

Institut für Elektronische Musik und Akustik

Toningenieur – Projekt:

**Entwurf und Durchführung eines Hörversuches zur
Ortung und Lokalisation
von Phantomschallquellen bei Elevation.**

Verfasser: Perkmann Michael

Matrikel – Nr.: 9930221

Studienkennzahl: F750

Datum: 26.04.2005

Betreuer: SONTACCHI Alois , Univ. Ass. DI Dr.

INHALTSVERZEICHNIS

Abstract	3
1. Einführung	3
2. Binaural und Transaurale Techniken	5
2.1 Binaural –Technik	6
3. Multichannel Auditory Displays	7
3.1 Vector Based Amplitude Panning (VBAP).....	8
3.1.1 2D – Amplitude Panning	8
3.1.2 Erweiterung auf den dreidimensionalen Raum: 3D - VBAP	10
3.2 Ambisonic Systeme	12
3.2.1 Allgemeines.....	12
3.2.2 2D-Ambisonic	12
3.2.3 2D → 3D Erweiterung von Ambisonic	15
3.3 Wellenfeldsynthese (WFS)	18
3.3.1 Mathematische Beschreibung	19
4. Bereits durchgeführte Hörversuche welche die Wahrnehmung von Schallquellen in der Elevation untersuchen	21
4.1 Hörversuch von John C. Middlebrooks (mit realen Schallquellen).....	21
4.1.1 Grundsätzliches	21
4.1.2 Versuchsaufbau und Durchführung	21
4.1.3 Directional Transfer Function (DTF).....	23
4.1.4 Lokalisation von Breitband – Signalen	23
4.1.5 Lokalisation von Schmalband – Signalen	24
4.1.6 Lokalisations – Modell.....	26
4.2 Hörversuch von Ville Pulkki und M. Karjalainen (mit VBAP).....	28
4.2.1 Methode.....	28
4.2.2 Versuchsanordnung	28
4.2.3 Panning auf der Medianebene	29
4.2.3.1 Testsignal: Rosa Rauschen.....	29
4.2.3.2 Testsignal: Oktav-Band rosa Rauschen.....	31
4.2.4 Panning in drei Dimensionen	33
4.2.4.1 Lautsprecher - Tripel I.....	33
4.2.4.2 Lautsprecher - Tripel II	34
4.2.5 Schluss und Erkenntnisse	36
5. Versuchsdesign und Durchführung	37
5.1 Ziel	37
5.2 Methode.....	37
5.2.1 Wiedergabeanordnung	37
5.2.2 Ablaufsteuerung	39
5.3 Versuchsanordnung.....	39
5.4 Hörversuchergebnisse:	40
5.4.1 Statistische Auswertung	40
5.4.2 Ergebnisse der einzelnen Testpersonen.....	42
5.4.3 Ergebnisse aller Testpersonen	47
6. Ausblick	48
7. Quellenverzeichnis	49

Abstract:

In der Literatur findet sich eine Unzahl von Hörversuchen und Versuchsanordnung zur Lokalisation, genauer gesagt, zur Lateralisation von Schallquellen. Bei diesen Hörversuchen wird überprüft unter welchen Bedingungen und wie exakt eine fiktive Schallquelle (Phantomschallquelle) bei Lautsprecherwiedergabe, die durch Überlagerung von Schallfelder (mindestens zwei) erzeugt wird, lokalisiert werden kann und welche Eigenschaften dieser Wahrnehmung zugeordnet werden können. Da im IEM Cube neben seitlichen Auslenken auch vertikale Auslenkungen von Quellen realisierbar sind, ist es von Interesse wie Phantomschallquellen mit Elevation lokalisiert werden können. Im Rahmen dieses Projektes sollte zu Beginn eine einführende und Überblick gebende Literatursuche über „Lokalisation von Phantomschallquellen mit Elevation“ erfolgen. Des Weiteren sollte ein Hörversuch entworfen und durchgeführt werden.

1. Einführung:

Aufgrund der Literaturrecherche werden zuerst die wichtigsten aller bisher bekannten Methoden zur Abbildung von Phantomschallquellen in der Medianebene vorgestellt und eventuell zusätzlich bereits durchgeführte Hörversuche zu den jeweils angesprochenen Techniken vorgestellt. Dies soll es in der späteren Versuchsdurchführung ermöglichen Vergleiche zwischen den erhaltenen Ergebnissen zu ziehen.

Jede bekannte Methode der akustischen Wiedergabe welche eine „Positionierung“ auf der Medianebene ermöglicht, ist eine 3D-Anwendung zur Darstellung eines akustischen Ereignisses im Raum. Man spricht hierbei meistens von einem „3D Auditory Display“. Durch die verschiedenen Wiedergabemethoden werden je nach Typ verschiedene Reize zur Lokalisierung der Phantomschallquellenposition im Raum, die sog. „localization cues“, erzeugt. Diese *localization cues* sind für die Abbildung in der horizontalen Ebene ausreichend bekannt: ITD (interaural time difference) bzw. ILD (interaural level difference). Anhand dieser *localization cues* ist es möglich die Lokalisation in der Horizontalen und alle dabei bestehenden Phänomene bei der Ortung von Phantomschallquellen zu erklären.

Bei Anwendungen welche auch die Bewegung bzw. die Positionierung in der Vertikalen enthalten reichen die herkömmliche Erklärungsversuche anhand der beiden Hauptreize jedoch nicht aus. Es besteht der begründete Verdacht, dass es zusätzlich noch andere Reize bzw. Informationen gibt welche die Lokalisation bei Elevation ermöglichen.

Ein einfaches Beispiel dafür ist z.B. der Versuch der Panoramisierung zwischen zwei Schallquellen welche sich auf der Medianebene befinden. In diesem Fall kann die ITD bzw. die ILD theoretisch keine Information darüber liefern wo sich die Phantomschallquelle befindet. Für jeden Punkt auf der Medianebene können immer nur dieselben ITD und ILD – Werte erhalten werden (angenommen: idealer symmetrischer Kopf der Versuchsperson). Der Bereich auf welchem sich mehrere Positionen bei gleich- bleibenden Zeit- bzw. Pegeldifferenzen einstellen können wird als „*cone of confusion*“ bezeichnet. Der einzige Reiz

welcher die Lokalisation in diesem Bereich ermöglicht sind spektrale Signalanteile. Dieser Einfluss der spektralen Wellenform wurde bereits 1974 von Blauert angesprochen. Diese spektralen Einflüsse werden später noch genauer beschrieben und deren Auswirkungen besprochen.

Zur Erzeugung des 3D - Sounds existieren verschiedene Verfahren die sich grob in zwei Kategorien einteilen lassen. Zum einen diejenigen die versuchen das originale Schallfeld in einem bestimmten Hörbereich zu reproduzieren (VBAP, Ambisonic, Wellenfeldsynthese):

Multichannel Auditory Displays.

Zum anderen Verfahren welche die Eingangsgrößen der originalen Schallwelle an den Trommelfellen zu realisieren versuchen: **Binaurale und Transaurale Verfahren.**

2. Binaural und Transaurale Techniken

Sind Techniken welche das Schallfeld direkt bei beiden Ohren des Hörers erzeugen um einen bestimmten Richtungseindruck zu kreieren der im optimalen Fall dem einer realen Schallquelle an genau dieser Position entspricht. Man kann hierbei zwischen zwei grundsätzlichen Wiedergabetechniken unterscheiden:

- Wiedergabe über Kopfhörer (Binaural-Technik)
- Wiedergabe über Lautsprecher ausgehend von binauralen Signalen (Transaural-Technik)

Diese genannten Techniken werden nur der Vollständigkeit halber kurz erwähnt, da sich dieses Projekt vorrangig mit der Lautsprecherwiedergabe (→ siehe Multichannel Auditory Displays) im speziellen bei Elevation beschäftigen soll.

Grundlage für binaurale bzw. transaurale Wiedergabetechniken bilden die sog. HRTFs. HRTF steht hierbei für den Ausdruck „Head Related Transfer Funktion“. Diese Transferfunktionen entsprechen dem Einfluss des menschlichen Körpers im Schallfeld. Die HRTFs sind definiert als das Verhältnis zwischen Schalldruck am Trommelfell und dem Punkt im Zentrum des Kopfes wenn der Hörer sich nicht dort befindet. Der Oberkörper (Torso), das Außenohr (Pinna) und der Schädel des Hörers haben im Schallfeld eine Art Filterfunktion zur Folge welche das sonst ungestörte Schallfeld im spektralen Sinne beeinflusst (siehe *Abb.1*). Die HRTF – Kurven enthalten dabei alle benötigten Informationen bzgl. der Schallausbreitung von einer bestimmten Raumrichtung aus zum Trommelfell des Hörers.

Ein Beispiel für die Beeinflussung des Testsignals durch die angesprochenen Körperteile ist z.B. die Zeitverzögerung von 0-300µsek (Batteau, 1968) die durch die Falten der Ohrmuschel am Trommelfell verursacht werden. Hier wird das Originalsignal mit dem zeitverzögerten überlagert was charakteristische Wellenformen in den Spektren zur Folge hat. Wie ein „akustischer Fingerabdruck“ beeinflussen die HRTFs je nach Einfallrichtung das Schallsignal.

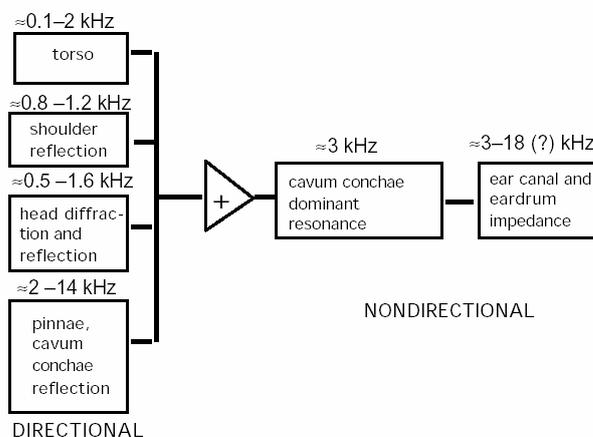


Abb.1: Beschreibung der richtungsabhängigen und richtungsunabhängigen Komponenten der HRTFs (nach Genuit, 1984; Gierlich, 1992) Die jeweils beeinflussten Frequenzbereiche je Körperteil können erkannt werden. [5]

Man kann HRTFs modellieren oder simulieren: spezielle Messungen in einem schalltoten Raum (um das Messergebnis möglichst wenig durch reflektierte Schallenergie zu beeinflussen) erhalten indem z.B. MLS – Signale (Maximum Length Sequences) aus möglichst allen Raumrichtungen mittels eines frei positionierbaren realen Quelle wiedergibt und das übertragene Signal möglichst nahe am Trommelfell der Testperson oder des Messobjekts (z.B. auch „Dummy heads“ → KEMAR – System) mit eine Mikrophon aufnimmt (Abb.2).

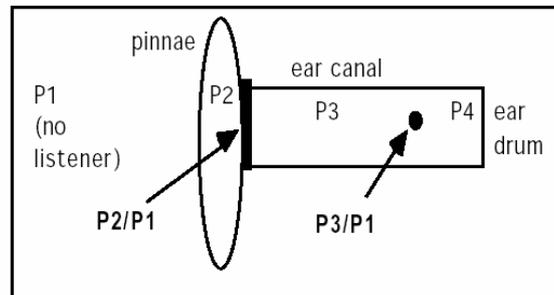


Abb.2: Messpunkte zur Aufnahme der HRTF (nach Møller, 1992) [5]

Beziehungen zwischen den verschiedenen Messpunkten für die HRTFs:

Bei Messungen an Punkt P3/P4 enthält die HRTF richtungsabhängige und richtungsunabhängige Einflüsse (→ Abb.1). An Punkt P1/P2 am Gehörgangseingang werden nur richtungsabhängige Einflüsse gemessen.

2.1 Binaural –Technik

Da ein schalltoter Raum theoretisch keinen Einfluss auf das übertragene Testsignal hat kann der Unterschied zwischen der gesendeten und empfangenen Sequenz auf die Einflussnahme der körperlichen Eigenschaften des Testobjekts zurückgeführt werden. Man erhält somit die Impulsantwort des Testobjekts oder der Testperson welche es gewährleistet durch Faltung eines beliebigen Signals einen bestimmten Raumrichtungseindruck und damit eine Spezialisierung zu erzeugen.

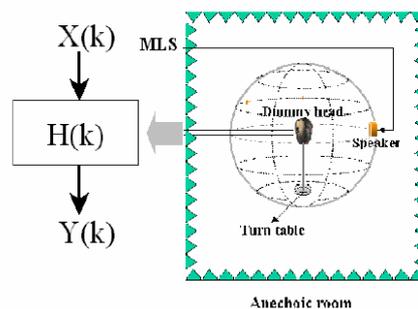


Abb.3: Messung von HRTF – Kurven [1]

Anders ausgedrückt werden dabei HRTFs als Filter verwendet um den entsprechenden 3D – Effekt zu erzielen. Wie in Abb.3 ersichtlich wird dabei einfach das Signalspektrum $X(k)$ mit jenem des HRTF $H(k)$ multipliziert.

Genauerer bzgl. Studien HRTF – Filterdesign: [2][3][1]

Es können mittels Binauraltechnik jedoch nicht nur statische Raumpositionen erzeugt sondern auch Bewegungen simuliert werden. Hierbei wird das zu spezialisierende Signal entsprechend der Position mit der jeweils zutreffenden HRTF gefiltert.

Um die sehr großen Datenmengen die gemessen werden müssen gibt es auch Ansätze HRTFs zu interpolieren [4].

2.2 Transaural – Technik

Ausgehend von der Binauralwiedergabe wurde der Versuch gestartet die HRTFs auch für die Lautsprecherwiedergabe auszunutzen. Binaurale Systeme, bei denen die Wiedergabe der Schallwellen über Lautsprecher erfolgt, werden als transaurale Systeme bezeichnet. In diesem Fall muss ein Übersprechen der Lautsprecher berücksichtigt werden. Im einfachsten Fall, bei welchem zwei Lautsprecher zum Einsatz kommen, wird die Übertragungsweginvertierung durchgeführt, indem die 2×2 Matrix, welche die Übertragungswege zwischen den Lautsprechern und den Ohren des Zuhörers beschreibt, invertiert wird. Die Übertragungsweginvertierung beinhaltet die Auslöschung des ungewollten Übersprechens (vom linken / rechten Lautsprecher zum rechten / linken Ohr) und wird daher auch als *crosstalk cancellation* bezeichnet (Abb.4).

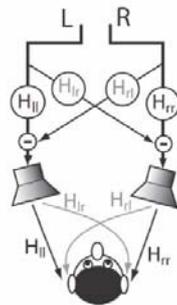


Abb.4: Transaurale Wiedergabe: das binaurale Signal wird über Lautsprecher wiedergegeben. Es muß der *crosstalk* zwischen den Lautsprechern und den Ohren des Hörers eliminiert werden (*crosstalk cancellation*). (Der Einfluß der Signale des rechten LS auf das linke Ohr und jene des linken LS auf das rechte Ohr muß kompensiert werden.)

Problematisch ist, daß der räumliche Bereich in dem *crosstalk cancellation* effektiv ausgeführt werden kann, relativ beschränkt ist d.h. dass der sog. „sweet spot“ sehr beschränkt ist.

Für Hörversuche die bzgl. Binauraltechnik durchgeführt wurden wird auf die weiterführende Quellen [6] verwiesen.

3. Multichannel Auditory Displays

Die zweiten Verfahren, welche neben der binauralen und transauralen Technik existieren, versuchen ein Schallfeld mit Hilfe von räumlich verteilten Lautsprecher – Anordnungen zu erzeugen. Hier werden Lautsprecher um den Hörer herum platziert um gerichtete

Schallwellen zu erzeugen. Bei diesen Verfahren können die *localization cues* ohne die aufwendige HRTF – Filterung erzeugt werden.

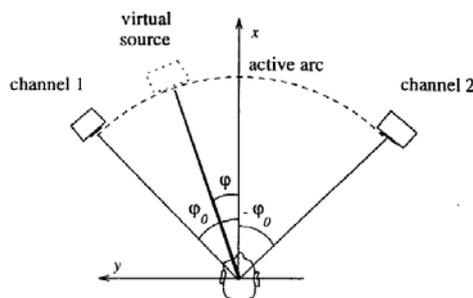
Es werden nun in der Folge die verschiedenen Techniken zur Erzeugung von dreidimensionalen Schallfeldern vorgestellt und qualitativ erklärt. Anschließend werden, wie bereits erwähnt, eventuell durchgeführte Hörversuche zu den einzelnen Wiedergabetechniken besprochen.

3.1 Vector Based Amplitude Panning (VBAP)

Eine einfache und effiziente Technik einer Multi-Lautsprecher Wiedergabetechnik ist das sog. VBAP. Wie bereits aus der Stereo-Wiedergabetechnik (*Abb.5*) bekannt handelt es sich hierbei um die weiterentwickelte Variante des Intensitäts-Pannings auf den 3D – Raum (natürlich kann dieses Verfahren auch für 2D Anwendungen ausgenutzt werden). Es gibt bzgl. der Lautsprecherzahl bei VBAP keine Beschränkung bei der Positionierung im Raum um den Hörer. Die verwendeten Lautsprecher sollten sich jedoch auf einer imaginären Kugeloberfläche befinden und auf die zentrale Hörerposition (Kugelmittelpunkt) gerichtet sein. Eventuelle Entfernung Unterschiede können jedoch mithilfe von Gains und Delays kompensiert werden. Zudem sollte der Abhörraum so wenig wie möglich Einfluss auf das synthetisierte Schallfeld haben (z.B. Hall). Zu jedem Zeitpunkt der Wiedergabe wird die minimal notwendige Zahl an Lautsprechern angesprochen. Das sind je nach Anwendung im zweidimensionalen Fall zwei und im dreidimensionalen Fall drei Lautsprecher. Die geometrischen Verhältnisse des Schallfeldes werden anhand von Vektoren dargestellt, was zu effizient berechenbaren linearen Gleichungssystemen für die Ermittlung der Signalamplituden führt.

3.1.1 2D – Amplitude Panning

Die Position der Phantomschallquelle (*virtual source*) ist von den Amplitudenbeziehungen zwischen den beiden Kanälen abhängig. Mit VBAP lassen sich Gain-Faktoren für Lautsprecher-Paare und später für die Lautsprecher -Tripel berechnen. Dabei wird vom „tangent law“ (1) ausgehend die entsprechende Darstellung mittels Vektoren vorgenommen.



$$\frac{\tan \varphi}{\tan \varphi_0} = \frac{g_1 - g_2}{g_1 + g_2} \quad (1)$$

Abb.5: Zweikanal – Stereophone Aufstellung und *tangent law* (1) (φ entspricht dabei dem Auslenkwinkel welcher durch das Amplitudenpanning erreicht wird. $g_1, g_2 \dots$ Gains der Kanäle 1 und 2) [7]

Bei der zweidimensionalen Amplitudenpanning – Methode wird die zweikanal-stereophone Aufstellung (Abb.5) anhand der Vektorendarstellung (Abb.6) beschrieben.

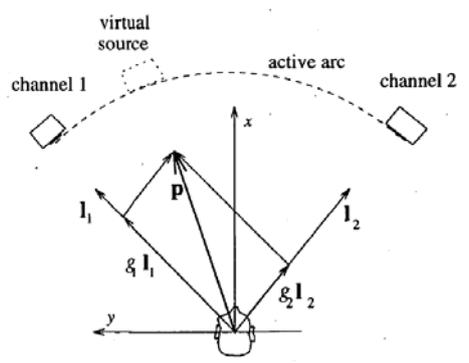


Abb.6: Stereophone Konfiguration formuliert mit Vektoren.[7]

Die Basis ist definiert durch die Einheitsvektoren $l_1=[l_{11} \ l_{12}]^T$ und $l_2=[l_{21} \ l_{22}]^T$, die in Richtung des jeweiligen Lautsprechers zeigen. Der Einheitsvektor, welcher direkt zur virtuellen Quelle weist, kann als Linearkombination der durch g_1 bzw. g_2 gewichteten Lautsprechervektoren aufgefasst werden (2).

$$p = g_1 l_1 + g_2 l_2 \quad (2)$$

In Matrixform, mit $g=[g_1 \ g_2]$ und $L_{12}=[l_1 \ l_2]^T$:

$$p^T = g L_{12} \quad (3)$$

Diese Gleichung kann für den Fall, dass $\varphi_0 \neq 0^\circ$ und $\varphi_0 \neq 90^\circ$, also dann wenn die Matrix L_{12}^{-1} existiert gelöst werden:

$$g = p^T L_{12}^{-1} = [p_1 \ p_2] \cdot \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}^{-1} \quad (4)$$

Die Ansteuerung der Lautsprecher funktioniert nach dem Prinzip des „Intensity Panning's“. Dieses beruht auf der Idee, die Summe der Wirkintensitäten am Punkt des Hörers für jede Position der Phantomschallquelle konstant zu halten (\rightarrow constant energy: $g_1^2 + g_2^2 = C = \text{konst.}$). Nun müssen die Gain-Faktoren noch normalisiert werden um der Forderung der constant energy zu genügen:

$$g^{\text{scaled}} = \frac{\sqrt{Cg}}{\sqrt{g_1^2 + g_2^2}} \quad (5)$$

In 2D-Wiedergabesystemen mit mehr als zwei Lautsprechern erfolgt nach der oben beschriebenen Vorgangsweise paarweises Panning (pairwise panning).

3.1.2 Erweiterung auf den dreidimensionalen Raum: 3D - VBAP

Die typische 2-Kanal-Stereoanordnung wird um einen dritten Lautsprecher erweitert. Bei einem 3D VBAP System soll die virtuelle Schallquelle beliebig auf einer Kugeloberfläche positionierbar sein. Die vektorielle Darstellung des Amplituden-Pannings wird dabei ins Dreidimensionale erweitert (pherifones Amplitude Panning). Für ein Kugelsegment werden drei Lautsprecher verwendet (Abb.7).

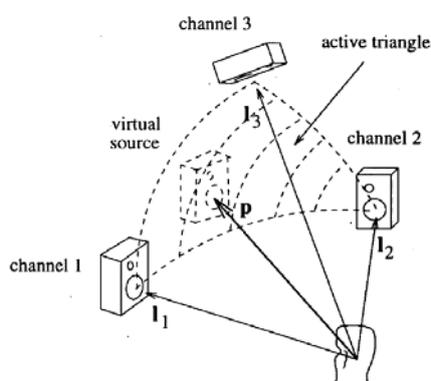


Abb.7: Beispiel für eine 3D-VBAP – Konfiguration. Ein dreieckiger Kugeloberflächenausschnitt dient als Panninggebiet für die Phantomschallquelle (aktive triangle).[7]

Wie bei der 2D-Anwendung zeigen die beteiligten Lautsprechervektoren $l_1=[l_{11} \ l_{12} \ l_{13}]^T$, $l_2=[l_{21} \ l_{22} \ l_{23}]^T$ und $l_3=[l_{31} \ l_{32} \ l_{33}]^T$ von Mittelpunkt der Kugelsphäre in Richtung des jeweiligen Lautsprechers. Die Position der Phantomschallquelle ist dann definiert als dreidimensionaler Einheitsvektor $p=[p_1 \ p_2 \ p_3]^T$ der durch die gewichteten Anteile (g_1, g_2, g_3) der einzelnen Lautsprecher im Tripel erzeugt wird. In vektorieller Form kann das durch Formeln (6) ausgedrückt werden.

$$p = g_1 l_1 + g_2 l_2 + g_3 l_3 \quad (6)$$

Durch Verwendung der Positionsmatrix der Lautsprecher L_{123} ergibt sich:

$$p^T = g L_{123} \quad (7)$$

L_{123} ... Lautsprecherkoordinaten – Matrix
 g ... Gain – Matrix ($g=[g_1 \ g_2 \ g_3]$)

Wenn die Lautsprecherkoordinatenmatrix existiert kann die Gleichung gelöst werden und durch Einsetzen aller Beziehung in das Gleichungssystem können die Faktoren für die Gewichtung der Pegel erhalten werden.

$$g = p^T L_{123}^{-1} = [p_1 \quad p_2 \quad p_3] \cdot \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \quad (8)$$

Um der Beziehung für *constant energie* ($g_1^2 + g_2^2 + g_3^2 = C = konst.$) zu genügen müssen die erhaltenen Faktoren durch (9) normiert werden.

$$g^{scaled} = \frac{\sqrt{Cg}}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \quad (9)$$

In 3D-Systemen mit mehr als drei Lautsprechern (*Abb.8*) erfolgt tripelweises Panning (*tripelwise panning*). Zu beachten ist die Tatsache, daß falls die virtuelle Schallquelle mit einer der Lautsprecherpositionen übereinstimmt, nur dieser zur Wiedergabe benutzt wird (\rightarrow höchstmögliche Lokalisationsschärfe bei VBAP). Die Lokalisationsschärfe der virtuellen Schallquelle ist bei den VBAP Systemen in der näheren Umgebung der Lautsprecher gut, während sie zwischen Lautsprechern schlechter ist. Durch Erhöhen der Lautsprecheranzahl kann die Dimension der einzelnen Kugelteiloberflächen und damit die empfundene Lokalisation verbessert werden. VBAP – Systeme sind dennoch in der Lage virtuelle Schallquellen an beliebigen Positionen zurealisieren. Allerdings können nicht alle Positionen mit konstanter Qualität wiedergeben werden.

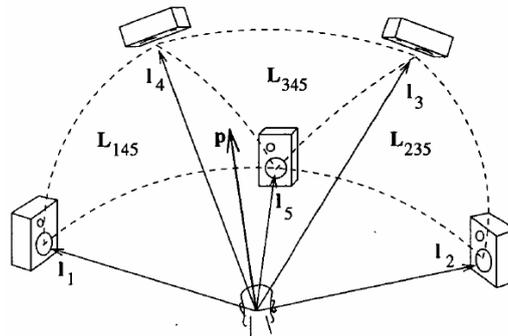


Abb.8: 3D-VBAP mit fünf Lautsprechern. Die Lautsprechervektoren weisen dabei auf den jeweiligen Lautsprecher. Die Kugelteiloberflächen (Panning - Regionen) sind durch die gestrichelten Linien dargestellt. [7]

Für eine gegebene Lautsprecheranordnung kann VBAP angewandt werden. Dazu muss das bestehende Lautsprechersetup in Tripel unterteilt werden [8]. Dazu sind folgende Schritte nötig (*Abb.9*)

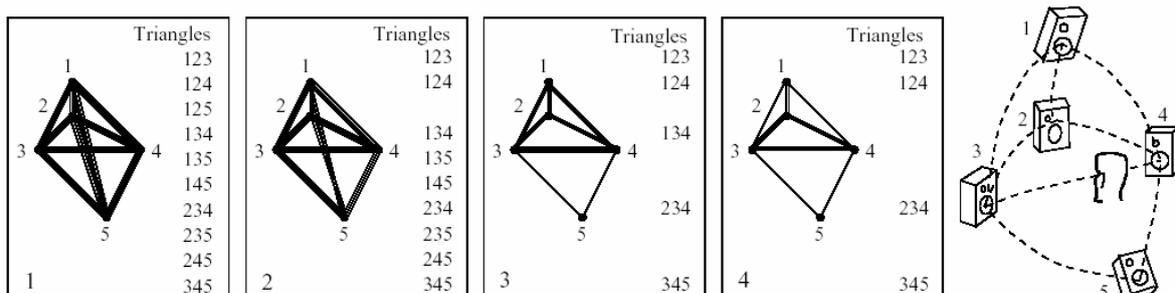


Abb.9: Prozess zur Bestimmung der bestmöglichen Lautsprechertripel. Die einzelnen Schritte der Vereinfachung bis zu Endprodukt können von links nach rechts nachvollzogen werden. [8]

1. Dreiecke aller Lautsprecherkombinationen werden gebildet.
2. Dreiecke welche im Verhältnis zu ihrem Umfang einen kleinen Flächeninhalt besitzen werden eliminiert.
3. Alle Überkreuzungen der beteiligten Dreiecksseiten werden gesucht und jeweils die längere der beiden Seiten eliminiert.
4. Alle übrig gebliebenen Dreiecke welche einen Lautsprecher enthalten werden entfernt. Zum Schluss werden alle Lautsprecher für jedes Dreieck getestet.

3.2 Ambisonic Systeme

3.2.1 Allgemeines

Ambisonic ist ein mehrkanaliges Aufnahme- und Wiedergabesystem. Mit dieser Technik ist es unter anderem möglich eine Phantomschallquelle in der Medianebene zu erzeugen [9] und somit ein Schallfeld dreidimensional über Lautsprecher zu reproduzieren. Bei der Aufnahme wird das Schallfeld in einem Punkt mit Mikrofonen unterschiedlicher Richtcharakteristik und Anordnung (z.B. Tetraeder, Dodekaeder, usw.) aufgenommen. Aus mathematischer Sicht wird bei der Aufnahme das im Raum existierende Schalldruckfeld durch die beteiligten Mikrophone in die sphärischen Harmonischen (*spherical harmonic* [9]) zerlegt. Bei der Wiedergabe werden die Lautsprechersignale in Abhängigkeit von den Lautsprecherpositionen entsprechend aus den zuvor gewonnenen aufgenommenen Signalen erzeugt bzw. zusammengesetzt.

Ambisonic verwendet ein um 90° gedrehtes Koordinatensystem (kartesisch). Dabei weist die x-Achse in die herkömmliche y-Richtung, entsprechend verschiebt sich die y-Richtung auf die herkömmliche negative x-Richtung. Wenn Ambisonic für einen 3D-Raum verwendet wird muss zusätzlich noch eine z-Achse eingeführt werden welche nach Oben gerichtet ist. Im Folgenden werden anhand des zweidimensionalen Falls die Grundgleichungen für das Aufnahme- und Wiedergabesystem betrachtet. Dabei müssen jedoch zwei Randbedingungen erfüllt sein, welche die Grundlage der Ambisonic-technik bilden:

- I. Das aufgenommene Schallfeld (Referenzwelle) muss eine ebene Welle sein (erfüllt für den Fall dass der Abstand zwischen Quelle zum Aufnahmepunkt groß genug ist).
- II. Die reproduzierte Schallwelle des Systems muss ebenfalls eine ebene Welle sein. (d.h. der Abstand zwischen Wiedergabe – System und Abhörpunkt muss wiederum genügend groß sein)

3.2.2 2D-Ambisonic

Ambisonic möchte ich hier ausgehend vom 2D-Fall erklären und auf den 3D-Fall erweitern. Im Folgenden Beispiel trifft eine ebene Schallwelle mit dem Winkel ψ bzgl. der x-Achse ein. Hierbei befindet sich der Abhörpunkt (nicht ideal) im Abstand vom Radius r und dem Winkel ϕ bzgl. der x-Achse.

Die ankommende Referenzwelle kann mathematisch durch Gleichung (10) beschrieben werden.

Referenzwelle:

$$S_{REFERENZ}(r, \phi) = P_{\psi} \cdot e^{jk\vec{r}} = P_{\psi} \cdot e^{jkr \cos(\phi - \psi)} \quad (10)$$

mit $k=2\pi/\lambda$...Wellenzahl; P_{ψ} ...Schalldruck d. ebenen Welle;

Unter der Voraussetzung der Annahme II. (siehe oben) kann das Schallfeld des n-ten Ambisonic –Lautsprechers (ϕ_n ... zugehöriger Winkel) ebenfalls wie folgt dargestellt werden:

$$S_n(r, \phi) = P_{\psi} \cdot e^{jk\vec{r}} = P_{\psi} \cdot e^{jkr \cos(\phi - \phi_n)} \quad (11)$$

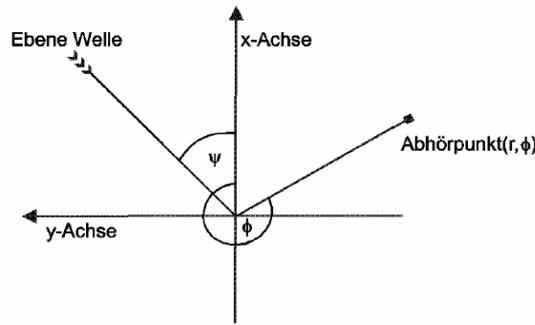


Abb.10: Achsen und Winkel bei der Aufnahme mit Ambisonic

Nach dem Modell von Ambisonic wird nun der Vergleich zwischen Referenzwelle (10) mit der Welle, die das Ambisonic-System letztendlich produzieren soll, durchgeführt. Diese resultierende ebene Welle aller Ambisonic-Lautsprecher kann durch die Summe über die einzelnen Lautsprecherwellen (vgl. (11)) formuliert werden (12).

$$S_{AMBISONIC}(r, \phi) = \sum_{n=1}^N S_n(r, \phi) = \sum_{n=1}^N P_n \cdot e^{jkr \cos(\phi - \phi_n)} \quad (12)$$

Der Vergleich durch die Äquivalenzbedingung (13):

$$S_{REFERENZ}(r, \phi) \cong S_{AMBISONIC}(r, \phi) \quad (13)$$

Die Referenzwelle kann anschließend, bei einer 2D-Anwendung (was hier gegeben ist), in eine Bessel-Fourierreihe entwickelt werden (14). Die Anzahl der m verwendeten Reihenelemente stellt die Ordnung des Systems dar. Hierbei stellt m auch den Grad des Ambisonicsystems dar.

$$S_{REFERENZ} = P_{\psi} J_0(kr) + 2P_{\psi} \sum_{m=1}^{\infty} i^m \cdot J_m(kr) \cos[m(\phi - \psi)] \quad (14)$$

$$S_{REFERENZ} = P_{\psi} \left\{ J_0(kr) + 2 \sum_{m=1}^{\infty} i^m \cdot J_m(kr) [\cos(m\phi) \cos(m\psi) + \sin(m\phi) \sin(m\psi)] \right\}$$

Denselben Schritt kann man für die Ambisonic-Systemwelle durchführen. Man erhält somit zwei Funktionen welche durch Koeffizientenvergleich (15) durch einsetzen in (13) die Forderung nach Äquivalenz erfüllen.

$$P_\psi = \sum_{n=1}^N P_n \quad (15)$$

$$P_\psi \cos(m\psi) = \sum_{n=1}^N P_n \cos(m\phi_n)$$

$$P_\psi \sin(m\psi) = \sum_{n=1}^N P_n \sin(m\phi_n)$$

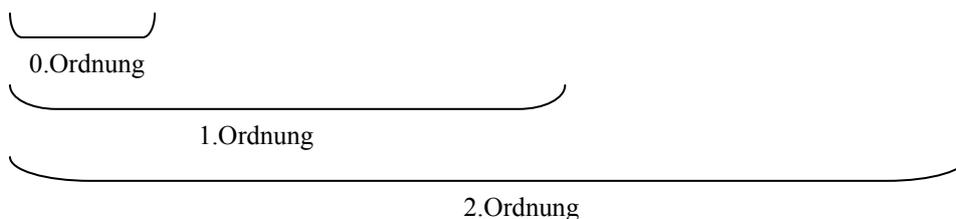
Theoretisch wäre es möglich m gegen Unendlich laufen zu lassen wodurch sich das identische Schallfeld zur Referenzwelle einstellen würde. Praktisch sind diesem Vorhaben jedoch Grenzen durch die begrenzte Rechenleistung und Lautsprecher- bzw. Kanalzahl gesetzt. Die Anzahl der benötigten Übertragungskanäle für ein zweidimensionales System kann durch die Formel $k=2m+1$ (16) berechnet werden. Im Dreidimensionalen Fall ergibt sich: $k=(m+1)^2$ (17).

Für ein System der 0ten Ordnung ergibt sich laut (16) ein Kanal. Beziehung (15) stimmt in dem Fall überein, d.h. Referenzwelle und Synthetisierte Welle stimmen nur bzgl. Amplitude (Schalldruck) überein. Bei der Kodierung (Aufnahme) der Signale hat man hierbei eine kugelförmige Richtcharakteristik. Diese Kodiervorschriften ergeben sich aus den Formeln (15). Mit ihnen lassen sich die beteiligten Signale durch spezielle Mikrophonarrays aufzeichnen (keine punktuelle Lösung für höhere Ordnungen. Synthetische Schallfelder können jedoch für beliebig viele Quellen und beliebig hoher Ordnung erzeugt werden.). Der große Vorteil von Ambisonic dabei liegt darin, dass die Kodierung der einzelnen Kanäle unabhängig von der Lautsprecheranordnung erfolgt.

Für Systeme höherer Ordnung ergeben sich immer komplexere Richtcharakteristiken (Abb. 11). Eine Quelle die ein Signal P_ψ aus der Richtung ψ ausstrahlt wird wie folgt in Ambisonic – Signale kodiert:

$$W = P_\psi \cdot 0.707 \quad X = P_\psi \cdot \cos(\psi) \quad Y = P_\psi \cdot \sin(\psi) \quad U = P_\psi \cdot \cos(2\psi) \quad V = P_\psi \cdot \sin(2\psi) \quad (18)$$

usw. ...



Dekodierung:

Die Dekodierung der gewonnenen Teilsignale kann durch umformen des Gleichungssystem (15) durchgeführt werden.

Matrixschreibweise (von (15)):

$$\vec{b} = A \cdot \vec{x} \quad (19)$$

b ... Übertragungskanäle;

x ... Vektor der gewichteten Lautsprecher signale (P_n);

A ... Lautsprecherpositionen entsprechend der Reihenentwicklung kodiert;

A ist eine allg. nichtquadratische Matrix, d.h. es muss die Pseudoinverse gebildet werden.

Man kann (19) umformen:

$$x = A^T \cdot (A \cdot A^T)^{-1} \cdot b \quad (20)$$

Die Matrix in welche die Positionen der Lautsprecher im Wiedergabesetup mit einfließen wird dabei durch A definiert:

$$A = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos(\varphi_1) & \cos(\varphi_2) & & \cos(\varphi_n) \\ \sin(\varphi_1) & \sin(\varphi_2) & & \sin(\varphi_n) \\ \vdots & \vdots & & \vdots \\ \sin(m\varphi_1) & \sin(m\varphi_2) & \dots & \sin(m\varphi_n) \end{bmatrix} \quad (21)$$

3.2.3 2D → 3D Erweiterung von Ambisonic

Wird das Ambisonic-Setup um eine dritte Dimension erweitert, müssen neue zusätzliche Signale bereitgestellt werden. Am Beispiel des Ambisonicsystems erster Ordnung bedeutet das, dass durch hinzufügen eines Signals Z , sowie zusätzlicher Winkelfunktionen in Abhängigkeit von φ , auch Information über die Elevation der synthetisierten Schallwelle bereitgestellt wird. Man erhält in diesem Fall vier Signale: W , X , Y und Z (20).

$$\begin{aligned} W &= P_\psi \cdot 0.707 \\ X &= P_\psi \cdot \cos(\psi) \cdot \cos(\varphi) \\ Y &= P_\psi \cdot \sin(\psi) \cdot \cos(\varphi) \\ Z &= P_\psi \cdot \cos(\varphi) \end{aligned} \quad (22)$$

Ψ ... Azimut

φ ... Elevation

Diese Methode ist die Einfachste um eine 3D-Umgebung akustisch zu synthetisieren. Spricht man von einem 3D-Setup bei Ambisonic geht man in der Regel jedoch von sog. HOA-Systemen (Higher Order Ambisonic) aus [19]. Hier wird nicht nur eine 2D-Referenzwelle betrachtet sondern ein Schalldruckfeld im 3D-Raum (Abb.11).

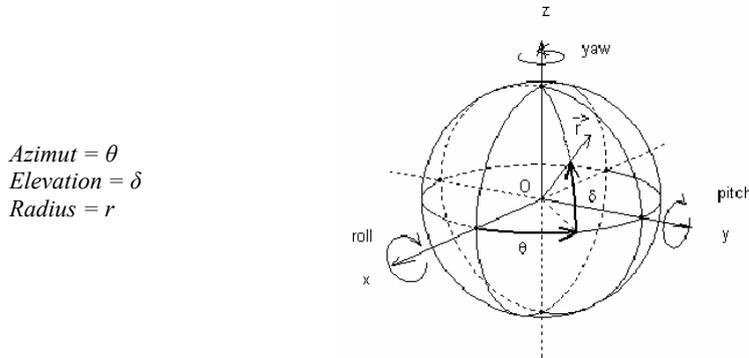


Abb.11: Kugel – Koordinaten – System. Jeder Punkt im Raum kann durch drei Parameter angegeben werden.[19]

Dieses Schalldruckfeld kann wiederum durch Zerlegung in seine Kugelflächenfunktionen ausgedrückt werden (23). Es wird das Schalldruckfeld für ein bestimmtes Raumvolumen definiert ($R_1 \leq r < R_2$) (Abb.12).

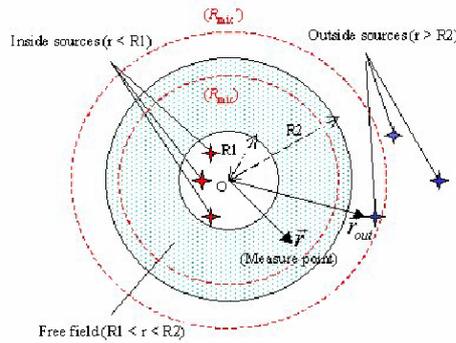


Abb.12: Gültigkeitsbereich für (23). Kugelvolumen zwischen zwei Kugeloberflächen. Außen – Quellen: blaue Sterne, Innen – Quellen: rote Sterne.

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0 \leq n < m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) + \sum_{m=0}^{\infty} j^m h_m^-(kr) \sum_{0 \leq n < m, \sigma = \pm 1} A_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (23)$$

In Formel (23) stellt das erste Polynom die auslaufende Welle welche durch außen stehende Quellen verursacht wird dar. Das zweite Polynom steht entsprechend für die Quellen aus dem Inneren. In Ambisonic nimmt man in der Regel eine zentrierte Position an: man befindet sich im Ursprung des Koordinatensystems, d.h. (21) kann vereinfacht werden zu (22).

$$p(\vec{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0 \leq n < m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \quad (24)$$

Hierbei stellt $j_m(kr)$ die sphärischen Besselfunktionen, B_{mn}^{σ} die Gewichtung der Signale (Ambisonic – Signale) und $Y_{mn}^{\sigma}(\theta, \delta)$ die sphärischen Harmonischen (Abb.13) dar. [19]

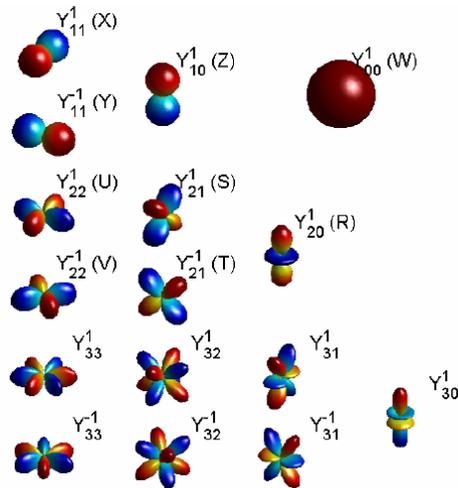


Abb.13: Dreidimensionale Ansicht der sphärischen Harmonischen mit ansteigender Ordnung.

Decodierung:

Prinzipiell geht man bei 3D-Anwendungen gleich vor wie im Beispiel aus 3.2.2.: Die bei der Aufnahme erhaltenen Ambisonic – Signale B_{mn}^σ können später zusammen mit der sphärischen Harmonischen – Matrix dazu verwendet werden die Lautsprechersignale zu berechnen (25) – (27).

$$c_i = \begin{bmatrix} Y_{00}^{+1}(\theta_i, \delta_i) \\ Y_{11}^{+1}(\theta_i, \delta_i) \\ Y_{11}^{-1}(\theta_i, \delta_i) \\ \dots \\ Y_{mn}^\sigma(\theta_i, \delta_i) \\ \dots \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} \tilde{B}_{00}^{+1} \\ \tilde{B}_{11}^{+1} \\ \tilde{B}_{11}^{-1} \\ \dots \\ \tilde{B}_{mn}^\sigma \\ \dots \end{bmatrix} \quad S = \begin{bmatrix} S_1 \\ S_2 \\ \dots \\ S_N \end{bmatrix} \quad (25)$$

\tilde{B} ... Ambisonicsignale; S ... Matrix der Lautsprechersignale; c_i ... Sphärische Harmonische;

Mit $C = [c_1 \dots c_N]$ kann wieder auf die allgemeine Matrix-Form übergegangen werden:

$$\tilde{B} = C \cdot S \quad (26)$$

$$S = D \cdot B \quad \text{mit} \quad D = C^T \cdot (C \cdot C^T)^{-1} \quad (27)$$

3.3 Wellenfeldsynthese (WFS)

Das Prinzip der WFS wurde Ende der 80er von A.J. Berkhout vorgestellt [21][22]. Hier wird die Abbildung einer virtuellen Quelle auf physikalischem Wege erzeugt. Das Prinzip der Wellenfeldsynthese beruht darauf, dass ein komplettes Schallfeld synthetisiert wird, das dem originalen Schallfeld physikalisch identisch ist. Dieses akustische Gegenstück zur optischen Holographie wird auch als „Holophonie“. Die Basis der WFS (*wave field synthesis*) bildet das Huygen'sche Prinzip:

„Wenn von einem Punkt S eines homogen isotropen Mediums eine Kugelwelle ausgeht, so kann man sich den Vorgang der Wellenfortpflanzung im einzelnen so vorstellen, dass ein durch äußere Kräfte in Schwingungen versetztes Teilchen diese seine Bewegung auf seine Nachbarpartikel überträgt. Dieser Vorgang setzt sich dann nach allen Seiten symmetrisch fort und gibt so Anlass zur Entstehung einer Kugelwelle. Ebenso gut könnte man nun auch jedes andere Teilchen, welches durch Auftreffen einer Wellenerregung in Schwingungen versetzt wird, als Zentrum einer Kugelwelle auffassen. ...“ [10]

Man kann dem Huygen'schen Prinzip nach ein Schallfeld, das z.B. durch die Schallquelle S erzeugt wird, aus einer Verteilung von sekundären Quellen auf einer Oberfläche O, die den Hörer umgibt, rekonstruieren. Diese sekundären Quellen bestehen dabei aus Monopol- und Dipolquellen. In *Abb.14* wird das Prinzip dieses Verfahrens grafisch dargestellt. Ist die Wellenfront aller Punkte auf der Fläche O bekannt, so ist es möglich die Schwingung in einem weiter entfernten Punkt P durch Summation aller von der Fläche O ausgehenden Huygens'schen Elementarwellen bestimmen.

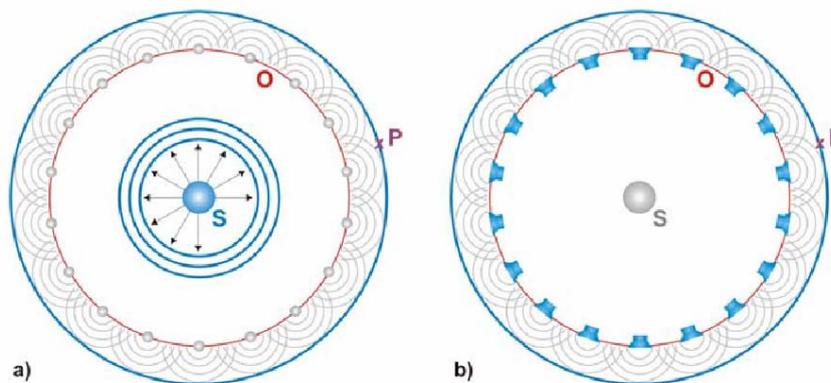


Abb.14

- a) S strahlt kugelförmige Wellenfronten ab = „primäre Quelle“. Der Schwingungszustand in einem beliebigen Punkt P kann durch Überlagerung aller beteiligten „sekundären Quellen“ auf der Oberfläche O, welche ebenso Kugelwellen ausstrahlen, bestimmt werden.
- b) Es können die die Punktquellen durch Lautsprecher ersetzt werden um das Schallfeld zu reproduzieren.

3.3.1 Mathematische Beschreibung:

Die mathematische Formulierung liefert das Kirchhoff-Helmholtz Integral (28).

$$p(\vec{r}) = \iint_{\delta A} \left[\vec{\nabla} p_0 \vec{n} - \frac{\vec{R}}{R} \vec{n} (1 + jkR) \frac{p_0}{R} \right] \frac{e^{-jkR}}{4\pi R} dS_0 \quad \text{mit } \forall \vec{r} \in A \quad (28)$$

In Gleichung (28) wird der Schalldruck p der in einem gegebenen Volumen A herrscht durch das Integral vom Schalldruck p_0 und seinem Gradienten $\vec{\nabla} p_0$ über die Oberfläche δA des betrachteten Volumens A berechnet. k beschreibt in Formel (28) die Wellenzahl und \vec{n} den normierten nach außen zeigenden Flächennormalvektor. Der Vektor \vec{R} beschreibt den Ausbreitungspfad zwischen einer sekundären Quelle und der Hörposition. (Abb.15)

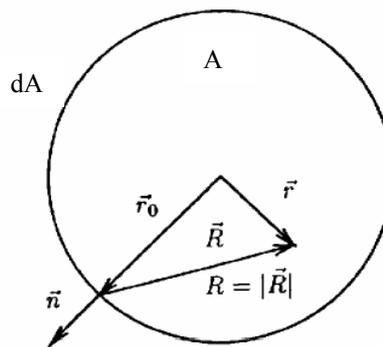


Abb.15: Geometrie für das Kirchhoff-Helmholtz Integral aus Gleichung (28)[11]

Jede sekundäre Quelle besteht, wie bereits erwähnt, aus Monopol- und Dipolquellen. Hierbei wird jede Monopolquelle durch das Schalldruck – Gradient – Signal und jede Dipolquelle durch das Schalldruck-Signal versorgt.

Es sollte erwähnt werden, dass das Kirchhoff-Helmholtz Integral im Gegensatz zum Huygen'schen Prinzip, nicht die Oberflächenbegrenzung als Wellenfront voraussetzt. Die Oberflächenbegrenzung kann jegliche Form besitzen und ist nicht von der Wellenfront abhängig. Diese Eigenschaft führt weiter zu einem weiteren erwähnenswerten Unterschied zwischen dem Kirchhoff-Helmholtz Integral und dem Huygen'schen Prinzip: Während ersteres wie bereits erwähnt die sekundären Quellen durch Schalldruck – Gradient – Signale bzw. Schalldruck-Signale ansteuert, findet beim Zweitem nur das Schalldruck-Signal Anwendung (es wird angenommen dass sich die sekundären Quellen alle in Phase befinden).

Bei der Aufzeichnung von Druckgradient und Druck entlang der begrenzenden Oberfläche greift man auf Schnelle- und Druckmikrofone zurück. Das Prinzip dieses Verfahrens ist in Abb.16 dargestellt. Man kann dabei zwischen zwei äquivalenten Aufnahmesituationen unterscheiden: Einmal kann die Hörposition durch Mikrophone umgeben sein, zu anderen kann auch um die primäre Schallquelle das Mikrophone – Array positioniert werden (Abb.17).

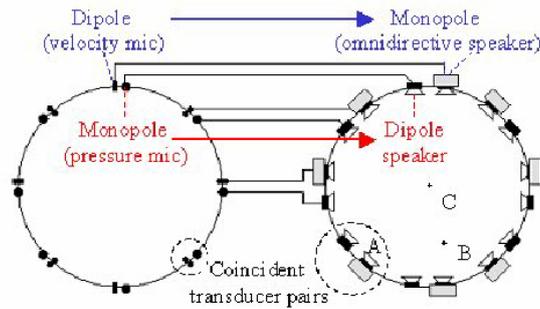


Abb.16: Anwendung des Kirchhoff – Helmholtz – Integrals zu holophonischen Schallfeldaufnahme bzw. Wiedergabe.[19]

Es muss darauf geachtet werden, dass sich die Geometrie zwischen Aufnahme (Mikrophon-Array) und Wiedergabesituation (Lautsprecher-Setup) nicht ändert um das Schallfeld exakt zu reproduzieren.

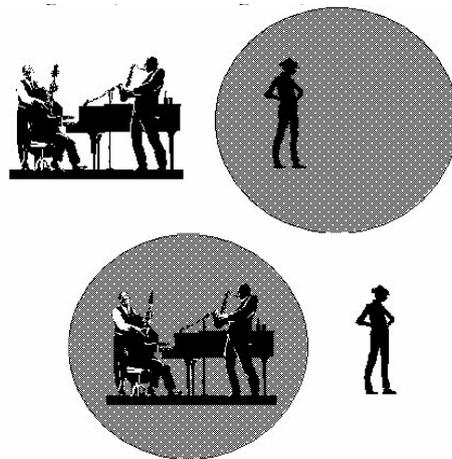


Abb.17: Zwei äquivalente holophonische Setups: es kann der Hörbereich eingehüllt (oben) oder auch die primäre Schallquelle mit Mikrofonen umgeben werden (unten).[19]

Für die Praxis wird eine Lösung angestrebt, welche die Zahl der sekundären Quellen so weit wie möglich reduziert. Dies wird meistens auf Kosten der Vertikalauslenkung realisiert. Die Lautsprecherzeilen werden im Normalfall lediglich in der horizontalen Ebene um den Hörer herum platziert (→ eigentlich kann man erst dann von WFS sprechen!). Deshalb wäre zwar die Methode der WFS theoretisch dazu geeignet Phantomschallquellen in der Elevation zu erzeugen in der Praxis jedoch wäre das Lautsprecher-setup sehr aufwendig.

Der große Vorteil der WFS liegt darin, dass man ausschließlich durch die Abhörordnung eine Begrenzung des Hörbereichs hat. Innerhalb des schallfeldreproduzierenden Bereichs ist jeder Abhörpunkt identisch zum originalen Schallfeld. Zudem wird theoretisch keine Signalverarbeitung zwischen Aufnahme und Wiedergabe benötigt. In der Realität hat man kaum die Möglichkeit ein Mikrofon – Array entsprechend der Wiedergabesituation zu verwenden (Bedingung aus Abb.16: Aufnahmesetup entspr. Wiedergabesetup). In diesem Fall muss ein sehr hoher Signalverarbeitungs – Aufwand betrieben werden.

WFS wird bei der Untersuchung nicht betrachtet, da es sich, wie bereits erwähnt wurde um einen 2D bzw. 2½-D Ansatz handelt und dadurch keine Elevation dadurch abgebildet werden kann.

4. Bereits durchgeführte Hörversuche welche die Wahrnehmung von Schallquellen in der Elevation untersuchen:

4.1 Hörversuch von John C. Middlebrooks [12] (mit realen Schallquellen)

4.1.1 Grundsätzliches

Der Versuch von J.C. Middlebrooks wurde in verschiedene Teilbereiche unterteilt. Zum Einen führte er verschiedene Versuchsdurchgänge bzgl. des Testsignals (Breitbandsignale und bandpassgefilterte Signale) an beeinträchtigungsfreien Hörern („binaural subjects“) durch, zum Anderen wurden anschließend Untersuchungen an - durch einseitigen Hörschaden bzw. Behinderung leidenden - Testpersonen („monaural subjects“) durchgeführt um Vergleiche zwischen den *binaural* und *monaural subjects* anstellen zu können.

Laut Middlebrooks findet Lokalisation in der Medianebene anhand spektraler Reize (*spectral cues*) statt. Für die Lokalisation in der horizontalen Ebene sind vorwiegend interaurale Reize (*interaural cues*) verantwortlich (ITD, ILD). Die spektralen Einflüsse die für die Lokalisation bei Elevation verantwortlich sind werden durch die sog. *directional transfer function* DTF (Middlebrooks und Green, 1990), welche hauptsächlich durch die Transferfunktion des Außenohres verursacht wird, beschrieben. Die räumliche Information welche aus den DTF gezogen werden können, werden von Middlebrooks als „*spectral shape cues*“ bezeichnet. Trotzdem haben *spectral shape cues* auch einen wenn auch geringen Einfluss auf die Lokalisation in der Horizontalen. Zu dieser Ansicht gelangt der Untersuchende aufgrund von Vergleichen zwischen den visuell dargestellten DTF's (*Abb.19*) und der dabei empfundenen horizontalen Hörereignisrichtung: die im Spektrum vorhandenen Wellenberge und Täler variieren bzgl. der Frequenz im gleichen Maße bei horizontaler (links/rechts) und vertikaler (hinten/vorne) Lokalisation. Zur gleichen Überzeugung kamen bereits 1980 R.A. Butler und R. Flannery [13]. Ihnen gelang es anhand von Versuchen zu beweisen, dass die horizontale Lokalisation durch Bandpassfilterung je nach Mittenfrequenz beeinflusst werden kann.

4.1.2 Versuchsaufbau und Durchführung

Der Versuch wurde in einem schalltoten Raum durchgeführt welcher durch akustische Dämmstoffe darauf ausgelegt war Reflexionen für Frequenzen $>500\text{Hz}$ zu unterdrücken. Die Ausmaße des Raumes betragen $2,6\text{m} \times 3,1\text{m} \times 4,3\text{m}$. In der Versuchsanordnung befand sich zudem eine halbkreisförmige Aufhängung (Radius= $1,2\text{m}$) auf welcher die Lautsprecher (insgesamt 16 Stück) in 10° - Schritten von $+90^\circ$ bis -60° angebracht wurden. Um eine Positionierung nicht nur in der vertikalen sondern auch in der horizontalen Ebene zu erlauben wurde die LS-Aufhängung drehbar gelagert und mit einem Schrittmotor angesteuert. Auf diese Weise wurde eine virtuelle Sphäre mit $2,4\text{m}$ Durchmesser erzeugt. Die verschiedenen Lautsprecher wurden über einen computergesteuerten Multiplexer angesprochen werden. Die Testperson wurde bei der hier beschriebenen Anordnung so positioniert dass deren interaurale Achse genau im Zentrum der virtuellen Sphäre zu liegen kam. Alle beteiligten

Signale wurden mit einer Samplingfrequenz von 50kHz generiert und aufgenommen. Die Mikrophonesignale wurden weiters durch eine FFT-Technik anhand eines DSP's weiterverarbeitet. Um Einflüsse durch die beteiligten Lautsprecher ausschließen zu können wurde deren Übertragungsverhalten linearisiert. Eine Skizze der gesamten Anordnung ist in *Abb.18* dargestellt.

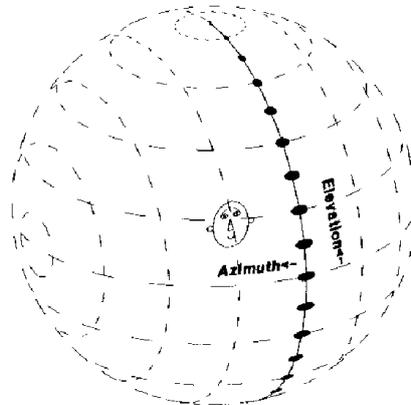


Abb.18: Lautsprecher-Array und Koordinatensystem. Die gefüllten Punkte stellen die Lautsprecher dar welche auf der halbkreisförmigen Aufhängung aufgebracht (10°-Schritte von 90° bis -60°) wurden. Die Sichtweise dieser Skizze ist aus 30° rechts und 30° von oben zu erkennen. Die gestrichelten Linien stellen jeweils 20° - Schritte bzgl. des Azimut- und des Elevationswinkels dar. Der Azimutwinkel wird im Uhrzeigersinn aus der Sicht der Testperson betrachtet. Der Elevationswinkel hat positive Werte für Positionen in der oberen Hemisphäre und entsprechend negative für die untere.[12]

Die DTF wurden bei obiger Versuchsanordnung anhand breitbandigen Rauschsignalen und Miniatur – Messmikrofonen welche an den Ohrkanälen angebracht wurden erhalten. Hierbei wurden die Mess-Signale aus 360 Positionen um die Testperson abgespielt und anschließend aufgezeichnet. Um den Einfluß der Messmikrophone und der stehenden Wellen im Gehörgang zu vermeiden bzw. zu korrigieren wurde ein „gemeinsames Spektrum“ (*common spectrum*) von dem Erhaltenen abgezogen. Dieses *common spectrum* wurde dadurch erhalten indem Middlebrooks den Durchschnitt aller log-Beträge der 360 beteiligten Spektren bildete. Das Ergebnis war nun eine Menge von 360 DTF's aus 360 Raumrichtungen. Um den Einfluss der Cochlea mit in die Funktionen einzubringen wurde zudem noch jede DTF mit Bandpassfilterbank (3dB-Bandbreite einer 0,22 Oktave, Mittenfrequenzen logarithmisch in 0,06-Oktav Schritten) gefaltet. Alle Experimente wurden entweder mit einem breitbandigem oder schmalbandigem gepulsten Rauschsignal mit der Dauer von 150ms bzw. 250ms durchgeführt. Die Positionserkennung durch die Versuchsperson wurde, indem die Kopfposition in Richtung der wahrgenommenen Schallquellenposition gerichtet werden sollte, durch eine elektromagnetische Vorrichtung durchgeführt mit welcher die Orientierung des Kopfes gemessen wurde (Polhemus ISOTRAK). Alle Versuche wurden in kompletter Dunkelheit durchgeführt so dass die Versuchsperson keinen visuellen Anhaltspunkt erhielt. Dabei war der Ablauf folgendermaßen geregelt: Zu Beginn des Versuchs wurde jeweils ein Signal aus 0° Azimut und 0° Elevation eingespielt welches die Testperson als Bezugspunkt für die späteren Testsignale verwenden konnte. Dieser Testimpuls wurde solange eingespielt bis ein „*response key*“ von der Testperson betätigt wurde. Nach einer variablen Wartezeit wurden dann die Rausch-Pulse aus unbekannter Position für die Testperson wiedergegeben. Durch Ausrichten des Kopfes und anschließender Betätigung des *response key* wurde das Messergebnis in die Testreihe aufgenommen.

4.1.3 Directional Transfer Function (DTF)

Als erstes Messergebnis wurden die sog. DFT's aufgenommen. Der Verlauf dieser Kennlinien für eine Testperson kann in *Abb.19* verfolgt werden. Man sieht die Berge und Täler in den spektralen Wellenformen je nach Schallquellenposition. Die in *Abb.19* ersichtlichen Kurvenverläufe ähneln einander auch bei verschiedenen Testpersonen. Zu beachten ist jedoch, dass sich die Positionen der Berge und Täler je Testsubjekt in ihrer Frequenz verschieben. Das ist durch unterschiedlichen Körperbau (Außenohr) zu erklären.

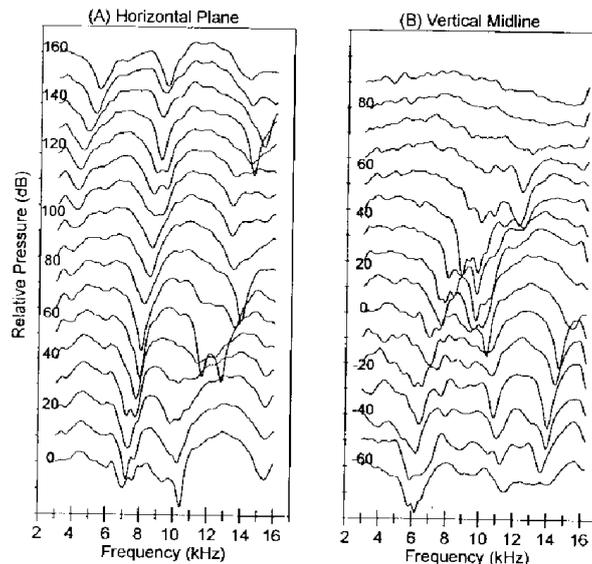


Abb.19: DFTs (directional transfer functions) vom rechten Ohr einer Testperson. (A) DFTs für die horizontale Ebene bei 0° Elevation. Man kann sehr gut die charakteristischen Berge (peaks) und Täler (notches) bei verschiedenen Schallquellenpositionen erkennen. (B) DFTs für die vertikale Ebene bei 0° Azimut. Auch hier sieht man den charakteristischen Verlauf der Kennlinien.[12]

4.1.4 Lokalisation von Breitband – Signalen

Dieser Versuch sollte vordergründig dazu dienen die Leistungsfähigkeit der menschlichen Lokalisationsfähigkeit im Raum unter nahezu idealen Bedingungen zu bewerten. Trotzdem liefen laut Middlebrooks Ungenauigkeiten Aufgrund der nichtidealen Kopfposition der Testpersonen in das Ergebnis mit ein. Testpersonen waren normale binaurale Hörer. Es wurden zwei verschiedene Testsignale in diesem Versuch verwendet.

- (a) 150ms langes gepulstes Rauschsignal mit zufälliger Phase und nahezu flachem Pegelverlauf im Bereich von 1,8kHz bis 16kHz.
- (b) 250ms langes gepulstes bandpaßgefiltertes Gauß'sches Rauschen von 2kHz bis 15kHz.

Man würde annehmen, dass vom flachen Spektrum des Signals (a) mehr spektrale Informationen zur Lokalisierung geliefert würden als von Signal (b). Es stellt sich jedoch heraus dass es keine gravierenden Unterschiede bei der Lokalisationsschärfe gab. Das Ergebnis des Hörversuchs mit breitbandigen Signalen stellt *Abb.20* dar.

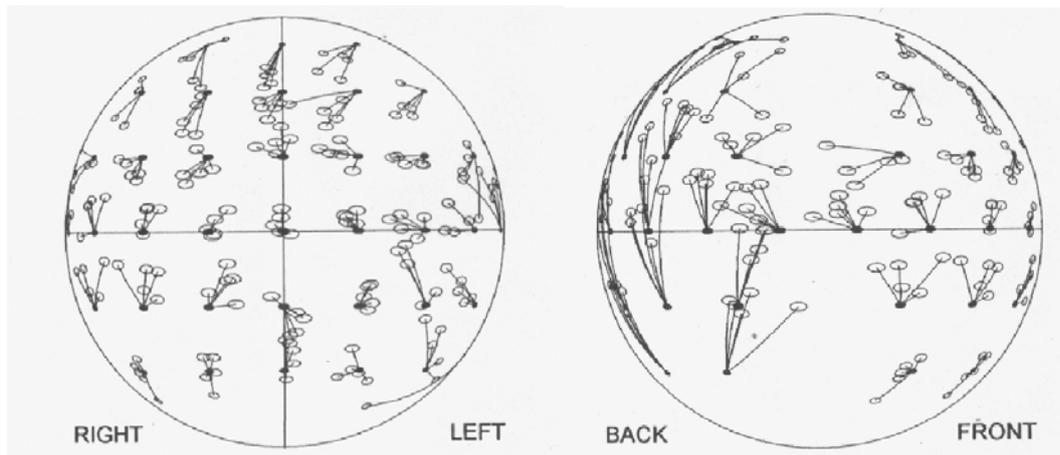


Abb.20: Breitband – Signal – Lokalisation. Die reale Position der Quellen ist durch die gefüllten Kreise dargestellt. Die Entsprechenden empfundenen Positionen sind durch die nicht-gefüllten Kreise welche mit Linien von den realen Quellen weggeführt abgebildet. Man kann erkennen, dass fünf Versuche je Quelle durchgeführt wurden. Die Sichtweisen der Diagramme sind einmal für die Position Frontal zur Testperson (links) und einmal aus $+90^\circ$ (rechts) gegeben. [12]

Wie angenommen ist die Lokalisation frontal zur Testperson am besten im Vergleich zum hinteren Bereich. Der gemittelte Lokalisationsfehler über alle Testpersonen betrug in der Horizontalen $5,8^\circ$ und in der Vertikalen $5,7^\circ$. Nach bisherigen Annahmen ging man laut Middlebrooks davon aus, dass die Lokalisation bzgl. des Azimutwinkels viel genauer möglich sei als in der Elevation. Für Bereiche die frontal zum Hörer liegen konnte diese Annahme bestätigt werden, jedoch für weiter lateral gelegene Quellen lieferte die Lokalisation in der Elevation genauere Ergebnisse.

Es wurde in 2-10% der Fälle eine Vorne/Hinten – Verwechslung (*front/back confusion*) festgestellt, bei welcher auffällig war, dass der Winkel zur vertikalen Mittelachse, zwischen realer und von der Testperson angegebener Position, meistens übereinstimmten obwohl jedoch die Elevationsangaben stark voneinander abwichen.

4.1.5 Lokalisation von Schmalband – Signalen

Bei dieser Versuchsreihe wurden Schmalband-Signale verwendet um den Einfluss von den spektralen Wellenformen (*spectral shape cues*) bei der Lokalisation zu untersuchen. Testpersonen waren auch hier normale binaurale Hörer welche gepulstes schmalbandgefiltertes Gaußsches Rauschen lokalisieren sollten. Die Schmalbandfilter hatten eine Bandbreite von 3dB einer 0,227 Oktave und Mittenfrequenzen von 6,8,10 und 12kHz. Das Ergebnis waren deutliche Fehler der Lokalisation in der Vertikalen und zudem gab es starke vorne/hinten Verwechslungen. Eine Versuchsperson lokalisierte dabei alle 6kHz-Stimuli in einem 20° breiten Bereich mit starker d.h. hoher Elevation (Abb.21). Hingegen wurden die meisten der 8kHz – Stimuli aus tiefen Elevationsgebieten wahrgenommen (Abb.22). Tendenziell lagen weiters alle 10kHz-Stimuli im hinteren Bereich und alle 12kHz-Stimuli auf der horizontalen Ebene vorne und hinten.

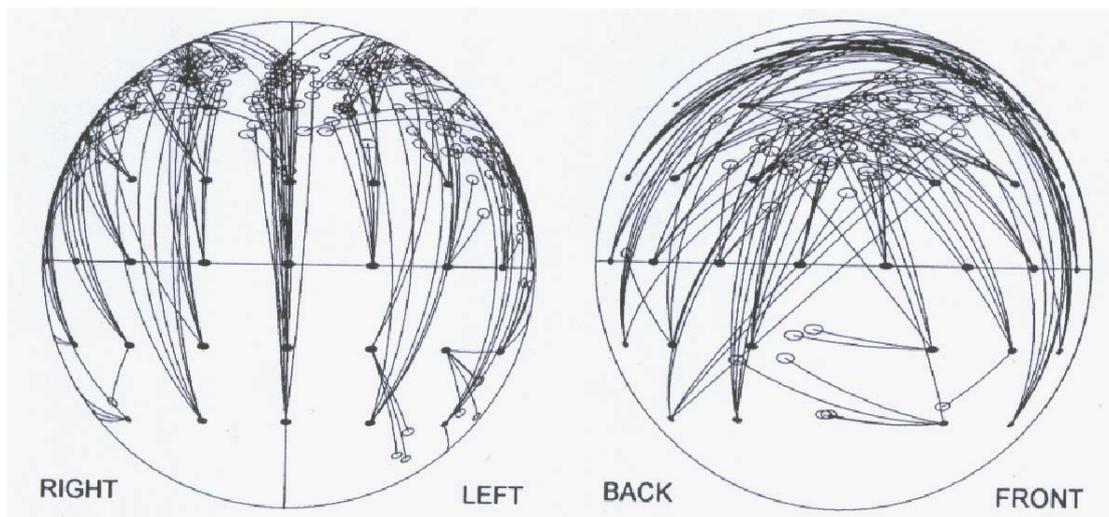


Abb.21: Schmalband – Signal – Lokalisation. Testsignal war bei diesem Versuchsdurchlauf ein Schmalbandsignal mit der Mittenfrequenz von 6kHz. Die Sichtweisen der Diagramme entsprechen jener aus Abb.14. [12]

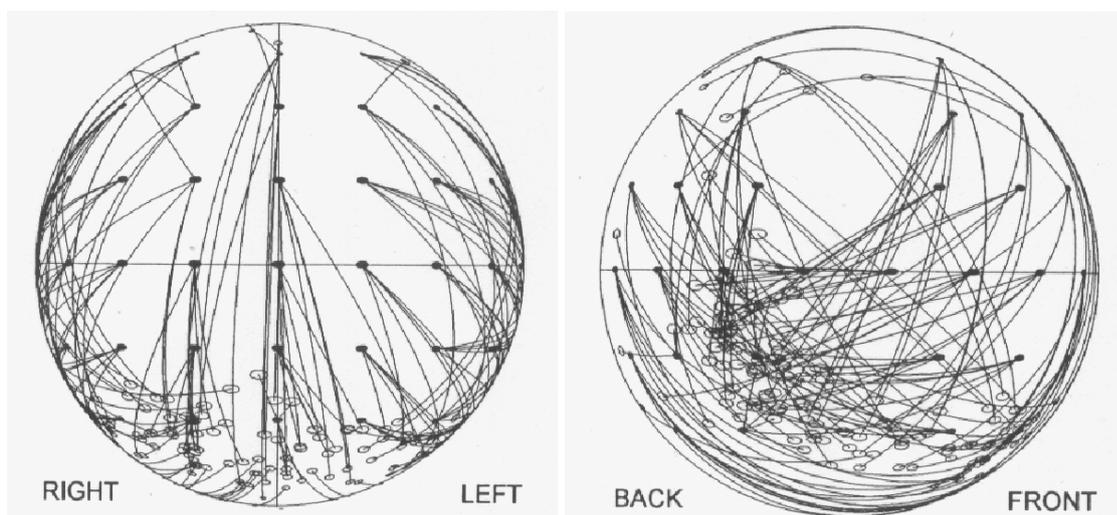


Abb.22: Schmalband – Signal – Lokalisation. Testsignal mit 8kHz Mittenfrequenz. [12]

Trotz des systematischen Fehlers bei der Lokalisation in der Vertikalen und im vorderen bzw. hinteren Bereich war die Lokalisation in der horizontalen nahezu gleich genau wie mit Breitbandsignalen. Diese Eigenschaft konnte auch mathematisch bewiesen werden indem man die wahre und die subjektive Position der Quellen durch die Winkelabweichungen bzgl. der Medianebene betrachtete. Hiervon wurde der Korrelationskoeffizient (zwischen Stimulus und subjektiver Schallquellenposition ohne Rücksicht auf vorne/hinten Lokalisation zu nehmen) gebildet. Es ergaben sich im Durchschnitt durch alle Testpersonen und Mittenfrequenzen Werte zwischen 0,932 und 0,970. Im Vergleich dazu war der Korrelationskoeffizient für die Lokalisation in der Horizontalen bei Breitbandsignalen, bei denselben Testpersonen, im Bereich von 0,973 und 0,985. Diese Ergebnisse ließen laut Middlebrooks darauf schließen, dass das Urteil der Testpersonen über die Quellenposition durch die schmalbandigen Stimuli unbeeinflusst blieb. Aufgrund der relativ hohen Mittenfrequenzen wurde weiters angenommen dass die Lokalisation in der Horizontalen fast ausschließlich auf ILD basierte (Diese Annahme wurde durch die Untersuchung der DTFs belegt).

4.1.6 Lokalisations – Modell

Aus den bisher gemachten Erfahrungen wurde von Middlebrooks (1992) [14] ein Modell basierend auf den *directional transfer functions* entworfen mit dem die Systematik, mit welcher der Mensch Lokalisation von Schallquellen betrieb, erklärt werden sollte. Das Modell wurde speziell für die Lokalisation von Schmalbandsignalen entworfen sollte aber auf einfache Art und Weise auf andere Signalarten anwendbar sein.

Middlebrooks kam zu dem Ergebnis, dass es einen Zusammenhang zwischen der DFTs der Testpersonen und dem spektralen Verlauf des Stimulussignals gab. Das stach besonders bei falscher vorne/hinten Lokalisation ins Auge. Die DFT aus der wahrgenommenen (in Abb.23 falschen) Richtung und das Stimulussignal hatten in diesen Fällen starke Ähnlichkeit in ihrer spektralen Wellenform (Abb.23).

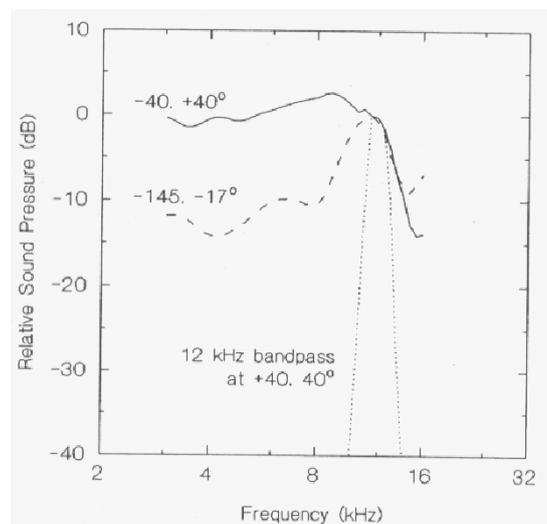


Abb.23: Ähnlichkeit zwischen der DTF aus der Richtung $-145^{\circ}, 17^{\circ}$ (linkshinten-unten) – durch die gestrichelte Linie dargestellt – und dem Erregersignal – Spektrum (punktierte Linie) mit Mittenfrequenz von 12kHz. Das Signal wurde aus hierbei aus $-40^{\circ}, 40^{\circ}$ (linksvorne-oben) abgestrahlt und ist hier durch die durchgehende Linie erkennbar.[12]

Es wurde daraufhin speziell die Ähnlichkeit der Wellenform zwischen Stimulus- und DTF – Spektrum anhand des Korrelationskoeffizienten betrachtet. Das Ergebnis dieser Berechnungen - aus 1652 äquidistanten Punkten auf der Wiedergabe-Sphäre - kann am Beispiel eines Schmalbandsignals mit Mittenfrequenz 12kHz in Abb.24 gesehen werden.

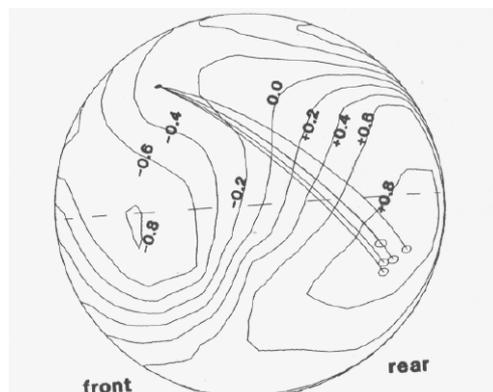


Abb.24: Spektrale Korrelation eines Schmalbandsignals mit Mittenfrequenz 12kHz und den DTFs der Versuchsperson. Die empfundene Lokalisation fällt mit dem Gebiet der größten Korrelation zusammen.[12]

Mann kann sehr gut die Gebiete hoher Ähnlichkeit erkennen in welchen die Wahrscheinlichkeit für die Lokalisation erhöht ist.

Durch Kombination dieser Korrelationsmessung mit der Messung des Unterschiedes der ILD zwischen Stimulus- und empfundener Quellposition wurde weiters ein „Ähnlichkeitsindex“ (*similarity index*) berechnet, der monoton in Richtung der empfundenen Quellenposition anstieg. Das Ergebnis dieser Berechnung kann am Beispiel eines Schmalbandsignals mit Mittenfrequenz 6kHz in *Abb.25* betrachtet werden.

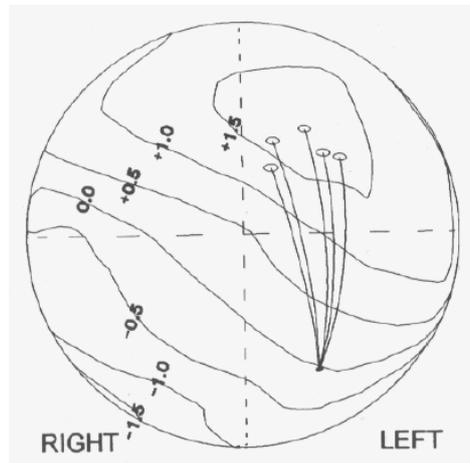


Abb.25: Räumliche Verteilung des Ähnlichkeitsindex („similarity index“) eines Schmalbandsignals mit Mittenfrequenz von 6kHz und den DTFs einer Testperson. Man kann sehr gut erkennen wie der „similarity index“ monoton in Richtung der empfundenen Quellenposition ansteigt. Die Stimulusquelle war an Position - 20°, -40° vorhanden. [12]

Durch Mittelung über alle Testpositionen und Probanden konnte basierend auf den Ergebnissen für alle Positionen und allen vier verschiedenen Mittenfrequenzen ein Diagramm (*Abb.26*) erstellt werden welches die Wahrscheinlichkeit der Lokalisation in einem bestimmten begrenzten Gebiet (durch % - Anteile angegeben) auf der Sphäre beschreibt.

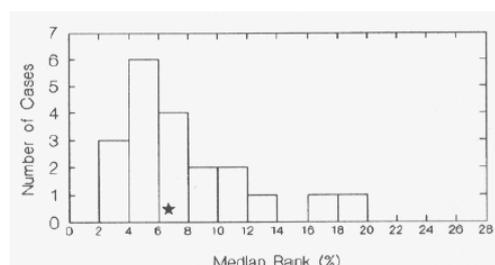


Abb.26: Wahrscheinlichkeit der Lokalisation in einem bestimmten begrenzten Gebiet. Der Stern gibt den Mittelwert der Verteilung im Diagramm an.[12]

Middlebrooks kam nach den bisherigen Testergebnissen zum Ergebnis, dass die horizontale Lokalisation vorwiegend von ILD cues und die Elevation (bzw. vorne/hinten) hauptsächlich durch *spectral cues* beeinflusst wurden.

Alle erhaltenen Ergebnisse, speziell jene durch die Korrelation zwischen Stimulus und empfundener Position erhaltenen, sind mit Vorsicht zu genießen, da die beschriebenen Methoden laut Middlebrooks mathematisch zwar bequem und zutreffend, biologisch jedoch, seiner Meinung nach, nicht plausibel sind.

4.2 Hörversuch von Ville Pulkki und M. Karjalainen [18] (mit VBAP)

4.2.1 Methode

Die Hörversucher wurden von Pulkki und Karjalainen anhand der „method of adjustment (MOA)“ durchgeführt; d.h. die Testperson musste die Schallquellen-Position (reale Bezugsquellen) durch eigenhändiges Panning so gut wie möglich einstellen. Diese Methode bietet den Vorteil, dass die Konzentration der Testperson über die gesamte Versuchsdauer hoch gehalten wird. Ein Nachteil der MOA-Methode hingegen sind systematische Fehler die z.B. durch gleich bleibende Anfangsbedingungen, undefinierte Regeln, undefinierte Versuchsdauer, usw. entstehen können. Diese negativen Effekte können jedoch durch zufällige Wahl der Stimuluspositionen oder Anfangsbedingungen minimiert werden. Trotzdem bleibt das Risiko, dass von der Testperson nicht die empfundene Position sondern nur der gleich „klingende“ Stimulus eingestellt wird (→gleichartiger systematischer Fehler).

4.2.2 Versuchsanordnung

Die Versuche wurden in einem großen schalltoten Raum durchgeführt. Dabei saßen die Testpersonen auf einem Stuhl welcher mit einer Kopflehne ausgestattet war. Sie waren dabei so in der Versuchsanordnung positioniert, dass sich ihre interaurale Achse genau auf derselben Ebene befand wie ein Lautsprecher mit dem Elevationswinkel $\phi_{cc}=0^\circ$. Während des Hörversuchs sollten die Probanden ihre Köpfe nicht bewegen und die Augen geschlossen halten um nicht durch visuelle Einflüsse abgelenkt zu werden. Die Testsignale waren ein breitbandiges Rosa Rauschen und ein Oktav-Band Rauschen welche zuerst als Referenzsignal und anschließend als virtuelle Quelle eingespielt wurde. Nach einer kurzen Pause wurde die Sequenz wiederholt (zeitlicher Verlauf siehe Abb.27). Details über den Hörversuch können aus folgender Tabelle entnommen werden:

Loudspeakers	Spherical: Ø 160 mm; Driver: Audax AT080M0
Loudspeaker distance	2m
Subjects	Ages 19–31; hearing threshold ≤ 20 dB
Familiarization	One virtual source adjustment
Pink noise	1.4-s sample of broad-band pink noise
Octave-band noise	Slope attenuation -25 dB/ octave; 10 center frequencies according to ERB scale
Loudness	60 dB(A) ± 4 dB

Tabelle 1: Hörtest-Details

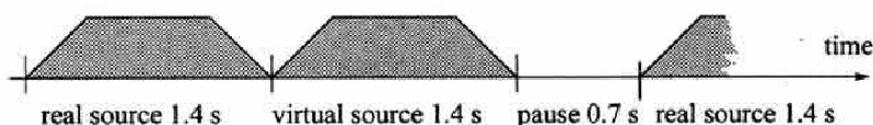


Abb.27: Zeitlicher Verlauf der Test-Signale

Um die Performance der Testpersonen besser interpretieren zu können wurde zur Modellierung bzw. Simulation der sog. „*auditory cues*“ die durch die Pantomoschallquellen erzeugt wurden ein persönliches binaurales Modell erstellt (d.h. die individuellen HRTFs wurden zuvor vermessen). Dabei wurden die ITD, ILD und die DLL (*directional loudness level* → entspricht der DTF aus den Versuchen von J. Middlebrooks, Kap. 4.1) Spektren für die virtuellen und die realen Schallquellen simuliert.

4.2.3 Panning auf der Medianebene

In diesem Experiment wurde die Lokalisation auf der Medianebene untersucht. Dabei wurde ein Panning zwischen zwei Lautsprechern welche vertikal angeordnet wurden ($\phi_{cc1}=-15^\circ$, $\phi_{cc2}=30^\circ$) anhand zweier realer Schallquellen betrachtet. Die Versuchsanordnung ist in *Abb.28* dargestellt; die beiden realen Quellen befinden sich hierbei an den Positionen $\phi_{cc}=0^\circ$ und $\phi_{cc}=15^\circ$.

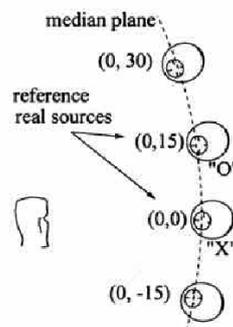
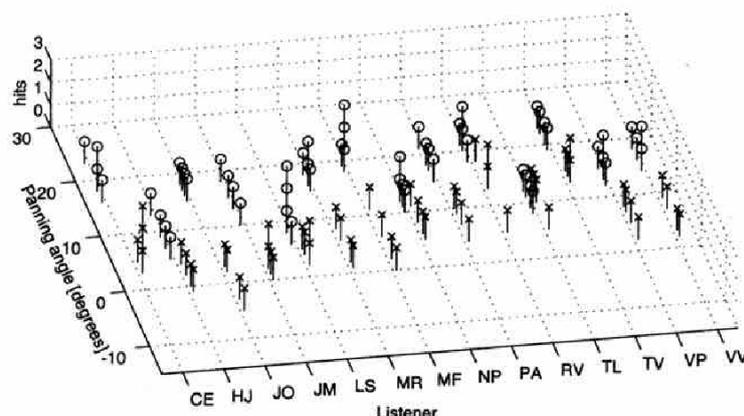


Abb.28: Setup für den Medianebenen-Hörtest. Die angegebenen Koordinaten bezeichnen hierbei den Azimut- und den Elevationswinkel.

4.2.3.1 Testsignal: Rosa Rauschen

Das Testergebnis kann für die 14 verschiedenen Probanden in *Abb.29* betrachtet werden. Die Eindrücke der versch. Hörer wichen jedoch einigermaßen voneinander ab.



*Abb.29: Empfundene Panning – Winkel nach Testperson geordnet. Testsignal: rosa Rauschen. x...gibt die empfundene Panningrichtung für die reale Quelle an $\phi_{cc}=0^\circ$ an, o...jene für die reale Quelle an $\phi_{cc}=15^\circ$. Vergleiche: *Abb.28* → Testsetup.*

Es wurde weiters für die Analyse der Ergebnisse die sog. ANOVA-Methode (*analysis of variance*) verwendet. Mit dieser sollte der Einfluss der beteiligten Variablen (Testpersonen (*subjects*), Quelle (*source*) und Wiederholungen (*repetition*)) untersucht werden. Hierbei stellte sich heraus dass es keine signifikanten Einflüsse von Wiederholungen und ebenfalls zwischen Wiederholungen und Quelle bzw. Testperson gab. Lernen hatte also keinen Einfluss auf das Testresultat.

Die Variable „Quelle“ hingegen hatte Auswirkungen auf das Testergebnis (Abb.30). Die Abweichungen waren bei $\phi_{cc}=0^\circ$ größer im Verhältnis zu $\phi_{cc}=15^\circ$.

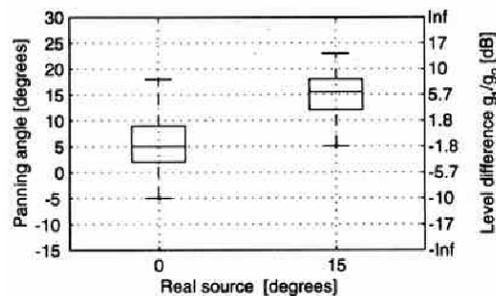


Abb.30: Zusammengefasstes Ergebnisses des Hörtests mit breitbandigem rosa Rauschen in der Medianebene.

In vergleichbaren Experimenten die die Lokalisationsschärfe in der Horizontalen untersuchten wurde eine durchschnittliche Abweichung von 3° festgestellt was im Verhältnis zum obigen Versuch (20° Abweichung) ein sehr viel besseres Ergebnis darstellt. Die Auswirkungen der der Variablen *subject* bzw. die Interaktion zwischen *subject* und *source* hatte laut ANOVA signifikante Auswirkungen. Es wurde festgestellt, dass sich der Panningwinkel bei gleichem Elevationswinkel subjektiv für jede Testperson unterschiedlich einstellte. Folglich konnte, laut Pulkki, kein einheitliches Gesetz für die Lokalisation in der Medianebene entdeckt werden.

Von acht teilnehmenden Versuchspersonen waren bereits HRTF – Kurven vorhanden so dass eine Analyse der theoretischen gewonnen Daten aus der HRTF-Simulation mit den DLL-Spektren der realen Quellen erfolgen konnte. Pulkki erwartete sich aus dieser Vorgangsweise, dass er eine Ähnlichkeit (speziell die Lage der sog. *peaks* und *notches*) zwischen den beiden Welleformen nachweisen könnte. Tatsächlich fand er jedoch bei keinem der Probanden nennenswerte Übereinstimmungen (siehe Abb.31).

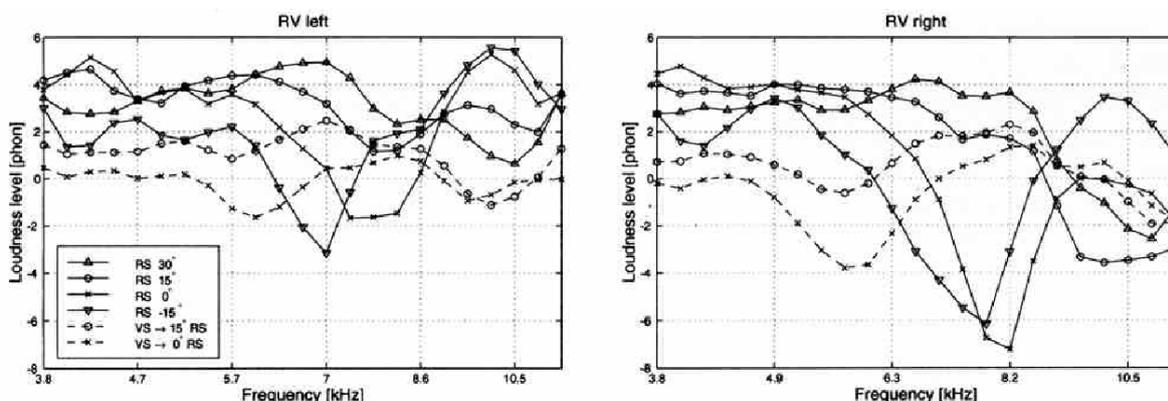


Abb.31: Monaurale DLL-Spektren für reale und virtuelle Quellen für beide Ohren der Testperson RV. RS...Reale Quellen, VS...Virtuelle Quellen;(Hörtest: Medianebene, Breitband – Rosa – Rauschen .[18]

4.2.3.2 Testsignal: Oktav-Band rosa Rauschen

Um die Lokalisationsfähigkeit bei schmalbandigeren Stimuli zu untersuchen wurde dieser Test mit Oktav-Band rosa Rauschen bei insgesamt 10 unterschiedlichen Mittenfrequenzen durchgeführt (gemäß ERB-Skala („Equivalent Rectangular Bandwidth“)). Die insgesamt zehn Versuchspersonen sollten die virtuelle Quelle entsprechend der realen Quellenposition (jeweils 0° und 15°) für jedes Stimulussignal einstellen. Bei gewissen Mittenfrequenzen war die Einstellung der Position für einige Versuchspersonen nicht möglich weshalb ein „not possible – button“ bei der Positionierungsvorrichtung hinzugefügt wurde. Die prozentuelle Anzahl der Fälle in denen die 10 Probanden den „not possible – button“ betätigten ist in *Abb.32* dargestellt.

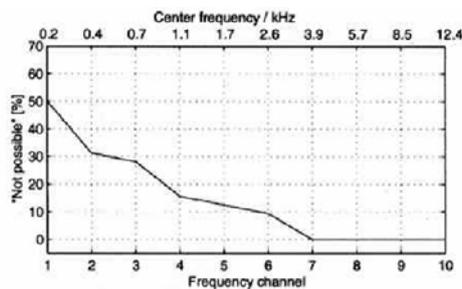


Abb.32: Prozentueller Anteil der „not possible“ – Wertungen (Hörtest für Medianeben-Panning mit Oktavband rosa Rauschen). [18]

Man kann in *Abb.32* die Tendenz erkennen, dass besonders tiefe Frequenzen schwer zu lokalisieren waren. Das deckt sich mit der bereits gewonnen Erkenntnis, dass besonders hohe Frequenzen für die Elevations-Wahrnehmung verantwortlich sind.

Eine Zusammenfassung des Testergebnisses ist in *Abb.33* dargestellt. Es kann festgestellt werden wie sich die Lokalisation bei tiefen Frequenzen (<3kHz) teilweise über den gesamten Panningbereich streut. Über dem 3kHz-Bereich hinaus nimmt die Lokalisationsschärfe im Vergleich dazu beträchtlich zu hat aber trotzdem eine beträchtliche Ungenauigkeit im Vergleich zu Versuchen mit realen Quellen. Wenn der „not possible – button“ betätigt wurde ein Zufallsergebnis mit in die Auswertung übergeben.

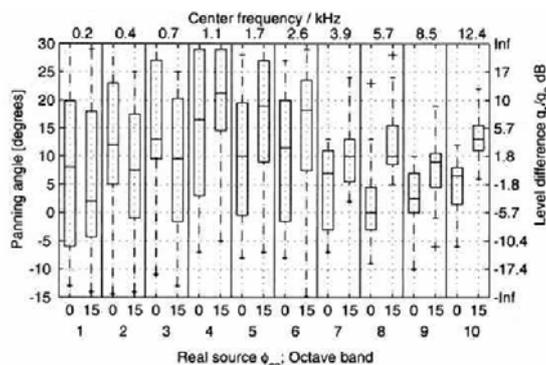


Abb.33: Hörtest in der Medianebene. Testsignal: Oktav-Band rosa Rauschen. Es sind die besten Ergebnisse für jedes Frequenzband und reale Quelle angegeben.[18]

Um herauszufinden ab welcher Frequenz verlässliche Lokalisation möglich ist wurde ein sog. Wilcoxon –Test durchgeführt der den Zusammenhang zwischen Panningwinkeln und oberer (15°) bzw. unterer (0°) Referenzquelle bei jeder Mittenfrequenz und jeder Testperson untersuchte. Dabei stellte sich heraus dass der Einfluss der Richtung der realen Quelle unter 3kHz keine signifikante Stellung einnahm, ganz im Gegensatz für Frequenzen > 3kHz. Hier gelang eine mehr oder weniger genaue Einstellung der virtuellen Quelle. Pulkki kommt Aufgrund des bisher Gehörten zur Überzeugung dass Amplitudenpanning in der Medianebene für Signale mit Frequenzen < 3kHz keine stabilen Phantomschallquellen erzeugen können.

Bei der Untersuchung der acht Probanden von denen HRTF – Kurven zur Verfügung standen wiederholte sich das Ergebnis aus 4.2.3.1. Zwischen den DLL – Spektren der realen und virtuellen Quellen bestand keine Korrelation. Trotzdem fand Pulkki eine sich wiederholende spektrale Charakteristik: Der Pegel in der Nähe von 6kHz steigt mit steigendem Panningwinkel (Abb.34).

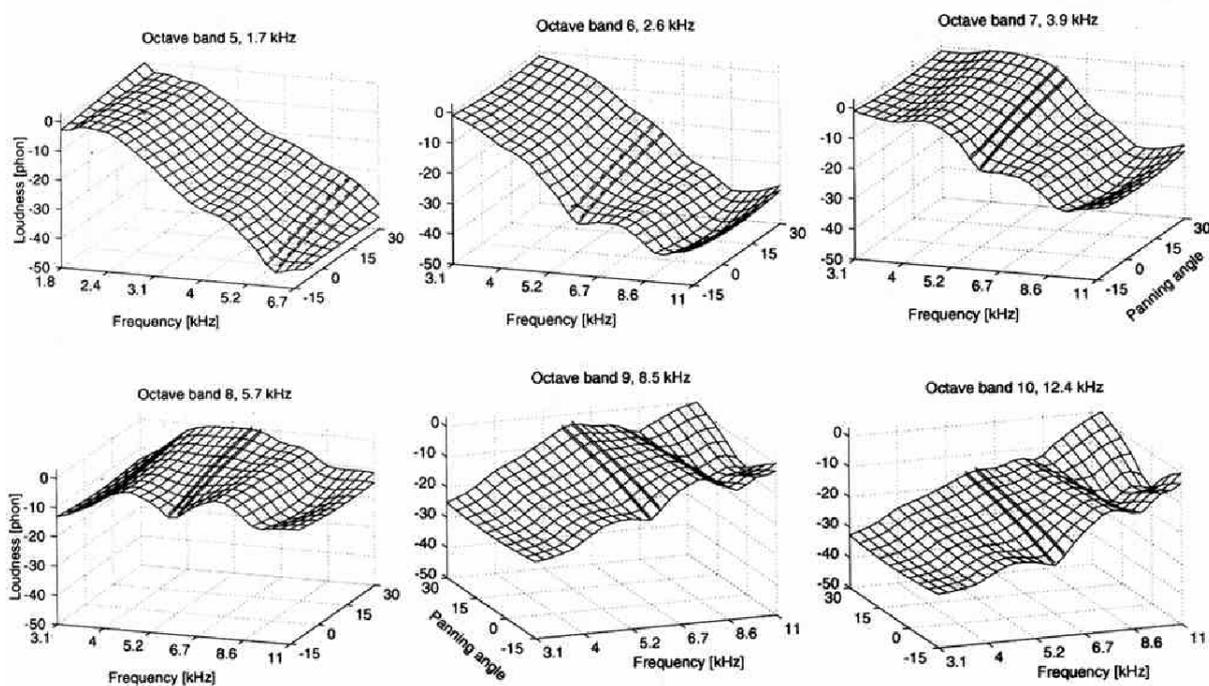


Abb.34: DLL-Spektrum für verschiedene Mittenfrequenzen. Panningbereich: -15° bis 30°. Testsignal: Oktav-Band rosa Rauschen. Eine spektrale Eigenheit welche möglicherweise die Wahrnehmung des Elevationswinkels beeinflusst ist durch die breiten Linien in der Nähe von 6kHz eingezeichnet. (gestrichelte Linien: keine sinnvolle Lokalisation möglich; durchgehende breite Linien: gute Lokalisation vorhanden) Testperson: HJ. [18]

Für Testsignale mit Mittenfrequenz < 3,9kHz ergeben sich Pegelunterschiede von ca. 20dB. Pulkki geht davon aus, dass diese Eigenschaft die ϕ_{cc} – Richtungsempfindung beeinflusst.

Hypothese von Ville Pulkki über die Wahrnehmung der ϕ_{cc} – Richtung:

Spektrale Wellenformen (*peaks and notches*) sind laut Pulkki für die Elevationswahrnehmung, besonders nach den durchgeführten Experimenten, keine befriedigende Grundlage. Deshalb wurde ein Versuch gestartet eine alternative, neue Hypothese aufzustellen:

Der Lautstärkepegel bei der Pinna – Modenfrequenz f_m wird mit $L(f_m)$ bezeichnet. Dieser Wert ändert sich bei Variation von ϕ_{cc} . Ist nun der Lautstärkepegel bei der Modenfrequenz $L_0(f_m)$ - der auftritt wenn es keine Pinna – Moden gäbe - bekannt, dann kann daraus die Elevationsrichtung ϕ_{cc} bestimmt werden. L_0 kann bestimmt werden für den Fall, dass der Pegelverlauf im Bereich um die Modenfrequenz flach verläuft. Der angenommene Wert wird dann mit $L_0^*(f_m)$ bezeichnet.

Jetzt kann der ϕ_{cc} – Richtungs – Reiz (ϕ_{cc} direction cue) $q(f_m)$ durch folgende mathematische Beziehung ausgedrückt und bei versch. Frequenzen berechnet werden:

$$q(f_m) = \frac{L(f_m)}{L_0^*(f_m)} \quad (29)$$

Die Richtigkeit dieser Formel begründet Pulkki mit der Überlegung, dass im Falle eines Schmalband – Signals, welches nahe an der Mittenfrequenz einer Mode – die Ihr Maximum bei hohen ϕ_{cc} hat – liegt, $q(f_m)$ entsprechend große Elevation voraussagt ($L(f_m) \rightarrow$ erhöht, $L_0^*(f_m) \rightarrow$ nicht erhöht).

Diese Betrachtungsweise stimmt mit den Ergebnissen frühere Schmalband - Experimente überein [14]. Zudem erklärt seine Hypothese teilweise qualitativ auch die Ergebnisse seiner Hörtest-Ergebnisse.

4.2.4 Panning in drei Dimensionen

Es wurden zweit Tests mit jeweils verschiedenen Lautsprecher – Tripeln durchgeführt. Acht Personen hatten wieder die Aufgabe virtuelle Quellen auf die Richtung von realen Quellen abzugleichen. Es wurden jeweils vier Wiederholungen durchgeführt. Als Testsignal dient auch hier rosa Rauschen.

4.2.3.1 Lautsprecher - Tripel I

Das erste Tripel bestand aus zwei Lautsprechern auf einer Linie welche um $\pm 28^\circ$ im Azimut und 17° in der Elevation bzgl. des Nullpunktes verschoben waren. Der dritte Lautsprecher hatte einen Azimut- von 0° und einen Elevationswinkel von 30° (siehe Abb.29). Die realen Bezugsquellen waren auf der Medianebene (Azimut 0°) mit den Elevationswinkeln 0° und 15° angeordnet (Abb.35).

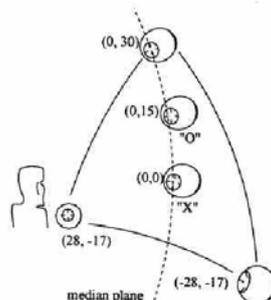
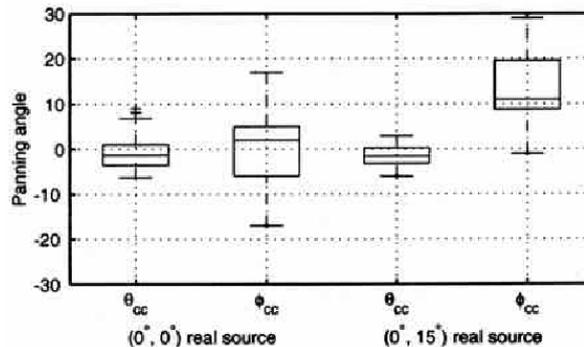


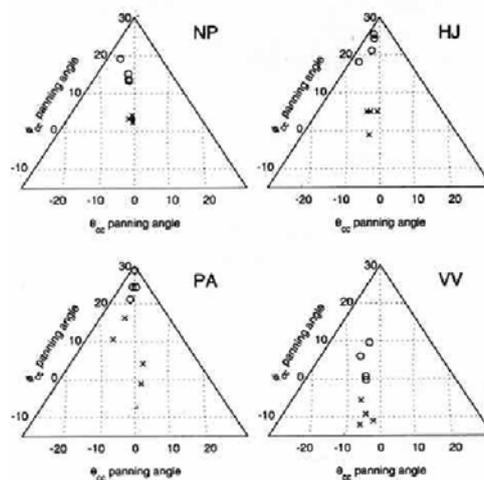
Abb.35: Testanordnung für Tripel I. Die beiden realen Quellen sind durch x und o angegeben. Schreibweise: (Azimut, Elevation).[18]

Das Testergebnis, gemittelt über alle Probanden, ist in *Abb.36* dargestellt. Obwohl die Azimutwinkel θ_{cc} sehr gute Lokalisationsergebnisse lieferten war die Wertstreuung beim Elevationswinkel ϕ_{cc} relativ hoch.



*Abb.36: Der gemittelte Wert der besten Testergebnisse der 8 Probanden für Tripel I. Testsignal: rosa Rauschen
Das Panning wurde insgesamt viermal durchgeführt.[18]*

Vier repräsentative Testergebnisse sind in *Abb.37* dargestellt. Auch hier wird deutlich, dass das Panning bzgl. θ_{cc} bessere Ergebnisse im Vergleich zu ϕ_{cc} liefert. Laut Pulkki müsste die ILD bzw. ITD für die realen Quellen den Wert „0“ annehmen, was eine Lokalisation auf der Medianebene nach sich ziehen würde; daß es trotzdem Abweichung gab führte er auf „nicht ideale“ Faktoren (keine genaue Spezifikation) im Versuchs – Setup zurück.



*Abb.37: Persönliche Ergebnisse für vier von Pulkki ausgewählte repräsentative Testpersonen.
 $x=(0^\circ, 0^\circ); o=(0^\circ, 15^\circ)$
Bei HJ und PA ein „Panning – Bias“ erkennbar. Gegensatz: NP sehr gute Lokalisationsperformance.[18]*

4.2.4.2 Lautsprecher - Tripel II

Das Tripel war durch zwei Lautsprecher in der Medianebene an den ϕ_{cc} - Positionen -15° und 30° und einem Dritten am Punkt $\theta_{cc} = -43^\circ$ und $\phi_{cc} = -20^\circ$ zusammengestellt. Die realen

Quellen befanden sich in der Horizontalebene an den Positionen $\theta_{cc} = -20^\circ$ und $\theta_{cc} = -10^\circ$. (Siehe Abb.38)

Das zusammengefasste Ergebnis ist in Abb.39 dargestellt. Die Azimutauslenkung wurde wiederum sehr gut festgestellt, doch auch die Elevationswinkel – Lokalisation lieferte bessere Ergebnisse als bei Tripel I. Es bestand jedoch die Tendenz, dass die Phantomschallquellen welche auf die reale Quelle bei $(-10^\circ, 0^\circ)$ eingestellt wurden, eine größere Elevation aufwiesen als jene die auf $(-20^\circ, 0^\circ)$ justiert werden sollten. Diese Phänomen konnte Pullki nicht erklären.

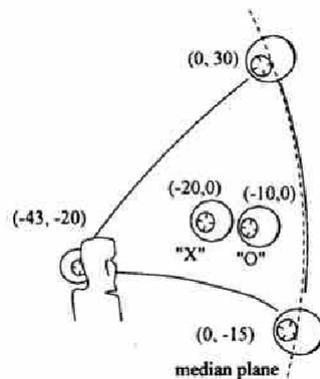


Abb.38: : Testanordnung für Tripel II. Die beiden realen Quellen sind durch x und o angegeben. Schreibweise: (Azimut, Elevation).[18]

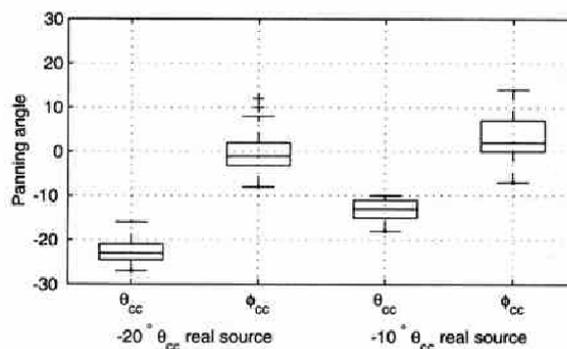


Abb.39: Der gemittelte Wert der besten Testergebnisse der 8 Probanden für Tripel II. Testsignal: rosa Rauschen Der Versuch wurde insgesamt vier mal durchgeführt.[18]

Vier ausgewählte repräsentative Testergebnisse von einzelnen Subjekten sind in Abb.40 dargestellt. θ_{cc} – Werte hatten zwischen allen Teilnehmern gute Übereinstimmungen. Wie schon im vorhergehenden Versuch waren die ϕ_{cc} – Werte zwar individueller verteilt, hatten jedoch je Proband auch überzeugende Übereinstimmungen.

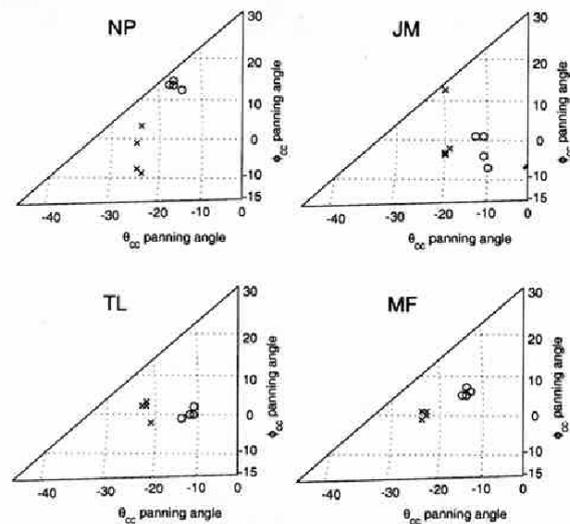


Abb.40: Persönliche Ergebnisse für vier ausgewählte repräsentative Testpersonen. $x=(-20^\circ, 0^\circ)$; $o=(-10^\circ, 0^\circ)$. [18]

4.2.5 Schluss und Erkenntnisse

Durch die durchgeführten Versuche kommt Pulkki zu dem Schluss, dass das ϕ_{cc} -Empfinden in der Medianebene, bei VBAP, eine höchst individuelle Eigenschaft der Probanden darstellt, da es große Streuungen bei den Experimenten zwischen den Leistungen Einzelner gab. Die Elevationsempfindung stieg jedoch monoton mit dem eingestellten ϕ_{cc} -Winkeln.

Wenn der Flächenmittelpunkt, der durch die Lautsprecher aufgespannten Fläche, nahe an der Medianebene liegt stimmen die Werte für ILDA und ITDA annähernd mit der gepannten θ_{cc} -Richtung überein. Diese Übereinstimmungen verschlechtern sich je mehr man sich von der Medianebene bzgl. des Azimuts entfernt.

Weiters wurde von Pulkki festgestellt, dass sich seine Ergebnisse mit denen von Middlebrooks (4.1) bzgl. des Einflusses der spektralen Wellenform der Testsignale auf die Lokalisation, nicht übereinstimmen. Daraufhin wurde von ihm eine alternative Hypothese, welche durch die Betrachtung der Pinna-Modenfrequenzen entwickelt wurde, aufgestellt.

5. Versuchsdesign und Durchführung

5.1 Ziel

Aufgrund der bisher kennen gelernten Wiedergabemethoden welche für die Abbildung von Phantomschallquellen in der Elevation geeignet sind wurde ein Hörversuch designt welcher die unterschiedlichen Qualitäten, bezogen auf den IEM CUBE, eben dieser Methoden untersuchen sollte.

Speziell wurde hierbei auf das VBAP und Ambisonic als Panningverfahren eingegangen.

5.2 Methode

Ausgehend von den Hörversuchen aus Kapitel 4 von Middlebrooks bzw. Ville Pulkki wurden die die Eigenschaften dieser beiden Hörtests zur Entwicklung eines neuen Hörversuchs herangezogen. Zur Messwerterfassung wurde, in Anlehnung an die Messung von J. Middlebrooks, der am Institut vorhandene Headtracker „Flock of Birds (FOB)“ von der Firma „AscensionTechnology“ verwendet welcher eine allgemeine Lösung in der Erfassung von Bewegung darstellt. Mit einer maximalen Reichweite von bis zu zwei Metern bietet der Flock of Birds einen großen Bewegungsspielraum vor der Projektionsfläche. Die Gleichstromtechnik dieses Systems sichert auch eine recht große Stabilität gegenüber Störungen des Magnetfeldes.

Weitere Eigenschaften des Head-Tracking-Systems:

- Erfassung von 6 Freiheitsgraden
- 20 bis 144 Messungen pro Sekunde
- Serielle Datenübertragung über RS-232C

Der Sensor des FOB wurde bei der Versuchsdurchführung durch einen handelsüblichen Haarreif am Kopf der Probanden befestigt. Die Sendeeinheit wurde in einem geeigneten Abstand auf einem Holzgestell positioniert.

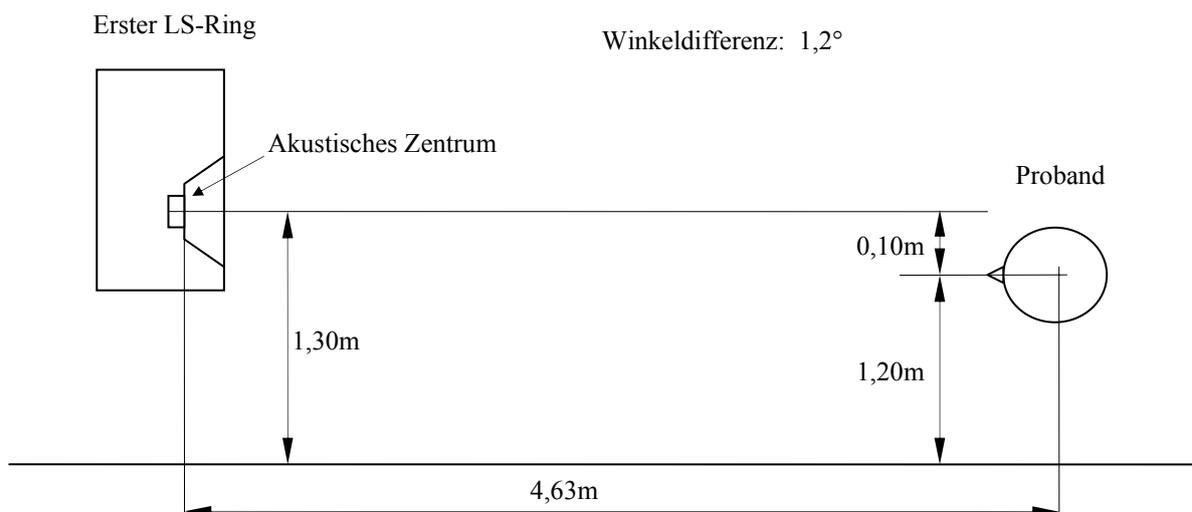
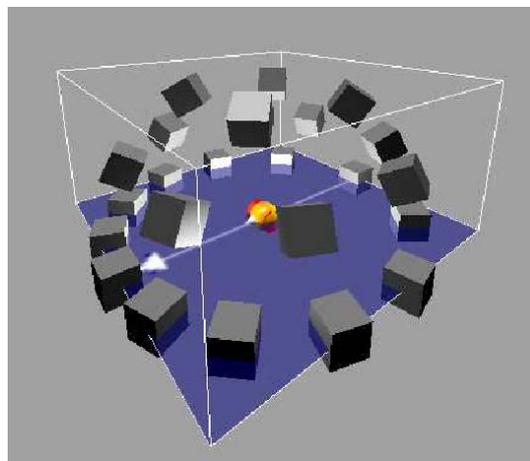
Im Laufe des Versuchs wurden pro Testperson eine Anzahl von 160 Panning - Positionen mit jeder der beiden Wiedergabemethoden wiedergegeben und die empfundene Schallereignisrichtung durch den FOB erfasst. Die dargebotenen Positionen erstreckten sich dabei in einem Elevationsbereich von 0° - 90° (10° - Schritte, mit jeweils zehn Wiederholungen bei 0° Azimut) und einem Azimutbereich von $\pm 15^{\circ}$ (15° - Schritte mit jeweils drei Whl. bei $\pm 15^{\circ}$ je Elevationsstufe). Das bedeutet dass für das VBAP als Wiedergabeverfahren nur die Lautsprecher im benötigten Bereich angesteuert wurden. Für Ambisonic als Panningverfahren wurde das für den Cube bereits implementierte Ambisonic 5ter – Ordnung verwendet.

5.2.1 Wiedergabeanordnung

Als Wiedergabeumgebung wurde der IEM CUBE gewählt. Insgesamt befinden sich im Cube 24 Lautsprecher welche in drei Ringen angeordnet sind, wobei der unterste Ring aus zwölf Lautsprechern besteht, um eine maximale Lokalisationsschärfe in der Horizontalebene zu

erreichen. Der mittlere Ring besteht aus acht Lautsprechern, und die restlichen vier Lautsprecher bilden den obersten Ring. Eine Schematische Darstellung der Lautsprecheranordnung ist in *Abb.41* dargestellt. Die Positionen der Lautsprecher sind folgender Tabelle angegeben:

Ring	Lautsprechernummer											
I	1	2	3	4	5	6	7	8	9	10	11	12
φ	0°	22°	45°	71,5°	102°	138,5°	180°	221°	258°	288°	315°	338°
θ	0°	0°	0°	0°	0°	0°	0°	0°	0°	0°	0°	0°
II	13	14	15	16	17	18	19	20				
φ	25°	69°	116°	155°	205°	244°	296°	335°				
θ	28°	28°	28°	28°	28°	28°	28°	28°				
III	21	22	23	24								
φ	45°	135°	225°	315°								
θ	59,5°	57°	57°	59,5°								



*Abb.41: oben: Schematische Darstellung der Lautsprecheranordnung im IEM CUBE [25].
unten: Position der Testperson bzgl. der Wiedergabeanordnung.*

5.2.2 Ablaufsteuerung:

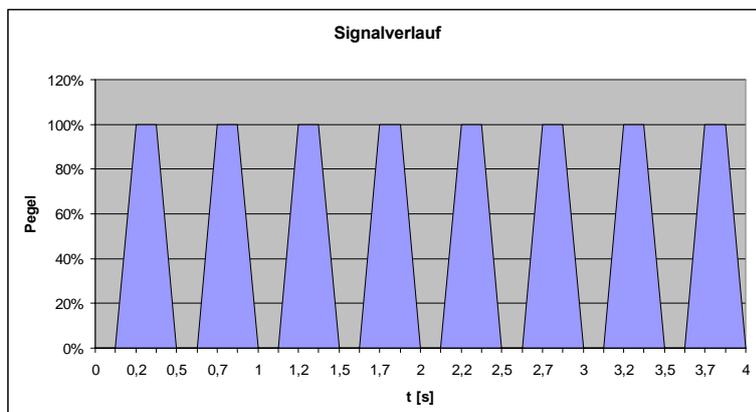
Die Ablaufsteuerung der Messung wurde mithilfe des Programms Pure Data (kurz PD) umgesetzt. PD ist eine grafische real-time (in Echtzeit ablaufende) Programmier-Umgebung für Audio- und Grafik-Verarbeitung. Es ähnelt dem Max/MSP-System, ist aber viel einfacher und portabler. Mithilfe von PD wurde das Durchlaufen und Abspielen einer zuvor generierten Zufallsliste von Panningpositionen realisiert. Zudem wurde die Messwertaufnahme anhand des FOB direkt im Programm implementiert, indem der Headtracker über die RS232 – Schnittstelle direkt gesteuert wurde. Hierbei war für jede Wiedergabe- bzw. Panningmethode ein eigenes Programm erforderlich. Die Steuerung der Patches konnte wahlweise über die Maus oder die Tastatur vorgenommen werden.

Die Testpersonen wurden zu Beginn des jeweiligen Versuchs dazu aufgefordert ihre Ruheposition also die Nullposition einzunehmen, und diese Position per Tastendruck zu bestätigen um den Headtracker zu kalibrieren und somit einen Bezugspunkt für die folgende Messung setzen. Anschließend wurde durch wechselweises betätigen der Wiedergabe bzw. Bestätigungstaste das entsprechende „Soundsample“ wiedergegeben bzw. nach Bestätigung durch die Testperson in einen Puffer geschrieben. Zur Erfassung der empfundenen Panningrichtung wurde die Testperson dazu aufgefordert ihre Nasenspitze auf die gehörte Position zu richten.

5.3 Versuchsanordnung

Testsignal:

Als Testsignal wurde aus den Erfahrungen welche von Pulkki bzw. Middlebrooks gemacht wurden gepulstes Rosa Rauschen gewählt. Dadurch wird auch gewährleistet dass die erhaltenen Ergebnisse mit jenen der besprochenen Hörversuche verglichen werden können. Das Audiosignal wurde mit der Dauer von 4 Sekunden erzeugt [23]. Der Signalverlauf bzw. die Signaleigenschaften sind in *Abb.42* dargestellt.



Signalart	Mono
Min Sample Value:	-25705
Max Sample Value:	21641
Peak Amplitude:	-2.04 dB
DC Offset:	-582
Minimum RMS Power:	-44.85 dB
Maximum RMS Power:	-12.95 dB
Average RMS Power:	-19.4 dB
Total RMS Power:	-17.96 dB
Actual Bit Depth:	16 Bits

Abb.42: Testsignal: Eigenschaften und Verlauf (Hüllkurve) (Full Scale).

Versuchssetup - Blockschaltbild:

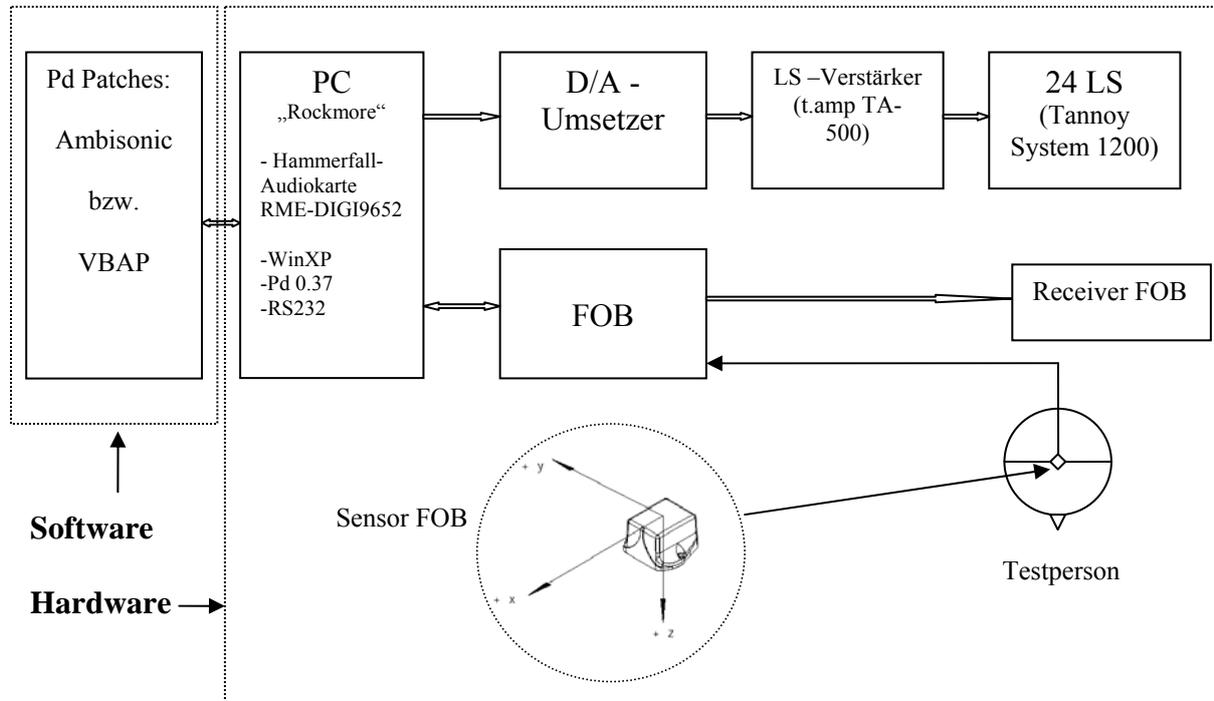


Abb.43: Blockschaltbild der Versuchsanordnung.

5.4 Hörversuchergebnisse:

5.4.1 Statistische Auswertung

Die Statistische Auswertung wurde mithilfe von Microsoft Excel 2003 und Matlab v6.5 durchgeführt. Die angezeigte horizontale Richtung φ sowie die vertikale Richtung θ wurde getrennt ausgewertet, allerdings mithilfe derselben statistischen Kriterien, wie folgt beschrieben. Um die Qualität der Lokalisation eines Wiedergabesystems quantitativ zu erfassen, entwickelte Hartmann 1983 [24] einige statistische Fehlergrößen, die auch in dieser Arbeit (an die Problemstellung angepasst) verwendet werden. Diese Fehlergrößen geben Aufschluss über den Lokalisationsfehler sowie auch die Lokalisationsunschärfe.

In einem Lokalisationsexperiment sind für eine Versuchsperson folgende Größen gegeben:

Anzahl der Richtungen:	L ,
Anzahl der Darbietungen pro Richtung:	M ,
Richtung der Quelle für Darbietung i :	S_i ,

Angezeigte Richtung für Darbietung i : R_i ,
 Fehler für Darbietung i : $e_i=R_i-S_i$

Laufvariable für die Richtungen bzw. Elevationsstufen k : $k \in \{1,2,\dots,L\}$

Es können folgende statistische Größen aus den gemessenen Daten errechnet werden:

Mittlere angezeigte Richtung (je Richtung):
$$\bar{R}(k) = \frac{1}{M} \sum_{i=1}^M R_i(k) \quad (30)$$

Mittlerer Lokalisationsfehler (je Richtung):
$$E(k) = \frac{1}{M} \sum_{i=1}^M e_i(k) \quad (31)$$

Standardabweichung:
$$s(k) = \sqrt{\frac{1}{M-1} \sum_{i=1}^M [R_i(k) - \bar{R}(k)]^2} \quad (32)$$

Mittlerer Lokalisationsfehler (je Testperson):
$$\bar{E} = \frac{1}{L} \sum_{k=1}^L E(k) \quad (33)$$

Da speziell die die Elevationslokalisierung betrachtet werden sollte wurde ausgehend von zahlreichen ANOVA – Analysen (Analysis Of Variance, Vergleich mehrerer Mittelwerte simultan auf Signifikanz ausgehend von Vergleich der Varianzen) eine Mittelung von Versuchsergebnissen gleicher Elevationsstufe (bei einem Signifikanzniveau von 5%) durchgeführt. Generell zeigten die durchgeführten Analysen, dass es keine bemerkenswerten signifikanten Unterschiede der Lokalisationsgenauigkeiten bzgl. der Elevation bei unterschiedlichen Azimutwinkeln gab.

5.4.2 Ergebnisse der einzelnen Testpersonen:

Bei allen folgenden Diagrammen ist ausschließlich die Medianebene (0° - Azimut) dargestellt.

Testperson „GE“:

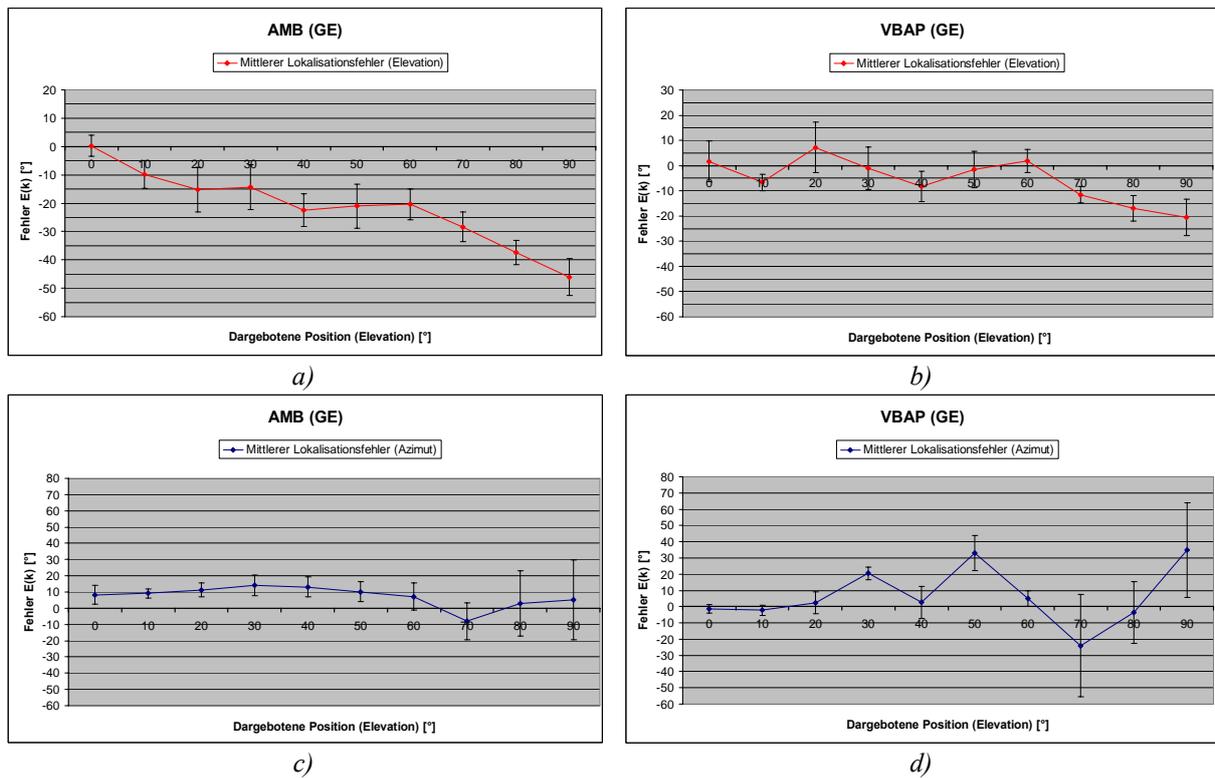
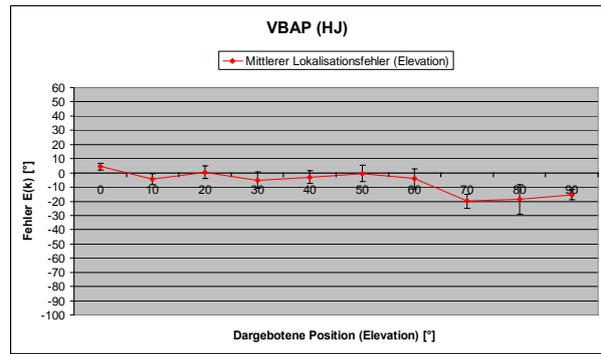
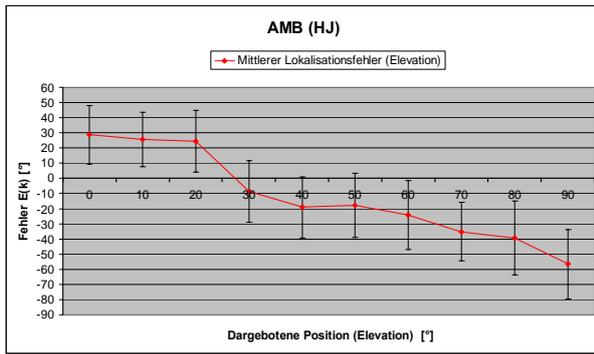


Abb.44: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

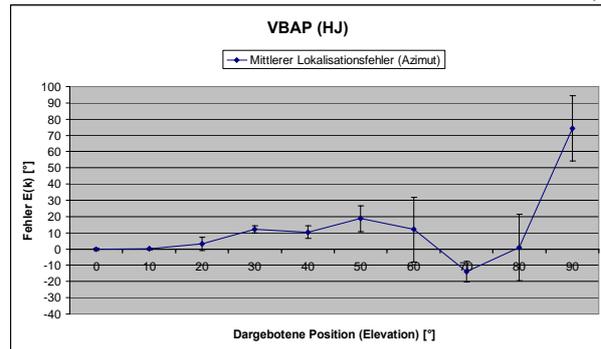
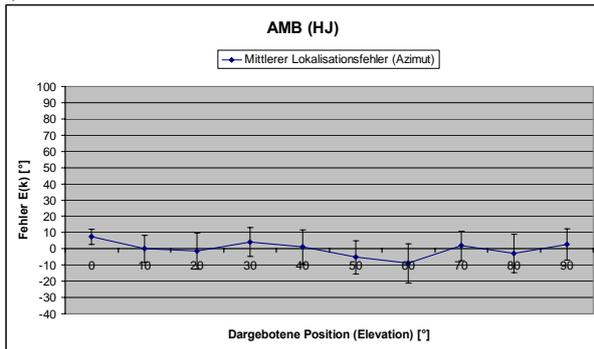
Die Testperson „GE“ erzielt im Bereich zwischen 0° - 60° einigermaßen homogene Ergebnisse bzgl. Elevation und Azimut bei beiden Wiedergabe mit beiden Panningvarianten. Der mittlere Lokalisationsfehler war aber augenscheinlich im angesprochenen Abschnitt bei VBAP (Fehler im Bereich von $-8,3^\circ$ bis $+7,2^\circ$) geringer im Vergleich zu Ambisonic (Fehler im Bereich von $-22,4^\circ$ bis $+0,2^\circ$). Speziell bei hohen Elevationsstufen zeigte sich ein markanter unterschied: Maximum bei VBAP = $-20,6^\circ$ (bei 90°) im Vergleich zu -46° (bei 90°) bei Ambisonic. Im Azimutbereich lieferte Ambisonic speziell bei hohen Elevationsstufen bessere Ergebnisse als VBAP.

Testperson „HJ“:



a)

b)

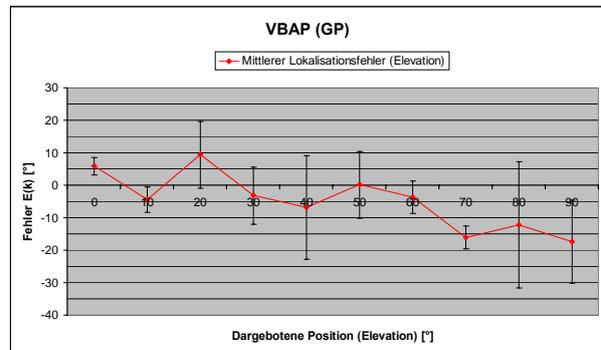
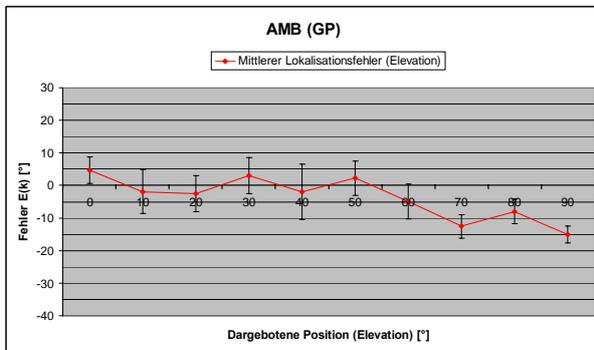


c)

d)

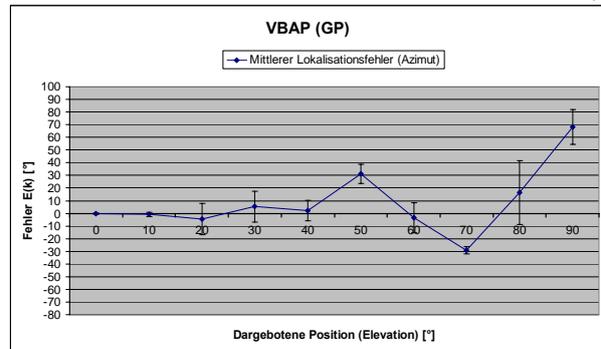
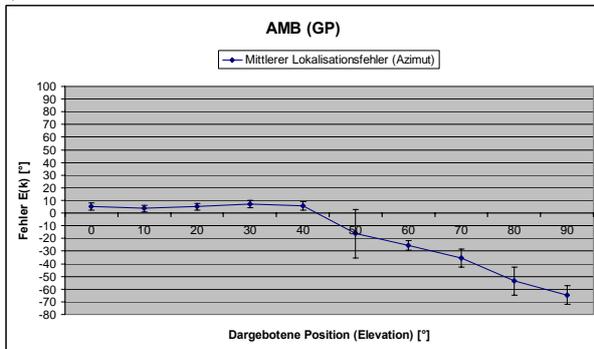
Abb.45: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

Testperson „GP“:



a)

b)

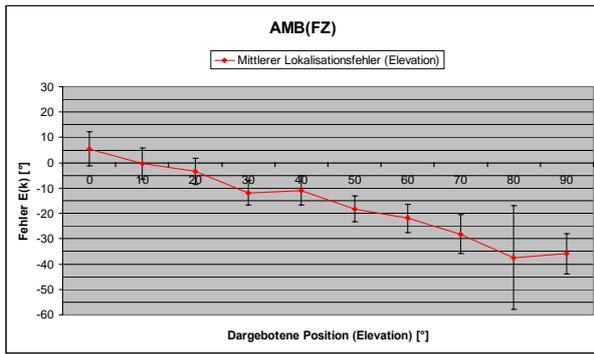


c)

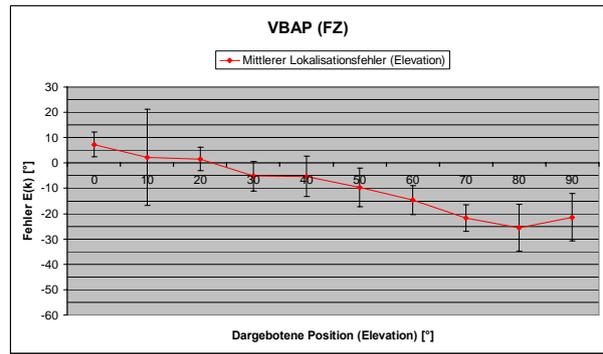
d)

Abb.46: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

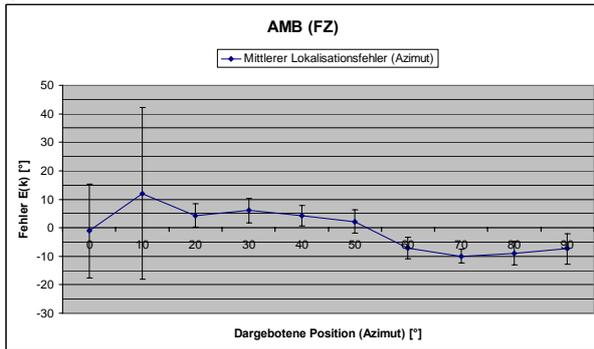
Testperson „FZ“:



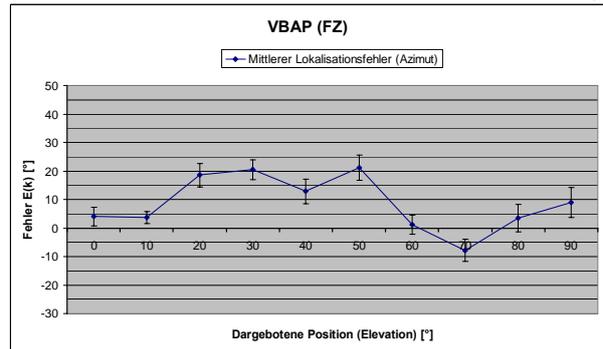
a)



b)



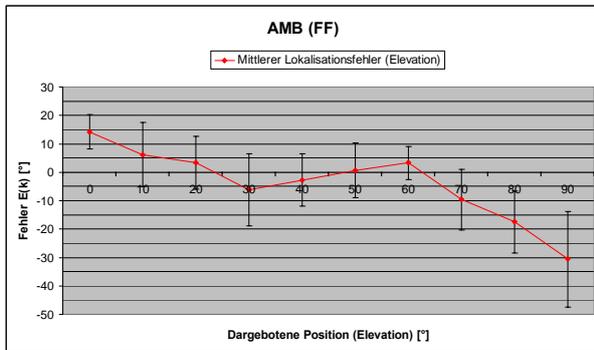
c)



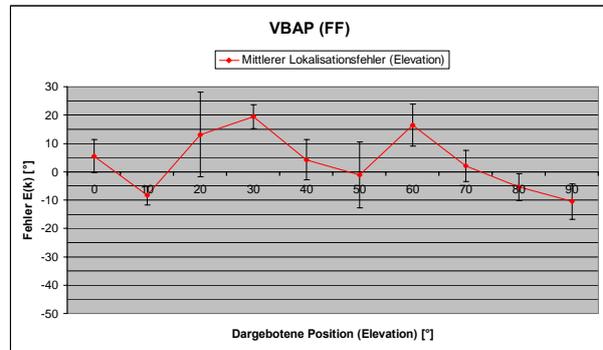
d)

Abb.47: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP

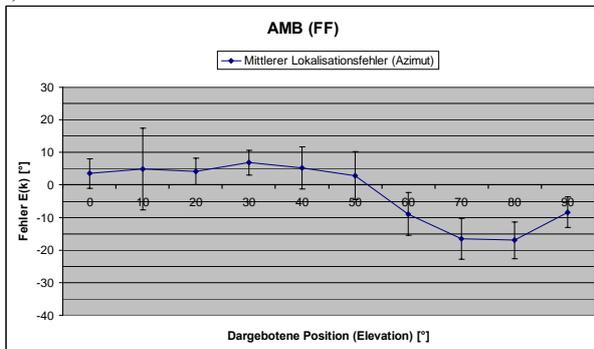
Testperson „FF“:



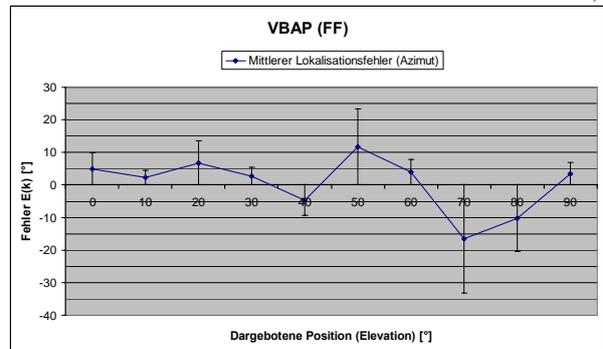
a)



b)



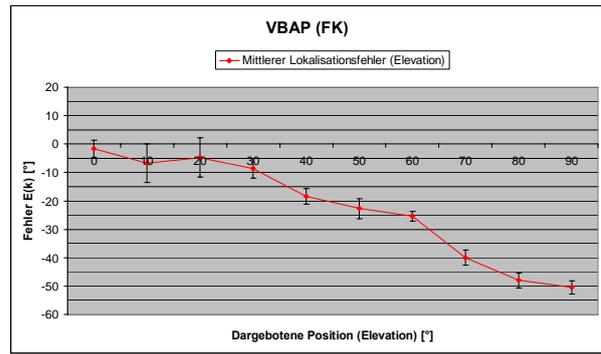
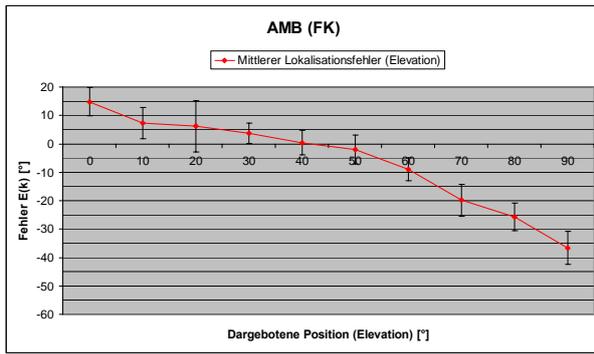
c)



d)

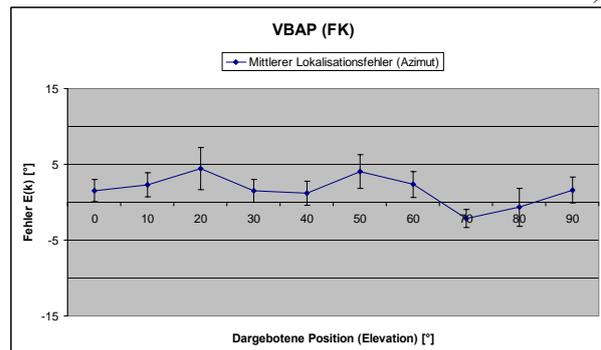
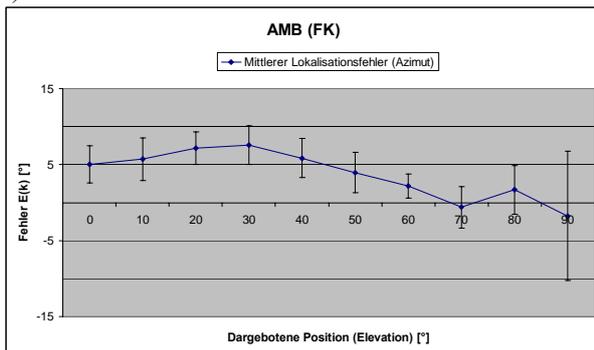
Abb.48: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

Testperson „FK“:



a)

b)

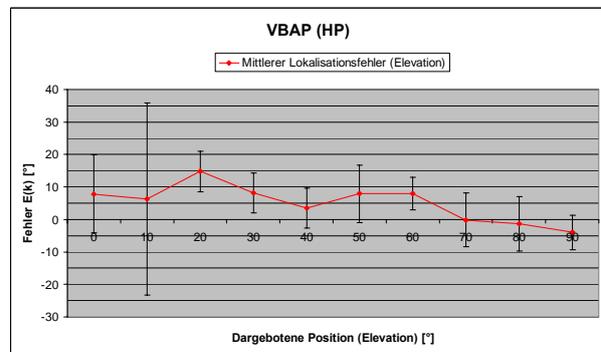
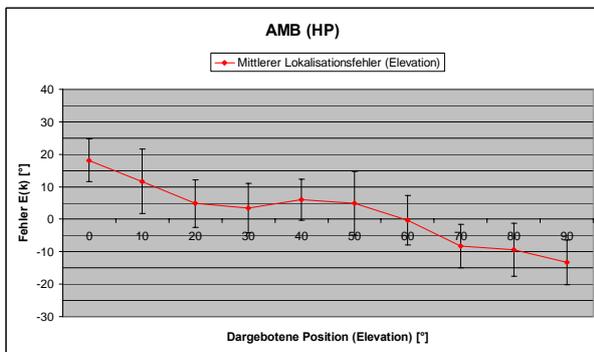


c)

d)

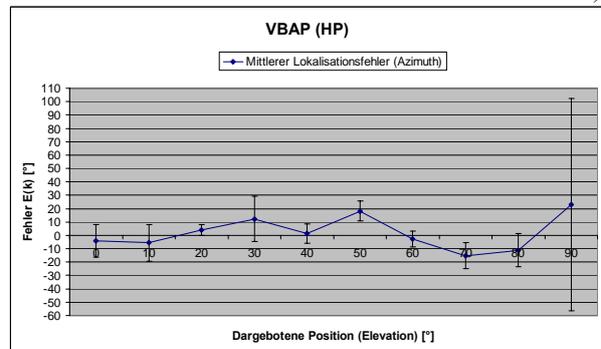
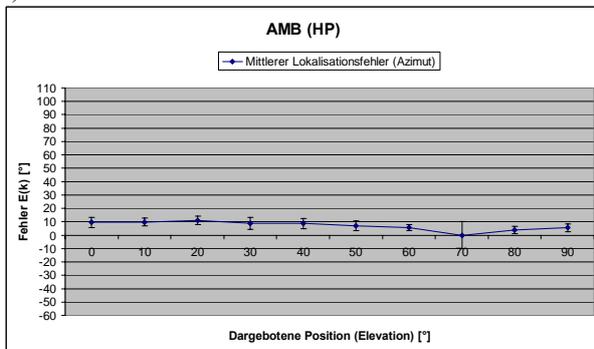
Abb.49: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

Testperson „HP“:



a)

b)



c)

d)

Abb.50: Testergebnis TP „GE“: a) Mittlerer Lokalisationsfehler Elevation bei Ambisonic. b) Mittlerer Lokalisationsfehler Elevation bei VBAP. c) Mittlerer Lokalisationsfehler Azimut bei Ambisonic. c) Mittlerer Lokalisationsfehler Azimut bei VBAP.

Eine andere recht anschauliche Darstellungsvariante kann mit Matlab umgesetzt werden (Hier nur für die Testperson mit im Mittel kleinsten Lokalisationsfehler über beide Wiedergabesysteme):

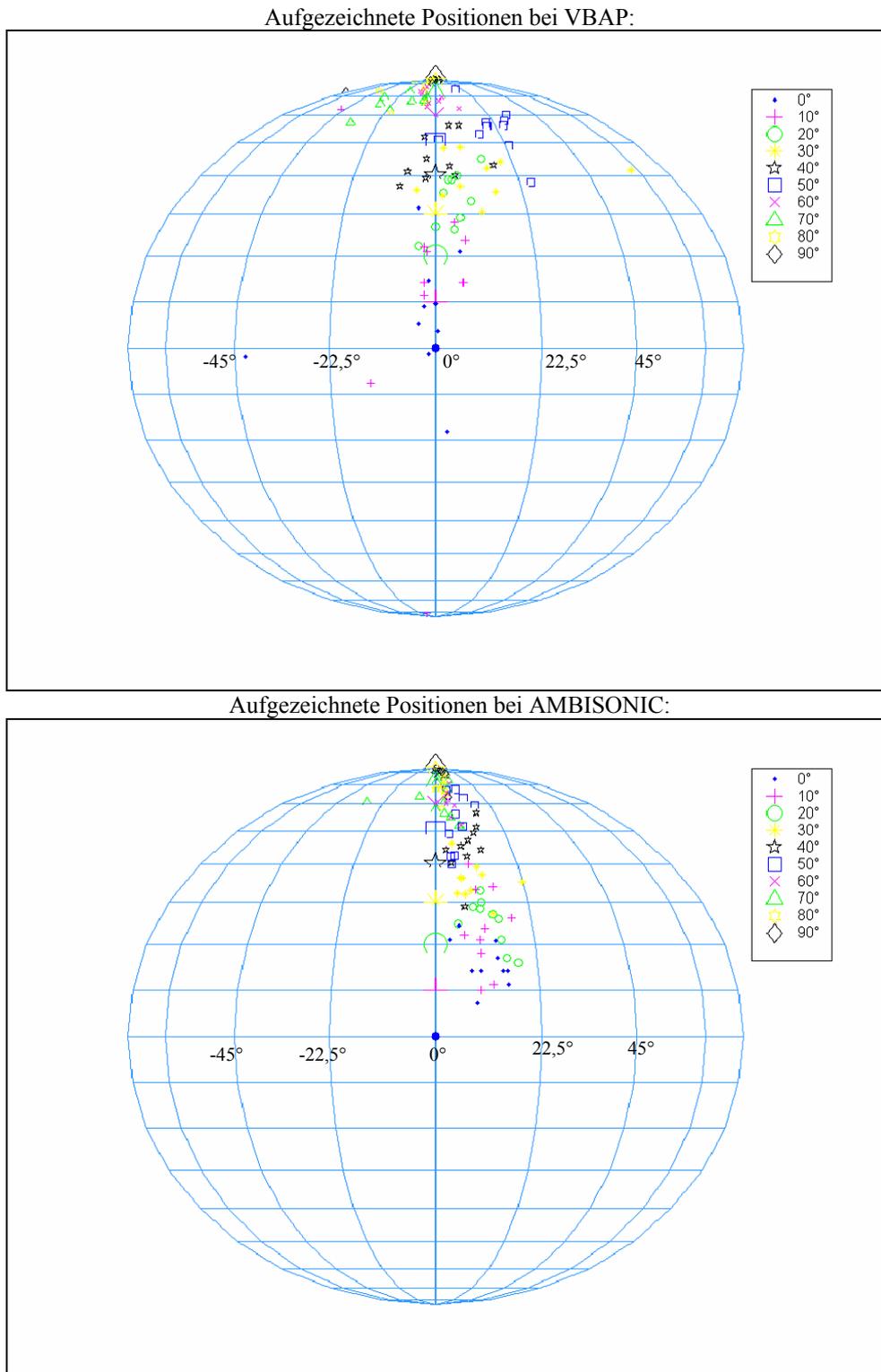


Abb.51: Abbildung der aufgenommenen Panningpositionen auf eine Sphäre.

5.4.3 Ergebnisse aller Testpersonen:

Es zeigten sich im Laufe des Versuchs signifikante Unterschiede zwischen den einzelnen Testpersonen. Durch Anwendung der ANOVA-Analyse (Signifikanzniveau 5%) wurde diese Annahme bestätigt. Besonders augenscheinlich war diese Feststellung bei Signifikanztests bzgl. der Ergebnisse die Ambisonic beim Versuch lieferte. Alle lokalisierten Positionen wichen zwischen den Testpersonen beträchtlich voneinander ab. Bei VBAP als Wiedergabemethode konnten im Vergleich dazu zwischen drei Testpersonen eine ähnliche Performance festgestellt werden.

Auffällig ist zudem, dass bei hohen Elevationsstufen durchwegs eine Unterschätzung der wahren Panningposition stattfand (ähnlich den Ergebnissen von J. Middlebrooks aus Kapitel 4). Dies ist jedoch nicht auf die falsche bzw. schlechte Lokalisationsfähigkeit der Probanden zurückzuführen sondern ist bedingt durch die Art der Messwerterfassung. Durch die Verwendung des FOB hat man bei der Aufzeichnung der Messdaten zwei entscheidende Nachteile:

- zum einen die Einschränkung, dass die Blickrichtung durch die Augen und die Ausrichtung der Nasenspitze je nach Augenstellung (primäres Orientierungsmittel der Testpersonen) mehr oder weniger stark variieren.
- zum andern sind der Richtungsangabe durch die Kopfdrehung, durch den körperlich bedingten beschränkten Bewegungsspielraum des Nackens der jeweiligen Testperson, Grenzen gesetzt.

Über alle Testpersonen ist tendenziell eine Kompression der Elevationswerte festzustellen (Überschätzung bei niedrigen und Unterschätzung bei hohen Elevationsstufen) Generell ist erkennbar, dass die Lokalisationsfähigkeit in der Elevation bei VBAP überwiegend bessere Ergebnisse liefert als das bei Ambisonic der Fall ist. Im Azimutbereich jedoch ist eine gegenläufige Tendenz feststellbar. Interessant ist dass, was die Lateralisation betrifft, bei Ambisonic augenscheinlich zutreffendere Positionsbestimmungen der Probanden möglich sind.

Obwohl die Mittelung der Lokalisationsfehler über eine Testperson aus statistischen Maßstäben (laut ANOVA) nicht zulässig ist, lässt folgendes Diagramm aus *Abb.52* trotzdem einen groben Überblick über die Performance der Probanden bzgl. der beiden Wiedergabemethoden zu.

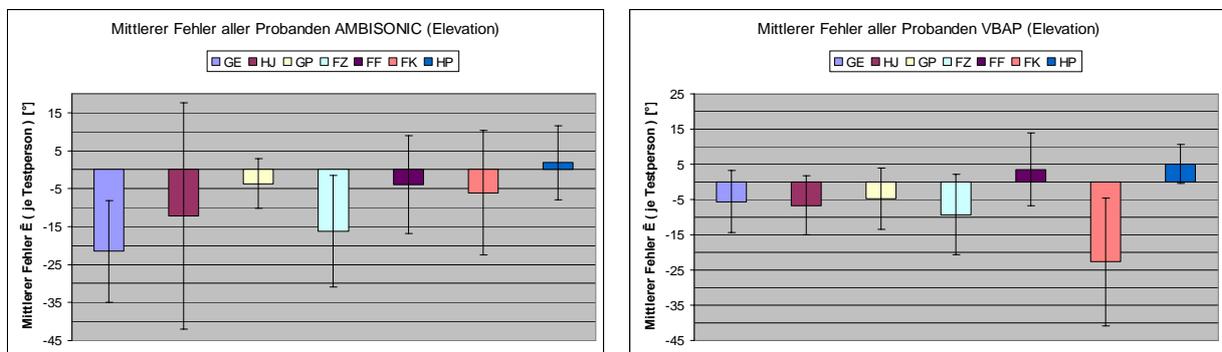


Abb.52: Mittlerer Lokalisationsfehler aller Testpersonen bei a) Ambisonic und b) VBAP

6. Ausblick:

Durch verschiedene konzeptionelle bzw. technische Verbesserungen beim Versuchs – Aufbau oder bei dessen Durchführung hätten mehr Informationen aus den Testergebnissen gezogen werden können. Leider wurden im Lauf der Planung und der Durchführung verschiedene kleine Fehler begangen welche in ihrer Summe eine mehr oder weniger große Auswirkung auf das Versuchsergebnis nach sich zogen. Bei einer erneuten Durchführung des Versuchs sollten deshalb diese Fehler möglichst vermieden werden.

Hierbei wäre primär die Messwerterfassung als Unsicherheitsfaktor anzusprechen. Wie in Punkt 5.4.3 bereits erwähnt ist der aus physiologischen Gegebenheiten eingeschränkte Bewegungsspielraum des menschlichen Nackens für eine Kompression der Messwerte bzgl. der Elevation verantwortlich. Eine Alternative Methode um diese Problem zu beheben wäre z.B. ein um einen festen Punkt drehbar gelagerter Laserpointer welcher mit dem Sensor des FOB versehen wird. Auf diese Art und Weise könnten auch die große Elevationsstufen (70°-90°) in derselben Qualität wie jene in niedrigeren Bereichen abgebildet werden.

Auch die Sitzposition der Testperson sollte nach Möglichkeit auf dieselbe Höhe wie das akustische Zentrum der Lautsprecher des ersten Ringes in der Wiedergabeordnung gebracht werden. Die in diesem Versuch gegebene Differenz von 10 cm hatte zwar nur einen relativ kleinen Winkelfehler von 1,2° zur Folge, dieser Fehler wäre jedoch auf sehr einfache Art und Weise vermeidbar gewesen.

Um die erhaltenen Ergebnisse nicht nur auf den IEM Cube beziehen zu können wäre ein Versuch bei optimalen Bedingungen, d.h. bei identischen Lautsprecherabständen (im Cube leider nicht der Fall → Differenzen mussten durch Pegel- und Laufzeitkorrekturen kompensiert werden) und bei optimaler Rasterung der Lautsprecher, nötig. Ein dazu prädestiniertes Messgestell ist bereits am IEM vorhanden und kann für diesen Zweck eingesetzt werden.

7. Quellenverzeichnis:

- [1] Hee Chong Lee, Hyun Bin Kim, Myung Jin Lee, Poong Min Kim, Sang Won Suh, Ki Hong Kim. *Development of 3D Sound Generation System for Multimedia Application*. 120-124. Third Asian Pacific Computer and Human Interaction, July 15-17, 1998, Kangawa, Japan, Proceedings. IEEE Computer Society 1998
- [2] J-M. Jot, V. Larcher, and O. Warusfel. Digital signal processing issues in the context of binaural and transaural stereophony. In *Presented at the 98th AES convention. An Audio Engineering Society preprint 3980 (16)*, Paris, February 1995.
- [3] W. Gardner. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, Boston, 1998.
- [4] V.Larcher, J.M. Jot, G. Guyard, and O.Warusfel. *Study and comparison of efficient methods for 3d audio spatialization based on linear decomposition of HRTF data*. Proc. 108th Audio Engineering Society Convention, 2000.
- [5] Durand R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Ames Research Center, Moffett Field, California, April 2000
- [6] Brown, C.P.; Duda, R.O.; *An efficient HRTF model for 3-D sound*. Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on , 19-22 Oct. 1997 Pages:4 pp
- [7] Ville Pulkki; *Virtual Sound Source Positioning Using Vector Base Amplitude Panning*; Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology; *J.AudioEng.Soc., Vol.45, No.6, 1997* Juni
- [8] V. Pulkki and Tapio Lokki. *Creating auditory displays to multiple loudspeakers using vbap: A case study with diva project*. In International Conference on Auditory Display, Glasgow, Scotland, 1998. ICAD.
- [9] M.A. Gerzon; *Periphony: With-Height Sound Reproduction*. J. Audio Eng. Soc., vol. 21, pp. 2-10 (1973 Jan./Feb.)
- [10] Geiger, H., Scheel, K.: *Handbuch der Physik, Band VIII*. Verlag von Julius Springer, Berlin (1926)
- [11] Boone, M.: *Acoustic rendering with wave field synthesis*. ACM SIGGRAPH AND EUROGRAPHICS CAMPFIRE: Acoustic Rendering for Virtual Environments; Snowbird, Utah, May 26–29, 2001
- [12] John C. Middlebrooks; *Spectral Shape Cues for sound Localization*. Binaural and Spatial Hearing in Real and Virtual Environments. R. Gilkey and T. Anderson, Erlbaum: 77-98. NY, 1997
- [13] Butler R.A., Flannery R.; *The spacial attributes of stimulus frequency and their role in monaural localization of sound in the horizontal plane*. Percept. Psychophys.28, 449-457, 1980
- [14] Middlebrooks, J.C.; *Narrow-band sound localisation related to external ear acoustics*. J.Acoust. Soc., Am 92, 2607-2624, 1992;
- [15] Fisher H.G., Freedman S.J.; *The role of the pinna in auditory localization*; J. Aud. res. 8, 15-26; 1968
- [16] Oldfield S.R., Parker P.A.; *Acuity of sound localization: a topography of auditory space. III. Monaural hearing conditions*. Perception 15, 67-81, 1986
- [17] Butler R.A., Humanski R.A., Musicant A.D.; *Binaural and monaural localization of sound in two-dimensional space*. Perception, 241-256, 1990
- [18] Ville Pulkki, M. Karjalainen; *Localization of amplitude-panned virtual sources II: two- and three-dimensional panning*. J. Audio Eng. Soc. Vol. 49 No.9. September 2001

- [19] Jérôme Daniell, Rozenn Nicol, Sébastien Moreau; *Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*. Convention Paper 5788, 114th Convention 2003 March 22–25 Amsterdam, Netherlands
- [20] Gardner, B. & Martin, K. *HRTF measurements of a KEMAR dummy-head microphone*. MIT Media Lab Perceptual Computing Technical Report #280, (1994).
- [21] A.J. Berkhout, *A Holographic Approach to Acoustic Control*. J. Audio Eng. Soc., pp. 977-995, 1988
- [22] A.J. Berkhout, D.d. Vries, and P. Vogel, *Acoustic Control by Wave Field Synthesis*. J. Acoust. Soc. Am., vol. 93, pp. 2764-2778, 1993
- [23] Fang Chen, *Localization of 3D-Sound Presented through Headphone – Duration of Sound Presentation and Localization Accuracy*. J. Audio Eng. Soc., Vol. 51, No. 12, Dec 2003
- [24] Hartmann, W. M.: *Localization of sound in rooms*. J. Acoust. Soc. Am. 74, 1380-1391 (1983)
- [25] Johannes M. Zmoelnig, Winfried, Ritsch, Alois Sontacchi: *the IEM-Cube*. Institute for Electronic Music and Acoustics, University of Music and Dramatic Arts, Graz, Austria.