Dynamische binaurale Auralisierung von Klassenraumakustik - Einfluss der Quellenrichtcharakteristik

Armin Budnik[1], Henning Steffens[2], Stephan D. Ewert[2,3], Steven van de Par[2,3], Matthias Blau[1,3]
[1] *Jade Hochschule Oldenburg, Institut für Hörtechnik und Audiologie, 26121 Oldenburg, Deutschland*
[2] *Carl von Ossietzky Universität Oldenburg, Dept. für medizinische Physik und Akustik, 26129 Oldenburg, Deutschland*
[3] *Exzellenzcluster Hearing4All*

Einleitung

In einer vorhergehenden Studie (Blau et al. [1]) wurden gemessene und simulierte binaurale Raumimpulsantworten (BRIRs) mit der realen Akustik eines Klassenraumes hinsichtlich verschiedener perzeptiver Attribute verglichen. Die innerhalb der simulierten BRIRs verwendeten Außenohrübertragungsfunktionen (HRTFs), bzw. Impulsantworten (HRIRs) und die genutzte dynamische Binauralsynthese gewährleisteten gute bis sehr gute Lokalisationsfähigkeiten für die Teilnehmer. Basierend auf den dortigen Ergebnissen und Probandenaussagen wurden jedoch mögliche „Schwächen“ in den simulierten BRIRs festgestellt, die speziell bei der Erzeugung früher Reflexionen vermutet wurden. Die dort verwendete Raumsimulationssoftware (RAZR - Wendt et al. [2]) betrachtete zum damaligen Zeitpunkt lediglich omnidirektionale Quellen. Steffens [3] hat in seiner Arbeit die Raumsimulationssoftware RAZR erweitert, sodass fortan die Abstrahlcharakteristik der Quelle bei der Erzeugung simulierter BRIRs berücksichtigt werden kann.

Basierend auf den Erkenntnissen von Blau et al. [1] und der Weiterentwicklung der dort verwendeten Raumsimulationssoftware wurde in der vorliegenden Arbeit eine neue Probandenstudie durchgeführt. Die Versuche fanden in demselben Seminarraum wie vorher statt.

Die Fragestellungen, die sich für die neuen Hörversuche ergaben, lauten:

1. Sind simulierte und gemessene BRIRs perzeptiv vergleichbar?
2. Sind simulierte BRIRs mit versus ohne Quellenabstrahlcharakteristik perzeptiv unterschiedlich?
3. Gibt es perzeptive Unterschiede bei der Nutzung individueller versus Kunstkopf HRTFs?
4. Führt die Nutzung synthetisierter HRTFs (über einen virtuellen Kunstkopf) im Vergleich zu original gemessenen HRTFs zu perzeptiven Unterschieden?

Neben diesen vier Fragestellungen wurde zusätzlich die intraindividuelle Probandenreliabilität überprüft. Andreopoulou und Katz [4] führten dazu Untersuchungen zur Wiederholbarkeit von subjektiven HRTF Bewertungen durch und haben Wiederholbarkeitsschwellen ermittelt. In der vorliegenden Studie wurde ein ähnlicher Ansatz gewählt.

Methoden

Das Probandenkollektiv bestand aus 20 Teilnehmern (9 weiblich), mit einem Durchschnittsalter von 29,1 Jahren, wobei die jüngste Teilnehmerin 19 Jahre und der älteste Teilnehmer 50 Jahre alt waren. Alle Teilnehmer waren normalhörend (Hörverlust ≤ 15 dB HL).

Versuchsaufbau

Abbildung 1: Versuchsaufbau während der Hörversuche.

In Abbildung 1 ist der Messraum und Versuchsaufbau während der Hörversuche dargestellt. Die Abmessungen des Raumes betrugen 11,94 x 7,12 x 2,98 m^3 und die Nachhallzeit belief sich auf T60 = 0,58 s. Die Decke des Raumes war mit einem breitbandigen Absorbermaterial (Lochplatten mit Mineralwolle - 30 cm Entfernung zur Betondecke) versehen, während die Wände aus verputztem Kalksandstein bestanden. Der Fußboden war mit einer Linoleumschicht versehen.

Drei Referenzlautsprecher (1x GENELEC 8030c, 2x GENELEC 8030b) befanden sich an drei Positionen im Raum verteilt. Der erste Lautsprecher (S1) befand sich leicht aus der Raummitte verschoben in 4,22 m Entfernung direkt vor dem Probanden auf einer Höhe von 1,6 m. Der zweite Lautsprecher (S2) war links, in einem Winkel von -90° und einem Abstand von 3,55 m zum Probanden platziert. Der dritte Lautsprecher (S3) war rechts hinter den Versuchsteilnehmern in einem Winkel von 117° in 1,64 m Entfernung positioniert. Die Lautsprecher S2 und S3 befanden sich jeweils in etwa auf der Höhe der Probandenohren (1,3 m) und sollten potentielle Mitschüler/Kommilitonen darstellen, während der vorde-

re Lautsprecher (S1) einen „stehenden“ Lehrer/Dozenten darstellt.

Testkonditionen

Sechs Testkonditionen/BRIR Datensätze wurden in den Hörversuchen verwendet:

1. **meas** - Gemessene Kunstkopf BRIRs des Messraumes
2. **simHats** - Simulierte Kunstkopf BRIRs
3. **simOwn** - Simulierte, individuelle BRIRs
4. **simOwnVIKK** - Simulierte, individuelle synthetisierte BRIRs (über VIKK)
5. **simHATS_noDir** - Simulierte Kunstkopf BRIRs (ohne Quellenrichtcharakteristik)
6. **simSphere** - Simulierte BRIRs eines Kugelkopfmodells

Der gemessene Datensatz (Testkondition 1, Kunstkopf G.R.A.S KEMAR 45BB) bestand aus 37 azimuthalen BRIRs (-90° bis +90° Kopfausrichtung in 5°-Schritten bei einer Elevation von 0°). Zur Messung der BRIRs kamen MEMS Mikrofone (Knowles SPV0840LR5H) zum Einsatz, die mit einem modifizierten Schaumstoffohrstöpsel am Eingang des Gehörganges platziert waren (vgl. Poppitz et al. [5]).

Die simulierten BRIR Datensätze wurden mit der Raumsimulationssoftware RAZR erstellt (vgl. Wendt et al. [2] und Steffens [3]). Zur Berechnung der frühen Reflexionen werden dabei Spiegelquellen genutzt, während die späten Reflexionen über ein Feedback Delay Network erzeugt werden. Für die Simulationen innerhalb dieser Arbeit wurden Spiegelquellen bis zur 3. Ordnung genutzt. Bei vier der fünf Simulationen wurden auf der Senderseite innerhalb von RAZR die Quellenrichtcharakteristiken der realen Lautsprecher berücksichtigt. Dabei war das Abstrahlmuster des Lautsprechertyps (GENELEC 8030c) zuvor in einem reflexionsarmen Raum mit einer Auflösung von 5° in horizontaler und vertikaler Richtung gemessen worden. Eine der fünf Simulationen wurde hingegen zum späteren perzeptiven Vergleich mit omnidirektionalen Quellen gerechnet.

Auf der Empfängerseite der BRIRs wurden fünf verschiedene HRIR Datensätze verwendet. Dabei wurden HRIRs des hier verwendeten Kunstkopfes, individuelle HRIRs jedes Probanden und ein Kugelkopfmodell von Duda und Martens [9] genutzt. Zusätzlich dazu wurde der Datensatz der individuellen HRIRs über einen virtuellen Kunstkopf (VIKK) synthetisiert (vgl. Rasumow et al. [6] und Fallahi et al. [7]). Die Messung der HRIRs erfolgte mit denselben MEMS Mikrofonen wie die Messung der BRIRs.

Die simulierten BRIR Datensätze beinhalteten ebenfalls 37 azimuthale Kopfausrichtungen (-90° bis +90° in 5°-Schritten), zusätzlich jedoch auch neun elevierte Kopfausrichtungen von -30° bis +30° in 7,5°-Schritten.

Durchführung

Jeder Versuchsteilnehmer hatte die Aufgabe, die Übereinstimmung zwischen Kopfhörersignal (Testkondition) und realem Lautsprechersignal im Klassenraum (s. Abb. 1 - S1, S2, S3) anhand von fünf perzeptiven Attributen (Halligkeit, Quellbreite, Quelldistanz, Schallquellenrichtung, Gesamtqualität) zu bewerten. Die Attribute wurden dabei so gewählt, dass die ersten drei eine Evaluation der Raumsimulationssoftware ermöglichen. Mit dem Attribut Schallquellenrichtung sollten die verwendeten HRTFs und die dynamische Binauralsynthesesoftware überprüft werden. Mit der Gesamtqualität gab es die Möglichkeit, außerdem klangliche Aspekte der verwendeten Testkonditionen im Vergleich zu den Referenzkonditionen zu bewerten.

Abbildung 2: Kopfhörer (Sennheiser HD800) mit angeschlossenem Head Tracker für die dynamische Binauralsynthese und Taster zum Umschalten zw. Test- und Referenzkondition.

Die Bewertung hatte innerhalb einer grafischen Nutzeroberfläche, anhand einer Skala (schlecht, dürftig, ordentlich, gut, sehr gut) mit jeweils einem Zwischenschritt zu erfolgen, sodass insg. eine neunteilige Bewertungsskala resultierte. Nachdem alle Testkonditionen/BRIR Datensätze bewertet wurden, konnten diese in absteigender Reihenfolge nach dem besten sortiert werden (Chevret und Parizet [8]). Dies erlaubte eine weitere Überprüfung der Bewertung. Während der Hörversuche konnte über einen Taster am Kopfhörer (s. Abb. 2) beliebig oft zwischen Test- und Referenzkondition umgeschaltet werden, sodass mit Ab-/Aufsetzen des Kopfhörers mit einer minimalen Verzögerung von 450 ms (Fade Out/In) das jeweilige Signal umgeschaltet wurde. Jeder Proband hatte sechs verschiedene Testkonditionen/BRIR Datensätze pro Referenzkondition dreimal zu bewerten (Neun Bewertungen pro Attribut). Bei fünf zu bewertenden Attributen resultieren daraus insg. 45 Messdurchläufe. Diese umfassten eine Dauer von ca. 90-180 Minuten und wurden daher an zwei separaten Terminen durchgeführt.

Stimulus

Als Stimulus wurde der erste Satz von „Nordwind und Sonne“, gesprochen von einer weiblichen Sprecherin ($f_g \approx 240\,\mathrm{Hz}$), verwendet. Dieses Audiobeispiel wurde hochpassgefiltert (Butterworth - 6. Ordnung) mit einer Grenzfrequenz von $f_c = 200\,\mathrm{Hz}$.

Intraindividuelle Probandenreliabilität

In Abbildung 3 sind zuerst die Ergebnisse der intraindividuellen Probandenreliabilität für jedes Attribut dargestellt. Durch die bereits angesprochene dreifache Wieder-

holung jeder Messkondition entstanden drei Paarvergleiche, zwischen denen eine Korrelationsanalyse nach Pearson durchgeführt werden konnte. Die daraus resultierenden drei Korrelationswerte wurden für jeden Probanden und jedes Attribut gemittelt und sind in Abbildung 3 als rote Kreuze dargestellt. Die zugehörige Standardabweichung der Werte ist in blauen Kreuzen eingezeichnet. Die von Andreopoulou und Katz [4] vorgeschlagenen Wiederholbarkeitsschwellen sind als durchgezogene schwarze Linie (Stabilitätskriterium) und gestrichelte schwarze Linie (Konsitenzkriterium) eingezeichnet.

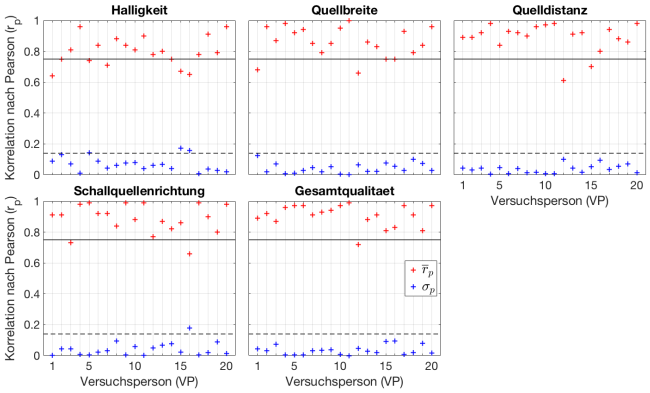

Abbildung 3: Ergebnisse der Korrelationsanalyse nach Pearson, mit eingezeichneten Wiederholbarkeitsschwellen nach Andreopoulou und Katz [4].

Tabelle 1: Attributspezifische Probandenausschlüsse

Attribut	Ausschluss
Halligkeit	VP1, VP15, VP16
Quellbreite	VP1, VP12
Quelldistanz	VP12, VP15
Schallquellenrichtung	VP16
Gesamtqualität	Keine

Basierend auf diesen Ergebnissen und einer zusätzlichen visuellen Betrachtung der Einzelergebnisse wurden einzelne Versuchspersonen von den finalen attributspezifischen Auswertungen ausgeschlossen (s. Tabelle 1).

Attributspezifische Ergebnisse

In Abbildung 4 sind die Ergebnisse der perzeptiven Probandenbewertungen für die drei Referenzkonditionen dargestellt. Bei allen drei Referenzkonditionen und allen Attributen erzeugen simulierte BRIRs mit Kugelkopfmodell eine schlechte perzeptive Übereinstimmung zwischen Kopfhörer- und Lautsprecherpräsentation. Ebenfalls werden simulierte BRIRs ohne Quellenrichtcharakteristik bei allen Attributen signifikant schlechter bewertet als simulierte BRIRs mit Berücksichtigung des Quellenabstrahlverhaltens. Zusätzlich dazu führen diese lediglich bei dem Attribut der Schallquellenrichtung zu ordentlichen bis guten perzeptiven Übereinstimmungen mit der Referenzkondition. Bei den vier anderen Attributen werden häufig lediglich dürftige Übereinstimmungen erreicht.

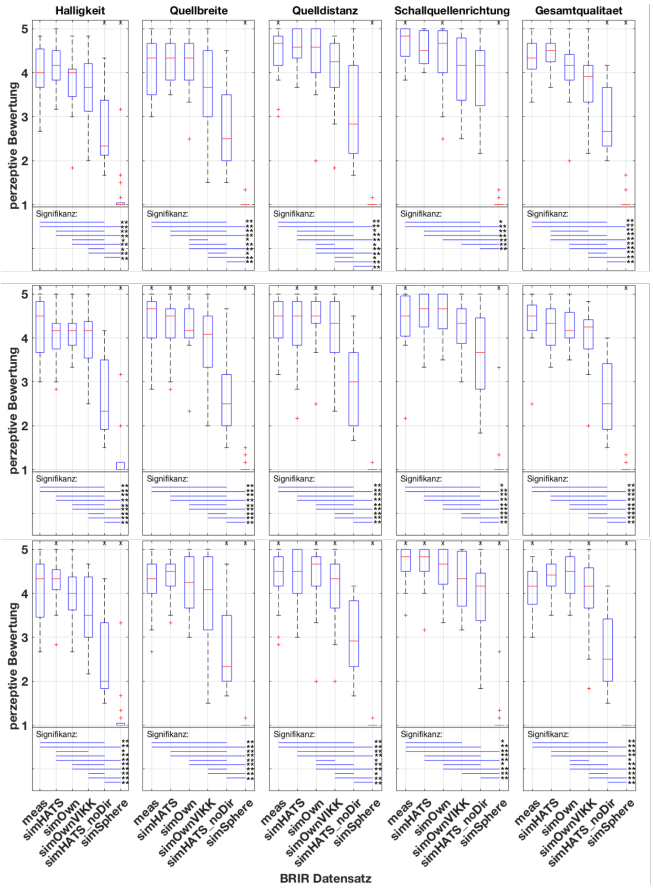

Abbildung 4: Perzeptive Probandenbewertungen pro Attribut, Testkondition und Referenzkondition (LS1 - Oben, LS2 - Mitte, LS3 - Unten). In jedem Boxplot sind die Mittelwerte der dreifachen Messwiederholungen pro Proband enthalten. Der Median ist als rote Linie dargestellt, während die blauen Boxen die Interquartilspannweite darstellen. Die Signifikanzniveaus der post-hoc Tests (Wilcoxon-Vorzeichen-Rang-Test) sind mit Sternen gekennzeichnet (* - alpha = $\frac{0,05}{15}$/** - alpha = $\frac{0,01}{15}$).

Allgemein werden simulierte BRIRs im Median sehr ähnlich bewertet wie original gemessene BRIRs im Raum. So liegen hierbei die Ergebnisse in den meisten Fällen zwischen guten bis sehr guten perzeptiven Übereinstimmungen zum Referenzsignal.

Bei Betrachtung der simulierten BRIRs mit individuellen, synthetisierten HRTFs wird deutlich, dass diese bei dem linken Referenzlautsprecher (-90°) keine signifikanten Unterschiede im Vergleich zu original gemessenen HRTFs ergeben. Bei dem vorderen und rechtsseitigen Lautsprecher führten diese jedoch bei jeweils zwei Attributen zu signifikanten Unterschieden.

Diskussion

BRIR Simulationen mit Berücksichtigung des Quellenabstrahlverhaltens führen im Vergleich zu gemessenen BRIRs des Originalraumes zu keinen signifikanten perzeptiven Unterschieden - bei allen Attributen und Referenzkonditionen. Wendt [10] stellte noch minimale, jedoch statistisch signifikante Unterschiede fest. Durch die Verwendung der Lautsprecherdirektivität wurden diese

perzeptiven Unterschiede jedoch vermutlich beseitigt - zumindest in der hier vorhandenen Messsituation. Die perzeptiven Übereinstimmungen simulierter BRIRs mit Quellenrichtcharakteristik werden bei allen Attributen und Referenzkonditionen signifikant besser bewertet als Simulationen mit omnidirektionalen Quellen. Speziell bei den Attributen Quellbreite und -distanz resultierten im Median nahezu sehr gute Übereinstimmungen. Die Betrachtung der Lautsprecherdirektivität führt an dieser Stelle dazu, dass virtuelle Schallquellen örtlich präziser und klanglich detaillierter im Raum dargestellt werden.

Weiterhin gab es innerhalb der BRIRs keine signifikanten Unterschiede zwischen der Nutzung individueller und Kunstkopf HRTFs. Blau et al. [1] kamen bereits zu ähnlichen Resultaten. Andererseits ist bekannt, dass die Nutzung individueller HRTFs im Freifeld zu einer verbesserten Lokalisationsfähigkeit führt, bzw. Vorne/Hinten Vertauschungen in der Lokalisation vermieden werden können (vgl. [11], [12]). Es ist also möglich, dass individuelle HRTFs in nicht-reflexionsarmer Umgebung, bei Verwendung von Sprachsignalen und inkl. visueller Stimuli nicht zwingend nötig sind.

Der letzte Punkt befasst sich mit der perzeptiven Qualität synthetisierter HRTFs (über VIKK) im Vergleich zu original gemessenen HRTFs. Hierbei ergaben sich bei der vorderen und rechtsseitigen Referenzkondition signifikante Unterschiede, bei dem linken Lautsprecher jedoch nicht. Dies könnte damit begründet sein, dass der Mensch in der Frontalebene bei 0 ° Elevation sensitiver auf mögliche Unterschiede oder Artefakte reagiert. Bei der rechten Kondition hingegen, könnte die Nähe zum Referenzlautsprecher und der damit verbundene, größere Direktschallanteil dazu beigetragen haben, dass mögliche Differenzen zwischen originalen und synthetisierten HRTFs hörbar wurden.

Zusammenfassung und Ausblick

Zusammenfassend kann gesagt werden, dass die Berücksichtigung des Quellenabstrahlverhaltens innerhalb von RAZR bei den hier vorgestellten Hörversuchen zu einer hoch signifikanten Verbesserung der wahrgenommenen virtuellen Akustik des Klassenraumes geführt hat. Mit VIKK synthetisierte, individuelle HRTFs führen im Vergleich zu original gemessenen HRTFs lediglich in einigen wenigen Fällen zu signifikanten perzeptiven Unterschieden.

Außerdem wurde bestätigt, dass die Auflösung der verwendeten HRTFs (5 ° horizontal) bei Raumsimulationen mit gewissem Nachhallanteil ausreichend genaue Lokalisationsfähigkeiten zulassen. Allerdings ist an dieser Stelle die visuelle Unterstützung zu berücksichtigen, wodurch mögliche geringe Abweichungen in der akustischen Darbietung verdeckt worden sein könnten.

Künftig sind weitere Hörversuche in anderen akustischen Umgebungen anzuraten. Speziell die Untersuchung des Einflusses anderer Audiosignale (z.B. Musik) stellt einen interessanten Aspekt dar. Des weiteren könnten Überprüfungen der perzeptiven Eigenschaften von VIKK Synthesen ebenfalls von Bedeutung sein, da diese innerhalb dieser Hörversuche teilweise gewisse perzeptive Unterschiede hervorgerufen haben.

Literatur

[1] Blau M., Budnik A., van de Par S.: Assessment of Perceptual Attributes of Classroom Acoustics: Real Versus Simulated Room. In: Proceedings of the Institute of Acoustics 40 (2018), S. 9.

[2] Wendt T., van de Par S., Ewert S. D.: A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation. In: Journal of the Audio Engineering Society 62 (2014), Dezember, Nr. 11, 748–766.

[3] Steffens H.: Perceptual relevance of speaker directivity modelling in virtual rooms, CvO Universität Oldenburg. 2019. – Master Thesis.

[4] Andreopoulou A., Katz B. F.: Investigation on Subjective HRTF Rating Repeatability, In: AES Convention Paper (2016), S. 10.

[5] Poppitz J., Blau M., Hansen M.: Entwicklung und Evaluation eines Systems zur Messung individueller HRTFs in privater Wohnumgebung. In: Fortschritte der Akustik - DAGA. Aachen, 2016

[6] Rasumow E., Blau M., Doclo S., van de Par S., Hansen M., Püschel D., Mellert V.: Perceptual Evaluation of Individualized Binaural Reproduction Using a Virtual Artificial Head. In: Journal of the Audio Engineering Society 65 (2017), Juni, Nr. 6, 448–459.

[7] Fallahi M., Blau M., Hansen M., Doclo S., van de Par S., Püschel D.: Constrained optimization for binaural sound reproduction using a virtual artificial head. In: DAGA (2018), S. 4.

[8] Chevret P., Parizet E.: An efficient alternative to the paired comparison method for the subjective evaluation of a large set of sounds. In: 19th International Congress on Acoustics. Madrid, Spain, September 2007.

[9] Duda R. O., Martens W. L.: Range dependence of the response of a spherical head model. In: The Journal of the Acoustical Society of America 104 (1998), November, Nr. 5, 3048–3058.

[10] Wendt T.: A computationally efficient and perceptually plausible room acoustics simulation method, CvO Oldenburg, Diss., 2018.

[11] Wenzel E. M., Arruda M., Kistler D. J., Wightman F. L.: Localization using nonindividualized head-related transfer functions. In: The Journal of the Acoustical Society of America (1993), No. 94

[12] Møller H., Sørensen M. F., Jensen C. B., Hammershøi D.: Binaural Technique: Do We Need Individual Recording?. In: J. Audio Eng. Soc. (1996), Vol. 44, No. 6