

Exposé zur Masterarbeit

Schnelle Erfassung von mehrkanaligen binauralen Raumimpulsantworten für sich kontinuierlich ändernde Kopf-über-Torso-Orientierungen

Daniel Kahlhöfer

Matr.-Nr.: 347949
Email: daniel.kahlhoefer@gmail.com
Datum: 29. Januar 2015

Zusammenfassung

Mithilfe der dynamischen Binauralsynthese können akustische Umgebungen überzeugend reproduziert werden. Ein wichtiger Faktor ist dabei das Nachführen der zur Auralisierung verwendeten Impulsantwort entsprechend der Kopf-über-Torso-Orientierung des Hörers, da Menschen zur besseren Lokalisierung von Schallquellen kleine Kopfbewegungen ausführen. Die Erfassung der dafür nötigen binauralen Raumimpulsantworten ist aufgrund ihrer großen Anzahl bisher ein zeitaufwendiges Unterfangen, insbesondere wenn mehrere Schallquellen simuliert werden sollen.

Im Rahmen dieser Arbeit wird ein Verfahren zur beschleunigten Messung von kopfbezogenen Übertragungsfunktionen auf die Messung von binauralen Raumimpulsantworten übertragen. Ziel ist es, anhand einer Computersimulation mit einem akustischen Raummodell die optimale Parameterisierung und die Messbeschleunigung des Verfahrens in Abhängigkeit von Raumakustik, Störgeräuschen, Quellenanzahl und Verzerrungen zu untersuchen. Die Untersuchungsergebnisse sollen anschließend mit einer exemplarischen Messung in einem realen Raum evaluiert werden.

1 Einleitung und Fragestellung

Die Basis der Binauraltechnik bilden kopfbezogene Übertragungsfunktionen (head-related transfer functions, HRTFs), die die Schallübertragung einer Quelle im Freifeld über Torso, Kopf, Außenohr (Pinna) bis zum geblockten Ohrkanaleingang abbilden. HRTFs kodieren alle Informationen, die für die Lokalisation der Schallquelle notwendig sind. Wird die komplette Übertragungstrecke von der Schallquelle durch den Raum bis zum Ohr gemessen, spricht man von binauralen Raumübertragungsfunktionen bzw. im Zeitbereich von binauralen Raumimpulsantworten (binaural room impulse responses, BRIRs). Diese beinhalten außer dem Direktschall auch den von den Raumbegrenzungen und anderen Oberflächen (z.B. Möbel) reflektierten Schall.

Für eine überzeugende Auralisierung einer akustischen Szene genügt eine einzige BRIR nicht, da die wahrgenommenen Quellpositionen mit der Kopfbewegung mitwandern würden. Vielmehr muss die zur Auralisierung verwendete Impulsantwort dynamisch je nach Kopforientierung ausgetauscht werden. Dieses Verfahren wird dynamische Binauralsynthese genannt. Für jede Kombination von Quellposition und Kopf-über-Torso-Orientierung wird eine BRIR benötigt, woraus eine große Anzahl von Messungen resultiert. Werden die Einzelmessungen sequentiell durchgeführt, ist dies mit einem großen Zeitaufwand verbunden. Dieser ist vielleicht für kunstkopfbasierte, aber nicht für personalisierte Messungen vertretbar.

Zur Beschleunigung von HRTF-Messungen existieren bereits mehrere Verfahren. Die von Majdak et al. [1] vorgestellte *Multiple Exponential Sweep Method* (MESM) erzielt eine Messbeschleunigung durch die zeitliche Überlappung und Verschachtelung der Anregungssignale für mehrere Systeme. Enzner [2] benutzt adaptive Filter (normalized least mean square, NLMS), um zeitvariante Impulsantworten mehrerer Systeme gleichzeitig zu erfassen.

Eine beschleunigte Erfassung von BRIRs würde die Vermessung von akustischen Mehrkanal-Umgebungen wesentlich erleichtern. So wäre beispielsweise die Evaluation von lautsprecherbasierten Auralisierungssystemen zur Erzeugung virtueller akustischer Umgebungen wie Wellenfeldsynthese (WFS) oder Ambisonics, sowie von Beschallungsanlagen oder mehrkanaligen Klangkunstinstallationen bzw. Lautsprecherorchestern mit weniger Zeitaufwand verbunden.

Diese Arbeit befasst sich damit, wie die Anwendung des NLMS-Algorithmus auf die Messung von BRIRs für mehrere Quellen und sich kontinuierlich verändernde Kopf-über-Torso-Orientierungen übertragen werden kann. Dabei soll in einer Simulation mit akustischen Raummodellen untersucht werden, mit welchen Parametern der NLMS-Algorithmus die maximale Messbeschleunigung gegenüber der sequentiellen Messung, abhängig von Raumeinfluss, Umgebungsgeräuschen, Quellenanzahl und nichtlinearen Verzerrungen, erreicht. Außerdem sollen geeignete Fehlermaße für die Evaluation der Messergebnisse gefunden werden.

Anschließend soll der entwickelte Algorithmus in die Software des Messsystems FABIAN (Fast and Automatic Binaural Impulse response AcquisitioN) [3] des Fachgebiets Audiokommunikation der Technischen Universität Berlin integriert werden, um damit die Simulationsergebnisse anhand einer exemplarischen Messung mit mehreren Schallquellen zu evaluieren.

2 Stand der Forschung

Farina [4] zeigt, wie mit logarithmischen Sweeps die Impulsantwort eines Systems und gleichzeitig die nichtlinearen Verzerrungen gemessen werden können. Dies ist möglich, da logarithmische Sweeps die Eigenschaft haben, dass bei der Entfaltung nichtlineare Verzerrungen höherer Ordnung als separate harmonische Impulsantworten (HIRs) erscheinen, die durch Fensterung voneinander getrennt werden können.

Die von Majdak et al. [1] vorgeschlagene *Multiple Exponential Sweep Method* (MESM) nutzt dies aus, um in der Lücke zwischen Impulsantwort und HIR zweiter Ordnung weitere Impulsantworten zu platzieren. Weiterhin muss nach einer so verschachtelten Gruppe von Sweeps mit dem Beginn der nächsten Gruppe nicht gewartet werden, bis die Sweeps der letzten Gruppe abgeklungen sind, solange die HIR höchster Ordnung des ersten Sweeps nicht die letzte IR der vorangegangenen Gruppe beeinflusst. Diese beiden Verfahren – zeitliches Verschachteln und Überlappen – können die Messung von mehreren Systemen stark beschleunigen, wobei die Messbeschleunigung für kurze Impulsantworten und geringe nichtlineare Verzerrungen am größten ist. Somit ist MESM gut für die beschleunigte Messung von HRTFs geeignet. Die Messungen müssen jedoch noch immer in diskreten Winkeln erfolgen und bei gewünschter höherer Auflösung interpoliert werden.

Weinzierl et al. [5] zeigen, wie MESM für Sweeps mit beliebigem Spektrum generalisiert werden kann. Dadurch kann das Spektrum des Anregungssignals an das Spektrum der vorhandenen Störgeräusche angepasst werden, was einen größeren Störabstand (signal to noise ratio, SNR) ermöglicht. Das zeitliche Verschachteln ist damit nicht mehr effizient, da die Anteile durch nichtlineare Verzerrungen zeitlich weiter ausgedehnt sind. Für lange Nachhallzeiten der Impulsantwort ist das Verschachteln im ursprünglichen Verfahren aber nicht mehr möglich. In diesem Fall kann das generalisierte Verfahren und damit der Vorteil eines höheren SNR genutzt werden.

Werden HRTFs für diskrete Azimutwinkel gemessen, muss zwischen ihnen interpoliert werden, wenn eine feinere Auflösung benötigt wird. Die von Enzner [6] beschriebene Methode zur azimut-kontinuierlichen Messung von HRTFs beseitigt diese Notwendigkeit. Dazu wird der Proband von einem Lautsprecher, der in einer bestimmten Elevation montiert ist, mit dem Anregungssignal beschallt und dabei einmal um die vertikale Achse gedreht. Mittels des adaptiven NLMS-Filteralgorithmus, der auf zeitvariante Systeme wie dieses zugeschnitten ist [7], werden die Impulsantworten der Übertragungstrecken zu den beiden, sich in den Ohrkanälen befindlichen Mikrofonen verfolgt. Durch die Zuordnung der Zeit zum entsprechenden Orientierungswinkel des Probanden lässt sich so aus dem aufgenommenen Signal eine HRTF für jeden beliebigen Winkel – die Winkelauflösung ist nur durch die Abtastrate begrenzt – extrahieren. In [2] wird gezeigt, wie dieses Verfahren auf den vollsphärischen Mehrkanal-Fall ausgeweitet werden kann, um mit mehreren Lautsprechern auf diskreten Elevationen gleichzeitig zu messen.

Als optimale Anregungssignale für den NLMS-Algorithmus im Sinne der schnellsten Konvergenz haben sich *perfect sequences* erwiesen. Mit dem Spezialfall *perfect sweep* lässt sich zusätzlich die Eigenschaft von Sweeps nutzen, dass sie trotz hoher Energieeffizienz weniger Verzerrungen bei der Wiedergabe erzeugen als binäre oder

ternäre Pseudo-Rauschsequenzen [4, 8, 9].

Die Abweichung der mit NLMS gemessenen von der echten Impulsantwort hängt vom SNR, von der Änderungsrate der Impulsantwort – hier also von der Rotationsgeschwindigkeit – und von der Schrittweite des NLMS-Algorithmus ab. Enzner [6] zeigt, wie aus diesen Parametern die Ungenauigkeit der Impulsantwort vorhergesagt werden kann.

HRTFs werden für gewöhnlich in reflexionsarmen Räumen gemessen. In dieser Arbeit soll das Verfahren auf die Messung von BRIRs übertragen werden – der Raum soll also „mitgemessen“ werden. Dies wurde in ähnlicher Weise schon für normale Raumimpulsantworten gemacht [10], dort allerdings mit dem Ziel, zeitvariante Impulsantworten zu messen und sie als Simulation für Vergleichstests von Echokompensationssystemen zu benutzen.

3 Methoden

Um das Verhalten des NLMS-Algorithmus in verschiedenen Situationen bezüglich Raumakustik, Verzerrungen und Störgeräuschen einfach untersuchen zu können, soll die Messung zunächst am Computer simuliert werden. Der zu simulierende Signalweg besteht dabei aus Verstärkern, Lautsprechern, der Schallausbreitung im Raum und der Beeinflussung des Schallfeldes durch Kopf, Torso und Außenohr. Als Programmierumgebung kommt MATLAB¹, eine matrizenbasierte Plattform für numerische Berechnungen, zum Einsatz.

Verstärker und Lautsprecher können zusammen als zeitinvariantes System modelliert werden, welches aber in der Praxis nicht linear ist, sondern harmonische Verzerrungen produziert. Ein allgemeines Modell, das die Nichtlinearitäten mithilfe des generalisierten Hammerstein-Modells miteinbezieht, existiert bereits als Teil einer vorangegangenen Masterarbeit [11].

Die BRIRs, die die Schallausbreitung im Raum bis zum Außenohr beschreiben, sollen mittels raumakustischer Modellierung simuliert werden. Aufgrund der Zeitvarianz des Systems ist für jeden Kopforientierungswinkel bzw. dem ihm zugeordneten Zeitpunkt und für jede Quellposition die Erzeugung einer BRIR nötig. Die Ohrsignale, die das Ausgangssignal der simulierten Messung darstellen, sollen dann mithilfe von zeitvarianter Faltung [12] des Lautsprechersignals mit den zuvor simulierten BRIRs erzeugt werden.

Für die Simulation der BRIRs müsste idealerweise ein vollsphärischer HRTF-Datensatz für jede zu simulierende Kopf-über-Torso-Orientierung oder bezüglich der Kopf-über-Torso-Orientierung geringer aufgelöste Datensätze, zwischen denen interpoliert wird, benutzt werden. Basierend auf der Annahme, dass sich der NLMS-Algorithmus sehr ähnlich verhält, wenn anstatt des Kopfes der gesamte Körper gedreht wird, soll hier nur ein einziger HRTF-Datensatz mit einer Kopf-über-Torso-Orientierung von 0° benutzt werden.

Die in jeder Messung vorhandenen Umgebungsstörgeräusche werden durch weißes Rauschen modelliert, welches spektral an das in typischen Räumen bestehende Grundrauschen angepasst wird. Das so generierte Rauschen wird zum Schluss auf die

¹siehe <http://www.mathworks.de/products/matlab/index.html>

zuvor berechneten Ohrsignale addiert, da es als unabhängig vom Anregungssignal angesehen werden kann.

Mit dem so modellierten Signalweg kann nun die Messung mit verschiedenen Bedingungen simuliert werden. So sollen Nachhallzeit, harmonische Verzerrungen, SNR, Quellenanzahl, Rotationsgeschwindigkeit und Schrittweite des Algorithmus variiert werden, um jeweils das Adaptionsverhalten des ebenfalls in MATLAB zu implementierenden NLMS-Algorithmus zu untersuchen. Das Ziel ist es, die maximale Rotationsgeschwindigkeit und damit die maximale Messbeschleunigung für eine gegebene maximale Ungenauigkeit der BRIRs zu finden.

Um die Ergebnisse der Simulation zu validieren, soll das Messverfahren in die Software des bestehenden Messsystems FABIAN integriert und eine exemplarische Messung in einem hallenden Raum mit mehreren Lautsprechern durchgeführt werden. Im Gegensatz zur Simulation soll dabei ausschließlich der Kopf rotiert werden anstatt des ganzen Körpers.

4 Arbeits- und Zeitplan

Januar 2015	Literaturrecherche Exposé Sichtung der Vorarbeiten
Februar 2015	Implementierung Messkettensimulation und NLMS-Algorithmus
März 2015	Implementierung Messkettensimulation und NLMS-Algorithmus Dokumentation
April 2015	Simulation der Messungen + Auswertung Dokumentation
Mai 2015	Integration NLMS-Algorithmus in die Messsoftware von FABIAN
Juni 2015	Integration NLMS-Algorithmus in die Messsoftware von FABIAN Exemplarische Messung Dokumentation
Juli 2015	Dokumentation

Literatur

- [1] Majdak, Piotr; Peter Balazs und Bernhard Laback (2007): „Multiple exponential sweep method for fast measurement of head-related transfer functions.” In: *Journal of the Audio Engineering Society*, 55(7/8) S. 623–637.
- [2] Enzner, Gerald (2009): „3D-continuous-azimuth acquisition of head-related impulse responses using multi-channel adaptive filtering.” In: *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*. IEEE, S. 325–328.

- [3] Lindau, Alexander; Torben Hohn und Stefan Weinzierl (2007): „Binaural resynthesis for comparative studies of acoustical environments.” In: *Audio Engineering Society Convention 122*. Audio Engineering Society.
- [4] Farina, Angelo (2000): „Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique.” In: *Audio Engineering Society Convention 108*.
- [5] Weinzierl, Stefan; Andre Giese und Alexander Lindau (2009): „Generalized multiple sweep measurement.” In: *Audio Engineering Society Convention 126*. Audio Engineering Society.
- [6] Enzner, Gerald (2008): „Analysis and optimal control of LMS-type adaptive filtering for continuous-azimuth acquisition of head related impulse responses.” In: *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, S. 393–396.
- [7] Haykin, Simon (1996): *Adaptive Filter Theory*. 3. Auflage. Upper Saddle River, New Jersey: Prentice-Hall.
- [8] Telle, Aulis; Christiane Antweiler und Peter Vary (2010): „Der perfekte Sweep—Ein neues Anregungssignal zur adaptiven Systemidentifikation zeitvarianter akustischer Systeme.” In: *Proc. of Deutsche Jahrestagung für Akustik (DAGA)*. S. 341–342.
- [9] Antweiler, Christiane; Aulis Telle; Peter Vary und Gerald Enzner (2011): „Perfect-sweep NLMS for time-variant acoustic system identification.” In: *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, S. 517–520.
- [10] Antweiler, Christiane und H.-G. Symanzik (1995): „Simulation of time variant room impulse responses.” In: *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, vol. 5. IEEE, S. 3031–3034.
- [11] Fallahi, Mina (2014): *Simulation and analysis of measurement techniques for the fast acquisition of individual head-related transfer functions*. Masterarbeit, Fachgebiet Audiokommunikation, Technische Universität Berlin.
- [12] Margrave, Gary F. (1998): „Theory of nonstationary linear filtering in the fourier domain with application to time-variant filtering.” In: *Geophysics*, 63(1) S. 244–259.