

Technische Universität Berlin
Fakultät 1 - Geisteswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation



Zur Qualität von binauraler Wiedergabe mit transauralen Wandlern

Expose zur Masterarbeit

vorgelegt von:

Gunar Schlenstedt

Matr.Nr.: 335396
Email: gschlenstedt@gmx.de

eingereicht am: 24. Oktober 2014
Studiengang: Audiokommunikation und -technologie

Erstgutachter: Prof. Dr. Stefan Weinzierl
Zweitgutachter: Fabian Brinkmann

Inhaltsverzeichnis

1	Einleitung und Fragestellung	1
2	Stand der Forschung	2
3	Methode und Quellen	4
4	Arbeits und Zeitplan	6
	Literatur	7
	Tabellenverzeichnis	9

1 Einleitung und Fragestellung

Die folgende Masterarbeit beschäftigt sich mit einem speziellen, kopfbezogenen Verfahren der akustischen 3D-Simulation, der dynamischen Binauralsynthese. Bei der dynamischen Binauralsynthese bleiben sämtliche virtuelle Quellen an der vorgegeben Position im Raum trotz Kopftrotation. Durch Kopfdrehungen ändern sich lediglich die kopfabhängigen Übertragungsfunktionen¹, welche bei der Wiedergabe mit den zwei Ohrsignalen gefaltet werden. Die Drehung des Kopfes bzw. die Bewegungen des Probanden werden durch einen Head-Tracker bestimmt und an die Audio-Software weitergegeben. Idealerweise ergibt sich ein vom Kopfhörer oder Lautsprecher losgelöstes authentisches Klangbild.

Besonderes Augenmerk liegt auf der Untersuchung der Lautsprecher-basierten Wiedergabe binauraler Aufnahmen, die sogenannte transaurale Binauralsynthese. Über ein spezielles Verfahren, die Crosstalk-Cancellation (CTC oder XTC) wird das Übersprechen der Lautsprecher auf das kontralaterale Ohr kompensiert. Zusammengefasst soll die CTC eine kopfhörerähnliche Abhörsituation mit Lautsprechern erzeugen.

Dieses Verfahren wird besonders bei Virtual Reality-Systemen eingesetzt in denen sich der Benutzer frei bewegen soll. Die CTC hat den Vorteil, dass der Anwender keine Kopfhörer oder zusätzliche, störende Geräte und Kabel am Körper tragen muss. Nichtsdestotrotz ergeben sich durch diese Methode auch Nachteile gegenüber einer konventionellen Binauralsynthese mit Kopfhörern. Das System beinhaltet die Laufzeiten der Lautsprecher zum Ohr, welche zusätzlich mit den Reflexionen des Raums überlagert werden. Zudem ist eine optimale Übersprechkompensation nur mit individualisierten HRTFs und bestimmten Öffnungswinkeln der Lautsprecher möglich, was eine Wiedergabe zusätzlich erschwert.

Besonders die systembedingten Artefakte der CTC und deren Auswirkung auf die Qualitätsempfindung der Rezipienten soll in der geplanten empirischen Studie analysiert werden. Es soll die Auswirkung verschiedener Störvariablen, welche während des Versuchs variiert werden, in Bezug auf eine Referenz ohne CTC untersucht werden. Die Studie stützt sich dabei auf bisherige Forschungserkenntnisse, Fragebögen und Skripte des Forschungsbereichs für Virtuelle Akustik des Fachgebiets Audiokommunikation der TU Berlin. Die aus dem Versuch gewonnenen Erkenntnisse sollen für weitere Versuche unterstützend sein und zudem Verbesserungsvorschläge für eine optimale Wiedergabe binauraler Signale mit Lautsprechern liefern.

¹Head Related Transfer Functions = HRTFs

2 Stand der Forschung

Mit der Wiedergabe von Binauralsynthese mit Lautsprechern ergeben sich systembedingt Probleme. Bekannte Störvariablen bei der Wiedergabe sind die Leistungsfähigkeit des Cancellation Algorithmus, die Latenz des Wiedergabesystems (bzw. der Abstand vom Lautsprecher bis zum Hörer), das Verhalten des Wiedergaberaums, die Genauigkeit der Anpassung der Filter und die Anzahl der Iterationen.

Voraussetzung für eine genaue Wiedergabe binauraler Signale und deren Cues ist es, dass der rechte und linke Kanal möglichst separat das jeweilige Ohr erreichen [7]. Während bei Kopfhörern die Separation äußerst hoch ist, muss beim Abspielen mit Lautsprechern eine Übersprechkompensation, CTC durchgeführt werden. Die Eigenschaften des CTC-Algorithmus sind fundamental für die Qualität der Wiedergabe.

Ein erster Vorschlag in Form einer elektrischen Schaltung stammt von Bauer [4]. Bei diesem Beispiel sind jedoch weder Filter noch mehrere Iterationen mit inbegriffen. Um eine Separation von 20-25 dB zu erreichen, sind mindestens 5 Iterationen notwendig [17, S.295]. Zusätzlich sollten die Abschattungseffekte des Kopfes durch Faltung der Signale mit passenden HRTFs berücksichtigt werden, um eine realistische Simulation zu gewährleisten. Erste Systeme mit Filtern lieferten Atal und Schröder im Jahre 1966 [2]. Ähnlich wie bei der dynamischen Binauralsynthese sind auch bei der CTC individualisierte HRTFs die beste Grundlage [16]. Besonders wichtig sind angepasste HRTFs, um eine im Pegel möglichst große Übersprechdämpfung zu erreichen [1, 15]. Dies steigert natürlich auch die Lokalisationsgenauigkeit von virtuellen Quellen bei der Wiedergabe. Weiterführend ist für die Genauigkeit auch die diskrete Winkelauflösung der CTC-Filter entscheidend. Eine Winkelauflösung der binauralen Impulsantworten von 2° in horizontaler und 1° in vertikaler Richtung garantieren bei einer Binauralsynthese ein plausible, kontinuierliche dynamische Binauralsynthese [14]. Ähnliche Werte sind auch für die CTC-Filter anzunehmen.

Ein weiteres Problem entsteht, wenn eine 360° Binauralsynthese in der horizontalen Ebene realisiert werden soll. Versuche haben gezeigt, dass eine dynamische CTC nur im aufgespannten Winkel des Lautsprecher setups realisiert werden kann. Außerhalb des Winkels sind die verwendeten Filter nicht mehr stabil und bringen das System zum Abstürzen [11]. Aufgrund dieser Erkenntnis hat die RHTW Aachen ein System mit vier Lautsprechern [10] entwickelt, um eine 360° -CTC zu ermöglichen. Grundsätzlich wird je nach Kopfposition ein Lautsprecherpaar aktiviert, welches im aufgespannten Winkel eine CTC durchführt.

Der Aufbau der CTC an der RHTW befindet sich in einer CAVE mit hart reflektierenden Wänden, was deutliche Einschränkungen mit sich bringt. Die Reflexionen an den Wänden führen dazu, dass die binauralen cues nicht mehr störungsfrei an das Ohr gelangen, bzw. verändert werden, was die virtuelle Umgebung „verschmutzt“. Kosmidis fand in seinen Studien heraus, dass frühe Reflexionen die Lokalisationsfähigkeit der Probanden besonders bei Winkeln in der Nähe von $\pm 30^\circ$ beeinflussen [9, S.103]. Auch die Soundqualität, wie Klarheit und Räumlichkeit werden stark durch Reflexionen beeinflusst [9, S.104]. Des Weiteren weist er darauf hin, dass die Positionierung der Lautsprecher, sowie die Geometrie des Raums entscheidend für die Lokalisierbarkeit von Quellen ist, was er anhand der Beeinträchtigung der IGD² durch Raumeinflüsse darstellt [9, S.103]. Dies ist auch der Grund dafür, dass die CTC hauptsächlich in akustisch trockenen Umgebungen, wie Aufnahmestudios oder reflexionsarmen Räumen, praktiziert wird. Bisherige Versuche lieferten noch keine aufschlussreichen Ergebnisse, welche Reflexionen dafür verantwortlich sind, dass die CTC nur noch bedingt funktioniert und perzeptiv nicht mehr authentisch wirkt. Auch Kosmidis weist darauf hin, dass es weitere Experimente nötig sind, um weitere Thesen zu bestätigen bzw.

²Interaural Group Delay, deutsch: Interaurale Gruppenlaufzeit

bisherige zu bekräftigen [9].

So soll diese Masterarbeit neue Erkenntnisse in Bezug auf Artefakte, wie Positionierung des Hörers, Raumreflexionen und Filteranpassung liefern. Spezielles Augenmerk liegt dabei auf der Wahrnehmung der Probanden, um herauszufinden inwiefern und ab welchem Grad Artefakte sich auf die Authentizität der virtuellen Umgebung auswirken. Ein genauer Fokus auf z.B. die Raumreflexionen soll im Laufe der Masterarbeit erarbeitet werden.

3 Methode und Quellen

Anders als bei bisherigen Versuchen soll die Wiedergabe mit Lautsprechern nur virtuell passieren. Es werden jeweils BRIRs³ von dem virtuellen Wiedergaberaum mit CTC und dem wiederzugebenden Raum der Binauralsynthese benötigt. Das erste Model für die CTC wird der CAVE⁴ der RHTW Aachen in Dimensionen, Lautsprecherpositionen und Wandmaterialien nachempfunden. Mit Ease 4.3 soll ein Reflektionsfile erzeugt werden, welches die Daten liefert, um durch ein bestehendes Matlab-Skript des Fachgebiets Akt⁵ die BRIRs für eine 360° CTC mit einer Winkelgenauigkeit von 1° zu generieren. Der Datensatz zur Erstellung der CTC-Filter basiert auf gemessenen HRTFs von FABIAN [5, 13]. Um den Raumeinfluss beurteilen zu können sollen mindestens zwei Modelle der CAVE mit unterschiedlichen Wandmaterialien simuliert werden. Als weitere Variation soll die Hörposition um mehrere cm aus dem Sweet Spot⁶ verschoben werden.

Die erzeugten CTC-Übertragungsfunktionen werden später mit den BRIRs der zu simulierenden Binauralsynthese gefaltet. Als Grundlage sollen entweder gemessene oder virtuell BRIRs aus verschiedenen Räumen benutzt werden. Die Raumantworten sollen sich in ihrem Volumen, ihren Wandmaterialien und somit in ihrem Nachhall deutlich voneinander unterscheiden, um eventuelle Effekte aufzudecken. Als Räume sind geplant: Ein mittelgroßes Studio mit vielen absorbierenden Wänden und ein Orchestersaal bzw. eine Kirche mit großen Raumvolumen und einer langen Nachhallzeit.

Die Impulsantworten werden über ein extraaurales Kopfhörersystem BK211 [6] wiedergegeben. Dieser Kopfhörer, dessen Membranen sich ca. 5cm vor dem Ohrkanal befinden, wurde speziell für die Wiedergabe von dynamischer Binauralsynthese und den Einsatz in VR-Systemen entwickelt. Niedrige Rausch-, Crosstalk- und Verzerrungspegel liefern gute Grundvoraussetzungen für die Studie. Zudem ist der Kopfhörer darauf konzipiert Head-Tracker (Polhemus Fastrak Sensors [6]) leicht am Gehäuse befestigen zu können, was den Aufwand des Versuch verringert.

Der generelle Vorteil an dem Verfahren mit Kopfhörern ist es, dass es dem Forscher möglich ist, Stimuli unabhängig von Raumeinflüssen zu betrachten. So können auch mehrere Raummodelle und deren Einfluss auf die Qualität von transauralen Wiedergabesystemen überprüft werden. Würde man die Wiedergabe mit Lautsprechern durchführen, müssten absorbierende Stellwände ständig im Wiedergaberaum umgestellt oder der Raum gewechselt werden, um den Einfluss von Raumreflexionen auf die Wiedergabe zu untersuchen.

Die Algorithmen zur Erzeugung der Ohrsignale sollen in Matlab geschrieben werden. Die erzeugten Impulsantworten und Filter von Wiedergaberaum mit CTC und wiederzugebender Binauralsynthese sollen erst bei der Wiedergabe in Echtzeit miteinander gefaltet und addiert werden. Dies erhöht die Flexibilität bei der Variation der einzelnen unabhängigen Variablen. Hier werden beispielsweise künstliche Latenzen erzeugt, um zu erkennen ab welcher Verzögerung die wahrgenommene Qualität deutlich abnimmt oder Raumimpulsantworten getauscht (evtl. auch Reflexionen kompensiert), um den Raumeinfluss bei der CTC beurteilen zu können. Zudem ist ein Vergleich verschiedener CTC Übertragungsfunktionen geplant, um eventuelle Vor- bzw. Nachteile der einzelnen HRTFs bzw. BRIRs zu ermitteln. Sämtliche CTC-Filter sollen beim Versuch mit einer Referenz, einer Binauralsynthese ohne CTC Simulation und Wiedergaberaum, verglichen werden.

Als Stimuli soll ein Testsignal, wie z.B. pulsierendes rosa Rauschen und ein alltägliches Signal, wie Sprache oder Musik verwendet werden.

³Binaural Room Impulse Response = Binaurale Raumimpulsantwort

⁴Cave Automatic Virtual Environment

⁵Audiokommunikation und -technologie

⁶optimale Hörposition in der Mitte des aufgespannten Winkels der beiden Lautsprecher

Zusammengefasst ergeben sich somit die unabhängigen Variablen aus Tabelle 1. Da die Anzahl an Ausprägungen eine große Stichprobenanzahl erfordert, soll im Rahmen von Vorversuchen und Absprachen noch Variablen aus dem Versuch ausgeschlossen werden.

Tabelle 1: Unabhängige Variablen

Art der unabh. Variable	Varianten (Beispielhaft)	Anzahl Ausprägungen
Audioinhalt Stimulus	RosaRauschen (gepulsed)	2
	Sprache/Musik	
Hörposition in VR	SweetSpot	2
	Versoben 1 m nach rechts	
Unterschiedliche CTC Übertragungsfunktionen	BRIR1	2
	BRIR2 mit Reflektionscancellation	
Binauralsynthese Räume	Studio	2
	Orchestersaal, Kirche	
CTC Räume	CAVE ungedämpft	2
	CAVE gedämpft	

Ausgeschlossen wurden in einem Gespräch mit dem Betreuer sowohl die unterschiedlichen CTC Übertragungsfunktionen besonders in Hinblick auf Reflexionscancellation, als auch die Hörpositionsverschiebung in der virtuellen Umgebung. Nach Rücksprachen mit Mitarbeitern der RHTW Aachen stellte sich heraus, dass die Unterdrückung von frühen Reflexionen in der CAVE nicht fehlerfrei funktioniert, weswegen dieser Stimulus ausgeschlossen wurde. Des weiteren liefern Kirkeby und Nelson [8], sowie Takeushi et al citeTakeushi.1997, Takeushi.2001, Takeushi.2002 bereits Aufschluss über die Größe des Sweet Spots der CTC, woraufhin sich eine Untersuchung der Verschiebung der Hörposition als redundant erwies. Mit höher werdenden Frequenzen wird empfohlen den Öffnungswinkel der Lautsprecher zu verringern, um eine robuste Wiedergabe zu garantieren [?]. Bai und Lee [3] bestätigen, dass mit kleinem Öffnungswinkel sich der Sweet Spot vergrößert, jedoch damit andere Probleme, wie hohe notwendige Pegel zu tiefen Frequenzen verursacht werden.

Zur Datenerhebung soll der bereits eingesetzte SAQi Fragebogen des Fachgebiets Audiokommunikation genutzt werden. Dieses Skript enthält 48 abhängige Variablen, welche Aufschluss über die perzeptive Qualitätsbeurteilung des Probanden geben. Damit sämtliche unabhängige Variablen getestet werden können, soll das Skript um einen ABC/HR-Test erweitert werden. Dieser Test ermöglicht es im Gegensatz zum ABX-Test, welcher nur Informationen über einen noch hörbaren Unterschied liefert, über standardisierte Ratingskalen den hörbaren Anteil mehrerer abhängiger Variablen gleichzeitig qualitativ zu bewerten [12, S.866]. Die Inferenz-statistische Auswertung soll in Matlab oder SPSS erfolgen. Im Idealfall werden die Daten schon während der Studie ausgewertet, um eventuelle Fehler bereinigen oder Anpassungen machen zu können.

4 Arbeits und Zeitplan

In der folgenden Tabelle wird der ungefähre zeitliche Ablauf der Arbeit aufgezeigt. Es soll ein grober Überblick über die Tätigkeiten und deren Beginn und Dauer geschaffen werden.

Tabelle 2: Zeitplan der Masterarbeit (Tätigkeiten können monatsübergreifend sein)

Monat	Art der Tätigkeit	Tätigkeit	Wochen
Mrz 2014	Vorbereitung	Literaturrecherche in Bibliotheken und Datenbanken der TU	12
		Zusammentragen bisheriger Erkenntnisse	4
		Vertraut machen mit den bisherigen Verfahren und Algorithmen	4
		Festlegung sämtlicher Parameter für Studie aus gewonnen Wissen	2
Jun 2014	Expose schreiben	Zusammentragen des bisherigen Wissens	4
Sept 2014	Versuchsplanung	Ermitteln der notwendigen Stichprobenanzahl	4
		Erstellung der notwendigen Stimuli (Impulsantworten) in Ease	2
		Weiteres Stimulidesign	2
		Semantisches Differential (SaQi) anfordern	1
		Test des Versuchsaufbaus	2
		Werbung von Probanden	8
Okt 2014	Versuchsdurchführung	Durchführung des Versuchs	8
		Erste Auswertungen der Messdaten	8
Okt 2014	Versuchsdurchführung	Weitere Auswertungen und erste Interpretation	4
Nov 2014	Schreibprozess	Sämtliches bisher geschriebenes Material organisieren	4
		Gesamten Daten aufbereiten und interpretieren	4
Dez 2014	Abgabe der Arbeit	Zusätzlicher Puffer bis zum 20.Dezember	4

Literatur

- [1] AKEROYD, Michael A. ; CHAMBERS, John ; BULLOCK, David ; PALMER, Alan R. ; SUMMERFIELD, A. Q. ; NELSON, Philip A. ; GATEHOUSE, Stuart: The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. In: *The Journal of the Acoustical Society of America* 121 (2007), Nr. 2, S. 1056–1069
- [2] ATAL, B. S. ; SCHRÖDER, M. R.: *Apparent sound source translator*. Version: 1966
- [3] BAI, Mingsian R. ; LEE, Chih-Chung: Comprehensive Analysis of Loudspeaker Span Effects on Crosstalk Cancellation in Spatial Sound Reproduction. In: AUDIO ENGINEERING SOCIETY (Hrsg.): *Audio Engineering Society Convention 120*, 2006
- [4] BAUER, Benjamin B.: Stereophonic Earphones and Binaural Loudspeakers. In: *J. Audio Eng. Soc* 9 (1961), Nr. 2, 148–151
- [5] BRINKMANN, Fabian ; LINDAU, Alexander ; WEINZIERL, Stefan ; GEISSLER, Gunnar ; VAN DE PAR, STEVEN: A high resolutional head-related transfer function database including different orientations of head above the torso. In: *Fortschritte der Akustik: Tagungsband d. 39. DAGA*. Merano and Italy, 2013, S. 596–599
- [6] ERBES, Vera ; SCHULTZ, Frank ; LINDAU, Alexander ; WEINZIERL, Stefan: An extraural headphone system for optimized binaural reproduction. In: DEUTSCHE GESELLSCHAFT FÜR AKUSTIK eV (Hrsg.): *Daga '12*, 2012
- [7] GARDNER, William G.: *3D Audio and Acoustic Environment Modeling*. (1999)
- [8] KIRKEBY, Ole ; NELSON, Philip A. ; HAMADA, Hareo: The "Stereo Dipole": Binaural Sound Reproduction using Two Closely Spaced Loudspeakers. In: AUDIO ENGINEERING SOCIETY (Hrsg.): *Audio Engineering Society Convention 102*, 1997
- [9] KOSMIDIS, Dimitrios: *The Influence of Early Reflections on the Interaural Time Difference in Crosstalk Cancellation Systems*. THESSALONIKI, ARISTOTLE UNIVERSITY OF THESSALONIKI, Diss., 2013
- [10] LENTZ, Tobias: Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments. In: *J. Audio Eng. Soc* 54 (2006), Nr. 4, 283–294
- [11] LENTZ, Tobias ; ASSENMACHER, Ingo ; VORLÄNDER, Michael ; KUHLEN, Torsten: Precise Near-to-Head Acoustics with Binaural Synthesis. In: *Journal of Virtual Reality and Broadcasting, Volume 3(2006), no. 2*, 2006
- [12] LERCH, Alexander: Bitratenreduktion. In: WEINZIERL, Stefan (Hrsg.): *Handbuch der Audio-technik*. Berlin and Heidelberg : Springer Berlin Heidelberg, 2008 (VDI-Buch)
- [13] LINDAU, Alexander ; WEINZIERL, Stefan: Fabian - Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden. In: *Fortschritte der Akustik: Tagungsband d. 33. DAGA*. Stuttgart, 2007
- [14] LINDAU, Alexander ; WEINZIERL, Stefan: Zur Wahrnehmbarkeit von Diskretisierungen in der dynamischen Binauralsynthese. In: VERBAND DEUTSCHER TONMEISTER (Hrsg.): *25. Tonmeister-tagung - VDT International Convention*, 2009, S. 546–559

-
- [15] MAJDAK, Piotr ; MASIERO, Bruno ; FELS, Janina: Sound localization in individualized and non-individualized crosstalk cancellation systems. In: *The Journal of the Acoustical Society of America* 133 (2013), Nr. 4, S. 2055
- [16] MØLLER, Henrik ; SØRENSEN, Michael F. ; JENSEN, Clemen B. ; HAMMERSHØI, Dorte: Binaural Technique: Do We Need Individual Recordings? In: *J. Audio Eng. Soc* 44 (1996), Nr. 6, 451–469
- [17] VORLÄNDER, Michael (Hrsg.): *Auralization: Fundamentals of acoustics, modelling, simulation, algorithms and acoustic virtual reality: Literaturverz.* S. [319] - 330. 1. ed. Berlin : Springer, 2008

Tabellenverzeichnis

1	Unabhängige Variablen	5
2	Zeitplan der Masterarbeit (Tätigkeiten können monatsübergreifend sein)	6