Dynamische Binauralsynthese auf Basis gemessener einkanaliger Raumimpulsantworten

Christoph Pörschmann, Stephan Wiefling

Fachhochschule Köln, Institut f. Nachrichtentechnik,50679 Köln, Email: Christoph.Poerschmann@fh-koeln.de

Einleitung

Für viele Anwendungen im Bereich auditiver virtueller Umgebungen werden Räume kopfhörerbasiert unter Nutzung der dynamischen Binauralsynthese auralisiert. Hierbei wird nicht immer angestrebt, eine authentische Darbietung zu erzielen, häufig reicht eine plausible Präsentation der Szene aus.

Die messtechnische Erfassung der hierfür erforderlichen Datensätze binauraler Raumimpulsantworten (BRIRs) erfolgt üblicherweise mit Hilfe eines Kunstkopfes, der in kleinen Winkelabständen im Raum gedreht wird und mit dem für jede Kopfdrehung eine separate BRIR erfasst wird. Solche Messungen sind jedoch hinsichtlich der erforderlichen Apparaturen und der Messdauer so aufwendig, dass sie in vielen Anwendungsbereichen kaum eingesetzt werden.

Es wird ein Verfahren vorgestellt, das aus einer einzelnen omnidirektionalen Raumimpulsantwort einen Datensatz synthetisierter BRIRs erzeugt. Obwohl wesentliche räumliche Informationen in der einkanaligen Impulsantwort nicht erfasst werden, zielt das Verfahren darauf ab, dass die mit dem synthetisierten Datensatz von BRIRs auralisierten Räume als gut identifizierbar und die Szenen als plausibel wahrgenommen werden. Hierbei sollen auch binaurale Merkmale synthetisiert werden, die bei einer dynamischen Binauralsynthese zur Lokalisierbarkeit und zur Externalisierung der Schallquellen führen. In diesem Verfahren werden Direktschall, frühe Reflexionen und Diffusschall separat voneinander behandelt und die unterschiedlichen Abschnitte der gemessenen Raumimpulsantwort in eine BRIR überführt. Dabei werden generische räumliche Informationen hinzugefügt.

Weiterhin werden die Ergebnisse psychoakustischer Experimente vorgestellt, die die perzeptive Ähnlichkeit zwischen synthetisierten und gemessenen BRIR-Sätzen am Beispiel von mehreren Räumen untersuchen.

Seitens des Benutzers werden lediglich Informationen über das Raumvolumen und die Richtung der Quelle vorgegeben. Es wird somit auf übliche Größen zurückgegriffen, die typischerweise im Rahmen von raumakustischen Messungen miterfasst werden. Das Verfahren kann z.B. für Ingenieurbüros und Architekten, aber auch für Virtual-Reality-Anwendungen interessant sein.

Grundidee des Verfahrens

Ziel des hier vorgestellten Verfahrens ist es, aus einer einzelnen gemessenen monauralen Raumimpulsantwort einen Satz binauraler Raumimpulsantworten zu synthetisieren. Dazu sollen vorhersagbare Größen aus dem Bereich der geometrischen Raumakustik auf die gemessene ommnidirektional erfasste Raumimpulsantwort geeignet übertragen werden. Hierfür werden die einzelnen frühen Anteile der Impulsantwort mit den Außenohrimpulsantworten (HRIRs) unterschiedlicher Einfallsrichtungen geeignet gefaltet. Zudem wird das Diffusschallfeld so nachgebildet, dass die zeitliche Struktur des Energieabfalls der omnidirektionalen Impulsantwort erhalten bleibt.

Die Eintreffrichtung des Direktschalls beim Hörer wird vom Benutzer vorgegeben. Das Zeitfenster, in dem sich in der omnidirektionalen Raumimpulsantwort der Direktschall befindet, wird mit der entsprechenden Außenohrübertragungsfunktion gefaltet. Die frühen Reflexionen werden statistisch als aus unterschiedlichen Richtungen eintreffend gewählt, die Richtungen dieser Reflexionen werden vorab und unabhängig vom spezifischen Raum festgelegt.

Der diffuse Nachhall wird als aus allen Raumrichtungen eintreffend statistisch verteilt aufgefasst und passend frequenzabhängig in seiner Hüllkurve geformt. Mehrere Studien [1,2] zeigen, dass der diffuse Nachhall oberhalb einer perzeptiven Mixing Time durch ein statistisches Signal (z.B. binaurales Rauschen) ersetzt werden kann. Eine exakte Nachbildung des Schallfeldes ist somit nicht mehr erforderlich. Der Wert der Mixing Time kann anhand verschiedener Näherungen in Abhängigkeit von der Raumgeometrie abgeschätzt werden [3].

Aufgrund der in der einkanaligen Raumimpulsantwort fehlenden Informationen, weist das Verfahren prinzipbedingt eine Reihe von Ungenauigkeiten auf. In dem synthetisierten Direktschallanteil kann die Ausdehnung der Schallquelle in keiner Weise nachgebildet werden. Hier wird immer von einer Punktschallquelle ausgegangen. Die Eintreffrichtungen der frühen Reflexionen werden nach einem generischen Reflexionsmuster abgeschätzt, das hinsichtlich der Eintreffrichtungen des Schalls nicht weiter an den zu synthetisierenden Raum angepasst wird. Auch kann die Diffusität der Reflexionen nicht adäquat nachgebildet werden. In psychoakustischen Untersuchungen wird evaluiert, welchen wahrnehmungsbezogenen Einfluss diese Näherungen haben.

Basierend auf rein statistischen Methoden wurden bereits in [4,5] Verfahren vorgestellt, die eine binaurale Synthese von Raumimpulsantworten durch eine geeignete Anpassung der Kohärenz aus gemessenen Raumimpulsantworten ermöglichen. Allerdings wurden hier nicht gezielt Direktschall, frühe Reflexionen und Diffusschall getrennt behandelt und auch keine daraus abgeleitete psychoakustisch motivierte Modellierung durchgeführt.

Implementierung

Die Implementierung erfolgte in Matlab und umfasste die folgenden Schritte:

- Dem Algorithmus werden die omnidirektional erfasste Raumimpulsantwort, das Volumen des Raumes und die Richtung der Schallquelle vorgegeben. Darüber hinaus greift das Verfahren auf eine Sequenz binauralen Rauschens und einen Satz von Außenohrimpulsantworten zu.
- Der Algorithmus wird nur für die Frequenzanteile oberhalb von 200 Hz eingesetzt. Unterhalb von 200 Hz wird die monaurale Impulsantwort unverändert übernommen. Damit wird der Eigenschaft binauraler Raumimpulsantworten Rechnung getragen, bei denen die Kohärenz unterhalb von 200 Hz nahezu bei 1 liegt und erst für höhere Frequenzen absinkt.
- Als Direktschall wird das erste relevante Maximum identifiziert. Der Zeitraum von t=0 ms bis zu 5 ms nach dem ersten Maximum wird gefenstert mit der für den Direktschall gewählten kopfbezogenen Außenohrimpulsantwort (HRIR) gefaltet.
- In einem nächsten Schritt werden im Bereich von t=5 ms bis zu maximal t=150 ms Abschnitte in der monauralen Raumimpulsantwort betrachtet (gleitende Fensterlänge mindestens 8 ms), in denen geometrische Reflexion identifiziert werden können. Hierzu werden energiereiche Abschnitte der Raumimpulsantwort bestimmt. Als Schätzer für eine geometrische Reflexion wird gefordert, dass die Energie um 6 dB über der mittleren Energie des gesamten Abschnitts liegen muss.
- Die Einfallsrichtungen der Reflexionen werden nach einem im vorab festgelegten Muster gewählt und die HRIR der entsprechenden Richtung mit dem gefensterten Abschnitt der einkanaligen Raumimpulsantwort gefaltet.
- Die verbliebenen Anteile der einkanalig gemessenen Raumimpulsantwort werden in 1/6 Oktavbänder aufgespalten. Hierzu wird eine „near perfect reconstruction filterbank“ eingesetzt [6].
- In jedem einzelnen Frequenzband wird der binaurale Diffusanteil des Schallfeldes nachgebildet. Dazu wird die Hüllkurve des binauralen Rauschsignals an die Hüllkurve der monauralen Impulsantwort angepasst. Das Verfahren hierzu wurde schon in [7] detailliert beschrieben.
- Die geometrisch ermittelten Reflexionen und der synthetisch erzeugte binaurale Diffusanteil werden geeignet gewichtet und überlagert, sodass die gesamte sich ergebende synthetische binaurale Raumimpulsantwort in allen Frequenzbändern den zeitlichen Verlauf der Energie der gemessenen monauralen Raumimpulsantwort möglichst gut nachbildet.

Abbildung 1 verdeutlicht die Funktionsweise des entwickelten Algorithmus.

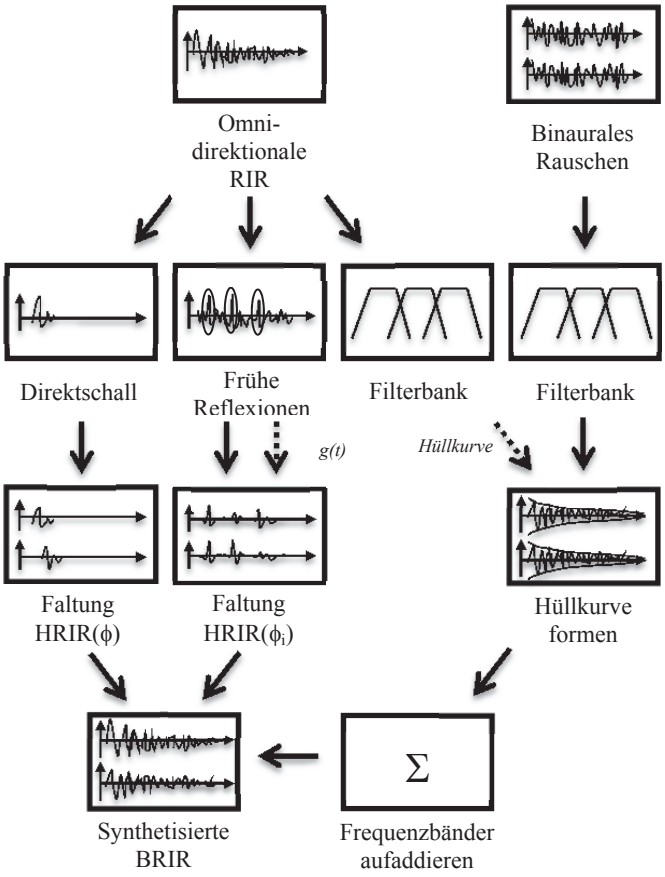

Abbildung 1: Blockdiagramm des Algorithmus zur Synthese einer binauralen Impulsantwort aus einer einzelnen omnidirektionalen Impulsantwort

Die Berechnung der synthetischen BRIR wird wiederholt durchgeführt und erfolgt in azimuthaler Verschiebung von gleichbleibender Schrittweite (z.B. 1°) für Direktschall und frühe Reflexionen. Somit entsteht ein gedrehter Satz von binauralen Raumimpulsantworten, der zur Verwendung in einem Binauralrenderer geeignet ist.

Für die verwendeten HRIRs wurde ein an der FH Köln gemessener Datensatz eines Neumann KU 100 Kunstkopfes verwendet [8].

Messungen

Zur Evaluierung des Algorithmus und zur Durchführung der psychoakustischen Untersuchungen wurden Messdaten von vier Räumen verwendet. Bei diesen Räumen wurde sowohl ein Referenzsatz von binauralen Raumimpulsantworten als auch eine omnidirektionale Raumimpulsantwort am Drehort des Kunstkopfes erfasst. Die binauralen Messungen erfolgten in 1°-Schritten in der Horizontalebene mit einem Neumann KU 100 Kunstkopf. Der für die Anregung ver-

wendete Lautsprecher und die Positionen von Sender und Empfänger blieben in einer Messreihe jeweils unverändert.

Es wurden die folgenden hinsichtlich ihrer akustischen Eigenschaften und ihrer Nutzung recht unterschiedlichen Räume ausgewählt:

- Abhörraum Regie 7 (WDR Köln):
  Volumen 168 m³, Grundfläche 60 m²,
  T_{60}(500/1000Hz) < 0,25 s
- Großer Sendesaal (WDR Köln):
  Volumen 6100 m³, Grundfläche 579 m²,
  T_{60}(500/1000Hz) = 1,8 s
- Kleiner Sendesaal (WDR Köln):
  Volumen 1247 m³, Grundfläche 220 m²,
  T_{60}(500/1000Hz) = 0,9 s
- TGC Übungsraum (Köln):
  Volumen 191 m³, Grundfläche 67 m²,
  T_{60}(500/1000Hz) = 2,3 s

Eine detaillierte Beschreibung der ersten drei Räume und des Messverfahrens finden sich in [9]. Die Messungen in dem TGC-Raum erfolgte nach den gleichen Verfahren.

Psychoakustischer Experimente

Zur perzeptiven Evaluierung des Verfahrens wurde ein Hörversuch durchgeführt. Hierbei wurde ein SAQI-Paradigma [10] gewählt, das es ermöglicht, den Unterschied der wahrgenommenen Ausprägungen spezifischer Attribute zwischen einem Stimulus und einer Referenz zu beurteilen. Ziel der psychoakustischen Untersuchungen war zum einen, die Stärke der Unterschiede zwischen dem synthetisch erzeugten Satz von binauralen Raumimpulsantworten und dem Referenzdatensatz zu beurteilen. Zum anderen sollten die Experimente dazu dienen, die maßgeblichen perzeptiven Einflussfaktoren für die Unterschiedlichkeit zu analysieren. Im Rahmen der hier vorgestellten Untersuchungen wird nur auf die Unterschiedlichkeit eingegangen, eine ausführlichere Darstellung aller Ergebnisse des SAQI-Tests ist für die Präsentation auf der ICSA 2015 in Graz vorgesehen.

Die Durchführung der Experimente erfolgte mit der Software SCALE [11,12]. Diese steuerte den Ablauf des psychoakustischen Experimentes und schaffte die Anbindung an den Sound Scape Renderer (SSR). Der SSR nimmt die Faltung der BRIR-Sätze mit den Audiosignalen vor und passt das Schallfeld an die Kopfdrehungen des Hörers an. Die Eingabe der Beurteilungen seitens der Probanden erfolgte über ein Touch-Screen-Tablet (iPad).

Aufgrund der Komplexität der Experimente wurde für den Hörversuch auf eine Gruppe von Expertenhörern zurückgegriffen. Diese bestand etwa zur Hälfte aus Toningenieuren des WDR, die anderen Teilnehmer waren mit psychoakustischen Experimenten seit langem vertraut und gewohnt, perzeptive Eigenschaften von virtuellen auditiven Umgebungen zu beurteilen. Die 11 Teilnehmer waren zwischen 28 und 63 Jahre alt, das Durchschnittsalter betrug 42 Jahre.

Die Audiosignale wurden über einen offenen Kopfhörer vom Typ AKG K-601 dargeboten. Die Ortung der Kopfbewegung erfolgte mithilfe des Headtrackingsystems Polhemus Fastrak, dessen Sensor mittig auf den Kopfhörerbügel montiert wurde. Die gefalteten Ausgangssignale wurden auf einen Pegel von 65dB(A) SPL normiert.

In 16 Vergleichen wurden die vier verschiedenen Räume durch eine Faltung mit der einkanaligen Raumimpulsantwort oder der synthetisch erzeugten binauralen Impulsantwort dargeboten. Der Vergleich erfolgte immer zu dem gemessenen Referenzdatensatz. Vergleichend wurde weiterhin die Unterschiedlichkeit zwischen der omnidirektionalen Raumimpulsantwort und dem Referenzdatensatz untersucht.

Die Bewertung der Unterschiedlichkeit erfolgte auf einer Skala von 0 (gar kein Unterschied) bis 3 (sehr großer Unterschied). Als Stimuli wurden dabei ein Schlagzeug- und ein Gitarrensignal in der Länge von vier bis acht Sekunden gewählt, die mit der jeweiligen Raumimpulsantwort gefaltet wurden. Zur Auswertung wurden 95%-Konfidenzintervalle mit dem Bootstrapping-Verfahren mit 2000 Samples gebildet [13].

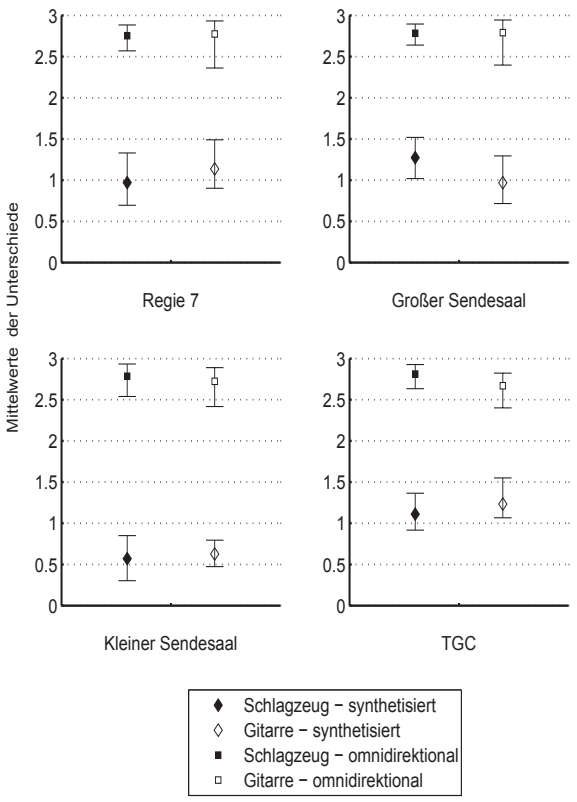

Abbildung 2: Unterschiedlichkeit (Mittelwerte und 95% Konfidenzintervalle) zwischen den Teststimuli und der Referenz (gemessene BRIRs) für die folgenden Räume: Abhörregie 7 des WDR, Großer Sendesaal WDR, Kleiner Sendesaal WDR, Übungsraum TGC Köln. In jeder Grafik sind die Beurteilungen der binaural synthetisierten Räume und die Bewertungen der einkanaligen synthetisierten Stimuli dargestellt. Der Vergleich erfolgte auf einer Skala von 0 (gar kein Unterschied) bis 3 (sehr großer Unterschied).

Die Ergebnisse zeigen, dass die erzeugten synthetischen BRIRs als deutlich weniger unterschiedlich zum Referenzdatensatz (BRIR) empfunden werden als es bei dem ebenfalls getesteten einkanaligen Ausgangsmaterial der Fall war. Besonders beim kleinen Sendesaal wurde die Darbietung der aus den synthetisierten binauralen Impulsantworten erzeugten Stimuli als sehr ähnlich bewertet. Die Tendenzen sind sowohl beim Gitarren- als auch beim Schlagzeugstimulus deutlich zu erkennen (vgl. Abbildung 2).

Die insgesamt recht geringen Unterschiede zwischen Referenz und Stimulus überraschen durchaus: So wurden z.B. Eigenschaften und Eintreffrichtungen der frühen Reflexionen nur basierend auf vorher festgelegten Reflexionsmustern angepasst. Auch wurde keine Information über diffus oder geometrisch reflektierende Wände des jeweiligen Raumes in die Synthese einbezogen.

Hinsichtlich der für die Unterschiede relevanten perzeptiven Merkmale zeigte sich, dass räumliche Attribute (z.B. Umhüllung, Lokalisierbarkeit, Ausdehnung der Schallquelle) maßgeblich für die veränderte Hörwahrnehmung der synthetischen Stimuli im Vergleich zur Referenz verantwortlich waren. Aber auch veränderte spektrale Eigenschaften und eine generell als etwas unnatürlicher empfundene Wahrnehmung der Räume spielten hier eine Rolle. Eine detaillierte Darstellung der Ergebnisse des SAQI-Tests ist zur Präsentation auf der ICSA 2015 geplant.

Statische Samples der präsentierten Stimuli sämtlicher im Hörversuch untersuchter Räume sind unter http://www.audiogroup.web.fh-koeln.de/DAGA2015.html abrufbar.

Zusammenfassung

Es wurde ein Verfahren entwickelt, implementiert und evaluiert, das es ermöglicht, aus omnidirektionalen gemessenen Raumimpulsantworten einen kompletten gedrehten Satz von binauralen Raumimpulsantworten zu synthetisieren. Dieser kann unter Nutzung der dynamischen Binauralsynthese zur Erzeugung von virtuellen auditiven Umgebungen genutzt werden. Das Verfahren ist für Anwendungen interessant, in denen eine plausible Darbietung hinreichend ist und bei denen eine aufwändige Vermessung der zu auralisierenden Räume nicht praktikabel ist.

Die Ergebnisse der hier vorgestellten psychoakustischen Untersuchungen belegen, dass eine durchaus hohe Ähnlichkeit der synthetisch erzeugten BRIRs zur Referenz erreicht wurde. Die Ähnlichkeiten sind abhängig von dem Raum und hängen auch davon ab, inwieweit willkürlich angenommene raumakustische Eigenschaften günstig gewählt wurden. Insgesamt waren die Unterschiede zur Referenz jedoch überraschend klein. Die Ergebnisse zeigen auch, dass für alle getesteten Räume eine gute Externalisierung gegeben war. Erwartungsgemäß wurden von den Probanden nur geringe Unterschiede im Bereich des diffusen Nachhalls festgestellt.

Die Untersuchungen wurden im Rahmen des vom BMBF in der Förderlinie Ingenieurnachwuchs geförderten Projekt 03FH005I3-MoNRa durchgeführt. Die Autoren danken für die Unterstützung.

Literatur

[1] Pörschmann, C., Bednarzyk, M. (1996). „Erzeugung eines synthetischen binauralen diffusen Nachhalls," in: Fortschritte der Akustik – DAGA '96, DEGA e.V., D – Oldenburg, pp. 400-401.

[2] Pörschmann, C., Zebisch, A. (2012). „Psychoakustische Untersuchungen zu synthetischem diffusen Nachhall," In: Proceedings of the VDT International Convention, Cologne, Germany.

[3] Lindau, A., Kosanke, L., Weinzierl, S. (2010). „Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses," Proc. of the 128th AES Convention, London.

[4] Menzer, F. (2010). „Binaural audio signal processing using interaural coherence matching," Thèse No. 4643, EPFL.

[5] Menzer, F., Faller, C., Lissek, H. (2011). „Obtaining Binaural Room Impulse Responses From B-Format Impulse Responses Using Frequency-Dependent Coherence Matching," IEEE Transactions on Audio, Speech and Language Processing, Vol. 19 (2), pp. 396-405.

[6] Lubberhuizen, W. (2007) „Near perfect reconstruction polyphase filterbank," Matlab Central. [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/15813, abgerufen am 26.3.2015.

[7] Pörschmann, C. , Schmitter, S., Jaritz, A. (2013). „Predictive Auralization of Room Modifications by the Adaptation of Measured Room Impulse Responses," In: Fortschritte der Akustik – AIA-DAGA 2013, DEGA e.V., D – Berlin, pp.1653-1656.

[8] Bernschütz, B. (2013). „A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU100," In: Fortschritte der Akustik – AIA-DAGA 2013, DEGA e.V., D – Berlin, pp.592-595.

[9] Stade, P., Bernschütz, B., Rühl, M. (2012). „A spatial audio impulse response compilation captured at the WDR broadcast studios," In: Proceedings of the VDT International Convention, Cologne, Germany.

[10] Lindau, A., Erbes, V., Lepa, S., Maempel, H.-J.; Brinkmann, F.; Weinzierl, S. (2014). „A Spatial Audio Quality Inventory (SAQI)," Acta Acustica united with Acustica vol. 100, pp. 984-994.

[11] Vazquez Giner, A. (2013). „Scale – A Software Tool for Listening Experiments," In: Fortschritte der Akustik – AIA-DAGA 2013, DEGA e.V., D – Berlin, pp.1316-1319.

[12] Vazquez Giner, A. (2015). „A Solution for Conducting Psychoacoustic Experiments with Real-time Dynamic Binaural Synthesis," In: Fortschritte der Akustik – DAGA 2015, DEGA e.V., D – Berlin.

[13] Bortz, J. (2005). „Statistik für Human- und Sozialwissenschaftler," 6. Auflage. Heidelberg: Springer Medizin.