

Alexander Fuß, Fabian Brinkmann, Thomas Jürgensohn, Stefan Weinzierl

Ein vollsphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen

Conference paper | Published version

This version is available at <https://doi.org/10.14279/depositonce-8801>



Fuß, Alexander; Brinkmann, Fabian; Jürgensohn, Thomas; Weinzierl, Stefan (2015): Ein vollsphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen. In: Fortschritte der Akustik - DAGA 2015: 41. Jahrestagung für Akustik, 16. - 19. März 2015 in Nürnberg. Berlin: Deutsche Gesellschaft für Akustik e.V. pp. 1114-1117.

Terms of Use

Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

Ein vollsphärisches Multikanalmesssystem zur schnellen Erfassung räumlich hochaufgelöster, individueller kopfbezogener Übertragungsfunktionen

Alexander Fuß¹, Fabian Brinkmann¹, Thomas Jürgensohn², Stefan Weinzierl¹

¹ Audio Communication Group TU Berlin, Einsteinufer 17c, D-10587 Berlin, Germany

Email: alexanderfuss@mailbox.tu-berlin.de, {fabian.brinkmann, stefan.weinzierl}@tu-berlin.de

² HFC Human-Factors-Consult GmbH, Köpenicker Straße 325, D-12555 Berlin, Germany

Email: juergensohn@human-factors-consult.de

Einleitung

Kopfbezogene Übertragungsfunktionen (head-related transfer functions - HRTFs) beschreiben die Freifeldschallübertragung von einer Quelle zu den Ohren einer Person und enthalten alle für das räumliche Hören verfügbaren Cues. Bedingt durch morphologische Unterschiede von Pinna, Kopf und Torso treten bei binauraler Wiedergabe nicht-individueller HRTFs unter anderem Lokalisations- und Klangfarbenfehler auf, weswegen die Verwendung individueller HRTFs diesbezüglich zu bevorzugen ist. Die Messung von HRTFs stellt jedoch in Bezug auf das Messsystem und die Signalverarbeitung eine Herausforderung dar. So existieren verschiedene Ansätze für deren Akquise, die sich in der Messmethode, der Auflösung der gewonnenen Daten, sowie der Messdauer unterscheiden. Letztere spielt bei der Akquise individueller Signale eine wichtige Rolle. Eine verkürzte Messzeit, die ungewollte Kopfbewegungen und Messartefakte verringert, kann sowohl genauer als auch für die Versuchspersonen komfortabler sein kann. Im Rahmen dieser Studie wird die Entwicklung und Evaluation eines Systems zur Messung vollsphärischer individueller HRTF Datensätze beschrieben. Bei simultaner Wiedergabe der Messsignale über 37 Lautsprecher und kontinuierlicher Drehung der Versuchsperson, konnte eine Messdauer von einer Minute bei hoher räumlicher HRTF-Auflösung erzielt werden. Zur Systemidentifikation wurden adaptive Normalized Least Mean Square (NLMS) Filter verwendet. Vorläufige Ergebnisse zeigten eine zufriedenstellende Übereinstimmung zu mit sequentiellen Sweeps gemessenen HRTFs.

Adaptive Systemidentifikation mit NLMS

Der adaptiven NLMS Algorithmus ermöglicht HRTF-Messungen für quasi-kontinuierliche Azimutwinkel durch kontinuierliche Drehung der zu vermessenden Person, wobei die Lautsprecher an diskreten Elevationen ϕ angeordnet sind [1]. Die kopfbezogenen Impulsantworten (head-related impulse responses - HRIRs) $\hat{\mathbf{h}}_{\phi}^{l/r}(n+1)$ mit einer Länge von N Samples werden dabei iterativ geschätzt

$$\hat{\mathbf{h}}_{\phi}^{l/r}(n+1) = \hat{\mathbf{h}}_{\phi}^{l/r}(n) + \mu \frac{e^{l/r}(n) \mathbf{x}_{\phi}^T(n)}{\sum_{\phi} \|\mathbf{x}_{\phi}\|_2^2}. \quad \forall \phi \quad (1)$$

Hierbei ist der Zeitpunkt $n+1$ Samples gleichbedeutend mit einem Azimuth von θ_{n+1} Grad. Die Schätzung

ist abhängig von der die Adaptionsgeschwindigkeit bestimmenden Schrittgröße μ , sowie vom Fehler $e^{l/r}(n)$, zwischen tatsächlichem und geschätztem Mikrofonsignal $y^{l/r}(n)$, bzw. $\hat{y}^{l/r}(n)$

$$e^{l/r}(n) = y^{l/r}(n) - \underbrace{\sum_{\phi} \hat{\mathbf{h}}_{\phi}^{l/r}(n) \mathbf{x}_{\phi}(n)}_{\hat{y}^{l/r}(n)}. \quad \forall \phi \quad (2)$$

$\hat{y}^{l/r}(n)$ ergibt sich dabei aus einer Skalarprodukt der in $\mathbf{x}_{\phi}(n)$ enthaltenen N letzten Samples des Anregesignals mit der aktuellen HRIR-Schätzung $\hat{\mathbf{h}}_{\phi}^{l/r}(n)$. Innerhalb der Stabilitätsbedingung $0 < \mu < 2$ ist die Schrittweite bei rauschbehafteten Messungen ein Kompromiss zwischen maximaler Konvergenzgeschwindigkeit für $\mu = 1$ und steigender Rauschunterdrückung des adaptiven Filters mit sinkendem μ [2].

Messlautsprecher

Die 37 verwendeten Lautsprecher wurden eigens für das Messsystem im Closed-Box-Design entwickelt. Die an sie gestellten Anforderungen waren (a) ein hoher Signalpegel bei möglichst niedrigen nichtlinearen Verzerrungen, um einen ausreichenden Signal-Rauschabstand zu erreichen, (b) eine breitbandige und zeitinvariante Übertragung, und (c) eine möglichst omnidirektionale Schallabstrahlung. Da letzteres unter realen Bedingungen nur angenähert werden kann, sollten zumindest die für die HRTF besonders relevanten Körperteile Kopf, Schultern und Torso innerhalb der Hauptkeule (-3dB-Bereich) des Lautsprechers liegen. Eine Simulation des Abstrahlverhaltens, angenähert durch eine kreisförmigen Kolbenmembran in einer schallharten Wand [3], ergab, dass dieses Kriterium für einen Messabstand von 1.5 m und eine 2-Zoll-Membran für Kopf und Schultern bis 20 kHz und für den Torso bis ca. 10 kHz erfüllt ist. Dabei stellt die Membrangröße in diesem Fall einen Kompromiss aus erreichbarem Signalpegel und Omnidirektionalität dar.

Außerdem wurde eine Abschätzung des Hochpassverhaltens mit Hilfe von Thiele-Small-Parametern [4] für 28 Treiber (1 – 3 Zoll) durchgeführt, um deren tieffrequentes Verhalten im Bereich um die Resonanzfrequenz in Abhängigkeit möglicher Gehäusevolumina zu simulieren. Anhand der Ergebnisse sowie von Herstellerangaben wurde eine Vorauswahl von fünf Treibern getroffen. Diese wurden – eingebaut in prototypischen Gehäusen

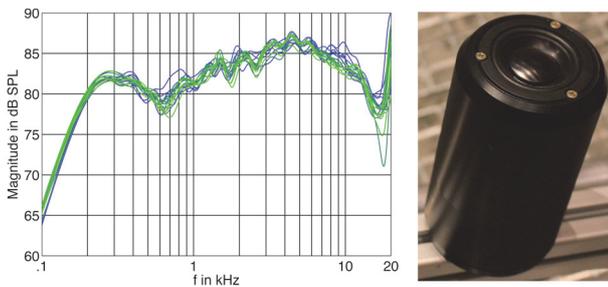


Abbildung 1: Links: Frequenzgang der Messlautsprecher im eingebauten Zustand (Sechstel-Oktav geglättet). Rechts: Beispielhafter Messlautsprecher im Gehäuse.

mit unterschiedlichen Volumina – hinsichtlich frequenzabhängiger Sensitivität, Impedanz und nichtlinearer Verzerrung vermessen. Zudem wurden 20-minütige Belastungstests durch wiederholte Messung des Frequenzgangs durchgeführt, um die Zeitvarianz der Treibermodelle zu analysieren.

Nach Begutachtung der Ergebnisse wurde der *Peerless NE65-04* 2-Zoll-Treiber ausgewählt, da er bei einer Leistungsaufnahme von 5 W eine ausreichende Sensitivität (82 dB SPL @ 1 kHz/1.3 m) bei vergleichsweise geringen nichtlinearen Verzerrungen (≤ -40 dB THD, $\forall f \geq 1$ kHz) und Zeitvarianzen (≤ 1 dB) aufwies. Zu Vermeidung von Gehäuseflexionen wurden zylinderförmige Gehäuse mit abgerundeten Kanten [5] aus Polyformaldehyd Copolymer-Hohlstäben (POM-C) hergestellt. Die Gehäuse umschließen ein Volumen von 0,2 Liter und wurden zur Bedämpfung von Resonanzen mit Polyesterwolle gefüllt. Die Messlautsprecher zeigten im eingebauten Zustand zwischen 170 Hz bis 20 kHz gut kompensierbare Frequenzgangsverzerrungen von ± 5 dB und eine geringe Varianz (≤ 3 dB @ 1 kHz, s. Abb. 1).

Messsystem

Das in Abb. 2 dargestellte Messsystem besteht aus einem Außen- und Innenkreis. Der aus 37 angewinkelten Konstruktionsprofilen zusammengesetzte und elastisch gelagerte Innenkreis trägt die Messlautsprecher, die in 10° Abständen aufgehängt und auf den Kreismittelpunkt ausgerichtet sind. Durch kleinere Profilstücke am Nord- und Südpol wurde hier der Abstand halbiert, wodurch eine 5° -Auflösung der Elevation nach einer kompletten Drehung der Versuchsperson erreicht wurde. Reflexionen durch die Kreiskonstruktion sowie benachbarte Lautsprecher wurden durch poröse Absorber bedämpft. Der Abstand der Lautsprechermembranen zum Kreismittelpunkt beträgt 1.47 m, wodurch Reflexionen von gegenüberliegenden Lautsprechern mit etwa 8.5 ms Verzögerung am Ohr eintreffen und ohne Beeinflussung der HRIR durch Fensterung eliminiert werden können. Um die Messung für HRIRs niedriger Elevationen möglichst wenig zu beeinflussen, wurden Drehteller und Messstuhl auf einem weitgehend schalldurchlässigen Gitterpodest montiert. Zur präzisen Positionierung der Versuchsperson im Mittelpunkt des Systems und über der vertikalen Drehachse des Drehtellers, ist der Messstuhl in Höhe und Position (vor-

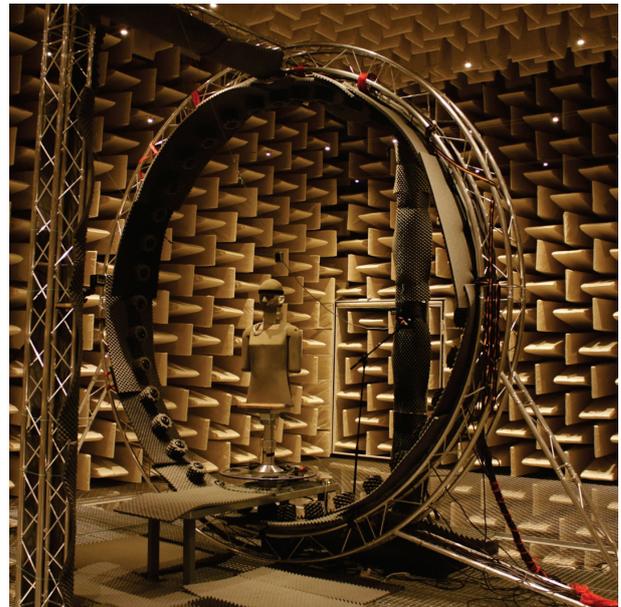


Abbildung 2: HRTF Messsystem mit FABIAN im reflexionsarmen Raum der TU Berlin.

ne/hinten) verstellbar. Die Positionierung kann dabei durch zwei auf die Arraymitte ausgerichtete Kreuzlinienlaser überprüft werden. Die Versuchsperson kann mit Umdrehungsgeschwindigkeiten von einer bis zehn Minuten rotiert werden. Aufgrund des Riemenantriebs durch einen Bürstenlosen DC-Motor ist die Drehung praktisch lautlos.

Das Messsystem wird über einen PC mit zwei kaskadierten Audiointerfaces (*RME HDSPe RayDAT*, *RME HDSPe AIO*) angesteuert. Die hier erzeugten 37-kanaligen Anregungssignale werden über fünf je achtkanalige DA-Wandler und Verstärker (*Behringer ADA8000*, *Cloud CXA 850*) an die Lautsprecher gesendet. Die Verstärker weisen vernachlässigbare Frequenzgangsschwankungen von $< \pm 0.5$ dB und nichtlineare Verzerrungen von ≤ -70 dB THD+N auf. Die Übersprechdämpfung benachbarter Kanäle beträgt ≤ 55 dB unterhalb und ≤ 45 dB oberhalb von 4 kHz. Als Mikrofonvorverstärker wurde eine *Lake People C360* eingesetzt.

Evaluation

Zur Validierung des Messaufbaus und der Signalverarbeitung wurden beispielhaft HRTF-Messungen mit dem Kopf und Torso Simulator FABIAN [6] durchgeführt, wozu die ITA-Toolbox [7] um eine Klasse zur Messung unter Verwendung von NLMS erweitert wurde.

Zunächst wurde hierfür der Einfluss der Filterlänge N auf das Ergebnis der Messung untersucht. Da sich die Adaptionsgeschwindigkeit der NLMS-Filter mit steigendem N verringert, sollte N möglichst klein gewählt werden, ohne dabei die Messung durch zeitliches Aliasing – das Auftreten, wenn die Filterlänge kürzer als die zu messende Impulsantwort ist [8] – zu beeinflussen. Eine im Mittelpunkt des Arrays gemessene Impulsantwort wies eine Länge von etwa 3000 Samples und einen SNR von

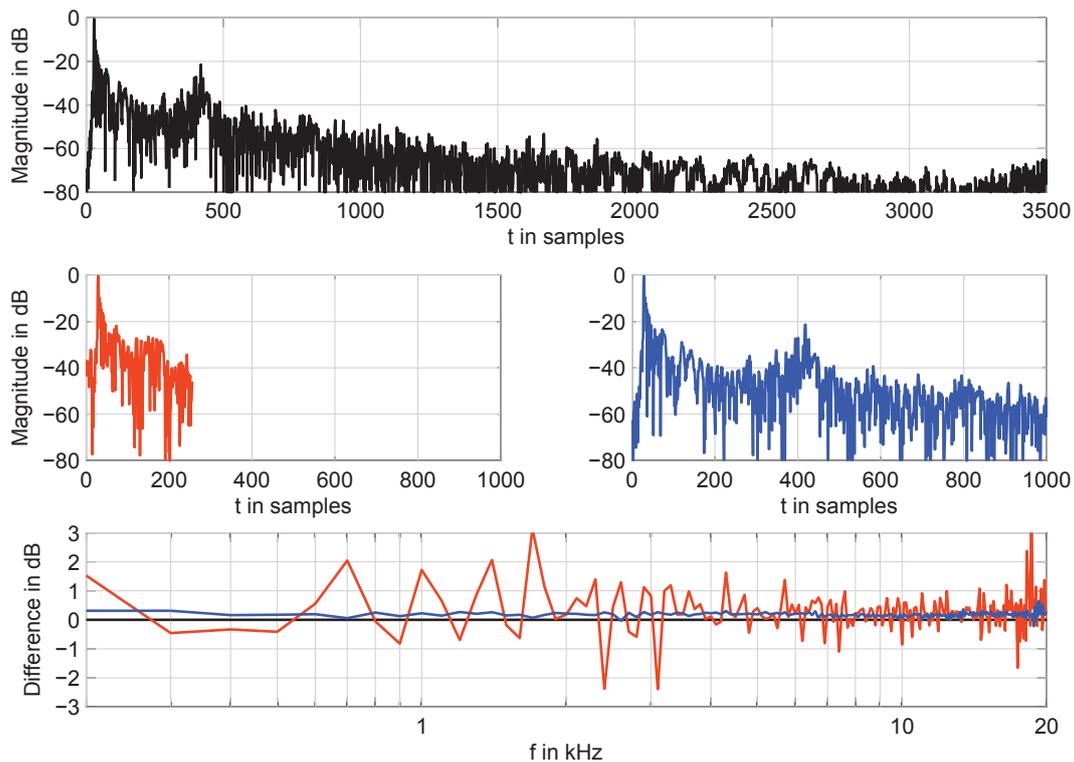


Abbildung 3: Einfluss der Filterlänge N auf das Messergebnis: In der Mitte des Messsystems gemessene Raumimpulsantworten (oben: $N = 3500$; mitte, links: $N = 256$; mitte, rechts: $N = 1024$) und aus der Begrenzung der Filterlänge resultierende spektrale Abweichungen (unten).

70 dB auf (Abb. 3, oben). Zudem zeigte sie eine durch gegenüberliegende Lautsprecher bedingte Reflexion ca. 350–400 Samples (8–9 ms) hinter dem Direktschall. Bei einer Filterlänge von $N = 256$ Samples verfälscht diese, bedingt durch zeitliches Aliasing, den Verlauf der Impulsantwort und führt zu einem Fehler von 3 dB im Betragsfrequenzgang (s. rote Kurven in Abb. 3). Erst bei einer Filterlänge von $N = 1024$ wird dieser Fehler auf unter 0.5 dB begrenzt, was als guter Kompromiss zwischen Adaptionsgeschwindigkeit und auftretendem Aliasing angesehen wurde (s. blaue Kurven in Abb. 3).

Für anschließende HRIR-Messungen mit einer Länge von 1024 Samples wurde FABIAN auf dem Messstuhl platziert und das interaurale Zentrum mit Hilfe der Kreuzlinien-Laser im Kreismittelpunkt positioniert. Zur Bestimmung der Azimutposition während der Drehung wurde ein elektromagnetischer Head-Tracker (*Polhemus PATRIOT*) verwendet. Als Anregesignale dienten perfekte Sweeps [9] der Länge $N \cdot C$, wobei C die Anzahl der Kanäle/Messlautsprecher bezeichnet. Die Anregung der Lautsprecher erfolgte dabei mit um $i \cdot N$ Samples verzögerten Signalen, mit $i = \{0, 1, \dots, C - 1\}$. Es wurden 37-kanalige Messungen mit Umdrehungsgeschwindigkeiten von $T_{360} = \{60, 120, 240\}$ Sekunden durchgeführt, bei denen FABIAN um 450° gedreht wurde. Die Messungen dauerten somit 75, 150 und 300 Sekunden. Die zusätzliche 90° -Drehung diente dabei der Adaption der Filter, sowie der Minimierung von Zeitvarianzen durch Erwärmung der Lautsprecher. Um durch Start- und Endpunkt der Messung entstehende Unstetigkeiten bei frontalen Quellpositionen zu vermeiden, wurde die Messung

bei einem Azimuth von 90° gestartet. Alle Ergebnisse wurden mit Schrittweiten von $\mu = \{0.25, 0.5, 1\}$ ausgewertet. Die Kompensation der in der Messstrecke involvierten elektroakustischen Übertrager erfolgte anhand zuvor durchgeführter Referenzmessungen, für die FABIANs Mikrofone im Zentrum des Lautsprecherarrays aufgebaut wurden. Um die Reflexion gegenüberliegender Lautsprecher zu entfernen, wurden die HRIRs anschließend auf eine Länge von 132 Samples (3 ms) gekürzt.

HRITFs in der Horizontalebene ($\phi = 0^\circ$) sind beispielhaft in Abb. 4 für das linke Ohr dargestellt. Ergebnisse für Umdrehungsgeschwindigkeiten ($T_{360} = 60$ s) weisen hochfrequent Unstetigkeiten über dem Azimutwinkel auf, die auf eine zu geringe Adaptionsgeschwindigkeit der Filter schließen lassen. Diese Unstetigkeiten nehmen mit steigender Umdrehungszeit ab und sind für $T_{360} = 240$ Sekunden nicht mehr sichtbar (vgl. Abb. 4, oben). Ein Vergleich der Schrittweite bei gleichbleibender Umdrehungszeit zeigt einen glättenden Effekt für kleine μ , der sich besonders hochfrequent zeigt (vgl. Abb. 4, unten).

Fazit

Ein 37-kanaliges Messsystem zur vollsphärischen und schnellen Erfassung individueller, kopfbezogener Übertragungsfunktionen wurde entwickelt und umgesetzt. Unter Verwendung von NLMS konnte mit diesem System ein HRITF-Datensatz mit einer quasi-kontinuierlichen Azimutaufösung und einer 5° -Auflösung der Elevation innerhalb von ein bis fünf Minuten gemessen werden. Erste Ergebnisse sind vielversprechend, ein Vergleich zu einer Referenz, bestehend

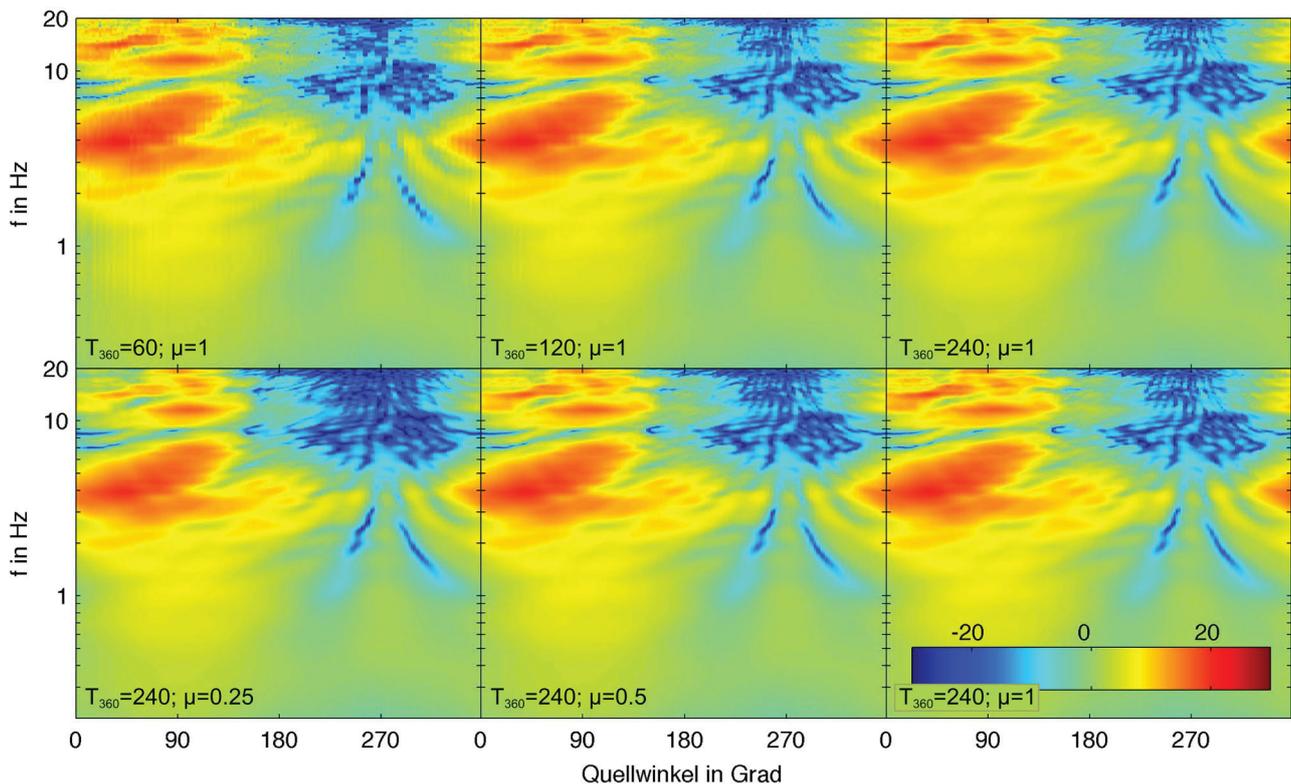


Abbildung 4: Betragsspektren von HRTFs des linken Ohres in der Horizontalebene. Umdrehungsgeschwindigkeiten T_{360} von 60, 120 und 240 Sekunden sind oben, Schrittweiten μ von 0.25, 0.5 und 1 unten dargestellt. Die Farbskala kennzeichnet die Amplitude in dB.

aus mit sequentiellen Sweeps gemessenen HRTFs anhand von physikalischen Fehlermaßen [10] steht noch aus.

In weiteren Untersuchungen soll zudem die Evaluation um HRTF-Messungen unter Verwendung zeitlich verschachtelter Sweepmessungen (*optimized MESM*) ergänzt werden, sowie die Erstellung einer umfangreichen HRTF-Datenbank inklusive anthropometrischer Daten erfolgen. Darüber hinaus soll der Einfluss optimierter adaptiver Filter auf Messergebnisse untersucht werden [11, 12].

Danksagung

Diese Studie wurde durch das DFG-Projekt *Simulation und Evaluation virtueller akustischer Umgebungen-SEACEN* (DFG WE 4057/3-2), sowie dem BMBF-Projekt *Orientieren, Informieren, Warnen. Orientierungshilfe für Blinde – OIWOB* gefördert. Wir danken Dr.-Ing. Roman Tschakert vom Institut für Strömungsmechanik und Technische Akustik der TU Berlin für die Unterstützung der Messreihe.

Literatur

- [1] Enzner., G.: 3D-continuous-azimuth acquisition of head related impuls responses using multi-channel adaptive filtering. *IEEE WASPAA* (2009), New Paltz, USA.
- [2] Paleologu, C., Ciochina, S., Benesty, J.: Variable Step-Size NLMS Algorithm for Under-Modeling Acoustic Echo Cancellation. *IEEE Signal Process. Lett.*, 15 (2008), 5-8.
- [3] Möser, M.: *Technische Akustik*. Springer, Berlin, Heidelberg, Germany, 8. Aufl. (2009).
- [4] Small., R.H.: Closed-Box Loudspeaker Systems Part I: Analysis. *J. Audio Eng. Soc.*, 20 (1972), 798-808.
- [5] Masiero, B., Pollow, M., Fels, J.: Design of a fast broadband individual head-related transfer function measurement system. *Forum Acusticum*, Aalborg, Denmark, (2011), 2197-2202.
- [6] Lindau, A., Hohn, T., Weinzierl, S.: Binaural resynthesis for comparative studies of acoustical environments. *122nd AES Convention*, Vienna, Austria, (2007).
- [7] ITA-Toolbox, Institute of Technical Acoustics, RWTH Aachen University: <http://ita-toolbox.org> (Aufgerufen am 27.3.2015).
- [8] Antweiler, C., Dörbecker, M.: Perfect sequence excitation of the NLMS algorithm and its application to acoustic echo control. *Ann. Télécommun.*, 49 (1994), 386-397.
- [9] Antweiler, C., Telle, A., Vary, P., Enzner, G.: Perfect-Sweep NLMS for time-variant acoustic system identification. *IEEE ICASSP* (2011).
- [10] Fallahi, M., Brinkmann, F., Weinzierl, S.: Simulation and analysis of measurement techniques for the fast acquisition of head-related transfer functions. *Fortschritte der Akustik - DAGA*, Nürnberg, (2015).
- [11] Antweiler, C., Kühl, S., Sauert, B., Vary, P.: System identification with perfect sequence excitation - Efficient NLMS vs. inverse cyclic convolution. *ITG-Fachbericht 252: Speech Communication*, VDE Verlag GmbH, Berlin, Offenbach, Germany (2014).
- [12] Hahn, N., Spors, S.: Identification of dynamic acoustic systems by orthogonal expansion of time-variant impulse responses. *IEEE 6th Int. Symposium on Communications, Control, and Signal Processing*, (2014), 161- 164.