Unterschiede durch Pinna, Torso usw. minimiert wird. Bei dem Mustervergleich werden die Muster entlang des cone-of-confusion als mögliche Richtung in der Lokalisationsmatrix angegeben. In Abbildungen 2.22 und 2.23 sieht man eine Verringerung der Mehrdeutigkeit in höheren Frequenzbändern gegenüber tiefen (Quelle bei 20° Azimutwinkel):

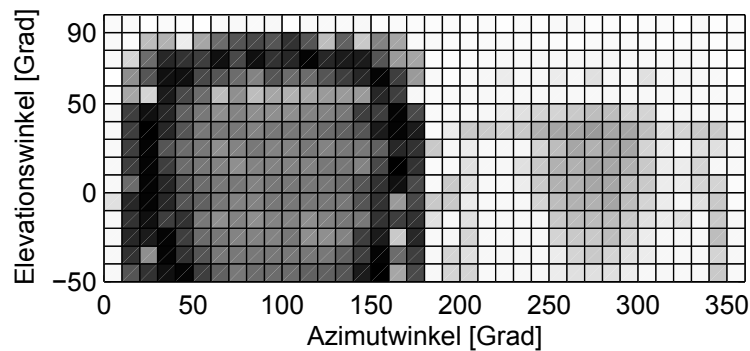


Abbildung 2.22: Lokalisationsmatrix der Barkbänder 4 bis 10 (Quelle bei 20° Azimutwinkel)

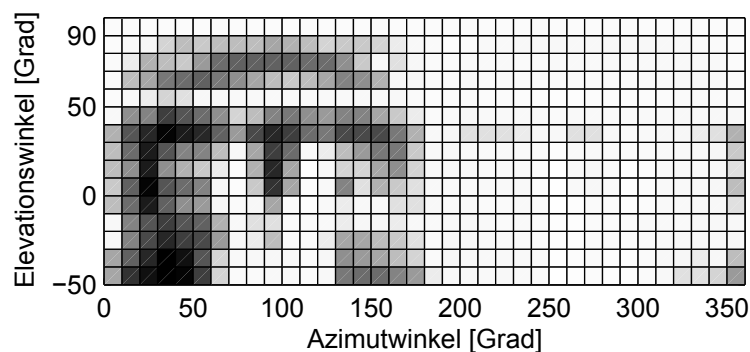


Abbildung 2.23: Lokalisationsmatrix der Barkbänder 12 bis 18 (Quelle bei 20° Azimutwinkel)

In Abschnitt 1.5 wird gezeigt, dass ein Mensch die Vorne-/Hinten-Mehrdeutigkeit überwinden kann, indem er den Kopf bewegt. Dies wird in diesem binauralen Lokalisationsmodell implementiert. Für die Beschreibung der Grundidee, wird der zweidimensionale Fall in der Azimutebene angenommen. Der Mensch benutzt ein Referenzkoordinatensystem, auf das er seine Hörwahrnehmung bezieht. Wird eine Schallquelle in der Azimutebene lokalisiert, werden eine richtige Schallquelle wie auch deren Verwechslung wahrgenommen. Die verwechselte Schallquelle ist eine an der interauralen Achse gespiegelte Quelle. Wird nun der Kopf bewegt, bleibt die richtige Quelle an dem selben physikalischen Ort, wie auch an der selben Stelle im Referenzkoordinatensystem, während sich die verwechselte Quelle in der Azimutebene verschiebt (siehe Abb. 2.24). Der Azimutwinkel der verwechselten Quelle verschiebt sich immer mit doppeltem Winkel der Kopfdrehung in der Richtung der Kopfdrehung. Dynamisch betrachtet bewegt sich die verwechselte Quelle immer mit doppelter Winkelgeschwindigkeit der Kopfdrehung in die selbe Richtung der Kopfdrehung.

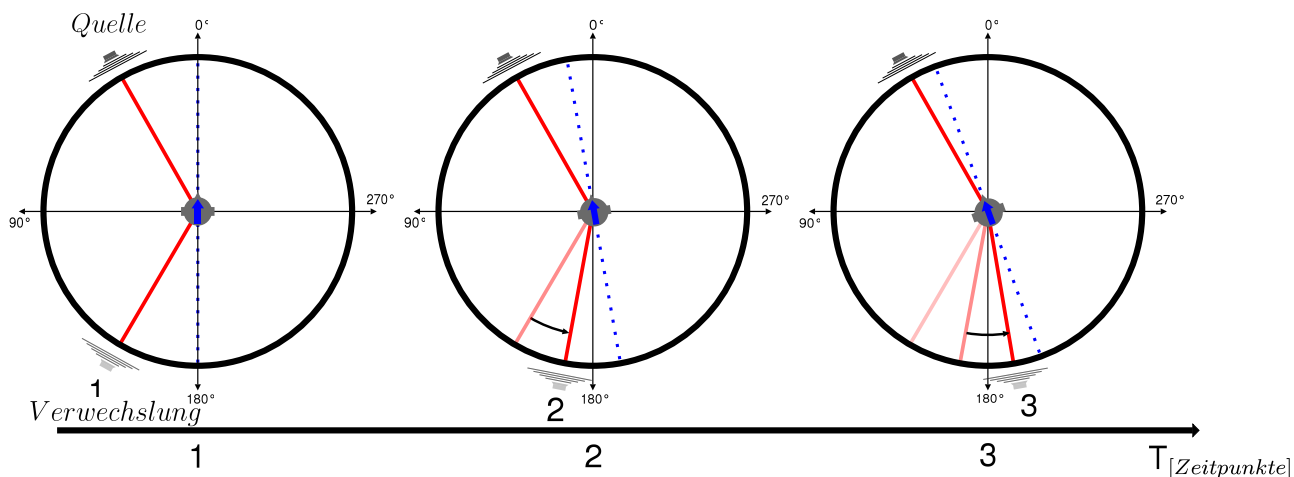


Abbildung 2.24: Bewegung einer verwechselten Quelle bei einer Kopfdrehung; die Quelle bewegt sich mit doppelter Winkelgeschwindigkeit der Kopfdrehung

In der Simulation in Matlab wird diese Gegebenheit genutzt, indem man einen HRIR Datensatz verwendet, der für mehrere Kopfdrehungen gemessen wurde. Das in den vorherigen Abschnitten vorgestellte Modell wird also für mehrere Kopfdrehungen durchgerechnet, bei denen die Quelle an dem selben physikalischen Ort verweilt. Als Referenzsystem wird das System ohne Kopfdrehung verwendet, auf das die Ergebnisse mit Kopfdrehung bezogen werden. Dies bedeutet, dass der Azimutwinkel der Lokalisationsmatrizen so verschoben wird, dass er mit dem Referenzsystem übereinstimmt. Werden nun die Lokalisationsmatrizen punktweise miteinander multipliziert, löschen sich die verwechselten Richtungen aus und die physikalisch richtige Richtung wird positiv überlagert.

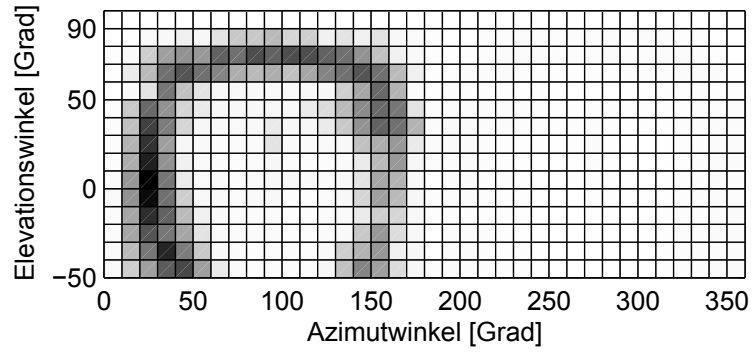


Abbildung 2.25: Lokalisationsmatrix mit Kopfdrehung = 0° und Schallquelle = 20°

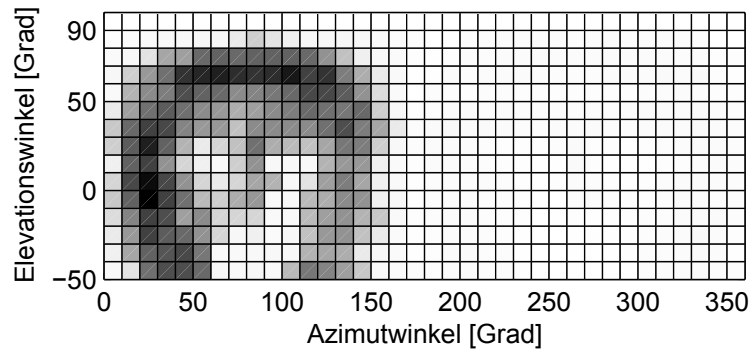


Abbildung 2.26: Lokalisationsmatrix mit Kopfdrehung = -10° und Schallquelle = 20° bezogen auf das Referenzsystem (Kopfdrehung = 0°)

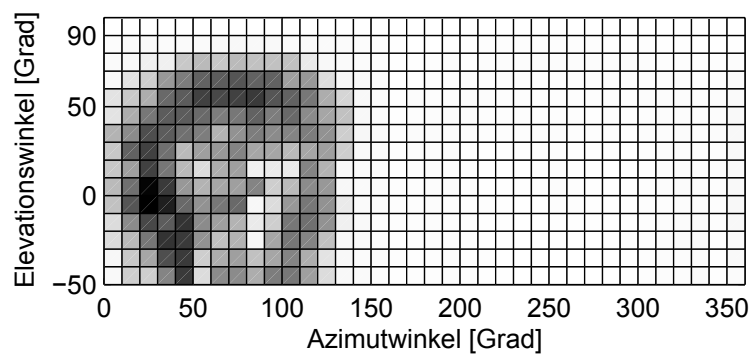


Abbildung 2.27: Lokalisationsmatrix mit Kopfdrehung = -20° und Schallquelle = 20° bezogen auf das Referenzsystem (Kopfdrehung = 0°)

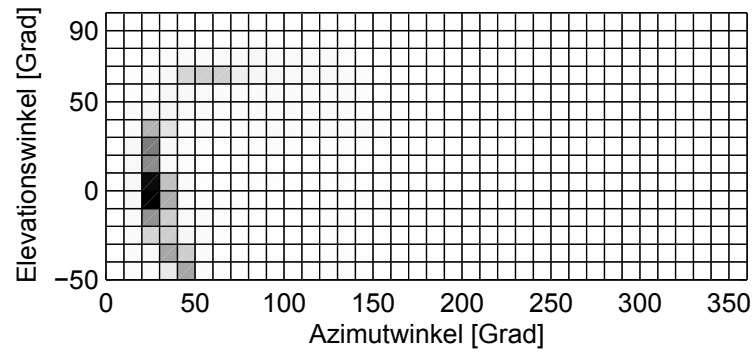


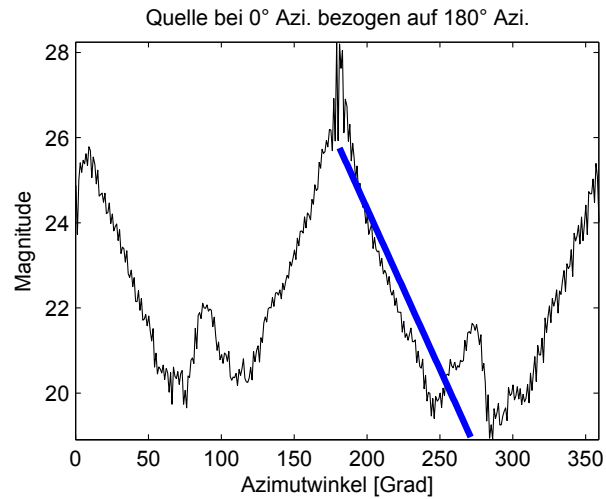
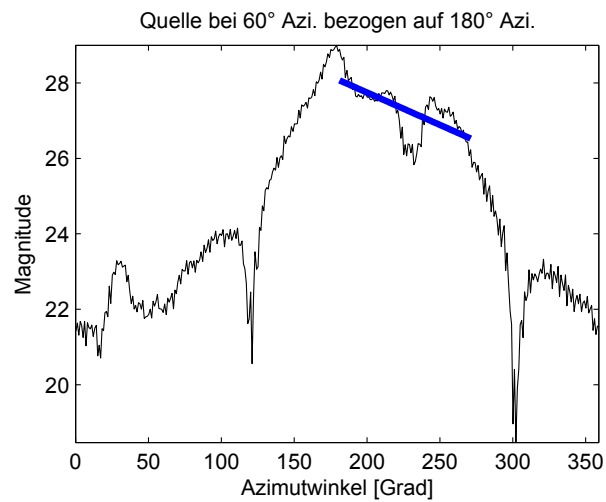
Abbildung 2.28: Ergebnis der Multiplikation der 3 Lokalisationsmatrizen; Schallquelle wird ohne Mehrdeutigkeit bei 20° lokalisiert)

In den Abbildungen 2.25 bis 2.27 sind 3 Lokalisationsmatrizen mit den Kopfdrehungen 0° , -10° und -20° dargestellt. Eine Schallquelle liegt bei 20° . Als Referenzsystem wird die Lokalisationsmatrix ohne Kopfdrehung verwendet. Die Lokalisationsmatrizen mit den Kopfdrehungen bei -10° und -20° sind auf das Referenzsystem bezogen. Werden diese 3 Lokalisationsmatrizen miteinander multipliziert, löschen sich die Mehrdeutigkeiten aus, da sie in allen 3 Matrizen an unterschiedlichen Orten vorkommen (Abb. 2.28). Der Mensch bewegt den Kopf bei der Lokalisation nur wenige Gradschritte. In der Simulation werden die Extremfälle aufgrund der groben Auflösung gewählt.

2.3 Berechnung der Quellenbreite und der Lokalisationsunschärfe

Da für eine Aussage über eine Quellenbreite und Lokalisationsunschärfe ein hochauflöser HRIR Datensatz benötigt wird, wird in diesem Abschnitt ein frei zugänglicher Datensatz des Spatial Audio Research Institutes der TU Berlin [TuB] verwendet, in dem nur HRIRs der Horizontalebene, aber dafür mit einer Auflösung von 1° Schritten, beinhaltet sind. Dies bedeutet, dass in diesem Abschnitt die lokalisierte Richtung durch das Maximum eines Lokalisationsvektors und nicht einer Lokalisationsmatrix beschrieben wird, da eine Dimension wegfällt. Das Maximum der Verwechslung wird hierbei außer Acht gelassen. Damit Quellenbreite und Lokalisationsunschärfe unterschiedlicher Richtungen miteinander vergleichbar sind, werden die Maxima in die 180° Richtung geschoben, dadurch befindet sich das Maximum jeder lokalisierten Richtung bei 180° .

Die Quellenbreite wird über die Steigung einer Geraden berechnet, welche an die Steigung einer Flanke des Maximums mit der Least Squares Methode angeglichen wird. Es wird immer die Flanke gewählt, welche dem Maximum der Hinten-Verwechslung näher ist. Die Gerade wird immer zwischen 180° und 270° bzw. 90° angesetzt. Dies ist bei einer Lokalisation einer frontal präsentierten Quelle, die auf 180° bezogen wird, genau die Mitte des Öffnungswinkels zwischen Quelle und der Verwechslung. Dieser Ansatz basiert auf der Idee, dass der Öffnungswinkel der Verwechslung einer Quelle, die zuerst frontal präsentiert wird und dann zur Seite wandert, kleiner wird. Das heißt das Maximum der Verwechslung schiebt sich in den Bereich der Geraden und damit wird die Steigung kleiner. Durch den Betrag der Steigung können nun die Quellenbreiten in Relation zueinander gesetzt werden. Eine größere Steigung bedeutet eine kleinere Quellenbreite.

Abbildung 2.29: Lokalisationsvektor einer Quelle bei 0° mit angeglicherer GeradeAbbildung 2.30: Lokalisationsvektor einer Quelle bei 60° mit angeglicherer Gerade

In den Abbildungen 2.29, 2.30 und 2.31 ist anhand der Steigungen der Geraden ersichtlich, dass die Breite einer Quelle bei $120^\circ > 60^\circ > 0^\circ$ ist.

Setzt man die Quellenbreiten (Beträge der Steigungen der Geraden) einer Quelle, die von 0° bis 180° wandert, in Relation zueinander, ergibt sich die Grafik siehe Abb. 2.32.

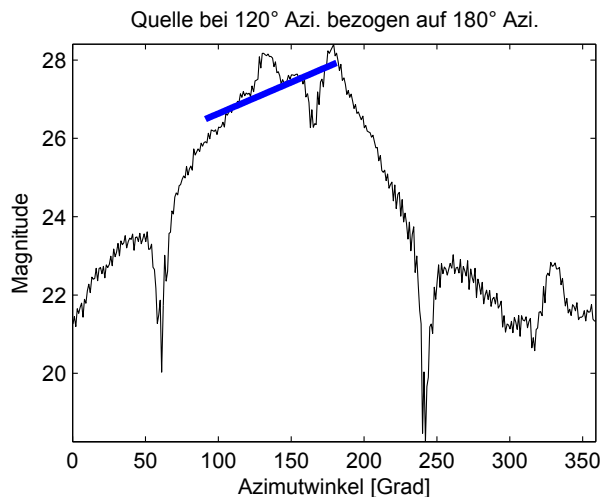


Abbildung 2.31: Lokalisationsvektor einer Quelle bei 120° mit angeglichenen Geraden

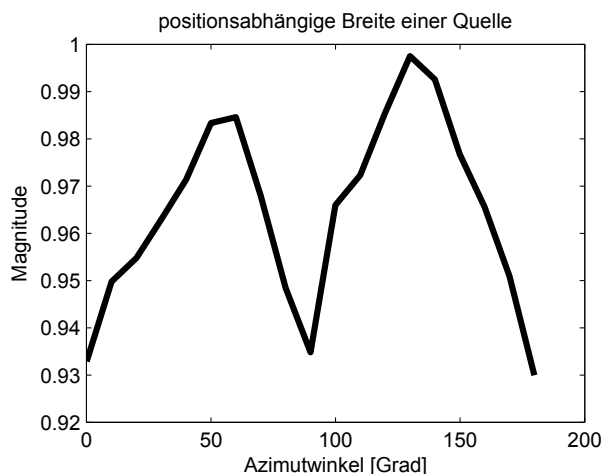


Abbildung 2.32: Quellenbreite einer Schallquelle bei unterschiedlichen Positionen

Invertiert man die Grafik, deckt sich das Ergebnis fast vollständig mit dem Ergebnis der IACC in Abschnitt 1.7. Die Breite einer Schallquelle ist abhängig von ihrer Position. Sie ist bei ca. 60° und 130° am größten, und bei 0° und 90° am kleinsten. Bringt man dies mit dem Öffnungswinkel zwischen Quelle und Verwechslung in Zusammenhang, so ist die Quelle bei den Extremfällen des Öffnungswinkel, bei 0° und 180° klein und wird im Zwischenbereich größer.

Die Steigung einer Geraden, die an die Flanke einer Phantomschallquelle bei 0° angeglichen wird, ist kleiner die einer realen Quelle bei 0° (siehe Abb. 2.33). Durch Hörversuche wird gezeigt, dass eine Phantomschallquelle größer als eine reale Quelle wahrgenommen wird. Dass eine kleinere Steigung der Geraden eine größere wahrgenommene Quelle beschreibt, wird dadurch zusätzlich bestätigt.

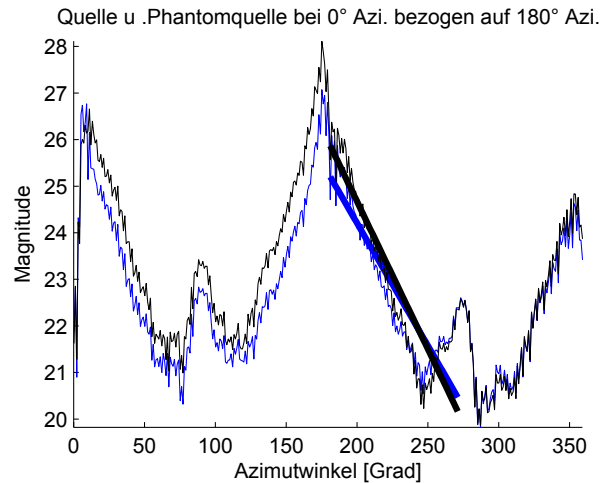


Abbildung 2.33: Phantomschallquelle blau; reale Quelle schwarz

Für die Berechnung der Lokalisationsunschärfe wird der Bereich $\pm 90^\circ$ um das Maximum genauer betrachtet (siehe Abb. 2.35). Das Maximum wird auf 1 normiert und anschließend wird der erwähnte Bereich aufaddiert. Es wird die Energie in diesem Bereich ermittelt. Durch die Normierung werden die Resultate unabhängig vom Maximum der Energie und abhängig von der Energieverteilung um das Maximum. Je mehr Energie im Bereich um das Maximum vorhanden ist, desto unschärfer wird eine Quelle lokalisiert.

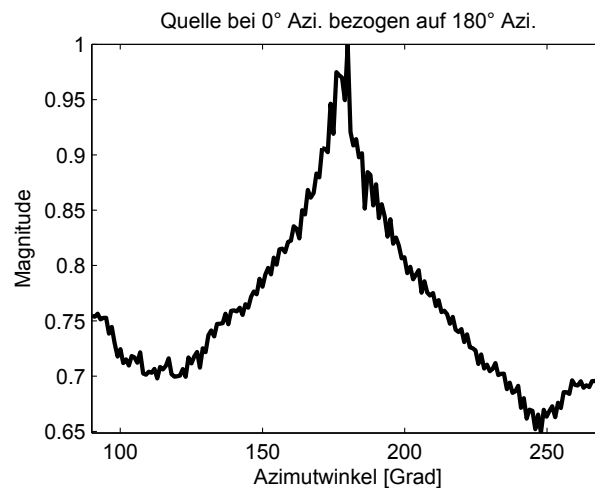


Abbildung 2.34: Bereich der für die Berechnung der Lokalisationsunschärfe aufaddiert wird

Es wird die Energie für reale Quellen bei 0° bis 180° Azimutwinkel berechnet. Das Ergebnis in Abb. 2.35 stimmt mit der von Blauert beschriebenen Lokalisationsunschärfe überein [Bla74]. Die Lokalisationsunschärfe ist vorne am kleinsten, nimmt zu 90° Azimutwinkel zu und nimmt zu 180° wieder ab.

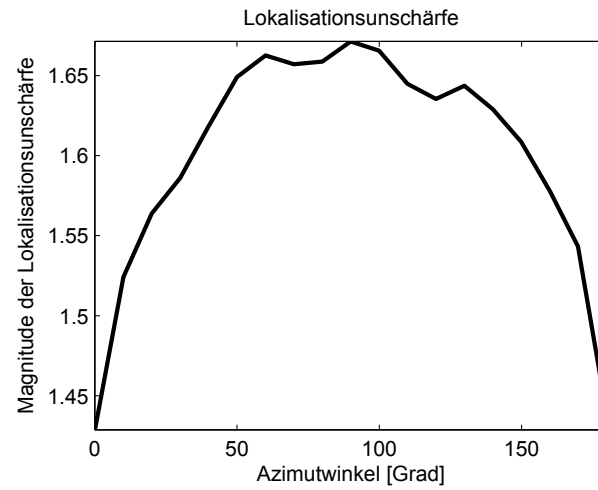


Abbildung 2.35: Lokalisationsunschärfe einer realen Quelle zwischen 0° und 180° Azimutwinkel

2.4 Echtzeitumsetzung des Lokalisationsmodells mit Pd



Bei dem Echtzeit-Lokalisationsmodell soll eine Schallquelle mittels eines Kunstkopfs lokalisiert werden. Dem Kunstkopf wird ein Signal mittels eines Lautsprechers zugespielt und die 2 Ohrsignale, die durch 2 Mikrofone in den Ohren des Kunstkopfs aufgezeichnet werden, werden in Echtzeit auf die Richtungsinformationen analysiert. Die Implementierung in Pd (eine grafische auf C basierende Echtzeit-Programmiersprache [pd]) ist der Umsetzung in Matlab sehr ähnlich. Aufgrund der beschränkten Rechenleistung durch eine Echtzeitanwendung wird aber der Lokalisationsbereich auf die Azimutebene verkleinert und die Anzahl der Barkbänder wird verringert. Auf die Reduzierung der Richtungsmehrdeutigkeiten durch die Berechnung einer zusätzlichen Kopfdrehung wird auch verzichtet. Es wird zusätzlich 2fach unterabgetastet. Als höchstes Barkband wird das 18. mit einer Mittenfrequenz bei 4380 Hz verwendet, wodurch es zu keinen Aliasing-Artefakten kommt.

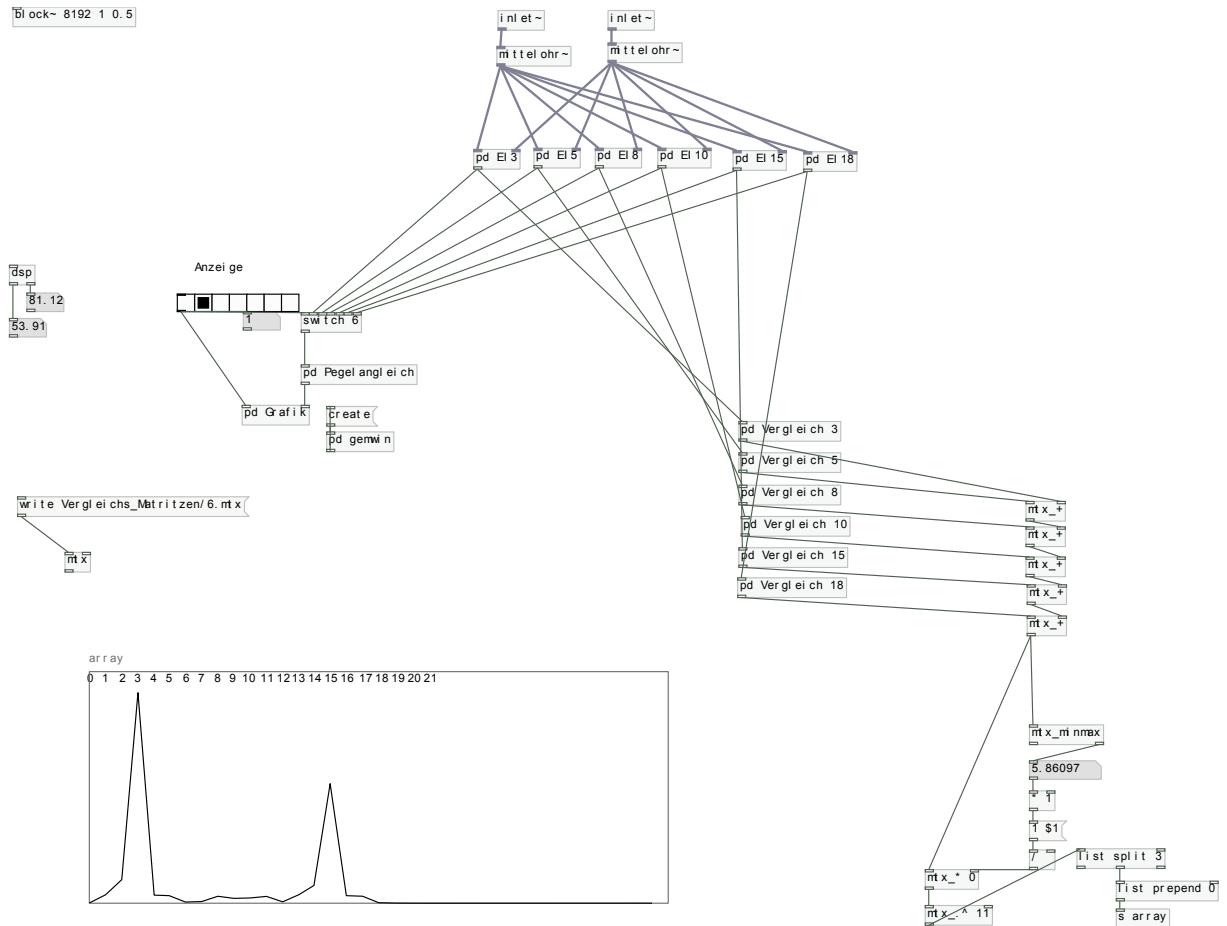


Abbildung 2.36: Ansicht eines Teils des Echtzeit-Lokalisationsmodells in Pd; links unten im Bild wird eine Schallquelle bei 30° lokalisiert

Für die Echtzeitimplementierung werden die Barkbänder 3, 5, 8, 10, 15 und 18 verwendet. Es wurden diese Barkbänder gewählt, damit trotz reduzierter Genauigkeit der gesamte zu untersuchende Frequenzbereich grob abgetastet wird. Da tiefe Frequenzen größere Unterschiede in der Zeitverzögerung und hohe Frequenzen im Pegel aufweisen, wird die Größe der EI-Muster optimal an das jeweilige Barkband angepasst, um die Rechenleistung zusätzlich zu optimieren. So werden zum Beispiel für das 3. Barkband 35 Verzögerungsstufen und 21 Dämpfungsschritte, und für das 18. Barkband 27 Verzögerungsstufen aber 23 Dämpfungsschritte verwendet.

Mehrere EI-Muster gleichzeitig zu berechnen stellte sich als zu rechenaufwändig heraus. Deshalb wurde von Franz Zotter ein eigenes Pd-External (eine in C programmierte Erweiterung für Pd), das ist eine in C programmierte Erweiterung, auf die in der grafischen Benutzeroberfläche von Pd zugegriffen werden kann, für das EI-Muster programmiert.

Kapitel 3

Evaluierung des Lokalisationsmodells

In diesem Kapitel wird die Funktionstüchtigkeit des Lokalisationsmodells evaluiert. Es werden unterschiedliche Eingangssignale unter unterschiedlichen Bedingungen getestet. Das Kapitel teilt sich in zwei Einheiten auf. In dem ersten Teil werden die Eingangssignale simuliert, im zweiten werden echte Kunstkopf-Aufnahmen als Eingangssignale verwendet. Als Offline-Modell wird der bereits vorher beschriebene HRIR-Datensatz mit Azimut, Elevation und einer Auflösung von 10° verwendet. Zusätzlich werden die Mehrdeutigkeiten durch die Kopfdrehung korrigiert.

3.1 Ergebnisse unter simulierten Bedingungen

Unter einer simulierten Bedingung werden simulierte Eingangssignale für das Modell verstanden. Die Eingangssignale werden mit einem vorhanden HRIR Datensatz gefaltet, wodurch die Richtungsinformation in das Signal eingepreßt wird. Die HRIRs wurden in einem trockenen Raum gemessen, wodurch keine Rauminformation vorhanden ist. In den nachfolgenden Ergebnissen werden die Richtungsmehrdeutigkeiten durch eine Kopfdrehung korrigiert.

3.1.1 Eine Quelle mit weißem Rauschen

Als Eingangssignal wird weißes Rauschen (1000 Samples) mit HRIRs gefaltet. Es werden dem Modell verschiedene Richtungen präsentiert. Einige Ergebnisse werden im Folgenden dargestellt:

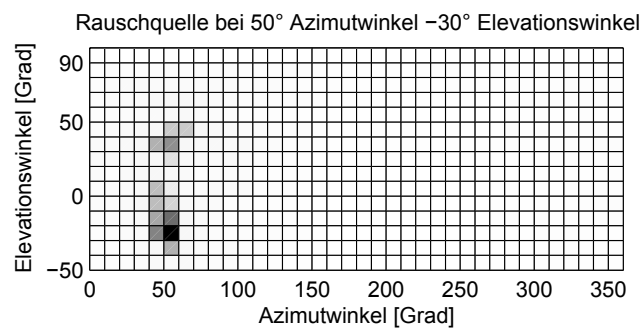
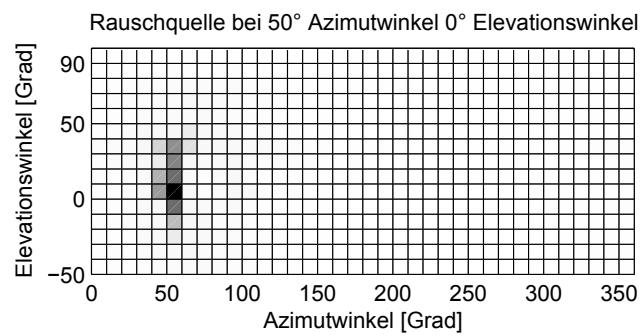
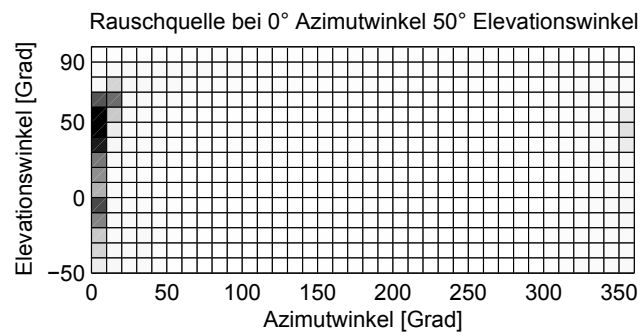
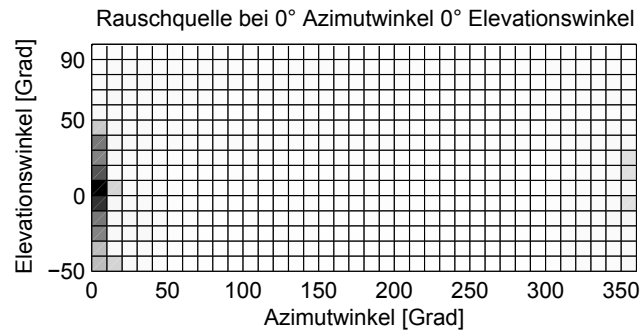


Abbildung 3.1: Lokalisationsmatrizen einer Rauschquelle bei unterschiedlichen Winkeln

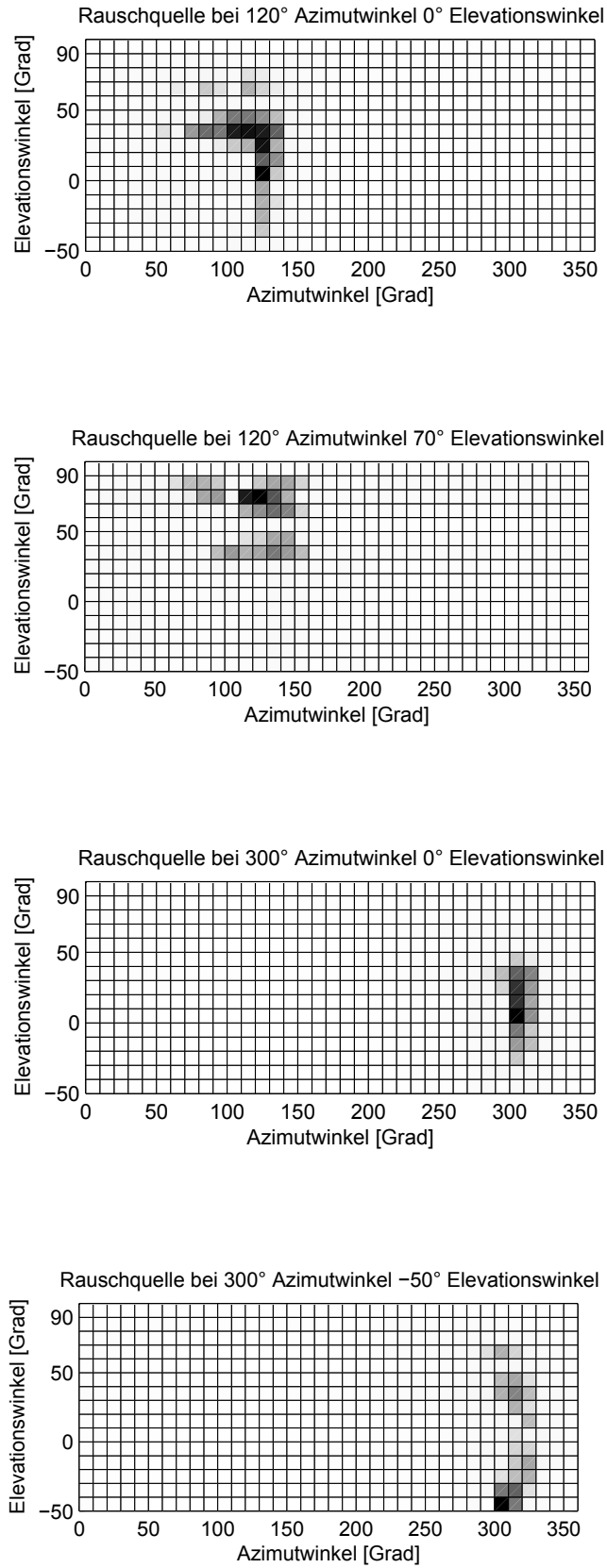


Abbildung 3.2: Lokalisationsmatrizen einer Rauschquelle bei unterschiedlichen Winkeln

In den Abbildungen 3.1.1 und 3.1.1 sind Lokalisationsmatrizen mit unterschiedlichen Lokalisationswinkeln dargestellt. Eine Rauschquelle in der Azimutebene wird immer eindeutig lokalisiert. Außerhalb der Azimutebene wird eine Quelle auch richtig lokalisiert, die Lokalisationsunschärfe wird aber größer.

3.1.2 Eine Quelle mit Schmalbandrauschen

Als Eingangssignal wird Schmalbandrauschen mit unterschiedlichen Mittenfrequenzen und 1 Bark Bandbreite verwendet.

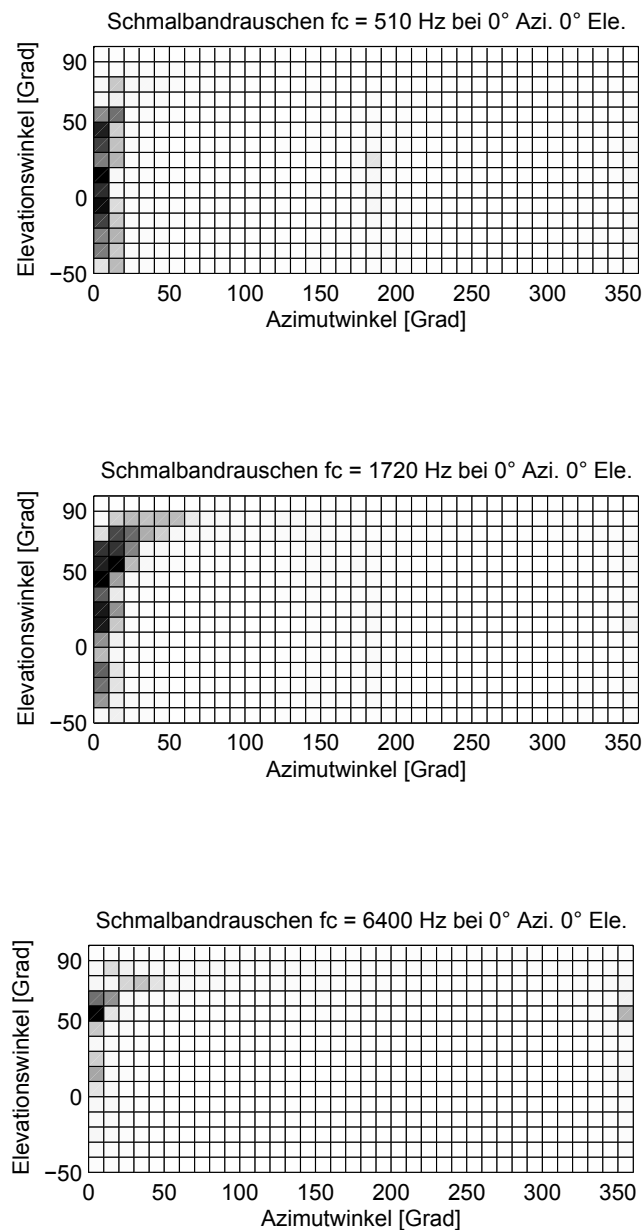


Abbildung 3.3: Schmalbandrauschquelle bei unterschiedlichen Mittenfrequ. und 0° Azimut

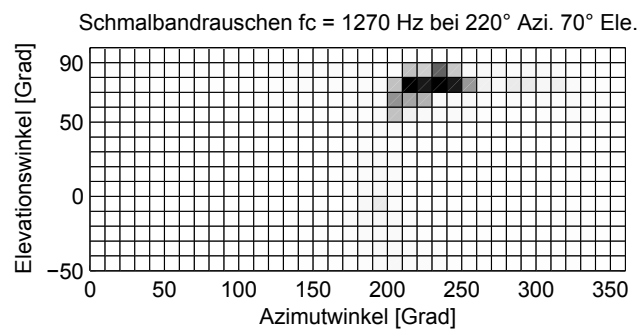
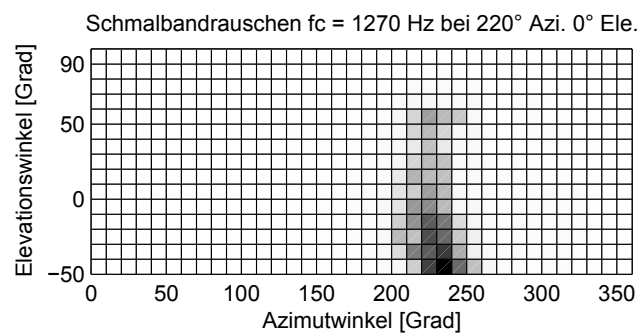
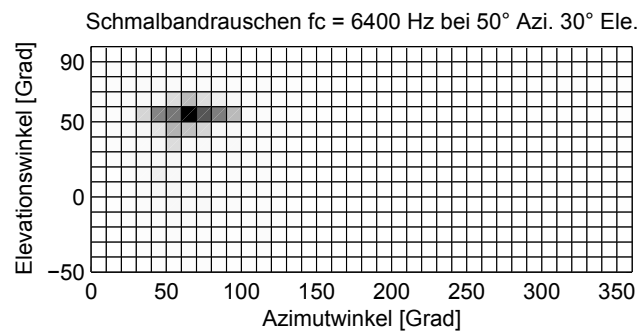
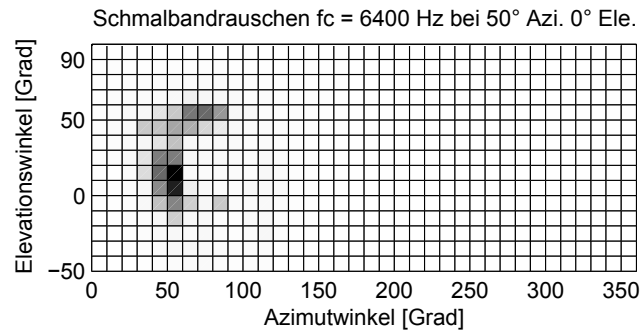


Abbildung 3.4: Schmalbandrauschquelle bei unterschiedlichen Mittenfreq. und Winkeln

Die Azimutwinkel der Schmalbandrauschquellen werden richtig lokalisiert. Die Elevation der Quelle wird sehr ungenau und meistens zu hoch lokalisiert. Wird eine Quelle bei einem größeren Elevationswinkel (z.B. 70°) dem Modell präsentiert, wird die Elevation genauer lokalisiert.

3.1.3 Eine Sinusquelle

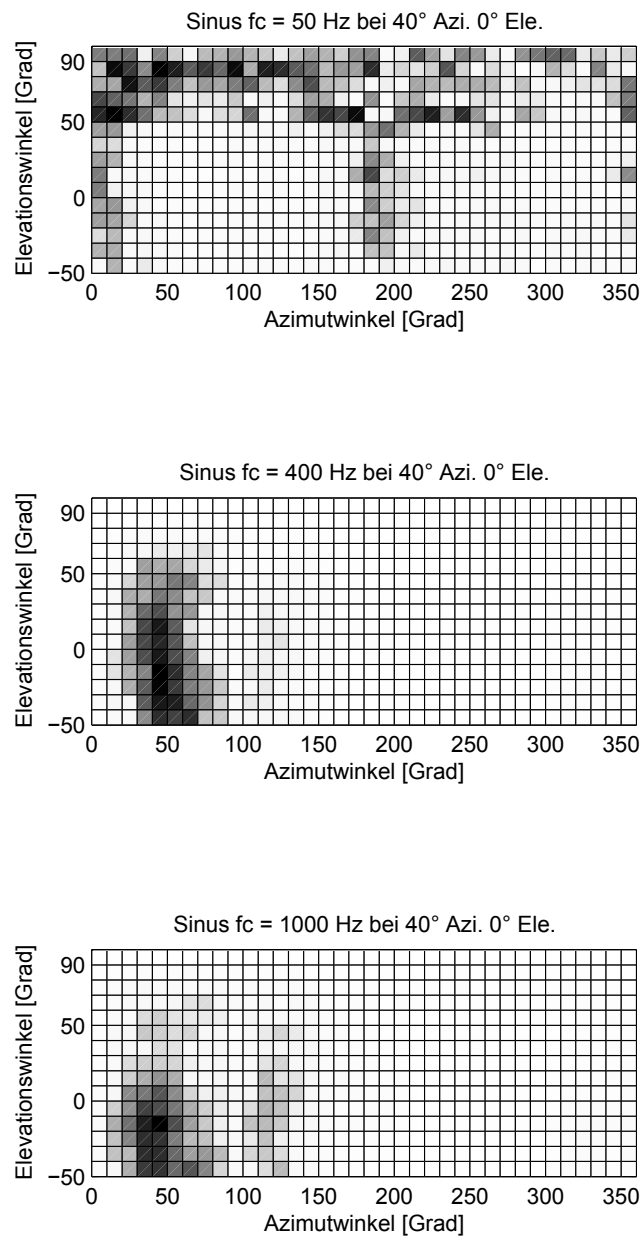


Abbildung 3.5: Sinusquelle mit 50Hz, 400 Hz und 1000 Hz bei 40° Azimut

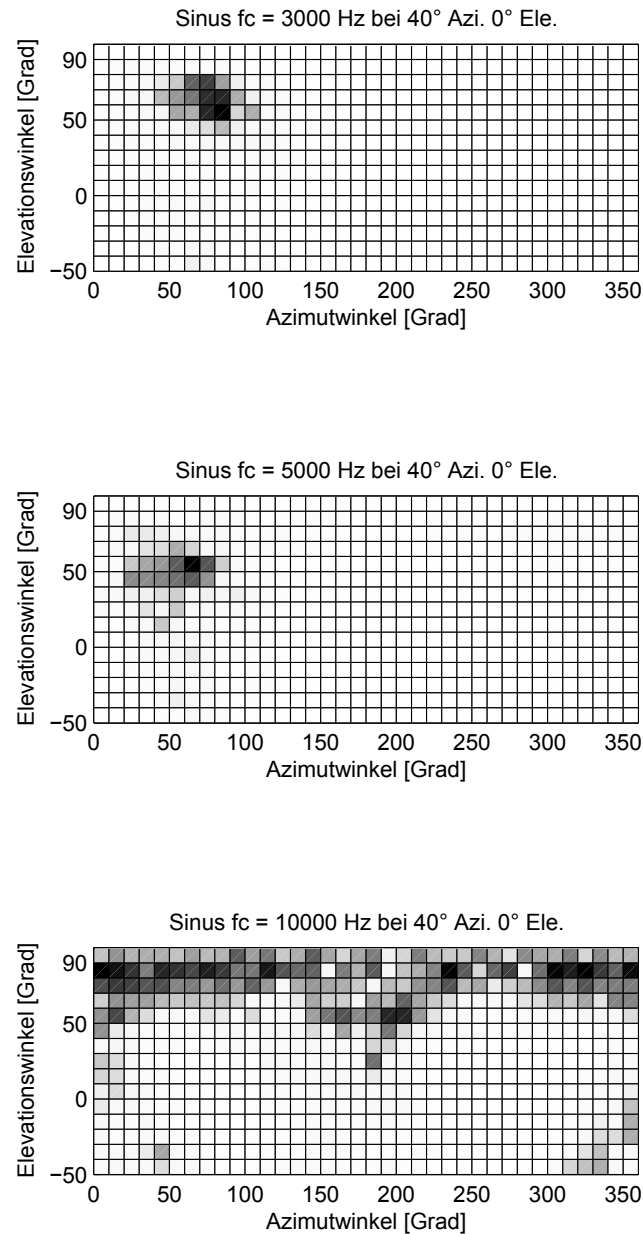


Abbildung 3.6: Sinusquelle mit 3000Hz, 5000 Hz und 10000 Hz bei 40° Azimut

Wie auch beim Schmalbandrauschen wird der Elevationswinkel der Quelle nicht exakt lokalisiert. In den Abbildungen 3.5 und 3.6 ist eine Abnahme der Lokalisationschärfe für hohe und tiefe Frequenzen sichtbar. Der Azimutwinkel einer Sinusquelle mit den Frequenzen bei 400 Hz und 1000 Hz wird richtig lokalisiert. Unter 100 Hz und ab 3000 Hz wird auch der Azimutwinkel nicht mehr erkannt. Wird ein sehr hoher oder ein sehr tiefer Sinuston von 2 Empfängern abgetastet geht die Phaseninformation verloren. Deshalb ist es auch für den Menschen sehr schwer Sinustöne im höheren und tieferen Frequenzbereich zu lokalisieren.

3.1.4 Eine Quelle mit menschlicher Stimme

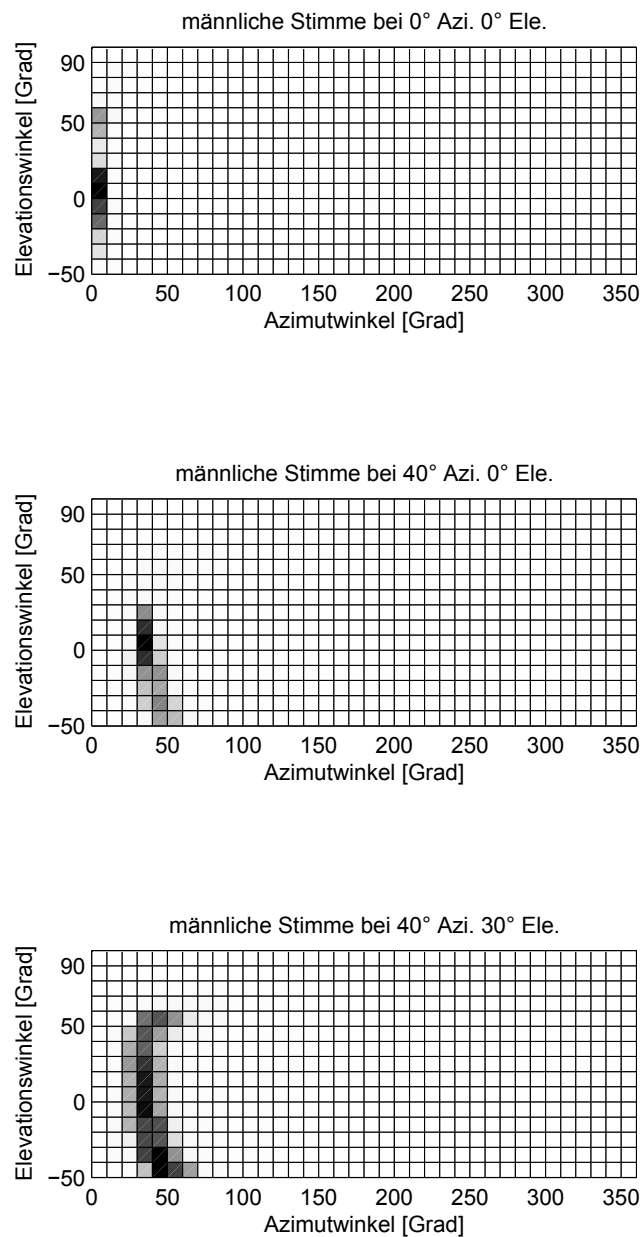


Abbildung 3.7: Männliche Stimme bei 0° Azimutwinkel 0° Elevationswinkel; 40° Azimutwinkel 0° Elevationswinkel; 40° Azimutwinkel 30° Elevationswinkel

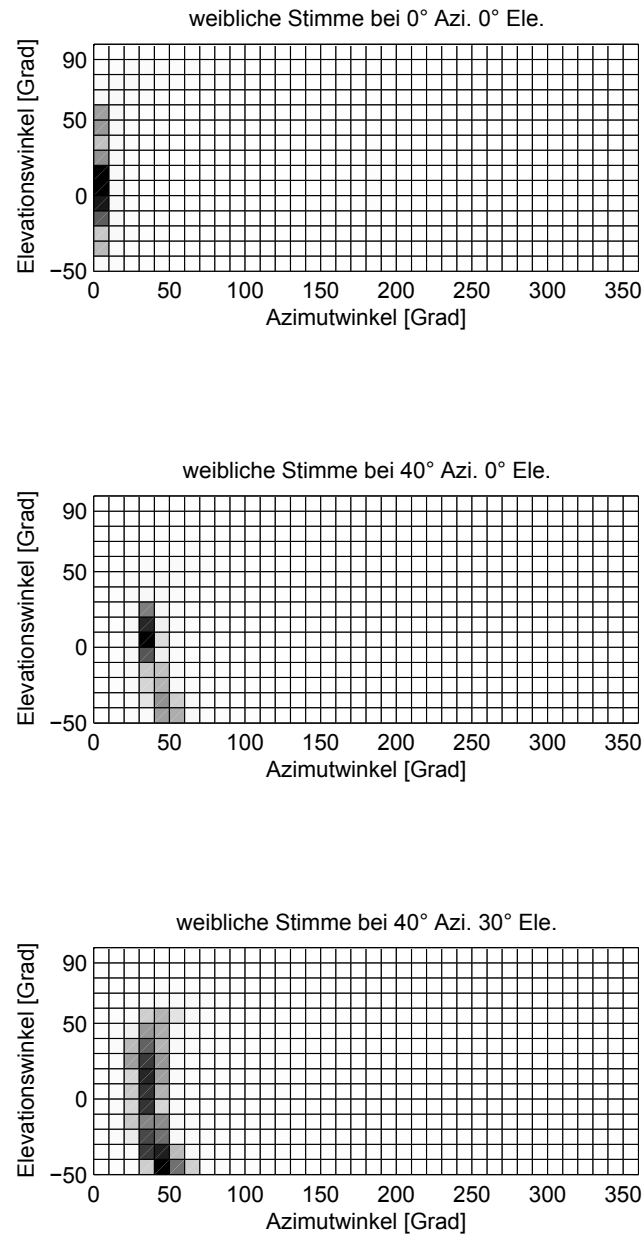


Abbildung 3.8: Weibliche Stimme bei 0° Azimutwinkel 0° Elevationswinkel; 40° Azimutwinkel 0° Elevationswinkel; 40° Azimutwinkel 30° Elevationswinkel

Eine männliche sowie eine weibliche Stimme wird in der Azimutebene richtig lokalisiert. Sobald sie außerhalb der Azimutebene liegen wird die Detektion der Elevation sehr unscharf. Gut sichtbar ist ein Teil des cone-of-confusion der durch die Kopfdrehung nicht ausgelöscht wird, da generell eine Unschärfe in der Lokalisation vorherrscht.

3.1.5 Zwei korrelierte Quellen

In diesem Abschnitt wird eine Phantomschallquelle simuliert, indem korreliertes weißes Rauschen für 2 Quellen verwendet wird. Die 2 Quellen sind wie bei der Stereophonie 60° voneinander entfernt und sollten eine Phantomschallquelle in der Mitte erzeugen.

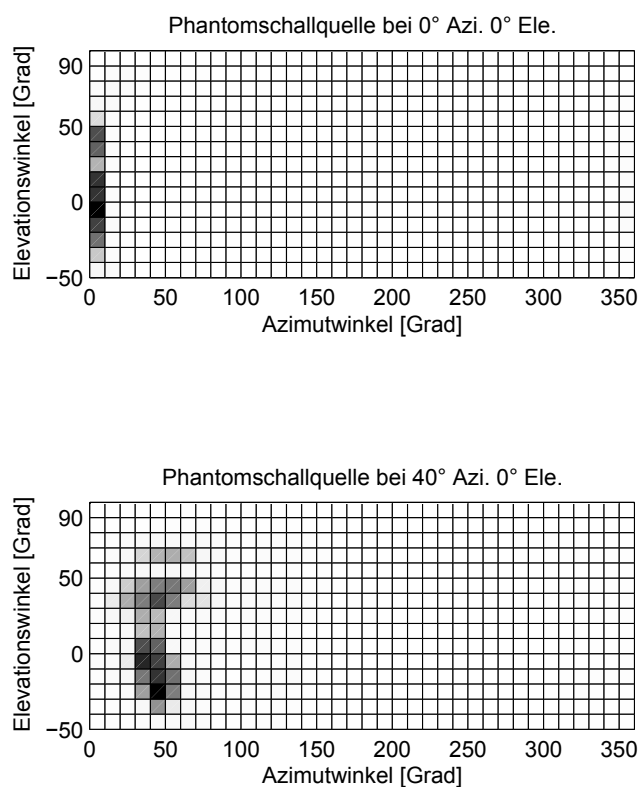


Abbildung 3.9: Phantomschallquellen bei 0° und 40° Azimutwinkel

2 korrelierte Signale werden zu einer Phantomschallquelle zusammengefasst. Die Phantomschallquelle wird mit richtigem Azimutwinkel und unscharfem Elevationswinkel lokalisiert.

Korrelierte Schallquellen bei 60° und 120° werden wie beim Menschen nicht als Phantomschallquelle bei 90° detektiert. 2 Quellen die symmetrisch um 90° positioniert werden, liegen immer exakt auf einem Öffnungswinkel des cone-of-confusion und werden deshalb als 2 Quellen lokalisiert. In Abbildung 3.10 wird eine Lokalisationsmatrix ohne herausgerechneter Mehrdeutigkeit dargestellt.

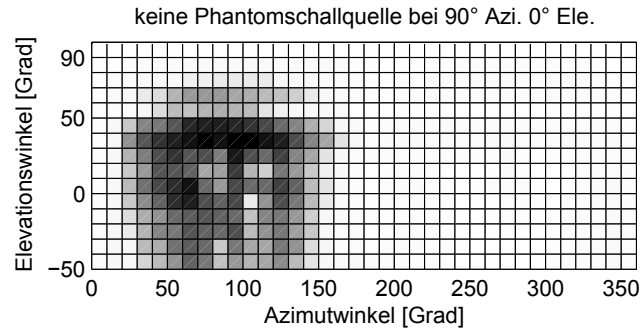


Abbildung 3.10: Korrelierte Quellen bei 60° und 120° Azimutwinkel ergeben keine Phantomschallquelle

3.1.6 Zwei nichtkorrelierte Quellen

Als Eingangssignale werden 2 nichtkorrelierte Rauschquellen oder eine weibliche und männliche Stimme an unterschiedlichen Orten verwendet.

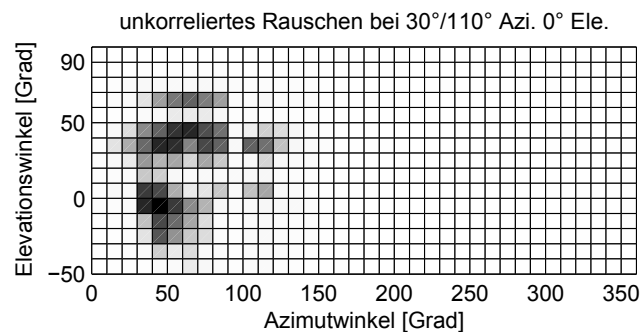
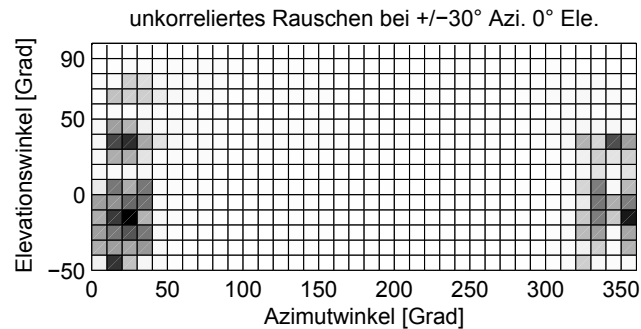


Abbildung 3.11: 2 nichtkorrelierte Rauschquellen bei $\pm 30^\circ$ Azimutwinkel; 30° und 110° Azimutwinkel

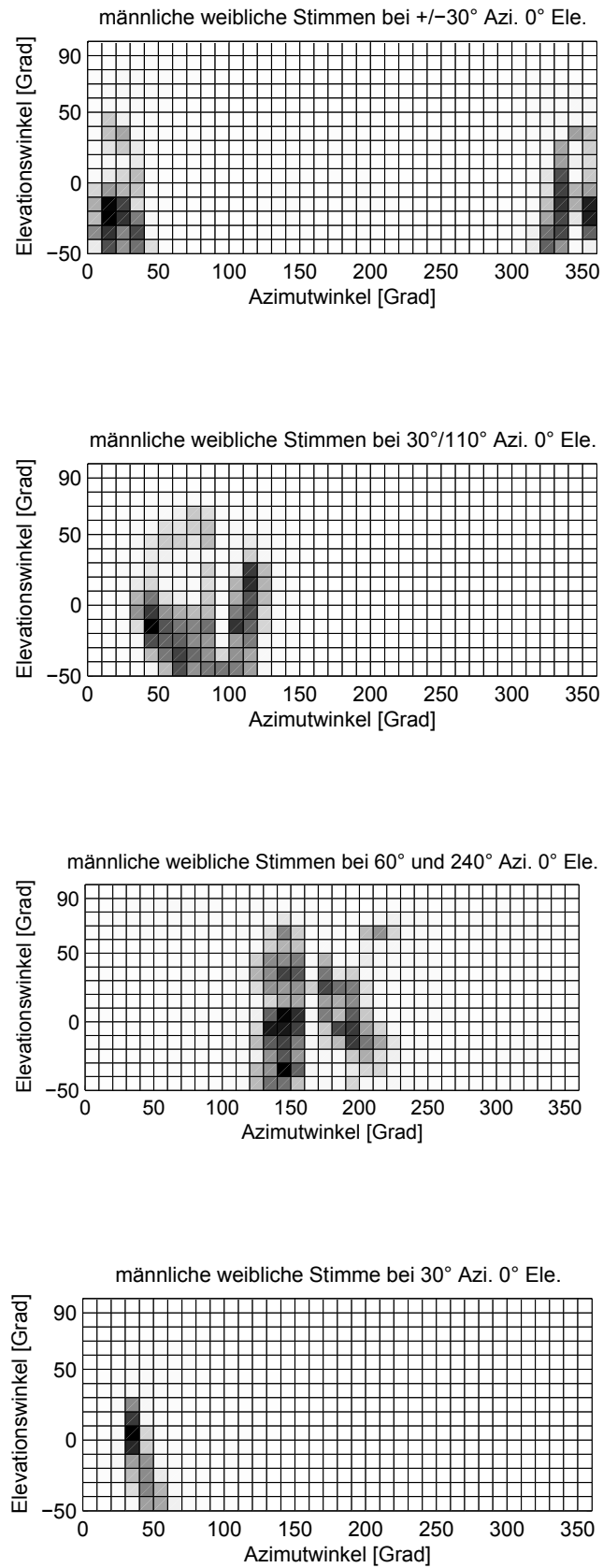


Abbildung 3.12: 2 nichtkorrelierte Stimmen an unterschiedlichen Orten

Das Modell lokalisiert bei den Versuchen mit Stimmen und Rauschen jeweils 2 separate Quellen. Diese werden unscharf im Zielbereich lokalisiert. Die Lokalisationsorte der beiden Quellen sind voneinander abhängig. So verschiebt sich der Ort der jeweiligen Quelle in Richtung der anderen. Männliche und weibliche Stimme an der selben Position, werden scharf als eine Quelle an einem Ort lokalisiert.

3.2 Ergebnisse in einem echten Raum

In diesem Abschnitt werden Messungen in einem echten Raum in Echtzeit und in der Simulation analysiert. Die Messungen werden im Cube des IEM durchgeführt. Es werden Lautsprechersignale mit weißem Rauschen durch einen Kunstkopf aufgezeichnet. Diese Signale werden mit Pd in Echtzeit analysiert und anschließend mit der Matlab Simulation verglichen. In den nachfolgenden Ergebnissen werden die Richtungsmehrdeutigkeiten nicht korrigiert.

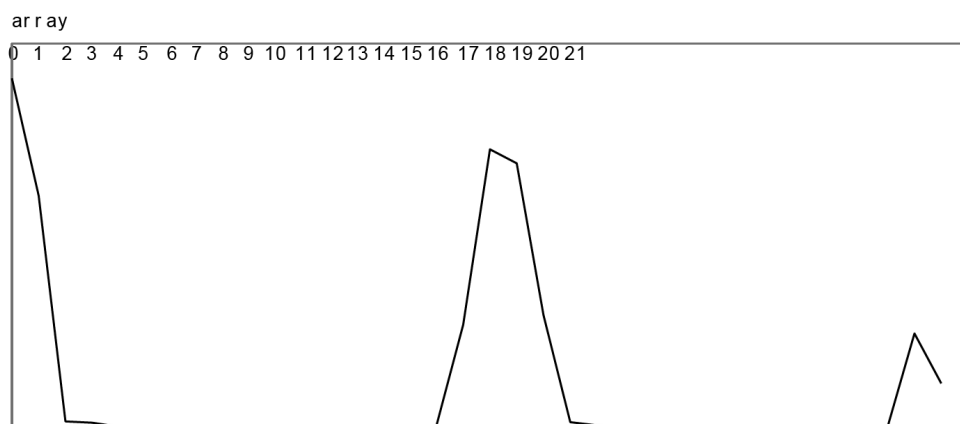


Abbildung 3.13: Echtzeitlokalisierung einer Schallquelle bei 0° Azimut

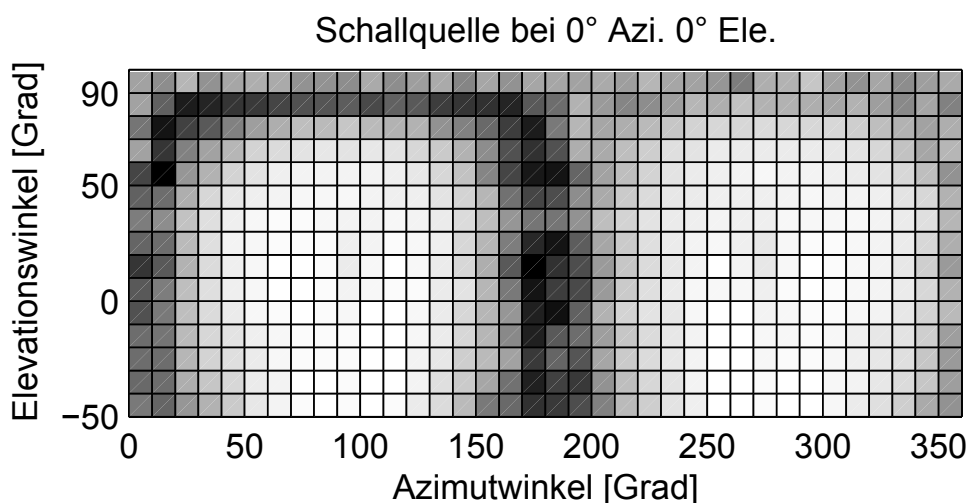


Abbildung 3.14: Eine Schallquelle bei 0° Azimut dargestellt im Offline-Modell

Eine Schallquelle bei 0° Azimutwinkel wird in Echtzeit an der richtigen Position lokalisiert.

Die selben Signale werden aufgezeichnet und in der 3D Offline-Simulation ohne Kopfdrehung verarbeitet. Der cone-of-confusion ist sehr gut ersichtlich, aber viel ungenauer als in der Simulation im trockenen Raum. Der Raumeinfluss macht sich stark bemerkbar, da das Modell nicht sehr stabil auf Reflexionen reagiert. Starke Reflexionen machen eine exakte Lokalisation sehr unwahrscheinlich. Die Verwechslungen im hinteren Bereich sind sehr stark. Die richtige Richtung wird nicht als stärkste Richtung am cone-of-confusion ersichtlich.

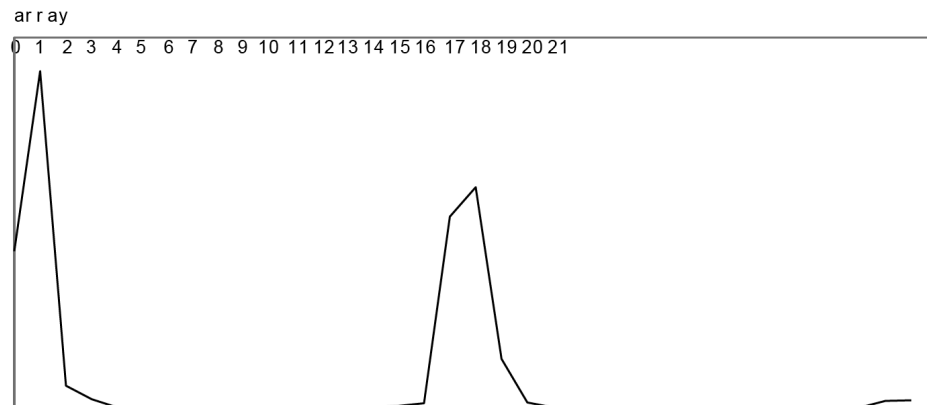


Abbildung 3.15: Echtzeitlokalisierung einer Schallquelle bei 10° Azimut

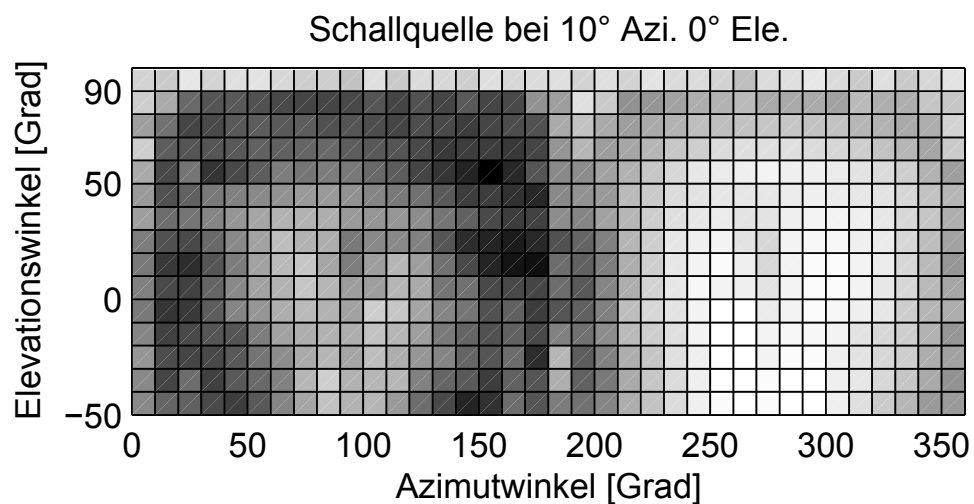


Abbildung 3.16: Eine Schallquelle bei 10° Azimut dargestellt im Offline-Modell

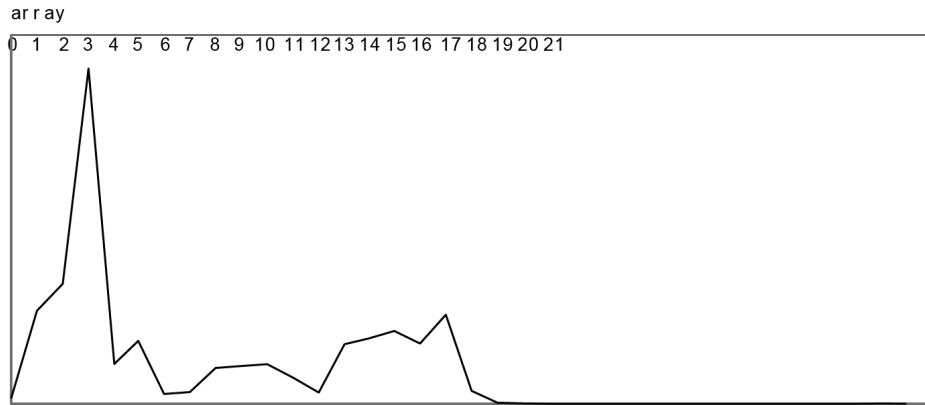


Abbildung 3.17: Echtzeitlokalisierung einer Schallquelle bei 30° Azimut

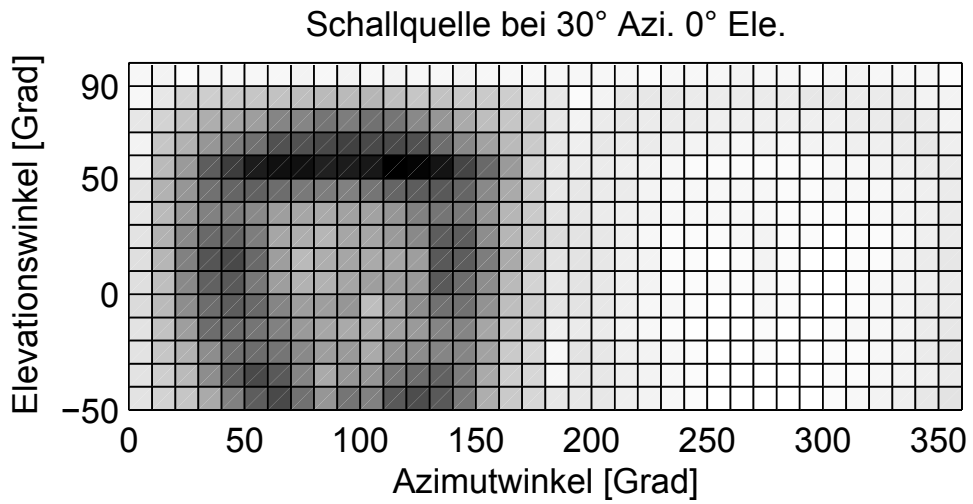


Abbildung 3.18: Eine Schallquelle bei 30° Azimut dargestellt im Offline-Modell

Eine Schallquelle bei 10° und 30° wird auch in Echtzeit richtig lokalisiert. Auch hier sind im Offline-Modell die durch die Reflexionen bedingten Verwechslungen sehr stark.

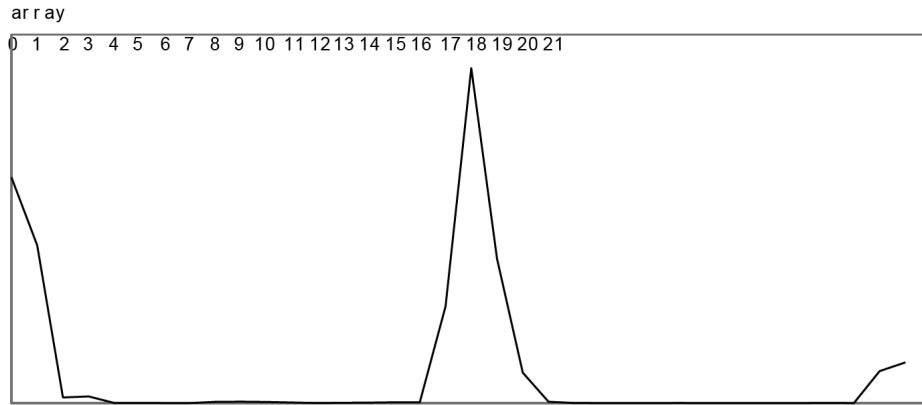


Abbildung 3.19: Echtzeitlokalisierung einer Phantomschallquelle bei 0° Azimut

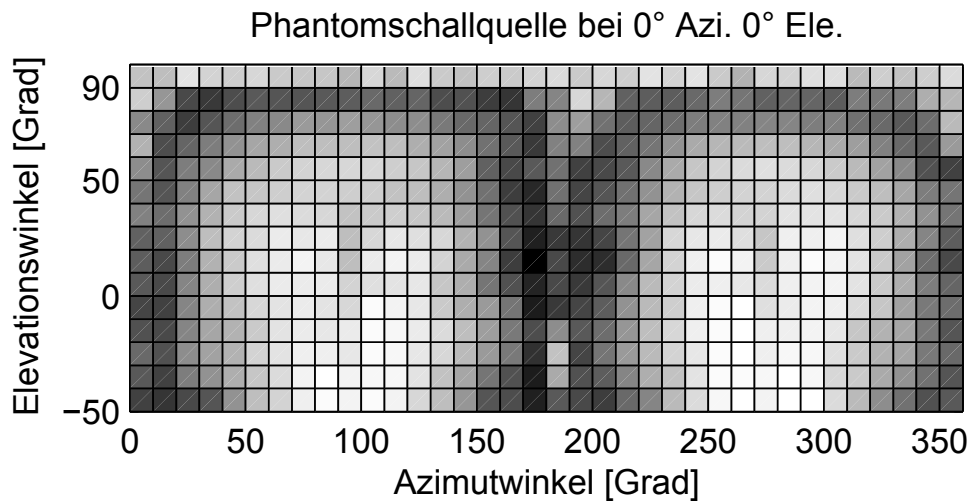


Abbildung 3.20: Eine Phantomschallquelle bei 0° Azimut dargestellt im Offline-Modell

Eine Phantomschallquelle wird in der Mitte bei 0° lokalisiert. Betrachtet man die Lokalisationsmatrix auf einer Kugel (siehe Abb. 3.21), sieht man, dass sich ein breiter cone-of-confusion genau in der Elevationsebene ergibt.

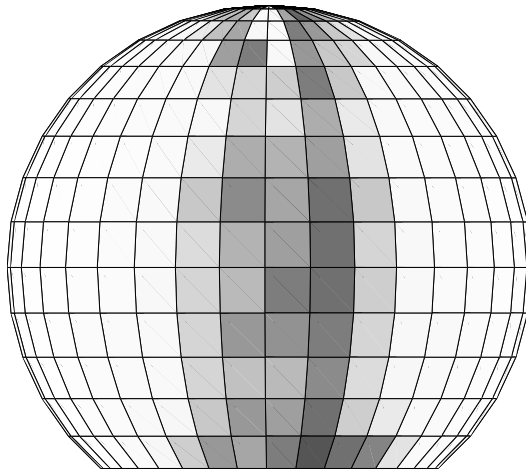


Abbildung 3.21: Phantomschallquelle bei 0° Azimut in frontaler Kugeldarstellung

Kapitel 4

Zusammenfassung

In dieser Diplomarbeit wurde ein Lokalisationshörmodell in einer Simulation und in einer Echtzeitanwendung umgesetzt. Das Breebaart-Modell wurde durch einen zusätzlichen Vergleich zwischen den Pegeldifferenzen erweitert. Dadurch wird ein genauerer Vergleich der zwei Eingangssignale ermöglicht. Das Modell kann Schallquellen mit allen Azimut- und Elevationswinkeln detektieren. Die resultierenden Richtungsmehrdeutigkeiten (cone-of-confusion) werden korrigiert und die richtige Richtung der zu lokalisierenden Quelle wird abgebildet. Dies wird durch eine mehrfache Analyse einer statischen Lokalisationszene erreicht. Dabei wurde eine Szene mit unterschiedlichen Azimutwinkeln des Kopfs in seiner Drehachse analysiert. Durch eine Kombination der Ergebnisse können die Mehrdeutigkeiten fast vollständig korrigiert werden.

Das Modell reagiert unterschiedlich auf Eingangssignale und den Raum in dem sich der Versuchsaufbau befindet. Ohne Raum wird weißes Rauschen in jeder Richtung eindeutig lokalisiert. Da je eine Analyse in jedem Barkband durchgeführt wird, kann auch Schmalbandrauschen lokalisiert werden. Der Azimutwinkel einer Quelle mit Schmalbandrauschen wird eindeutig detektiert. Im Elevationswinkel reagiert das Modell ungenauer. Die Mehrdeutigkeiten werden nicht exakt korrigiert. Das Modell verhält sich für eine menschliche Stimme ähnlich. Zwei korrelierte Quellen werden als eine Phantomschallquelle und zwei nicht korrelierte Quellen als zwei separate Quellen detektiert. Die Ergebnisse werden durch den Einfluss von Reflexionen ungenauer, wenn die Versuche in einem echten Raum durchgeführt werden.

Dem Modell ist es in der Horizontalebene möglich eine Quellenbreite durch die Steigung einer Geraden anzugeben, die in den aus dem Modell resultierenden Lokalisationsvektor gelegt wird. Je geringer die Steigung ist, desto breiter ist die Quelle. Präsentiert man dem Modell Schallquellen zwischen 0° und 180° und vergleicht die resultierenden Steigungen miteinander ist ersichtlich, dass die Schallquellengröße mit der Position der Quelle variiert und bei ca. 60° und 130° am breitesten ist. Auch wird im Hörmodell eine Phantomschallquelle bei 0° breiter als eine reale Quelle bei 0° angegeben. Wird die wahrscheinlichste

Richtung im Lokalisationsvektor auf 1 normiert und die Energie im umgebenden Bereich aufaddiert, deckt sich das Ergebnis mit den Ergebnissen der Lokalisationsschärfe von Blauert. Je mehr Energie in der Umgebung der wahrscheinlichsten Richtung vorhanden ist, desto größer ist die Lokalisationsunschärfe.

In weitere Folge sollte das Model stabiler auf Reflexionen reagieren. Hierbei müsste aber auch auf die Darstellung der Quellenbreite Rücksicht genommen werden, die im direkten Zusammenhang mit den Reflexionen steht. Eine Gewichtung der Analyseergebnisse der einzelnen Barkbänder vor ihrer Summation und ein exakterer Verlauf der Dämpfungsschritte, der an den Pegelabfall rund um den Kopf angeglichen werden könnte, sollte zusätzlich untersucht werden.

Literaturverzeichnis

- [Bla74] J. Blauert, *Räumliches Hören*. Hirzel, 1974, no. Bd. 1.
- [BvdPK01] J. Breebaart, S. van de Par, and A. Kohlrausch, “Binaural processing model based on contralateral inhibition. i. model structure,” *Journal of the Acoustical Society of America*, vol. 110, no. 2, pp. 1074–1088, 2001.
- [BvdPT99] L. R. Bernstein, S. van de Par, and C. Trahiotis, “The normalized correlation: Accounting for nospi thresholds with gaussian and “low-noise” masking noise,” *Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 1342–1343, 1999.
- [FMS11] M. Frank, G. Marentakis, and A. Sontacchi, “A simple technical measure for the perceived source width,” *Deutsche Gesellschaft für Akustik*, 2011.
- [Gel97] S. Gelfand, in *Essentials of audiology*, 1997.
- [GGF89] W. Gulick, G. Gescheider, and R. Frisina, *Hearing: physiological acoustics, neural coding, and psychoacoustics*. Oxford University Press, 1989.
- [Jef48] L. A. Jeffress, “A place theory of sound localization,” *Journal of Comparative and Physiological Psychology*, vol. 41, no. 1, pp. 35–39, 1948.
- [KKD08] J. Kießling, B. Kollmeier, and G. Diller, *Versorgung und Rehabilitation mit Hörgeräten*. Thieme, 2008.
- [Lab06] B. Laback, “Skriptum zur Vorlesung Psychoakustik 2,” 2006.
- [LAH02] L. Lin, E. Ambikairajah, and W. H. Holmes, “perceptual domain based speech and audio coder,” *Proc. of the third international symposium DSPCS 2002*, 2002.
- [Maj10] P. Majdak, “Räumliches Hören 1,” in *Skriptum zur Vorlesung Algorithmen in Akustik und Computermusik*, 2010.
- [Mata] [Online]. Available: <http://www.mathworks.de>

- [Matb] [Online]. Available: <http://mathworld.wolfram.com/SigmoidFunction.html>
- [Moo95] B. Moore, in *Hearing*. Academic Press, 1995.
- [MRT98] P. Mackensen, K. Reichenauer, and G. Theile, "Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören," *20. Tonmeistertagung*, pp. 229–237, 1998.
- [pd] [Online]. Available: <http://puredata.info>
- [Pfl00] M. Pflüger, "Skriptum zur Vorlesung Psychoakustik," 2000.
- [SB95] G. A. Souludre and J. S. Bradley, "The influence of late arriving energy on spatial impression," *Journal of the Acoustical Society of America*, vol. 97, no. 4, pp. 2263–2271, 1995.
- [Tra90] H. Traunmüller, "Analytical expressions for the tonotopic sensory scale," *Journal of the Acoustical Society of America*, vol. 88, no. 1, pp. 97–100, 1990.
- [TuB] [Online]. Available: <http://audio.qu.tu-berlin.de>
- [WB06] D. Wang and G. Brown, *Computational auditory scene analysis: principles, algorithms and applications*. Wiley interscience, 2006.
- [Wup] [Online]. Available: http://www.dasp.uni-wuppertal.de/ars_auditus/ueberfunktion/funktion4.htm
- [ZA02] U. Zölzer and X. Amatriain, *DAFX: digital audio effects*. Wiley, 2002.
- [Zot04] F. Zotter, "Unterdrückung hörbarer Störgeräuschen in Echtzeitsystemen," 2004.