

Soundscapes in der dynamischen Binauralsynthese

Exposé zur Magisterarbeit

Sebastian Roos

12.10.2009

Zusammenfassung

Die virtuelle Darstellung von Klangquellen in modernen Auralisationsumgebungen ist technisch weit fortgeschritten und erreicht ein hohes Mass an Überzeugungskraft. Die Erzeugung virtueller Objekte konzentriert sich dabei vor allem auf die primär interessierenden Klangquellen, wie z.B. den Instrumenten in einer Konzertaufführung. Ein immersives Erlebnis der Aufführungssituation wird jedoch auch durch weitgehend unbewusste Eindrücke anhand der Beschaffenheit der Umgebung, wie z.B. dem Vorhandensein anderer Besucher, gestützt.

Mit dem „Motion-Tracked Binaural“ -Verfahren [Alg04] wurde eine Technologie vorgestellt, die sich dazu eignet, sozio-akustische Komponenten wie diese zu erzeugen. Im Rahmen der Magisterarbeit soll ein MTB-System gebaut werden und, z.B. auch im Zusammenhang mit einer bestehenden Visualisierungsumgebung, in ihren Auswirkungen auf die Empfindung des Präsenzeindrucks eines Konzertbesuchers untersucht werden.

1 Einleitung und Motivation

Zur richtungstreuen Reproduktion räumlicher Schallfelder stehen mit Technologien wie Wellenfeldsynthese, Ambisonics oder der Binauraltechnik verschiedene Lösungen zur Verfügung, die in der praktischen Ausführung jeweils spezifische prinzipbedingte Vor- und Nachteile mit sich bringen.

Im Bereich der binauralen Schallreproduktion erlangte zunächst die Kunstkopfstereophonie in den 70er Jahren grössere Aufmerksamkeit [Wei08]. Das zweikanalige Aufnahmeverfahren, bei dem sich die Mikrofone in den Gehörgängen einer Nachbildung bzw. in der eines echten menschlichen Kopfes befinden, ermöglicht eine beeindruckende räumliche Abbildung der Aufnahmesituation, vor allem durch die bei herkömmlicher Kopfhörerwiedergabe nicht vorhandene Externalisierung von Klangquellen ausserhalb des Kopfes. Aufgrund der festgelegten anatomischen Form des Kunstkopfs im Schallfeld sind die Einfallswinkel der Schallquellen nachträglich jedoch nicht mehr veränderbar. Dies macht sich bei Kopfbewegungen während der Wiedergabe bemerkbar, indem sich die dargebotenen Schallquellen entsprechend mitbewegen.

Die Entwicklung binauraler Technologien erweiterte sich, nicht zuletzt auch wegen der stetig wachsenden Leistung moderner Prozessoren, um computergestützte Verfahren wie der faltungsbasierten dynamischen Binauralsynthese, ein Verfahren, das am Fachgebiet Audiokommunikation einen Forschungsschwerpunkt bildet. Dabei werden Klangquellsignale mit richtungsabhängigen binauralen Raumimpulsantworten (BRIRs) gefaltet, um einem frei wählbaren Quellsignal eine Raum- und Ortsinformationen aufzuprägen. Anders als bei der Kunstkopfstereophonie, ist eine Vorzugsrichtung der gesamten akustischen Szene zum Aufnahmezeitpunkt nicht vorgegeben. Im Zusammenspiel mit einem 'Motion-Tracking'-System, das die Kopfbewegungen des Hörers erfasst, können die Einfallsrichtungen der Quellen den Ohrpositionen des Hörers entsprechend dynamisch angepasst werden. So wird die Wiedergabe von Schallquellen mit einem hohem Plausibilitätsgrad ermöglicht [Lin07]. Die Erzeugung der Ohrsignale ist jedoch mit nicht unerheblichem Aufwand verbunden, und zwar einerseits auf der Ebene der Bereitstellung geeigneter Impulsantworten, andererseits durch die Last der Signalverarbeitung, die pro Schallquelle anfällt.

Mit der Zielsetzung einer validen und plausiblen Auralisation kommen, neben den technisch-perzeptiven Mechanismen, auch soziale und interaktive Faktoren zur Geltung. Die Beziehung des Hörers zur dargestellten Umgebung und die Anwesenheit sich darin bewegend Menschen kann sich wesentlich auf das Gefühl, am (virtuellen) Ort zu sein, also der (virtuellen) Präsenz, auswirken [Thi98][Sch01].

Im Kontext des Forschungsvorhabens Konzertsaal versus mediale Rezeption des Fachgebiets Audiokommunikation bedeutet dies, soziale Aspekte in die Simulationen von Aufführungssituationen gezielt zu integrieren. Dies betrifft beispielsweise die in einer Konzertaufführung zusätzlich vorhandenen Schallereignisse, wie Geräusche durch Bewegungen anderer Besucher, sowie subtile Raumumgebungsgeräusche und -eindrücke.

Unter dem Namen 'Motion-Tracked Binaural' (MTB) wurde 2004 von Algazi et al. [Alg04] nun ein binaurales Verfahren vorgestellt, das sich dazu eignet, eben diese Schallfeldkomponenten zu erzeugen. Dabei nehmen auf einer Kugel befindliche Mikrofone die Schallereignisse des Aufnahmeraums gleichzeitig an mehreren möglichen Ohrpositionen auf. Unter Berücksichtigung von prinzipbedingten Approximationen und der Anwendung geeigneter Interpolationsmechanismen werden richtungsflexible Ohrsignale erzeugt, die der Position des Hörers bei der Wiedergabe dynamisch angepasst sind.

Neben der Fähigkeit, sekundäre Quellen im Auralisationskontext darzustellen, eignet sich das System für verschiedene weitere Anwendungszwecke :

- Dokumentation und Archivierung von Klangkunst und Installationen
- Aufzeichnung von Naturklanglandschaften
- Musikübertragung bei gleichzeitiger Erhaltung der Raumakustik
- Konferenz- und Gesprächsrundenübertragung
- Sound für Computerspiele

Im Rahmen der Magisterarbeit soll ein MTB-System gebaut und in der am Fachgebiet entwickelten StereoViewer-Visualisierungsumgebung [Bau09] zum Einsatz gebracht werden. Zusätzlich zum mechanischen Aufbau des Arrays wird eine Software geschrieben, die die Signale prozessiert und den Zugriff auf verschiedene Systemparameter erlaubt. Die tatsächliche Eignung, rezeptive Parameter wie Präsenz und Immersion innerhalb der Visualisierungsumgebung zu erhöhen, soll abschliessend unter technischen und perzeptiven Gesichtspunkten evaluativ untersucht werden.

2 Technologie und Forschungsstand

Die der Wirkungsweise der Verfahren zugrunde liegenden psychoakustischen Eigenschaften des menschlichen Gehörs und ihrer Signalverarbeitung wie

- die Interaurale Laufzeitdifferenz (engl. ITD)
- die Interaurale Pegeldifferenz (engl. ILD)
- Spektrale Einflüsse der Aussenohrübertragungsfunktion (engl. HRTF)
- der Einfluss von Kopfbewegungen auf die Lokalisationsfähigkeit
- entfernungsbestimmende Signalmerkmale

werden in der Grundlagenliteratur der menschlichen Wahrnehmung und der Psychoakustik eingehend behandelt [Bla97] [Bla08] [Dic97] [Gel04].

Das Motion-Tracked Binaural Verfahren wird in [Alg04] erstmals beschrieben. Es besteht aufnahmeseitig aus einer Anzahl von Mikrofonen, die äquatorial gleichabständig auf einer Kugel angeordnet sind, und ermöglicht somit die simultane Aufnahme von Schallquellen in verschiedenen Einfallswinkeln und Entfernungen (Abb. 3.1 links). Die kopfgrosse Kugel befindet sich als Objekt im Schallfeld und nimmt auf diese Weise Einfluss auf richtungsbestimmende Merkmale im Signal.

Da die Zahl der Mikrofone begrenzt ist, erfolgt die Wiedergabe durch Interpolation der Mikrofonensignale, die entsprechend der Position der Ohren des Hörers im Raum ausgewählt werden. Die Position des Hörerkopfes wird über ein Motiontracking-System kontinuierlich bestimmt, so dass sich auch lokalisationsunterstützende geringe Kopfbewegungen in der Wiedergabe niederschlagen können. Mackensen et al. [Mac98] zeigten, dass die Möglichkeit, Kopfdrehungen beim binauralen Hören durchführen zu können, typische Fehler wie die Vorn-Hinten-Vertauschung fast vollständig eliminieren kann. So argumentieren Algazi et al. [Alg04] auch, dass die Kopfbewegungen beim MTB-System grundsätzlich andere Lokalisationsfehler auszugleichen vermag.

Algazi et al. [Alg04] und weiterführend Hom et al. [Hom06] stellen verschiedene Methoden zur Interpolation vor, die in der signalverarbeitenden Software dieses MTB-Systems wählbar sein sollen.

Ähnlich wie bei der Kunstkopfstereophonie sind die physikalischen Eigenschaften des Aufbaus für die Abbildungstreue des MTB Verfahrens entscheidend. Neben dem fehlenden Torso sind hier die Merkmale, die eine akustische Richtwirkung verursachen - wie die Ohrmuscheln und die allgemeine Kopfform - zugunsten der Richtungsflexibilität vernachlässigt. Zwar hat dies zur Folge, dass wesentliche, z.B. für die Wahrnehmung von elevierten Quellen notwendige Merkmale fehlen, es erlaubt jedoch, dass der Schalleinfallswinkel relativ zum Hörer erst zur Wiedergabezeit festgelegt werden muss. In Folge dessen können aus den Daten eines MTB-Outputs theoretisch beliebig viele dynamische Ohrsignale erzeugt werden, sodass verschiedene Hörer mit jeweils individuellen Ausrichtungen gleichzeitig am Klanggeschehen teilnehmen können.

Eine Schwierigkeit, die sich bei binauralen Systemen häufig zeigt, ist die fehlende exakte Übereinstimmung der Aussenohrübertragungsfunktionen von Aufnahmesystem und Hörer [Dan07]. Nicht nur sind Ohrmuscheln verschiedener Menschen unterschiedlich geformt, auch ist die Kopfgrösse von Mensch zu Mensch variabel [Alg01a]. Vor allem letzterer Umstand schlägt sich bei MTB, aufgrund von Abweichungen von der persönlichen ITD, in Form von Lokalisationsungenauigkeiten und, bei Bewegung des Hörerkopfes, in Verschiebungen der Quellpositionen nieder. Auswirkungen auf richtungsbestimmende Parameter wie der ITD sind bei Brown et al. [Bro05] und Duda et al. [Dud99] beschrieben.

Umsetzungen zur individuellen Anpassung bzw. zur Umgehung verschiedener prinzipieller Probleme des MTB Systems werden in [Mel04] thematisiert.

3 Methode

Die Funktionsweise und die dafür benötigten Bestandteile des MTB-Systems sind in Abb. 3.1 (links) verdeutlicht.

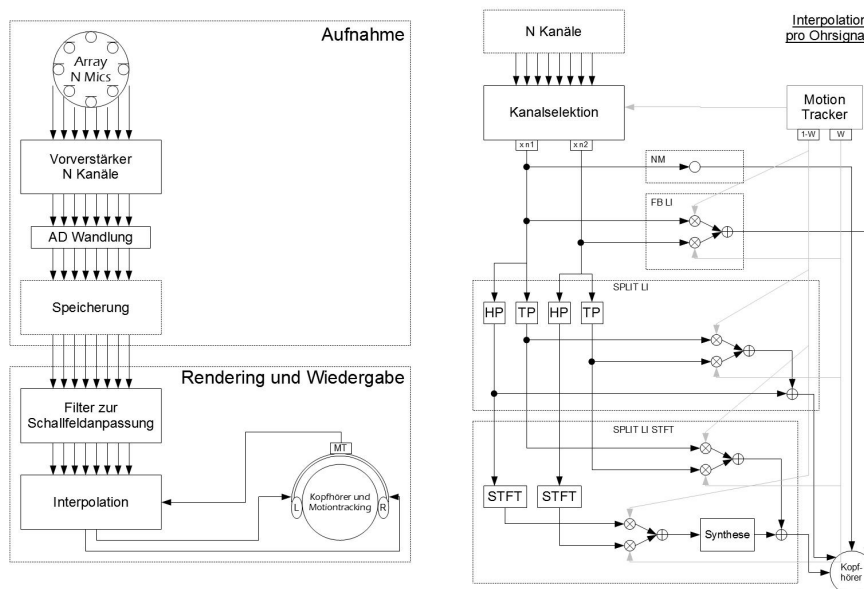


Abbildung 3.1: Links: Prinzipieller Aufbau des MTB-Signalverlaufs für 8 Mikrofone; Rechts: Schaltbild möglicher Interpolationsarten für eins von zwei Ohrsignalen, nach [Hom06]

3.1 Hardwaredesign

Der im Schallfeld wirksam werdende Teil des MTB-Systems besteht aus einer trennbaren Kugel, die mit Löchern als Fassungen für die Mikrofone besetzt ist, und die auf einem Stativ befestigt werden kann (Beispielmodell Abb. 3.2). Zur Ermittlung

der geeigneten Grösse der Kugel werden verschiedene öffentlich zugängliche Messreihen menschlicher Köpfe ausgewertet [Din05][Alg01b][Dre60] und in einen optimalen Durchmesser der Arraykugel umgesetzt [Alg01a].

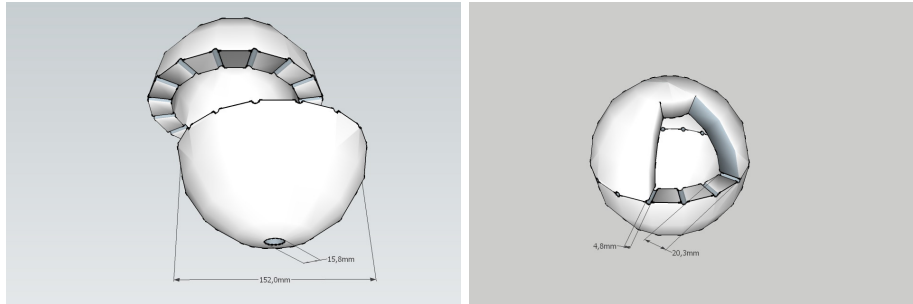


Abbildung 3.2: Beispielmodell der MTB-Kugel

Es werden Panasonic WM-61a Elektret-Mikrofone benutzt, deren Kenndaten es erlauben, die Betriebsspannung aus der konventionellen Phantomspeisung eines Mikrofon-Vorverstärkers zu beziehen. Dazu wird eine Schaltung entwickelt und jedem Mikrofon vorgeschaltet.

Zunächst wird ein Prototyp mit einem einzelnen Mikrofon gebaut, um in einem Vorabtest eine generelle Aussage über die letztendlich benötigte Anzahl der Mikrofone machen zu können, und um Rohdaten zum Gebrauch während der Softwareentwicklung zu erzeugen. Sind alle wesentlichen Parameter festgelegt, wird ein finales MTB-Array im Rapid Prototyping Verfahren angefertigt.

3.2 Softwaredesign

Die zu entwickelnde Software wird als Jack-Client in C++ realisiert. Sie leistet, neben zuschaltbaren Vorfilterungen (siehe Abb. 3.1 links), die korrekte Aufbereitung der Signale nach verschiedenen selektierbaren Interpolationsmechanismen (Abb. 3.1 rechts). Eine OSC-Schnittstelle zum Empfang der Motiontracker-Daten sowie zur Steuerung von internen Parametern wird, ebenso wie eine graphische Benutzeroberfläche, Bestandteil der Software sein.

4 Evaluation

4.1 Vorversuch

Mithilfe des Einzelmikrofon-Arrays wird zunächst ein virtuelles MTB-System erzeugt. Dazu werden Lautsprecher in einem Raum verschieden positioniert und mit Impulsen belegt. Durch sukzessive Drehung des Arrays auf die entsprechenden 8 bis 32 Mikrofonpositionen wird die Positionsbeziehung der Quellen im Raum aufgenommen und als datenbasierte IR-Sammlung abrufbar gemacht. Mittels Faltung der Impulsantworten mit Quellsignalen werden so die Kanäle des virtuellen MTB-Systems erzeugt.

Die maximal möglich Anzahl der Mikrofone des virtuellen MBT-Arrays soll 32 betragen. Dieser Kanalaufbau soll im Vortest als Referenz dienen. Im Vergleich mit weiteren virtuellen Arrays mit 8, 16 und 24 Kanälen soll dann der Schwellwert der Wahrnehmung gefunden werden, der eine unterschiedsfreie Wiedergabe bestimmt. Dabei sollen verschiedene Schallarten verwendet werden (menschliche Geräusche, Sprache, Musik). Auch können, je nach Aufwand und Umfang, weitere Systemparameter wie Diffusfeldentzerrung oder Interpolationsalgorithmen evaluiert werden.

4.2 Anwendungsbezogene Evaluation

Nachdem das finale MTB-System aufgrund der im Vortest festgestellten Mikrofonanzahl angefertigt wurde, werden abschliessend die damit möglichen Erweiterungen virtueller Szenen in ihrer Wirkung auf den Rezipienten untersucht.

Im Speziellen heisst das, dass das MTB-System als sozioakustische Komponente in der Auralisierungs- und Visualisierungsumgebung des Fachgebiets integriert wird. Geräusche in verschiedenen Variationen und Intensitätsgraden werden einer Konzertaufführung zugespielt, und mit einem fragebogenbasierten Verfahren evaluiert, das auf sinnesmodale Rahmenbedingungen dieses Mediums abgestimmt ist [Sch01].

Der Kern der Untersuchung bezieht sich auf die Fragestellung, ob die empfundene Präsenz des Rezipienten mithilfe der neu geschaffenen Möglichkeiten erhöht werden kann. Weiterhin kann durch variable Reize festgestellt werden, ob bestimmte Einflüsse mehr oder weniger entscheidend sind. Beispielparameter hierzu sind

- Beschaffenheit des Aufnahmeraums (Nachhallzeit)
- Art der Geräusche (Husten, Unterhaltungen oder Kommentare, ortsfremde Geräusche wie Strassenlärm)
- Vorhandensein von Teilaspekten (Stimmung der Instrumente, Applaus)
- Funktionalität des Motion-Trackings (vorhanden/nicht vorhanden)
- Individualisierungstechniken nach [Mel04] (Pinna-Ersatz, Anwendung als frontale Applikation)

5 Zeitplan

Monat	Aufgabe	Kal.woche	Wochen
Oktober	Anfertigung Einkanal-Array	41	2
	Software Implementierung	42	4
November	Vorversuch Einkanal-Array	46	1
	Anfertigung Multikanal-Array	47	2
Dezember	Erstellung Hörversuch	49	2
	Hörversuch	51	1
Januar 2010	Auswertung Hörversuch	1	1
	Schriftliche Ausarbeitung	1	16

Literaturverzeichnis

- [Alg01a] ALGAZI, V. R. ; AVENDANO, C. ; DUDA, R. O.: Estimation of a Spherical-Head Model from Anthropometry. In: *J.Audio Eng. Soc.* 49 (2001), June, Nr. 6 [2](#), [3.1](#)
- [Alg01b] ALGAZI, V. R. ; DUDA, R. O. ; THOMPSON, D. M. ; AVENDANO, C.: The CIPIC HRTF database. In: *Proc. IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, 2001, S. 99–102 [3.1](#)
- [Alg04] ALGAZI, V. R. ; DUDA, R. O. ; THOMPSON, D. M.: Motion-Tracked Binaural Sound. In: *J.Audio Eng. Soc.* 52 (2004), Nr. 11, S. 1142–1156 ([document](#)), [1](#), [2](#)
- [Bau09] BAUMBACH, Ralf: *Implementierung einer netzwerkfähigen und interaktiven stereoskopischen Visualisierungsumgebung*. Magisterarbeit, TU Berlin, 2009 [1](#)
- [Bla97] BLAUERT, Jens ; ALLEN, John S.: *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, 1997 [2](#)
- [Bla08] BLAUERT, J. ; BRAASCH, J.: Räumliches Hören. In: WEINZIERL, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin/Heidelberg : Springer-Verlag, 2008 [2](#)
- [Bro05] BROWN, C.H. ; MAY, B.J.: Comparative mammalian sound localization. In: POPPER, A.N. (Hrsg.) ; FAY, R.R. (Hrsg.): *Sound Source Localization*. Springer, 2005 [2](#)
- [Dan07] DANIEL, P. et a.: Kunstkopftechnik–Eine Bestandsaufnahme. Nuntius Acusticus in: *Acta Acustica*. In: *Acustica* 93 (2007), Nr. 1 [2](#)
- [Dic97] DICKREITER, M. ; SFRT (Hrsg.): *Handbuch der Tonstudioteknik*. Bd. Band 1. München : K.G. Saur, 1997 [2](#)
- [Din05] *DIN 33402-2 : Ergonomie - Körpermaße des Menschen - Teil 2: Werte*. Berlin : Deutsches Institut für Normung, 2005 [3.1](#)
- [Dre60] DREYFUSS, H.: *The measure of man*. Whitney Library, 1960 [3.1](#)

- [Dud99] DUDA, RO ; AVENDANO, C. ; ALGAZI, VR: An adaptable ellipsoidal head model for the interaural timedifference. In: *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. Proceedings*. Bd. 2, 1999 2
- [Fas99] FASTL, H. ; ZWICKER, E.: *Psychoacoustics - Facts and Models*. Berlin/Heidelberg : Springer-Verlag, 1999
- [Gel04] GELFAND, S.A.: *Hearing: An Introduction to Psychological and Physiological Acoustics*. Marcel Dekker, 2004 2
- [Hom06] HOM, R. ; DUDA, R.O. ; ALGAZI, V.R.: High-Frequency Interpolation for Motion-Tracked Binaural Sound. In: *121th Convention of the Audio Engineering Society*, 2006 2, 3.1
- [Lin07] LINDAU, A. ; HOHN, T. ; WEINZIERL, S.: Binaural resynthesis for comparative studies of acoustical environments. In: *Audio Engineering Society, 122nd Convention*, 2007 1
- [Mac98] MACKENSEN, P. ; REICHENAUER, K. ; THEILE, G.: Einfluß der spontanen Kopfdrehungen auf die Lokalisation beim binauralen Hören. In: *Tonmeis-tertagungsbände*, 1998 2
- [Man99] MANTOVANI, G. ; RIVA, G.: Real Presence: How Different Ontologies Generate Different Criteria for Presence, Telepresence, and Virtual Presence. In: *Presence* 8 (1999), Nr. 5, S. 540–550
- [Mel04] MELICK, J. B. ; ALGAZI, V. R. ; DUDA, R. O. ; THOMPSON, D. M.: Customization for personalized rendering of motion-tracked binaural sound. In: *117th Convention of the Audio Engineering Society*, 2004 2, 4.2
- [Møl92] MØLLER, Henrik: Fundamentals of Binaural Technology. In: *Applied Acoustics* 36 (1992), S. 171–218
- [Sch01] SCHUEMIE, M.J. ; VAN DER STRAATEN, P. ; KRIJN, M. ; MAST, C. Van d.: Research on Presence in VR: a Survey. In: *Cyberpsychology and Behavior* 4 (2001), Nr. 2, S. 183–202 1, 4.2
- [Thi98] THIE, S. ; VAN WIJK, J.: A general theory on presence: Experimental evaluation of social virtual presence in a decision making task. In: *Presence in Shared Virtual Environments Workshop*, 1998, S. 10–11 1
- [Wei08] WEINZIERL, S.: Aufnahmeverfahren. In: WEINZIERL, S. (Hrsg.): *Handbuch der Audiotechnik*. Berlin/Heidelberg : Springer-Verlag, 2008 1