Entwicklung und Evaluation eines Mikrofonarrays für die Aufnahme von räumlichen Schallfeldern nach dem Motion-Tracked Binaural (MTB) Verfahren

Felicitas Fiedler[1], David Ackermann[1], Fabian Brinkmann[1], Martin Schneider[2], Stefan Weinzierl[1]
[1]*Fachgebiet Audiokommunikation, TU Berlin,*
felicitas.fiedler@gmx.net, david.ackermann@tu-berlin.de, fabian.brinkmann@tu-berlin.de,
stefan.weinzierl@tu-berlin.de
[2]*Georg Neumann GmbH Berlin, martin.schneider@neumann.com*

Einleitung

Das Motion-Tracked Binaural (MTB) Verfahren erlaubt die pseudobinaurale Aufnahme und Wiedergabe von räumlichen Schallereignissen. Im Gegensatz zu synthetischen Szenen, die durch binaurale Impulsantworten generiert werden, ist auch eine dynamische, auf die Kopfbewegung des Hörers nachgeführte Wiedergabe von realen Schallfeldern möglich. Umgesetzt wird dies durch ein 16-kanaliges Mikrofonarray, wie es in einer Kooperation der TU Berlin mit der Sennheiser GmbH gebaut wurde. In zwei Hörversuchen zeigte sich, dass das MTB-Verfahren eine überraschend plausible Reproduktion räumlich akustischer Szenen mit nur geringfügigen perzeptiven Einbußen erlaubt.

Technischer Aufbau und Signalverarbeitung

Das MTB-Verfahren basiert auf einem Mikrofonarray, bei dem die Kapseln auf dem Äquator einer Kugel angebracht sind, deren Größe sich am Ohrabstand eines menschlichen Kopfes orientiert [1]. So beträgt ihr Durchmesser 17,6 cm. Der Körper des Arrays wurde mit dem SLS (Selective Laser Sintering) Verfahren gefertigt und besitzt aufgrund des als Material verwendeten Kunststoffes schallharte Eigenschaften. Er besteht aus einer Ober- und Unterschale sowie einem Ring, der zwischen den Schalen sitzt und in den 16 omnidirektionale Elektretkapseln (Sennheiser KE 14) eingebracht sind. Im Inneren befindet sich eine Halterung für die Vorverstärker-Platinen, welche durch ein Stahlblech-Gehäuse geschirmt wird. Aufgrund der geringen Betriebsspannung der Kapseln wurde ein Spannungsteiler auf den Platinen verbaut, sodass das Array mit einer Phantomspannung von 48 V gespeist werden kann. Zugleich wird das Signal auf den Platinen symmetriert und geht in ein 16-kanaliges Multicorkabel über, an dessen Ende eine entsprechende Anzahl an XLR-Steckern angebracht ist. Die Selektion der Mikrofonkapseln erfolgte über einen Vergleich ihrer im uneingebauten Zustand gemessenen Amplitudengänge. Diese unterschieden sich über einen Frequenzbereich von 20 Hz bis 15 kHz nicht mehr als 1 dB in ihrer Amplitude.

Bei der binauralen Wiedergabe werden zwei gegenüberliegende Mikrofone benutzt, deren Auswahl durch Head-Tracking auf die Kopfposition des Hörers nachgeführt wird. Liegt die Position zwischen zwei Mikrofonen, werden die Signale aus den benachbarten Kapseln interpoliert. Die akustische Szene kann somit dynamisch erlebt werden. Für die Richtungslokalisation auf der horizontalen Ebene sind hierbei insbesondere die interaurale Pegeldifferenz (interaural level difference, ILD) im tieffrequenten und die interaurale Laufzeitdifferenz (interaural time difference, ITD) im hochfrequenten Bereich verantwortlich [1].

Abbildung 1: Außen- und Innenansicht des MTB-Mikrofonarrays

Für die Interpolation kommt ein Algorithmus zum Einsatz, der unter Nutzung der Short Time Fourier Transform (STFT) im tieffrequenten Bereich linear im Zeitbereich und im hochfrequenten Bereich spektral interpoliert. Diese Methode erwies sich in einer Vor-Studie von Lindau & Roos als qualitativ überlegen [2].

Hörversuch

In einem ersten Hörversuch (N = 20, Ø = 29 Jahre, musikalisch erfahrene Hörer) wurde die Wiedergabequalität des MTB-Arrays durch das Rating von acht Qualitäten des Spatial Audio Quality Inventory (SAQI) getestet [3], die sich in einem Vorversuch als kritisch erwiesen haben. Testbedingungen waren zwei diffusfeldentzerrte MTB-Aufnahmen mit 8 und 16 Kanälen. Als Referenz diente eine diffusfeldentzerrte dynamische Binauralsynthese mit den BRIRs des Kunstkopfes FABIAN [4], welche für zwei Lautsprecherpositionen aufgenommen wurden. Alle Signale wurden mit einer Grenzfrequenz von 50 Hz hochpassgefiltert. Getestet wurden somit 8 Bedingungsvariationen in einem vollständigen Versuchsdesign mit Messwiederholungen und den Faktoren Inhalt (Sprache/Rauschen), Raum (kurze/lange Nachhallzeit) und Quellposition (LS 1: frontal, LS 2: 70° Azimut und 17° Elevation). Für die Wiedergabe wurden Sennheiser HD 800 Kopfhörer mit Head-Tracking-System (Polhemus Patriot) eingesetzt. Um eine bestmögliche Auralisation zu erreichen, wurde außerdem eine Entzerrung der Kopfhörer vorgenommen [5] und die interaurale Laufzeitdifferenz durch ITD-Extraktion und -Manipulation individuell angepasst [6]. So wurde bei den Teilnehmern im Vorfeld die Intertragusdistanz gemessen. Entsprechend der Mixing Time wurden außerdem die dynamischen und die statischen Teile der Raumimpulsantorten voneinander getrennt [7]. Diese

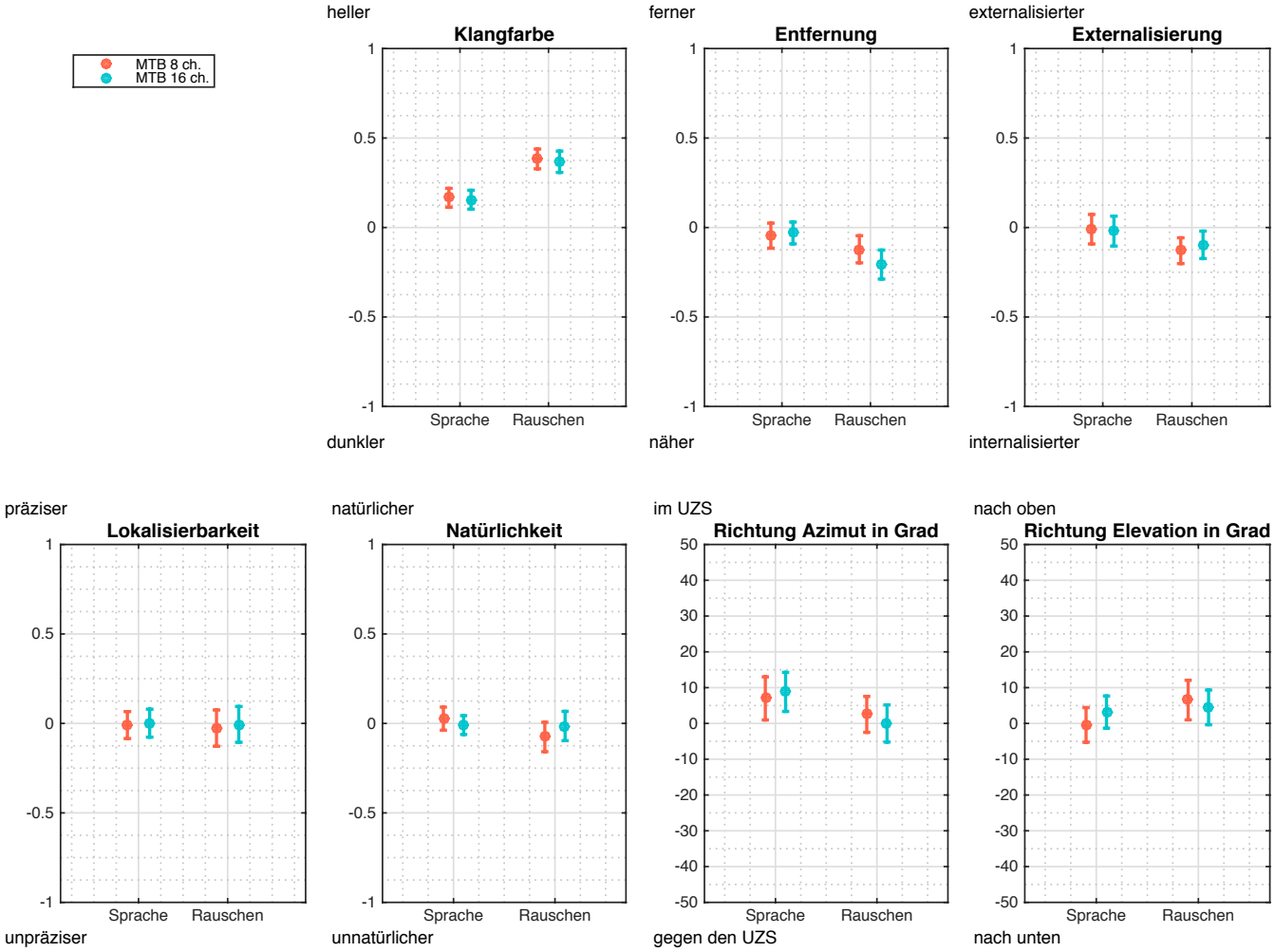

Abbildung 2: Mittelwerte der Bewertungen aus dem SAQI Hörversuch mit 95%-Konfidenzintervall

Vorgehensweise lehnt sich an den Versuchsaufbau von Lindau & Weinzierl zur Beurteilung der Plausibilität virtueller Umgebungen an [8]. Als Benutzeroberfläche für den Versuch wurde eine auf Matlab basierende Software verwendet, mit welcher es möglich war, den Unterschied hinsichtlich des jeweiligen SAQI-Itmes zwischen Testsignal und Referenz auf einer Skala anzugeben.

In einem zweiten Hörversuch (N = 11, Ø = 29 Jahre, musikalisch erfahrene Hörer) wurde die ‚Plausibilität' der Wiedergabe in zwei verschiedenen Gruppen für eine 8- und 16-kanalige Aufnahme getestet. Den Teilnehmern wurden 100 durch das MTB Verfahren reproduzierte und 100 reale Stimuli mit unterschiedlichem Audioinhalt (Sprache, Solo-Instrument, Popularmusik) in randomisierter Reihenfolge und mit einer Länge von etwa 5 s vorgespielt. Die Hörer mussten in einem Ja/Nein-Paradigma nach jedem Stimulus entscheiden, ob das Signal von einem realen Lautsprecher abgespielt wurde. Für die Wiedergabe wurde ein extra-auraler Kopfhörer (BK2/11, [9]) mit einem Polhemus Patriot Head-Tracking-System verwendet, der während des Hörversuchs nicht abgesetzt werden musste. Die MTB-Signale wurden mit einer Grenzfrequenz von 50 Hz hochpassgefiltert, diffusfeldentzerrt, mit dem Entzerrungsfilter des Kopfhörers und dem Diffusfeldentzerrungsfilter von FABIAN versehen.

Ergebnisse & Diskussion

Im Hinblick auf die mit dem SAQI erhobenen perzeptiven Qualitäten zeigten sich signifikante Unterschiede zwischen dem MTB Signal und der binauralen Referenz für den Sprachstimulus nur in einer geringfügigen Veränderung der Klangfarbe und der azimutalen Richtung, während etwa der Grad an Externalisierung, die Lokalisierbarkeit und die Natürlichkeit des Klangeindrucks auf gleichem Niveau wie die binaurale Referenz lagen. Etwas größer waren die Unterschiede für das Rauschsignal, wo eine geringfügige aber signifikante Abnahme der Externalisierung auftrat.

Da die Positionen von FABIAN und dem MTB-Mikrofonarray bei der Aufnahme auf azimutaler Ebene 6° versetzt waren, wurden die Daten im Nachhinein um diesen Wert korrigiert. Dass im Ergebnis trotzdem ein Unterschied bestehen bleibt, könnte an der Schwierigkeit gelegen haben, einen exakten Wert für die wahrgenommene Positionsverschiebung der Schallquelle anzugeben.

Im Hinblick auf die Plausibilität wurden die Ja/Nein-Entscheidungen des Tests mit Hilfe der Signalentdeckungstheorie (SDT) analysiert. Hierbei ergab sich eine Sensitivität von $d' = 0.57$ für die MTB-Wiedergabe mit 8 Kanälen, eine Sensitivität von $d' = 0.43$ für die Wiedergabe mit 16 Kanälen. Wandelt man diese Werte in entsprechende Erkennungsraten

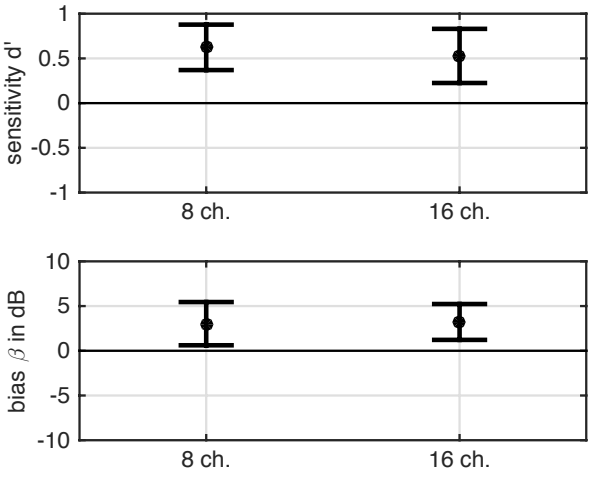

Abbildung 3: Mittelwerte von Sensitivität und Bias

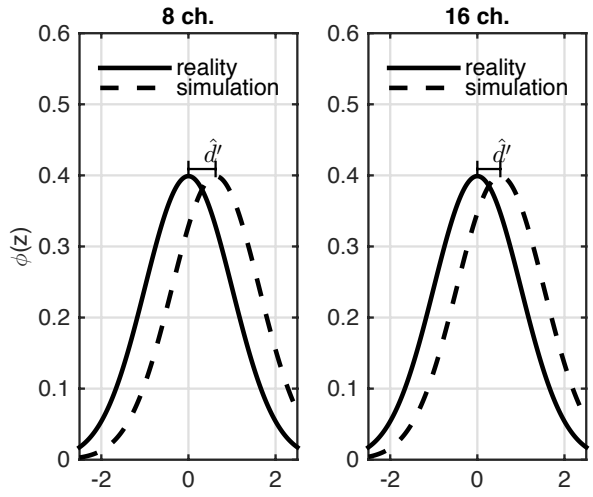

Abbildung 4: Dichteverteilungen des Antwortverhaltens

für ein 2AFC-Paradigma um, ergeben sich Werte von *Pc* = 0,66 (8 Kanäle) bzw. *Pc* = 0,62 (16 Kanäle). Sind Hörer also aufgefordert, eine durch das MTB-Verfahren übertragene, akustische Szene als ‚nicht real' zu erkennen, so liegt die Erkennungsrate gegenüber einem Zufallsniveau von 50 % nur überraschend geringfügig erhöht, bei 62 %, sofern mit ausreichender Kanalzahl und optimaler Interpolation bei dynamischer Wiedergabe gearbeitet wird. Da sich eine auf binauralen Raumimpulsantworten beruhende Resynthese solcher Szenen, mit der Lindau & Weinzierl eine Erkennungsrate von 51 % erreichen konnten [8], für die Live-Aufnahme und Live-Übertragung von räumlichen Schallfeldern nicht einsetzen lässt, bietet das MTB-Mikrofon somit eine vielversprechende Alternative.

Überraschenderweise konnte von den Probanden trotz fehlender Pinna- und Torsoeinflüsse außerdem die Elevation der Schallquellen wahrgenommen werden, wie bereits bei Algazi et al. [10] untersucht wurde. Eine mögliche Erklärung könnte in den Veränderungen von ITDs und ILDs durch die Kopfbewegungen auf horizontaler Ebene liegen [11].

Literatur

[1] Algazi, V.R.; Duda, R.O.; Thompson, D.M. (2004). Motion-Tracked Binaural Sound. *J. Audio Eng. Soc.*, 52(11), 1142–1156.

[2] Lindau, A.; Roos, S. (2010). Perceptual evaluation of discretization and interpolation for motion-tracked binaural (MTB) recordings. *Proc. of the 26th Tonmeistertagung,* Leipzig, 680-701.

[3] Lindau, A.; Erbes, V.; Lepa, S.; Maempel, H. J.; Brinkman, F. & Weinzierl, S. (2014). A spatial audio quality inventory (SAQI). *Acta Acustica united with Acustica*, *100*(5), 984-994.

[4] Lindau, A.; Hohn, T.; Weinzierl, S. (2007). Binaural resynthesis for comparative studies of acoustical environments. *Proc. of the 122nd AES Convention.*

[5] Lindau, A.; Brinkmann, F. (2010). Perceptual evaluation of individual headphone compensation in binaural synthesis based on non-individual recordings. *Journal of the Audio Engineering Society, 60*(1/2), 54-62.

[6] Lindau, A.; Estrella, J. & Weinzierl, S. (2010). Individualization of dynamic binaural synthesis by real time manipulation of ITD. *Audio Engineering Society Convention 128*.

[7] Lindau, A.; Kosanke, L. & Weinzierl, S. (2010). Perceptual evaluation of physical predictors of the mixing time in binaural room impulse responses. *Audio Engineering Society Convention 128*.

[8] Lindau, A.; Weinzierl, S. (2012). Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica 98*(5), 804–810.

[9] Schultz, F.; Lindau, A.; Makarski, M. & Weinzierl, S. (2011). An extraaural headphone for optimized binaural reproduction. *Proc. of the 26th Tonmeistertagung,* Leipzig, 702-714.

[10] Algazi, V. R.; Avendano, C. & Duda, R. O. (2001). Elevation localization and head-related transfer function analysis at low frequencies. *The Journal of the Acoustical Society of America*, *109*(3), 1110-1122.

[11] McAnally, K. I. & Martin, R. L. (2014). Sound localization with head movement: implications for 3-d audio displays. *Frontiers in neuroscience* 8, 210.