

# FABIAN – Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden

Alexander Lindau<sup>1</sup>, Stefan Weinzierl<sup>1</sup>

<sup>1</sup> *Fachgebiet Audiokommunikation, Technische Universität Berlin, Einsteinufer 17, Sekr. EN-08, 10587 Berlin, Deutschland*  
*Email: tonalex@gmx.de, stefan.weinzierl@tu-berlin.de*

## Einleitung

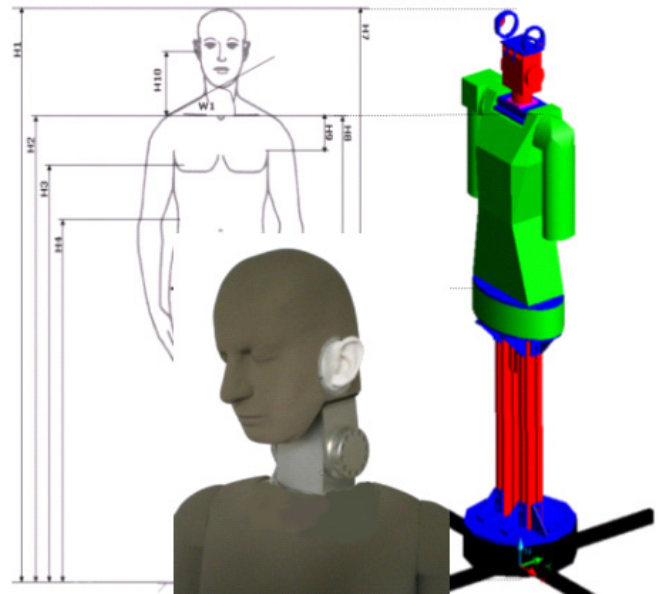
Zur Messung kopfbezogener Impulsantworten als Grundlage für binaurale Simulationen gibt es eine Reihe von eingeführten Systemen, deren Leistungsfähigkeit in der Regel durch Hörversuche zur Lokalisation virtueller, auralisierter Schallquellen belegt wird [1]. Zwei neuere Systeme ermöglichen eine automatische Kopffrotation in der Horizontalebene [2] [3], da die von Hand nachgeführte Vermessung kompletter Datensätze von binauralen Raumimpulsantworten (BRIRs) zeitraubend und tendenziell mechanisch ungenau ist. Um vergleichende Auralisationen für mehrere Quellpositionen, verschiedene Hörerplätze und mehrere Freiheitsgrade der Bewegung von Kopf und Torso durchführen zu können, ist neben den mechanischen Freiheitsgraden auch die Geschwindigkeit und Automatisierbarkeit des Systems von Bedeutung. Für derartige Anwendungen wurde ein bestehendes System [2] durch eine Software-gesteuerte Mechanik und ein integriertes Messsystem erweitert und durch einen Hörversuch perzeptiv evaluiert.

## System

Der neu entwickelte Kunstkopf-Torso-Simulator FABIAN (Fast and Automatic Binaural Impulse response AcquisitioN) ermöglicht die Rotation und Elevation des Kopfs über dem Torso mittels eines servomotorischen Halsgelenks. Der dritte rotatorische Freiheitsgrad (laterale Flexion) kann durch inverse Montage des Halsgelenks ermöglicht werden. Die Rotation des gesamten Torsos auf einem Drehteller ermöglicht somit beliebige Orientierungen von Kopf und Schulter, wie sie spätestens dann relevant werden, wenn dem Hörer eine freie Bewegung im Raum ermöglicht werden soll. Das verwendete Halsgelenk erlaubt eine hochgenaue ( $0,002^\circ$ ) und schnelle Positionierung des Kopfes. Der Torso wurde auf Grundlage anthropometrischer Daten entworfen und repräsentiert den Median der deutschen Bevölkerung im Alter zwischen 18-65 Jahren (DIN 33402-2 E), während der Kopf ein bereits in [2] beschriebener, individueller Abguss ist (Abbildung 1). Zur Messung in sitzender Haltung kann der Torso ohne Drehteller und Unterbau benutzt werden, sämtliche Zuleitungen (Steuerung, Energie und Mikrofonsignale) verlaufen innerhalb der Torsos. Die Außenohren sind individuelle Silikon-Abgüsse, als Mikrofone kommen sehr rauscharme Miniaturkondensatormikrofone vom Type DPA 4060 ( $\varnothing$  5.6 mm) zum Einsatz, die bündig und körperschallentkoppelt in den Boden der Ohrmuschel eingelassen sind.

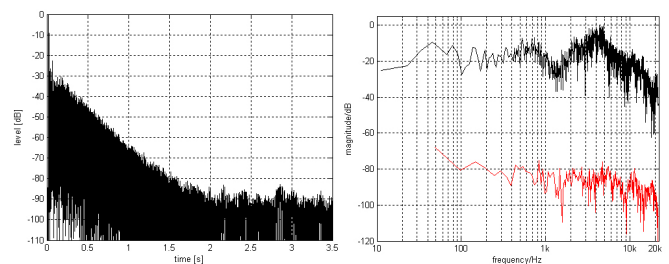
Eine spezielle Matlab®-Applikation kontrolliert die mechanische Steuerung und die Messung. Die verwendeten ASIO-Treiber unterstützen 44.1-96 kHz Samplerate und 32 bit floating point Dateiformat. Die Messung erfolgt mit kontinuierlichen, frequenzgewichteten Sinussweeps konstanter Amplitude und anschließender spektraler Entfaltung [4]. Die wählbare sukzessive Anregung von bis zu 8 Quellen vor jeder Neuausrichtung des Kopfes beschleunigt den Messablauf. Für die Messroutine kann ein beliebiges Raster an Winkelschritten, jeweils unabhängig für die Horizontal- und Vertikalebene vorgegeben werden. Zur

Qualitätssicherung der Messautomation werden Eingangssignalübersteuerungen und Unterschreitungen eines vorgegebenen Mindest-SNR automatisch berücksichtigt.



**Abbildung 1:** Bemaßte anthropometrische Entwurfsvorlage und CAD-Entwurf des modularen Torsos (Farben kennzeichnen unterschiedliche Materialien/Baugruppen)

Für den im Weiteren beschriebenen Hörversuch wurden im Audimax der TU Berlin BRIR-Datensätze an einer mittigen Sitzposition in der 3. Reihe für je eine Quelle (Messlautsprecher) frontal auf der Bühne (hor./ver.  $0^\circ/0^\circ$ ) sowie eine weitere in größerer Entfernung ( $120^\circ/30^\circ$ ) gemessen. Bei einer Winkelauflösung von  $1^\circ/5^\circ$  in einem Bereich von  $\pm 75^\circ/\pm 30^\circ$  wurden 3926 BRIRs von je 3.5 s Dauer in einer Messdauer von etwa 24 h akquiriert. Mit nur drei Mittelungen wurde ein durchschnittlicher SNR von 85 dB erreicht (Abbildung 2).



**Abbildung 2:** am linken Ohr gemessene BRIR, Kopf frontal ausgerichtet (links: ETC, rechts: Raumübertragungsfunktion und Grundstörpektrum)

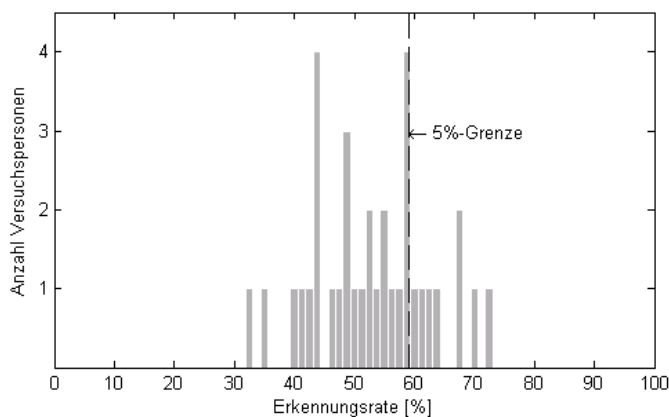
Für Auralisationen mit einem elektrostatischen Kopfhörer (Stax SR 202) wurden die BRIRs durch ein die gemessene Kopfhörerübertragungsfunktion approximierendes und mittels eines frequenzgewichteten LMS-Kriteriums optimiertes FIR-Filter [5] entzerrt.

## Hörversuch

Um eine integrale, nicht nur am Kriterium der Lokalisation orientierte Bewertung der mit FABIAN gemessenen Impulsantworten zu erhalten, wurde ein direkter AB-Vergleich von Realität und binauraler Simulation durchgeführt. Damit ein akustisch besonders transparenter Kopfhörer (Stax SR 202) während des Versuchs aufgesetzt bleiben konnte, trug ihn auch der Messroboter während der Messung. Für den Hörversuch kam ein einfaches AB-Forced-Choice-Testdesign zur Anwendung, bei dem jeder Versuchsperson eine randomisierte Folge von 80 Stimuluspaaren aus je einer Simulation und einer „realen“ Darbietung präsentiert wurde. Bei einer Stimulsdauer von 6 s wurden folgende Bedingungen variiert: Stimulustyp, Quellposition (vorn/hinten), sowie die Anzahl der Freiheitsgrade durch Abschalten des vertikalen Kopftrackings. Aus der EBU SQUAM CD bzw. der Archimedes Sammlung wurden fünf Stimuli (männl. u. weibl. Sprecher, akust. Gitarre, Schlagzeug, Trompete) ausgewählt. Am Hörversuch nahmen 35 Personen zwischen 21 und 40 Jahren teil, welche aufgrund der vorangehenden Teilnahme an einer Lehrveranstaltung zur virtuellen Akustik als fachlich geschulte Hörer angesehen werden können.

## Ergebnisse

Die Erkennungsrate (ER) über alle 2800 Einzelentscheidungen betrug 52,9%. Dies ist ein geringer, jedoch – getestet an einer  $\chi^2$ -Verteilung – statistisch signifikanter Unterschied gegenüber der Ratewahrscheinlichkeit von 50%. Aufgeschlüsselt nach einzelnen Versuchspersonen (VP) erhält man eine relativ symmetrische Verteilung der ER im Intervall von 32% bis 72% (Abbildung 3). Getestet gegen eine  $\chi^2$ -verteilung für 80 Einzelmessungen ergab sich für 8 von 35 VP eine überzufällig ( $\alpha = 5\%$ ) hohe ER von mehr als 59%.



**Abbildung 3:** Häufigkeitsverteilung der Erkennungsraten der 35 Versuchspersonen im AB-Unterschiedstest mit 5%-Grenze für einseitigen  $\chi^2$ -Test

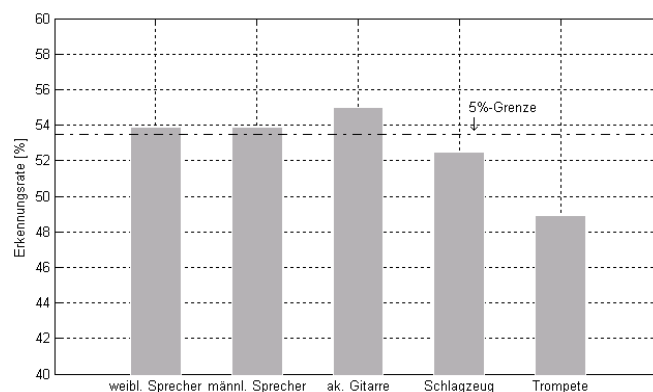
Die Tatsache, dass 5 dieser 8 VP im anschließend ausgefüllten Fragebogen von sich aus angaben, dass sie nur geraten hätten, lässt die Vermutung zu, dass sich diese VP zwar Unterschiede besser einprägen und diese relativ konsistent den beiden Alternativen Realität bzw. Simulation zuordnen konnten, dass die korrekte Zuordnung aber weitgehend zufallsbedingt war. Dies wird auch durch die Symmetrie des Histogramms in Abbildung 3 nahe gelegt.

Auf die Frage, welche Merkmale für die Unterscheidung zwischen Realität und Simulation herangezogen wurden, wurden von den 9 VP, die zwischen beiden Situationen konsistent unterscheiden konnten (nur VP außerhalb der entsprechenden zweiseitigen 5%  $\chi^2$ -Werte von 39% bzw. 61% ER in Abbildung 3), folgende Qualitäten genannt:

1. Spektrale Unterschiede (5x)
2. Lokalisationsunterschiede (4x)
3. Unterschiede im Nachhallverhalten (2x)
4. Energieunterschiede am quellabgewandten Ohr (2x)

Trotz aufwändiger Entzerrung der Kopfhörer-Mikrofonstrecke blieben spektrale Unterschiede somit das offensichtlichste Unterscheidungsmerkmal. Da Lokalisationsunterschiede hauptsächlich für die  $130^\circ/30^\circ$ -Quelle angegeben wurden, sind diese höchstwahrscheinlich auf geringe Fehlanpassungen der interauralen Laufzeitdifferenzen durch Verwendung nichtindividueller HRTFs zurückzuführen. Bei der Beurteilung der Gesamtpformance der Simulation ist allerdings festzuhalten, dass die genannten Merkmale keiner VP ermöglichten, die Simulation mit einer Sicherheit von mehr als 72% zu erkennen.

Im Hinblick auf zukünftige Hörversuche wurde die ER auch bezüglich der verschiedenen Audioinhalte ausgewertet.



**Abbildung 4:** Erkennungsrate in Abhängigkeit vom Stimulustyp

Hier zeigte sich, dass die akustische Gitarre sowie die Sprachsamples am besten zur Aufdeckung potentieller Artefakte der Simulation geeignet waren. Weiterhin war bei der frontalen Quelle eine geringfügig höhere Erkennungsrate im Vergleich zur weiter entfernten Quelle bei  $130^\circ/30^\circ$  zu beobachten (53,9% gegen 51,8%). Die Relevanz des zweiten Freiheitsgrads der genutzten BRIRs wird Thema weiterer Untersuchungen sein.

## Literatur

- [1] Minaar P, Olesen SK, Christensen F, Møller H (2001) Localization with Binaural Recordings from Artificial and Human Heads, *J. Audio Eng. Soc.* 49/5:323-336
- [2] Moldrzyk C, Ahnert W, Feistel S, Lentz T, Weinzierl S (2004) Head-tracked Auralization of Acoustical Simulation. 117<sup>th</sup> AES Convention, San Francisco, Preprint 6275
- [3] Christensen F, Martin G, Minaar P, Song WK, Pedersen B, Lydolf M (2005) A Listening Test System for Automotive Audio – Part 1: System Description. 118<sup>th</sup> AES Convention, Barcelona
- [4] Müller S, Massarani P (2001) Transfer-Function Measurement with sweeps. Director's Cut including previously unreleased material and some corrections. Originally in: *J. Audio Eng. Soc.* 49:443-471
- [5] Kirkeby O, Nelson PA (1999) Digital Filter Design for Inversion Problems in Sound Reproduction. *J. Audio Eng. Soc.* 47/7/8:583-595