

Perzeptive Evaluierung von lautsprecher- und kopfhörerbasierter Auralisierung von Aufnahmen kugelförmiger Mikrofon-Arrays

Jens Ahrens, Carl Andersson

Division of Applied Acoustics, Chalmers University of Technology, SE-412 96 Gothenburg

{jens.ahrens, carl.andersson}@chalmers.se

Einleitung

Kugelförmige Mikrofon-Arrays ermöglichen die Zerlegung des einfallenden Schalldruckfeldes in Kugelflächenfunktionen. Diese stellen orthonormale Basisfunktionen der räumlichen Struktur des Schallfeldes dar. Mit Hilfe von Außenohrübertragungsfunktionen können im Rahmen der möglichen physikalischen Genauigkeit jene Signale bestimmt werden, die an den Ohren einer Hörerin/eines Hörers auftreten, die/der dem erfassten Schallfeld ausgesetzt ist [1, 2, 3]. Dabei spricht man von *Rendern* oder *Auralisieren* des Schallfeldes.

Derartige Schallfeldrepräsentationen können auch mit Hilfe von Lautsprecher-Arrays gerendert werden, wobei weitere substantielle physikalische Ungenauigkeiten auftreten. Die Auswirkungen dieser Ungenauigkeiten auf die Wahrnehmung ist weitgehend ungeklärt. Bzgl. der perzeptiven Evaluierung der zentralen Hörposition wurden einige Studien veröffentlicht [4, 5, 6, 7], wobei der Fokus meist auf der Lokalisationsgenauigkeit lag. Weitergehende Studien sind z.B. [8, 9, 10, 11, 12, 13].

Wir stellen in der vorliegenden Arbeit die Ergebnisse zweier Hörversuche bzgl. der wahrgenommenen Räumlichkeit sowie des Timbre von lautsprecher- und von kopfhörerbasierter Auralisierung vor.

Verwendete Daten

Die Stimuli wurden auf Basis der Messdaten aus [14]¹ erstellt. Dabei handelt es sich um Raumimpulsantworten, die sowohl mit einem kugelförmigen Mikrofon-Array mit Streukörper mit einem Radius von 8,75 cm sowie mit einem Neumann KU100 für verschiedene Kopforientierungen gemessen wurden. Somit lassen sich die gemessenen Räume sowohl direkt über die Kunstkopfdaten als auch indirekt über die Mikrofon-Array-Daten auralisieren. Im vorliegenden Fall wurden die Daten des Tonstudioraumes (Control Room 1, CR1, RT60 \approx 0.2 s) sowie des Kleinen Sendestudios (Small Broadcasting Studio, SBS, RT60 \approx 1 s) verwendet.

Die robuste Berechnung der Zerlegung des einfallenden Schalldruckfeldes in Kugelflächenfunktionen erfordert eine Beschränkung der angewendeten Verstärkung bei tiefen Frequenzen. Die hier gewählte Beschränkung ist 0 dB, so wie sie in [2, Fig. 3.21, p. 96] illustriert ist. Die Wahl liegt eher auf der konservativen Seite, und Verstärkungen bis zu 40 dB sind ebenso vorstellbar. Die vorliegende Wahl wurde getroffen, um die Ergebnisse zu jenen aus [2]

vergleichbar zu machen, wo ebenfalls eine Beschränkung von 0 dB verwendet wurde.

Kopfhörerbasiertes Rendering

Beim kopfhörerbasierten Rendering wird ein Kunstkopf virtuell in das Schallfeld gesetzt, das vom Mikrofon-Array erfasst wurde [2]. Dabei werden Kugelflächenfunktionskoeffizienten der Außenohrübertragungsfunktionen des Kunstkopfes (Daten aus [14]¹ mit einer Auflösung von 1°) mit den Kugelflächenfunktionskoeffizienten des erfassten Schallfeldes gefaltet [3]. Die direkte Auralisierung der gemessenen Kunstkopfdaten kann somit als Referenz angesehen werden. Jegliche Abweichung der Ohrsignale, die beim kopfhörerbasierten Rendering der Array-Aufnahmen auftreten, können als Artefakte identifiziert werden.

Die Daten wurden mit [15]² verarbeitet. Jede Testkondition wurde durch ein Set von Ohrimpulsantworten für verschiedene Kopforientierungen mit einer Auflösung von 1° repräsentiert, die in Echtzeit mit einem Testsignal entsprechend der momentanen Kopforientierung der Hörerin/des Hörers gefaltet wurde. Alle Stimuli wurden von Hand derart entzerrt, dass der gleiche Timbreindruck wie für die Kunstkopfdaten für eine Schallquelle geradeaus vorne entstand.

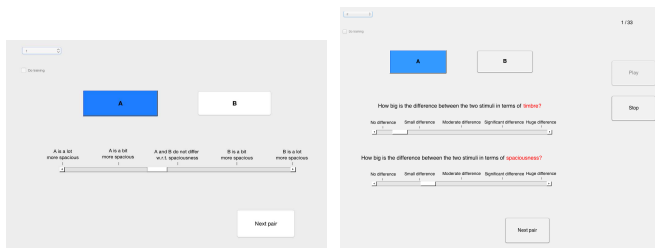
Die untersuchten Ordnungen der Kugelflächenfunktionszerlegung waren 1, 3, 5 (mit jeweils 50 Mikrofonen auf der Kugeloberfläche) sowie 8 (mit 110 Mikrofonen auf der Kugeloberfläche). Eine Schallquelle geradeaus vor der Hörerin/dem Hörer sowie ein Schallquelle bei 90° auf der rechten Seite der Hörerin/des Hörers wurde verwendet.

Lautsprecherbasiertes Rendering

Zur Evaluierung des lautsprecherbasierten Renderings wurde ein 56-kanaliges kreisförmiges Array mit 3 m Durchmesser verwendet. Die binauralen Raumimpulsantworten jedes Lautsprechers wurden für verschiedene Kopforientierungen des FABIAN Kunstkopfes [16] mit einer Auflösung von 1° für die zentrale Hörposition sowie 1 m außerhalb des Zentrums gemessen. Auf Basis dieser Daten wurde das Lautsprecher-Array über Kopfhörer emuliert. Dadurch kann auf Kopfdruck zwischen den beiden Hörpositionen umgeschaltet werden sowie das laut-

¹http://audiogroup.web.th-koeln.de/wdr_irc.html

²https://github.com/QLab/sound_field_analysis-py/



(a) Experiment zum Lautsprecherrendering (b) Experiment zum Lautsprecherrendering

Abbildung 1: Bildschirmfotos der GUIs

sprecherbasiertes Rendering gegen die direkte Auralisierung der Kunstkopfdaten verglichen werden.

Es wurden 2 Resyntheseverfahren verwendet: 1) Ambisonics [17, Gl. (20)] sowie 2) Wellenfeldsynthese [18, Gl. (21)]. Die akustische Antwort des Raumes, in dem das Lautsprecher-System aufgebaut ist, wurde durch Fensterung der binauralen Impulsantworten der Lautsprecher unterdrückt.

Die untersuchte Ordnung der Kugelflächenfunktionszerlegung war 8 (mit 110 Mikrofonen auf der Kugeloberfläche). Wie auch beim kopfhörerbasierten Rendering wurde die Quellposition vorne sowie lateral untersucht.

Zusätzlich zu den akustischen Umgebungen CR1 und SBS, die beim kopfhörerbasierten Rendering getestet wurden, wurde hier auch eine synthetische Kugelwelle im Freifeld untersucht, um ggf. Eigenschaften der Resyntheseverfahren offenzulegen.

Die getesteten Rendering-Konditionen werden in Tabelle 1 aufgelistet.

Tabelle 1: Liste der Paare die verglichen werden

| Kürzel | Kondition |
|----------|--|
| Ambi (c) | in Ambisonics gerendert, zentrale Hörposition |
| Ambi (l) | in Ambisonics gerendert, seitliche Hörposition |
| WFS (c) | in Wellenfeldsynthese gerendert, zentrale Hörposition |
| WFS (l) | in Wellenfeldsynthese gerendert, seitliche Hörposition |
| KK | binaurale Resynthese mittels Daten des Neumann KU100 Kunstkopfes |

Versuchsaufbau

Kopfhörer- und lautsprecherbasiertes Rendering wurden getrennt voneinander untersucht. 11 bzw. 12 Probanden beider Geschlechter im Alter von 25-46 Jahren mit nach eigener Aussage unbeeinträchtigtem Gehör nahmen teil. Die Experimente fanden in einer ruhigen Büroumgebung statt. Die Probanden saßen vor einem Computerbildschirm und trugen entzerrte Kopfhörer des Modells AKG

K702. Die Antworten wurden mit einer Computermaus eingegeben.

Das Nachführen der momentanen Kopforientierung (Headtracking) erfolgte mit einem Polhemus Patriot, die Echtzeitfaltung der Impulsantworten mit dem Testsignal erfolgte durch den SoundScape Renderer im BRS-Modus³ [19]. Als Testsignal wurde eine Schleife einer Rock-Schlagzeug-Aufnahme gewählt, da sie eine große Frequenzbandbreite sowie starke Transiente aufweist. Die Kommunikation zwischen SSR und der graphischen Schnittstelle des Experiments erfolgte über die TCP/IP-Schnittstelle des SSR.

Die Experimente wurden als direkter Paarvergleich mit vorherigem Training durchgeführt. Die Reihenfolge der Paare sowie die Tasterbelegung wurden randomisiert.

Kopfhörerbasiertes Rendering

Beim kopfhörerbasierten Rendering wurde immer die Array-Auralisierung gegen die direkte Auralisierung der Kunstkopfdaten verglichen. Die Aufgabe der Probanden bestand darin, die wahrgenommenen Unterschiede bzgl. der Räumlichkeit auf einer kontinuierlichen Skala von „Stimulus A ist deutlich räumlicher als Stimulus B“ bis „Stimulus B ist deutlich räumlicher als Stimulus A“ anzugeben (siehe Abb. 1a). Die Probanden wurden instruiert, bei Auftreten eines oder mehrerer der folgenden Unterschiede, den betreffenden Stimulus als räumlicher einzustufen: 1) Die Schallquelle klingt weiter entfernt, 2) der Hall klingt stärker, 3) die Umgebung klingt größer. Bei widersprüchlichen Unterschieden sollte eine beliebige Antwort gegeben werden.

Lautsprecherbasiertes Rendering

Die Paare aus Tabelle 1 wurden für alle drei Umgebungen (CR1, SBS, synthetisches Freifeld) verglichen. In diesem Experiment bestand die Aufgabe der Probanden darin, den vorzeichenlosen Unterschied zwischen den beiden Stimuli eines Paares bzgl. der Räumlichkeit (instruiert wie oben beschrieben) sowie des Timbre getrennt voneinander auf einer kontinuierlichen Skala von „Kein Unterschied“ bis „Sehr großer Unterschied“ anzugeben (siehe Abb. 1b).

Ergebnisse und Diskussion

Kopfhörerbasiertes Rendering

Die Antworten der Probanden sind in Abb. 2 als Kastengraphiken dargestellt. Diese wurden bereits in [20] vorgestellt. Die Hauptbeobachtungen sind:

- CR1 Bei der Schallquelle vorne ist die durchschnittliche Unterschied in der Räumlichkeit ungefähr null. Die Varianz ist sehr hoch für Array-Ordnung 1 und wird stetig geringer zu Ordnung 8 hin, wo sie sehr gering ist.

³<http://spatialaudio.net/ssr/>

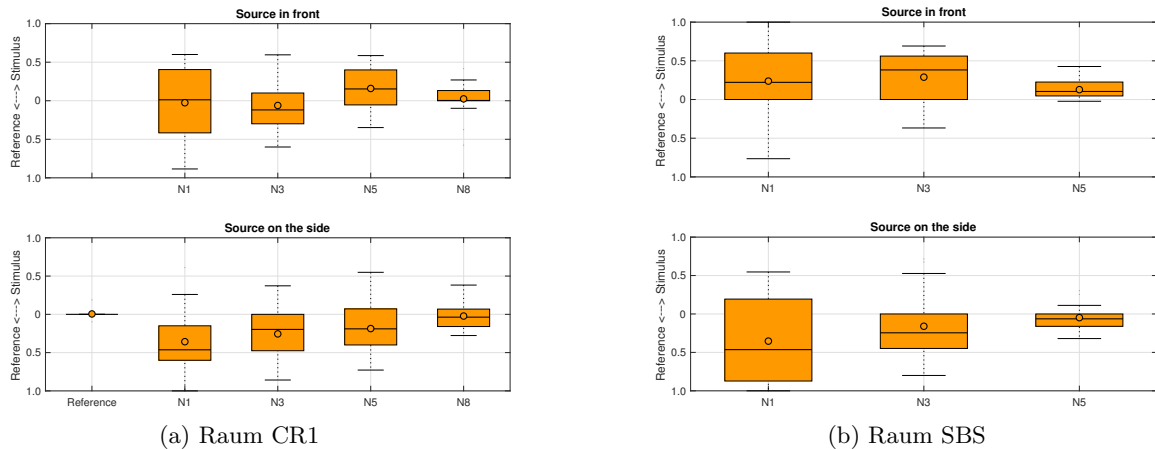


Abbildung 2: Ergebnisse des Experiments zum Kopfhörer-Rendering; Werte oberhalb der Mittellinie sind zugunsten der Array-Auralisierung, Werte unterhalb zugunsten der direkten Auralisierung der Kunstkopfdaten

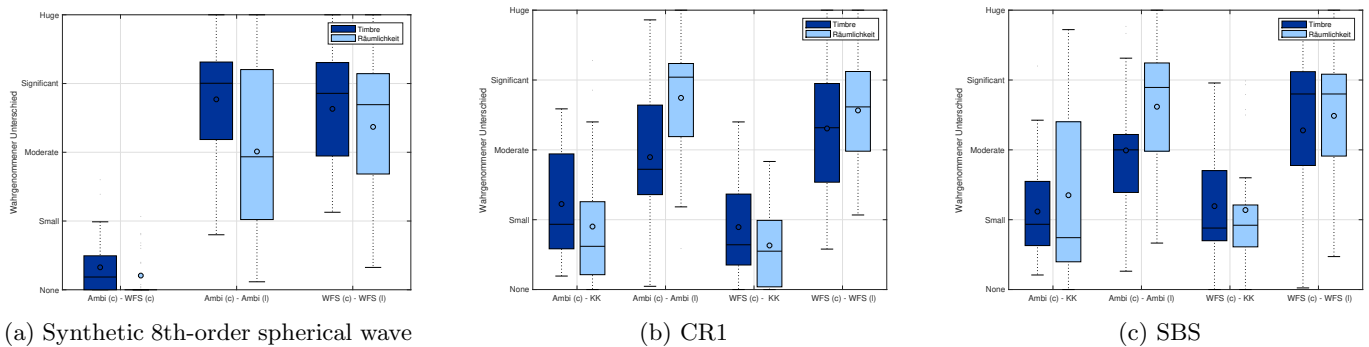


Abbildung 3: Ergebnisse des Experiments zum Lautsprecher-Rendering

CR1 Bei der lateralen Schallquelle ist die Räumlichkeit der Array-basierten Auralisationen für niedrige Ordnungen deutlich eingeschränkt und steigt stetig zu Ordnung 8 hin. Dort ist sie vergleichbar mit der Räumlichkeit der kunstkopfbasierten Auralisierung. Es stellt sich heraus, dass die Schallquelle bei niedrigeren Ordnungen deutlich näher empfunden wird als für höhere, wo die Entfernung mit jener der Quelle in den Kunstkopfauralisationen übereinstimmt.

SBS Hier gelten die gleichen Beobachtungen wie für Raum CR1, jedoch erfolgen die Beobachtungen bei geringeren Ordnungen.

SBS Die Referenz (Kunstkopf vs. Kunstkopf) wurde zuverlässig erkannt.

Informelles Hören legt nahe, dass die Array-Auralisierung des akustisch trockeneren Raumes CR1 bei Ordnung 8 von der direkten Auralisierung der Kunstkopfaufnahmen kaum unterscheidbar ist. Dasselbe gilt für den halligeren Raum SBS für Ordnung 5.

Lautsprecherbasiertes Rendering

Die Ergebnisse sind in Abb. 3 dargestellt. Die Hauptbeobachtungen sind:

- Es bestehen keine signifikanten Unterschiede zwischen Ambisonics und Wellenfeldsynthese.

- Es bestehen keine signifikanten Unterschiede zwischen den Räumen CR1 und SBS.
- Es bestehen keine signifikanten Unterschiede zwischen den Bewertungen der Räumlichkeit und den Bewertungen des Timbre.
- Die Unterschiede zwischen der zentralen und der lateralen Hörposition werden unabhängig von den übrigen Parametern als am größten angegeben.

Die großen Unterschiede zwischen der zentralen und der lateralen Hörposition lassen sich über die physikalischen Eigenschaften von Schallfeldern mit begrenzter Ordnung der Kugelflächenfunktionszerlegung erklären. Wohlgermerkt können kugelförmige Mikrofon-Arrays nur eine ordnungsbegrenzte Approximation eines Schallfeldes erfassen (im vorliegenden Fall Ordnung 8). Wie Abb. 4a illustriert, kann die Amplitude frequenzabhängig lokal sehr niedrig sein, sodass starke Beeinträchtigungen des Timbre auftreten (vgl. Betragsspektren in Abb. 4b). Das zusätzlich auftretende räumliche Aliasing führt zu unterschiedlichen lokalen Fortschrittingsrichtungen des Schallfeldes [21, 13], was die großen berichteten räumlichen Unterschiede erklärt.

Nennenswerte Beeinträchtigungen der Lokalisation wurde bereits in [6] für Ordnungen bis 3 berichtet. Allerdings ist die Literatur nicht konsistent bzgl. der Beeinträchtigungen an nichtzentralen Hörpositionen bei Dar-

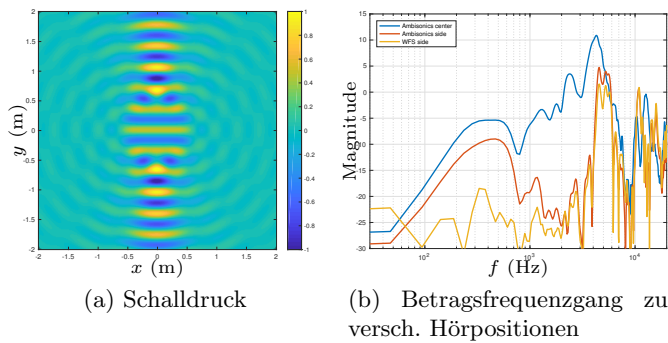


Abbildung 4: Schallfeld 8. Ordnung, $f = 1000$ Hz

bietung mittels traditionellem Ambisonics (geringe Beeinträchtigung in [8] vs. nennenswerte Beeinträchtigung in [6]). Es scheint jedoch, als sei traditionelles Ambisonics gutmütiger gegenüber nichtzentralen Hörpositionen.

Die Verwendung von unterschiedlichen Kunstköpfen bei der direkten Aufnahme und dem lautsprecherbasierten Rendering im vorliegenden Experiment scheint nicht ausschlaggebend für die Bewertungen gewesen zu sein. Nach sorgfältiger Entzerrung klingen die Außenohrübertragungsfunktionen der beiden verwendeten Modelle sehr ähnlich. Vor diesem Hintergrund sind fielen die Angaben zu den Unterschieden zwischen den zentralen Hörposition und der direkten Kunstkopfaufnahme (d.b. der Referenz) überraschend hoch aus. Die Daten aus Abb. 3 spiegeln nicht den Eindruck der Autoren wider. Ein Grund hierfür mag sein, dass die Probanden wenig bis keine Erfahrung mit kritischem Hören und Hörversuchen hatten.

Literatur

- [1] Larry S. Davis, Ramani Duraiswami, Elena Grassi, Nail A. Gumerov, Zhiyun Li, and Dmitry N. Zotkin, “High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues,” in *119th Convention of the AES*, Oct. 2005, p. 6540.
- [2] Benjamin Bernschütz, “Microphone arrays and sound field decomposition for dynamic binaural recording,” PhD thesis, Technische Universität Berlin, 2016.
- [3] Carl Andersson, “Headphone auralization of acoustic spaces recorded with spherical microphone arrays,” master’s thesis, Chalmers University of Technology, 2017.
- [4] Sebastian Braun and Matthias Frank, “Localization of 3D ambisonic recordings and ambisonic virtual sources,” in *Proceedings of ICASA*, Detmold, Germany, 2011.
- [5] Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, and Olivier Warusfel, “Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources,” *Acta Acustica united with Acustica*, vol. 99, pp. 642–657, 2013.
- [6] Peter Stitt, Stéphanie Bertet, and Maarten van Walstijn, “Off-centre localisation performance of Ambisonics and HOA for large and small loudspeaker array radii,” *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 937–944, 2014.
- [7] Lewis Thresh, Calum Armstrong, and Gavin Kearney, “A direct comparison of localization performance when using first, third, and fifth ambisonics order for real loudspeaker and virtual loudspeaker rendering,” in *143rd Convention of the AES*, New York, NY, USA, 2017, p. paper 9864.
- [8] Matthias Frank, Franz Zotter, and Alois Sontacchi, “Localization experiments using different 2d ambisonics decoders,” in *25th Tonmeistertagung*, Leipzig, Germany, Nov. 2008, VDE, pp. 1–9.
- [9] Gyan Vardhan Singh, “Psychoacoustic investigation on the auralization of spherical microphone array processing with wave field synthesis,” in *138th Convention of the AES*, Warsaw, Poland, May 2015, p. paper 9244.
- [10] Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, and Olivier Warusfel, “Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field,” in *Ambisonics Symposium*, Graz, Austria, 2009, pp. 1–9.
- [11] Enda Bates, Sean Dooney, Marcin Gorzel, Luke O’Dwyer, Hugh Ferguson, and Francis M. Boland, “Comparing ambisonic microphones – part 2,” in *142nd Convention of the AES*, Berlin, Germany, 2017, p. 9730.
- [12] Fiete Winter, Christoph Hold, Hagen Wierstorf, Alexander Raake, and Sascha Spors, “Colouration in 2.5d local wave field synthesis using spatial bandwidth-limitation,” in *Proceedings of IEEE WASPAA*, New Paltz, NY, USA, 2017, IEEE, pp. 1–5.
- [13] Hagen Wierstorf, “Perceptual assessment of sound field synthesis,” doctoral dissertation, Technische Universität Berlin, 2014.
- [14] Philipp Stade, Benjamin Bernschütz, and Maximilian Rühl, “A spatial audio impulse response compilation captured at the wdr broadcast studios,” in *27th Tonmeistertagung – VDT International Convention*, 2012.
- [15] Christoph Hohnerlein and Jens Ahrens, “Spherical microphone array processing in python with the sound_field_analysis-py toolbox,” in *Proceedings of DAGA*. DEGA, 2017.
- [16] Alexander Lindau and Stefan Weinzierl, “FABIAN – Schnelle Erfassung binauraler Raumimpulsantworten in mehreren Freiheitsgraden,” in *Proceedings of DAGA 2007*, Munich, Germany, 2007, DEGA, pp. 633–634.
- [17] Jens Ahrens and Sascha Spors, “An analytical approach to sound field reproduction using circular and spherical loudspeaker distributions,” *Acta Acustica united with Acustica*, vol. 94, pp. 988–999, 2008.
- [18] Jens Ahrens and Sascha Spors, “Wave field synthesis of a sound field described by spherical harmonics expansion coefficients,” *JASA*, vol. 131, no. 3, Mar. 2012.
- [19] Matthias Geier, Sascha Spors, and Jens Ahrens, “The soundscape renderer: A unified spatial audio reproduction framework for arbitrary rendering methods,” in *124th Convention of the AES*, May 2008.
- [20] Jens Ahrens, Christoph Hohnerlein, and Carl Andersson, “Auralization of acoustic spaces based on spherical microphone array recordings,” in *Proceedings of Acoustics ’17*, Boston, MA, USA, 2017, ASA/EAA, pp. 1–4.
- [21] Jens Ahrens, *Analytic Methods of Sound Field Synthesis*, Springer, Berlin/Heidelberg, 2012.