

Stefan LEITNER, Alois SONTACCHI, Robert HÖLDRICH

Institut für Elektronische Musik und Akustik

Universität für Musik und Darstellende Kunst Graz

KOPFPOSITIONSBEZOGENE WIEDERGABE FÜR BINAURALSIGNALE MIT AMBISONIC

HEAD POSITION RELATED BINAURAL SOUND REPRODUCTION - THE AMBISONIC APPROACH

Zusammenfassung

Um für Kopfhörerwiedergabe Monoquellen zu spatialisieren oder die originalgetreue Wiedergabe eines aufgezeichneten Schallfeldes zu erreichen, ist die Auswirkung der Aussenohrübertragungsfunktionen (head related transfer function, HRTF) zu berücksichtigen. Ein weiterer psychoakustischer Effekt ist die Verbesserung der Lokalisation durch kleine Kopfbewegungen des Hörers. Im Folgenden wird ein Ansatz vorgestellt, mit dem es möglich ist kopfpositionsbezogene Binauralsignale recheneffizient in Echtzeit zu erhalten. Zur Anwendung kommt die Rekonstruktion eines Schallfeldes mit dem mathematischen Modell von Ambisonic.

1 Spatialisation von Monoquellen

Richtungshören im natürlichen Schallfeld zeigt, dass Schalle aus verschiedenen Richtungen mit unterschiedlichen Pegeln, Laufzeiten und Frequenzgewichtungen auf die Trommelfelle beider Ohren treffen. Mathematisch drücken sich diese Signalunterschiede in den Aussenohrübertragungsfunktionen (HRTF) aus [1,2]. Die Spatialisation für Kopfhörerwiedergabe erfordert daher die Filterung mit diesen richtungsspezifischen HRTF's.

Es besteht die Möglichkeit individuelle HRTF's eines Hörers, die Mittelung von HRTF's vieler Hörer, mit einem mathematischen Modell berechnete oder an einem Kunstkopf gemessene Aussenohrübertragungsfunktionen zu verwenden. Letztere kommen in dieser Implementation zur Anwendung.

2 Auswirkung von Kopfbewegungen auf das räumliche Hören

Kopfbewegungen führen beim natürlichen Hören zu einer deutlichen Verbesserung der Lokalisation.[3].

Um Informationen aus den Kopfbewegungen bei Kopfhörerwiedergabe verfügbar zu machen, muss eine im Raum statische Schallquelle mit unterschiedlichen HRTF's in Abhängigkeit der Kopfposition gefiltert werden (Abbildung 1). Wird ein dynamisches System (Berücksichtigung der Kopfbewegung mittels Head Tracker) konstruiert, führt das zum Problem der Interpolation einer grossen Anzahl von Impulsantworten.

Umgangen wird das Problem durch Berechnung statischer Filter, aber kopfpositionsbezogener Veränderung der Eingangssignale durch Anwendung der Ambisonic-Technik.

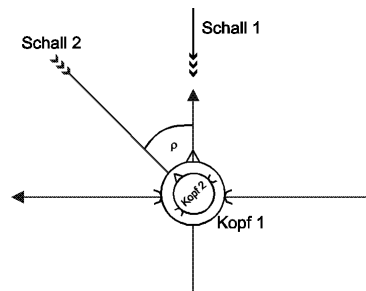


Abbildung 1: Auswirkungen einer Kopfdrehung auf die Filtercharakteristik des Ohres. Bewegt sich der Kopf nach links, ändern sich die Signalunterschiede und damit die zu berechnenden Filter an den beiden Ohren.

3 Ambisonic

3.1 Grundlagen und Begriffsdefinition

Ambisonic bietet die Möglichkeit ein Schallfeld dreidimensional über Lautsprecher zu reproduzieren.[5,8,9]. Im Folgenden werden anhand des zweidimensionalen Falls die Grundgleichungen für dieses Aufnahme- und Wiedergabesystem betrachtet [4].

Die Anzahl der benötigten Übertragungs-Kanäle für ein zweidimensionales System ist $(2m+1)$, für ein dreidimensionales $(m+1)^2$, wobei m die Ordnung des Systems ist.

Das Modell basiert auf einem Vergleich einer Referenzwelle mit der Welle, die das Ambisonic-System produziert. Eine Voraussetzung für die folgenden Überlegungen ist, dass die Referenzwelle eine ebene Welle ist. Das gilt für Schallquellen, die ausreichend weit vom Aufnahmepunkt entfernt sind. Die zweite Voraussetzung ist, dass die von den Lautsprechern abgestrahlten Wellen auch ebene Wellen sind. Dazu muss sich der Hörer in ausreichendem Abstand zu den Lautsprechern befinden.

Bei Ambisonic ist es üblich das Koordinatensystem um 90° gedreht darzustellen. Die x -Achse zeigt daher zur Raumvorderseite, die y -Achse zeigt nach links. Die z -Achse für den dreidimensionalen Fall zeigt nach oben. Der Ursprung des Koordinatensystems bildet das Zentrum der Aufnahmeanordnung.

Wir nehmen an, dass die ebene Welle aus einer Richtung ψ in Bezug auf die x -Achse eintrifft. Der Aufpunkt (=Abhörpunkt) befindet sich in einer Entfernung r und Winkel ϕ .

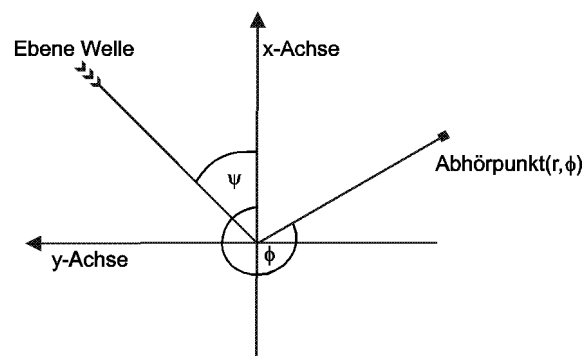


Abbildung 2: Achsen und Winkeldefinition bei Ambisonic.

Eine ebene Welle im Aufpunkt \vec{r} lässt sich ausdrücken durch:

$$P_y e^{i\vec{k}\vec{r}}. \quad (1)$$

P ist der Druck der ebenen Welle, k ist die Wellenzahl $2\pi/\lambda$ und λ die Wellenlänge. Der Vektor \vec{k} repräsentiert eine Welle mit der Wellenzahl k, die sich in eine bestimmte Richtung bewegt. Die Gleichung kann auch geschrieben werden als

$$S_y = P_y e^{ikr \cos(\mathbf{f}-\mathbf{y})}. \quad (2)$$

S_ψ ist hier die originale ebene Welle. Das Ziel des Ambisonic-Systems ist, diese ebene Welle im Zentrum der Hörzone zu reproduzieren.

Gleichung (3) beschreibt den Zusammenhang zwischen einer ebenen Welle und einer Reihenentwicklung aus Cosinus und Zylindrischen Besselfunktionen.

$$S_y = P_y J_0(kr) + 2P_y \sum_{m=1}^{\infty} i^m J_m(kr) \cos m(\mathbf{f}-\mathbf{y}). \quad (3)$$

Der Cosinus Term kann umgeschrieben werden, und es folgt:

$$S_y = P_y \left(J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\mathbf{y}) \cos(m\mathbf{f}) + \sin(m\mathbf{y}) \sin(m\mathbf{f})] \right). \quad (4)$$

Das ursprüngliche Schallfeld soll im ersten Schritt mittels Lautsprechern rekonstruiert werden. Wenn jeder Lautsprecher eine ebene Welle abstrahlt, ist der Output des n-ten Lautsprechers:

$$S_n = P_n \left(J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) [\cos(m\mathbf{f}) \cos(m\mathbf{f}_n) + \sin(m\mathbf{f}) \sin(m\mathbf{f}_n)] \right). \quad (5)$$

ϕ_n ist hier der Winkel des n-ten Lautsprechers. Jeder Lautsprecher trägt zur resultierenden Welle bei, daher ergibt sich die resultierende Welle als Summe der einzelnen Lautsprecherwellen. Die Gesamtwelle ist

$$S = \sum_{n=1}^N P_n J_0(kr) + \sum_{m=1}^{\infty} 2i^m J_m(kr) \left(\sum_{n=1}^N P_n \cos(m\mathbf{f}_n) \cos(m\mathbf{f}) + \sum_{n=1}^N P_n \sin(m\mathbf{f}_n) \sin(m\mathbf{f}) \right). \quad (6)$$

Damit die Lautsprecher die originale Welle nachbilden, müssen folgende Bedingungen eingehalten werden. Der Koeffizientenvergleich mit Gleichung (4) ergibt:

$$P_y = \sum_{n=1}^N P_n \quad (7)$$

$$P_y \cos(m\mathbf{y}) = \sum_{n=1}^N P_n \cos(m\mathbf{f}_n) \quad (8)$$

$$P_y \sin(m\mathbf{y}) = \sum_{n=1}^N P_n \sin(m\mathbf{f}_n) \quad (9)$$

Jede Ordnung m entspricht der sphärischen Harmonischen m -ter Ordnung. Wenn eine unendliche Anzahl an Kanälen zur Verfügung stehen würde, könnte die originale Welle exakt in jedem Punkt \vec{r} reproduziert werden. Eine begrenzte Anzahl an Übertragungskanälen und Lautsprechern bedeutet, dass die Reihe bei einer bestimmten Ordnung m abgebrochen wird.

Die Bedingungen (7) bis (9) beschreiben die Gleichungen, aus denen die Informationen für das Ambisonic-Signal bezogen werden. Bei einem System nullter Ordnung muss nur (7) übereinstimmen. Wenn (7) eingehalten wird, stimmt der Druck der beiden Wellen überein, die originale und reproduzierte Welle haben gleiche Amplitude. Das richtungsunabhängige Signal aus Gleichung (7) heißt W-Signal. Für Systeme erster Ordnung wird zusätzlich die Schnelle im Ursprung in Übereinstimmung gebracht, das heißt, die originale und reproduzierte Welle bewegen sich in die gleiche Richtung. Für $m=1$ ergibt sich aus den Gleichungen (8) und (9) das X und Y Ambisonic-Signal.

$$\begin{aligned} X &= P_y \cos(\mathbf{y}) \\ Y &= P_y \sin(\mathbf{y}) \end{aligned} \quad (10)$$

Diese drei Komponenten bilden ein 2D-Ambisonic-System erster Ordnung. Für $m=2$ in Gleichung (8) und (9) folgen die Ambisonic-Signale U und V.

$$\begin{aligned} U &= P_y \cos(2\mathbf{y}) \\ V &= P_y \sin(2\mathbf{y}) \end{aligned} \quad (11)$$

Ein 2D-System zweiter Ordnung besteht aus folgenden fünf Signalen: W;X;Y;U;V.

Im Folgenden werden die Eingangssignale für die einzelnen Lautsprecher berechnet. Die Frage lautet: Welche Signale P_n müssen die Lautsprecher erhalten damit die Gleichungen (7) bis (9) eingehalten werden. Ein System erster Ordnung berücksichtigt die Signale W,X,Y. Die Lautsprechersignale ergeben sich zu

$$P_n = \frac{1}{N} (W + 2X \cos \mathbf{f}_n + 2Y \sin \mathbf{f}_n). \quad (12)$$

Für ein System zweiter Ordnung ergeben sich zwei zusätzliche Terme:

$$P_n = \frac{1}{N} (W + 2X \cos \mathbf{f}_n + 2Y \sin \mathbf{f}_n + 2U \cos 2\mathbf{f}_n + 2V \sin 2\mathbf{f}_n). \quad (13)$$

mit $(\phi_n, \gamma_n) =$ Azimuth und Elevation des n-ten Lautsprechers.

Die eigentliche Wiedergabe wird nicht mit realen Lautsprechern durchgeführt. Vielmehr wird durch eine Filterung mit Aussenohrübertragungsfunktionen ein binaurales Kopfhörersignal berechnet.

3.2 Berechnung eines binauralen Stereosignals

Betrachtet wird ein 2D-Ambisonic-System 2. Ordnung. Um ein binaurales Stereosignal zu erhalten, ist es notwendig, dass eine einfallende Schallwelle aus einer bestimmten Raumrichtung mit einer entsprechenden Aussenohrübertragungsfunktion gefiltert wird. Als einfallende Schallwellen werden in diesem Fall die Signale der Lautsprecher aus dem Ambisonic-System betrachtet. Jedes Lautsprechersignal dieser virtuellen Lautsprecher wird mit den der Lautsprecherrichtung \mathbf{f}_n entsprechenden Aussenohrübertragungsfunktionen der KEMAR Impulsantworten [7] gefiltert, und man erhält durch Superposition ein Kopfhörersignal (L,R).

$$\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} H_{1,L} & H_{2,L} & \cdots & H_{N,L} \\ H_{1,R} & H_{2,R} & \cdots & H_{N,R} \end{pmatrix} * \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} \quad (14)$$

Die Lautsprechersignale berechnen sich aus den Ambisonic-Signalen und W,X,Y,U und V für ein System 2.Ordnung:

$$\begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_N \end{pmatrix} = \frac{1}{N} \begin{pmatrix} \mathbf{a}_1 & 2\mathbf{b} \cos \mathbf{f}_1 & 2\mathbf{g} \sin \mathbf{f}_1 & 2\mathbf{d} \cos 2\mathbf{f}_1 & 2\mathbf{e} \sin 2\mathbf{f}_1 \\ \mathbf{a}_2 & 2\mathbf{b} \cos \mathbf{f}_2 & 2\mathbf{g} \sin \mathbf{f}_2 & 2\mathbf{d} \cos 2\mathbf{f}_2 & 2\mathbf{e} \sin 2\mathbf{f}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_N & 2\mathbf{b} \cos \mathbf{f}_N & 2\mathbf{g} \sin \mathbf{f}_N & 2\mathbf{d} \cos 2\mathbf{f}_N & 2\mathbf{e} \sin 2\mathbf{f}_N \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix} \quad (15)$$

Die Faktoren α , β , γ , δ und ϵ sind für regelmäßige Lautsprecheranordnungen gleich eins [4].

3.3 Rotation des Schallfeldes

Das Schallfeld, das in den Ambisonic-Signalen W, X, Y, U und V abgebildet ist, kann beliebig gedreht werden; das heißt, der Schall wird bei der Wiedergabe aus einer anderen Richtung wahrgenommen als das originale Schallfeld. Die unrotierten Signale lauten:

$$W = \frac{P_y}{\sqrt{2}} \quad X = P_y \cos(\mathbf{y}) \quad U = P_y \cos(2\mathbf{y}) \\ Y = P_y \sin(\mathbf{y}) \quad V = P_y \sin(2\mathbf{y}) \quad (16)$$

Bei einer Rotation gilt: $\cos(\mathbf{y}) \rightarrow \cos(\mathbf{y} + \mathbf{r})$, $\sin(\mathbf{y}) \rightarrow \sin(\mathbf{y} + \mathbf{r})$, $\cos(2\mathbf{y}) \rightarrow \cos(2(\mathbf{y} + \mathbf{r}))$ und $\sin(2\mathbf{y}) \rightarrow \sin(2(\mathbf{y} + \mathbf{r}))$. Mit den Summensätzen für die Sinus und Cosinusfunktionen folgt für die um einen Winkel \mathbf{r} rotierten Signale W', X', Y', U', V' :

$$\begin{pmatrix} W' \\ X' \\ Y' \\ U' \\ V' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos \mathbf{r} & -\sin \mathbf{r} & 0 & 0 \\ 0 & \sin \mathbf{r} & \cos \mathbf{r} & 0 & 0 \\ 0 & 0 & 0 & \cos 2\mathbf{r} & -\sin 2\mathbf{r} \\ 0 & 0 & 0 & \sin 2\mathbf{r} & \cos 2\mathbf{r} \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix} \quad (17)$$

Aus den Gleichungen (14), (15), und (17) berechnet sich ein binaurales Stereosignal für das rotierte Schallfeld aus den Ambisonic-Signalen W, X, Y, U und V

Gleichung (18):

$$\begin{pmatrix} L \\ R \end{pmatrix} = \begin{pmatrix} H_{1,L} & H_{2,L} & \cdots & H_{N,L} \\ H_{1,R} & H_{2,R} & \cdots & H_{N,R} \end{pmatrix} \cdot \frac{1}{N} \cdot \begin{pmatrix} \mathbf{a}_1 & 2\mathbf{b} \cos f_1 & 2\mathbf{g} \sin f_1 & 2\mathbf{d} \cos 2f_1 & 2\mathbf{e} \sin 2f_1 \\ \mathbf{a}_2 & 2\mathbf{b} \cos f_2 & 2\mathbf{g} \sin f_2 & 2\mathbf{d} \cos 2f_2 & 2\mathbf{e} \sin 2f_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_N & 2\mathbf{b} \cos f_N & 2\mathbf{g} \sin f_N & 2\mathbf{d} \cos 2f_N & 2\mathbf{e} \sin 2f_N \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \cos \mathbf{r} & -\sin \mathbf{r} & 0 & 0 \\ 0 & \sin \mathbf{r} & \cos \mathbf{r} & 0 & 0 \\ 0 & 0 & 0 & \cos 2\mathbf{r} & -\sin 2\mathbf{r} \\ 0 & 0 & 0 & \sin 2\mathbf{r} & \cos 2\mathbf{r} \end{pmatrix} \cdot \begin{pmatrix} W \\ X \\ Y \\ U \\ V \end{pmatrix}$$

Der einzige Teil, der sich in Abhängigkeit von der Kopfposition ändert, ist die Rotationsmatrix für die Ambisonic-Signale. Gefiltert werden die rotierten Ambisonic-Signale, deren Filter statisch und daher leicht zu implementieren sind.

Der Vorteil liegt in der Tatsache, dass die Rotation auf Ebene der Ambisonicsignale und nicht in der HRTF-Domäne (Interpolation der Filter) erfolgt.

Nicht nur die künstliche Spatialisation von Monoquellen, auch die kopfpositionsbezogene Kopfhörerwiedergabe von Aufnahmen wird im Gegensatz zu Kunstkopfaufnahmen möglich. Die Aufnahme muss die Ambisonicsignale liefern. Das erreicht man entweder mit dem Soundfieldmikrophon [10,11] (bis Ambisonic 1.Ordnung) oder mit einer Mikrophongruppe und anschließender Berechnung der Ambisonic-Signale.[8].

4 Implementation

Die Messungen von Gardner und Martin (1994) [7] an einem KEMAR Kunstkopf bilden das Basismaterial für die zur Filterberechnung verwendeten Aussenohrübertragungsfunktionen. Dabei handelt es sich um 512 Punkte Impulsantworten aus 710 Richtungen, die bei Elevationen von -40° bis 90° an unterschiedlich vielen Azimuthstellen pro Elevation gemessen wurden. Um kürzere Filter realisieren zu können, wurden diese Originalimpulsantworten auf 256 gekürzt.

Die Berechnung der Filter erfolgte mit MATLAB 5.1. Das Echtzeitsystem wurde auf einer SGI O2 unter MAX/FTS implementiert. Als Headtracker kam ein Fbck of birds (Ascension) zum Einsatz.

Berechnet werden können Kopfhörersignale für unterschiedliche Anzahl virtueller Lautsprecher und unterschiedliche Ordnung der Ambisonic-Signale.

Blockschaltbild der Versuchsanordnung:

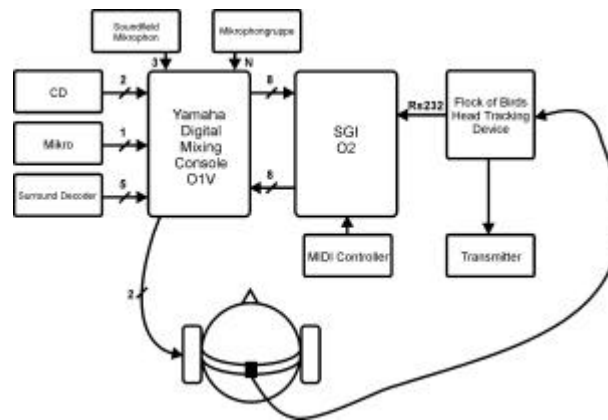


Abbildung 3: Blockdiagramm der Versuchsanordnung zur Berechnung von dynamischen Binauralsignalen für Kopfhörerwiedergabe.

5 Literatur

- [1] Robert H. Gilkey, Timothy R. Anderson: "Binaural and Spatial Hearing in Real and Virtual Environments"; Lawrence Erlbaum Associates, Publishers; Mahwah, New Jersey; 1997.
- [2] Jens Blauert: "Spatial Hearing, The Psychophysics of Human Sound Localization"; The MIT Press; Revised Edition 1997.
- [3] Philip Mackensen, Klaus Reichenauer, Günther Theile: "Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören"; Tagungsband, 20. Tonmeistertagung 1998.
- [4] J.S. Bamford: "An Analysis of Ambisonic Sound Systems of First and Second Order", Thesis presented to the University of Waterloo, Waterloo, Ontario, Canada; 1995.
- [5] David G. Malham, Anthony Myatt: "3-D Sound Spatialization using Ambisonic Techniques"; Computer Music Journal, 19:4, pp.58-70, 1995 Massachusetts Institute of Technology; Winter 1995.
- [6] M. Poletti: "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems"; J. Acoust. Soc. Am., Vol. 44, No. 11; November 1996.
- [7] B. Gardner, K. Martin: "HRTF Measurements of a KEMAR Dummy-Head Microphone"; <http://sound.media.mit.edu/KEMAR.html>; 1994.

- [8] Rozenn Nicol, Marc Emerit: “3D-Sound Reproduction Over An Extensive Listening Area, A Hybrid Method Derived From Holophony And Ambisonic”; AES 16th International Conference, pp.436-453; 1999.
- [9] Rozenn Nicol, Marc Emerit: “Reproducing 3D-Sound for Videoconferencing: a Comparison between Holophony and Ambisonic”; DAFX98 Proceedings pp.17-20; 1998.
- [10] Michael A. Gerzon: “ The Design of Precisley Coincident Microphone Arrays for Stereo and Surround Sound”; Preprint No.20 50th Convention of the Audio Engineering Society, London, March 1975.
- [11] K. Farrah: “The Soundfield Microphone”; Wireless World, pp.99-103; November 1979.