

Vergleich unterschiedlicher Ton-Wiedergabesysteme von Virtual Reality-Brillen

Jelena Kitanovic¹, Dieter Leckschat² und Christian Epe³

Fachbereich Medien, Hochschule Düsseldorf, 40476 Düsseldorf, Deutschland

E-Mails: ¹ jelena.ktnvc@gmail.com, ² dieter.leckschat@hs-duesseldorf.de, ³ christian.epe@hs-duesseldorf.de

Einleitung

Die virtuelle Realität, kurz VR, hat sich in wenigen Jahren aus einer Zukunftsvision zu einer breit zugänglichen Technologie entwickelt. Seit dem Erscheinen erster massenmarktauglicher Produkte, insbesondere VR-Brillen, ist das Interesse an dem Medium stetig gewachsen. Während die visuelle Qualität bei VR-Anwendungen kontinuierlich gestiegen ist, blieben solche Entwicklungen hinsichtlich der Ton-Wiedergabesysteme jedoch weitestgehend aus. Viele der derzeit in Kombination mit VR-Headsets verwendeten Kopfhörer oder Lautsprecher sind nicht geeignet, um virtuelle akustische Umgebungen räumlich korrekt wiederzugeben. Verschiedene Forschungsergebnisse legen nahe, dass die dafür erforderlichen Übertragungseigenschaften am ehesten mit einer extraauralen Kopfhörer-Bauart realisiert werden können [1], [2], [3]. Diese wird derzeit allerdings bis auf eine Ausnahme kommerziell nicht genutzt.

In diesem Beitrag geht es um die Konzeption, Vorbereitung, Durchführung und Auswertung eines Hörversuchs, bei dem untersucht wurde, wie sich ein extraauraler Kopfhörer von einem circumauralen und intraauralen hinsichtlich seines Einflusses auf die Ortung virtueller Schallquellen unterscheidet. Durchgeführt wurde der Versuch im Rahmen einer Bachelorarbeit zum Thema Ton-Wiedergabesysteme von Virtual Reality-Brillen [4]. Da sich die Fragestellung konkret auf VR-Anwendungen bezog, hat der Versuch auch entsprechend in einer virtuellen Umgebung stattgefunden.

Hypothesen

Mithilfe des Versuchs sollte überprüft werden, ob folgende Hypothesen zutreffen:

1. Mit dem extraauralen Kopfhörer ist bei der Lokalisation virtueller Schallquellen die Fehlerquote geringer als bei den Vergleichsmodellen.
2. Mit dem extraauralen Kopfhörer ist bei der Lokalisation virtueller Schallquellen die Winkelabweichung geringer als bei den Vergleichsmodellen.
3. Der extraaurale Kopfhörer ermöglicht eine schnellere Lokalisation virtueller Schallquellen als die Vergleichsmodelle.
4. Beim extraauralen Kopfhörer ist der Tragekomfort höher als bei den Vergleichsmodellen.

Methode

Versuchsaufbau

Für die Erhebung der erforderlichen Messdaten wurde folgender Versuchsaufbau entworfen: Nach Aufsetzen eines Head-Mounted-Displays (HMD) sollten die Probanden einen geschlossenen Raum mit visuell dargestellten virtuellen Schallquellen sehen. Während des Versuchs sollte über diese

nacheinander der gleiche Stimulus abgespielt werden. Die Probanden sollten dann angeben, welche Schallquelle zu den jeweiligen Zeitpunkten ihrer Einschätzung nach aktiv sind. Dieser Vorgang sollte mit allen Kopfhörer-Modellen wiederholt werden.

Um einen Einfluss möglicher Gewöhnungs- oder Lernmuster zu verhindern, sollte die Reihenfolge der Kopfhörer nach jedem Probanden rotieren. Bei der Aktivierung der virtuellen Schallquellen sollte die Reihenfolge zufällig sein, allerdings mit der Bedingung, dass in jedem Durchgang jede Quelle genau einmal drankommt.

Virtuelle Versuchsumgebung

Für die Erstellung der virtuellen Versuchsumgebung wurde die Spiele-Engine *Unity* des US-Herstellers *Unity Technologies* verwendet und für die akustische Gestaltung das HRTF-basierte Open Source SDK *Resonance Audio*.

Für die visuelle Darstellung der virtuellen Schallquellen wurden mittig in einem 8 m x 8 m x 3 m großen virtuellen Raum auf einer Höhe von 1,70 m, 36 freischwebende Kugeln in Abständen von jeweils 10° horizontal angeordnet (Abb. 1). Der Radius des somit entstandenen Kreises betrug 3 m. Die Kugeln selbst hatten einen Radius von 20 cm.

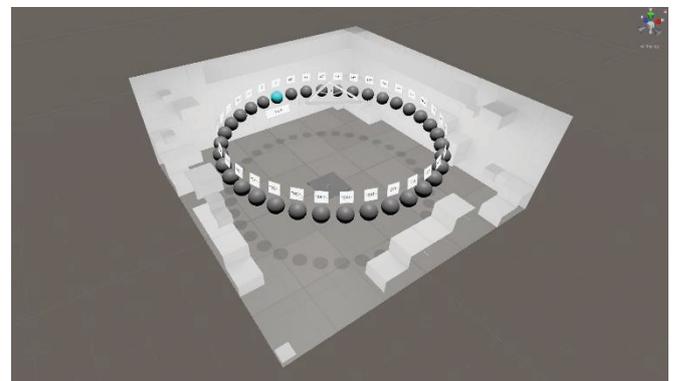


Abbildung 1: Virtuelle Versuchsumgebung

Oberhalb jeder Kugel wurde ein Button eingefügt, der mit der jeweiligen Gradzahl beschriftet wurde. Während des Versuchs sollten darüber die Eingaben getätigt werden. Beschriftet wurden die Buttons oberhalb der Kugeln von 0° ausgehend mit 10° bis 180° im Uhrzeigersinn und -10° bis -170° gegen den Uhrzeigersinn. Das sollte die Orientierung erleichtern und die Teilnehmer dazu verleiten, nach jeder Eingabe zur Ausgangsposition (0°) zurückzukehren.

Da der virtuelle Raum keinerlei schallabsorbierende Einrichtungsgegenstände enthielt, wurden mithilfe *Resonance Audio* den Wänden Absorptionseigenschaften zugewiesen um eine unerwünschte „Badezimmerakustik“ zu verhindern.

Reale Versuchsumgebung

Der Versuch wurde im Labor für *Mixed Reality and Visualization* (MIREVI) der Hochschule Düsseldorf durchgeführt.

Das Labor bietet konstante Rahmenbedingungen und verfügt über einen fest installierten Tracking-Bereich von ca. 3 m x 3 m.

HMD und Tracking-System

Für die visuelle Wiedergabe kam im Versuch das *HTC Vive*-Headset zum Einsatz. Der im rechnergebundenen Headset verbaute Bildschirm hat eine Auflösung von 1080 x 1200 Pixel pro Auge, eine Bildwiederholrate von 90 Hz und ein Sichtfeld von 110 Grad. Das Tracking erfolgte mit einer Kombination aus einem Gyroskop, Beschleunigungssensor, Nähesensor und dem *SteamVR-Tracking* (ehem. „*Lighthouse*“) von *Valve*.

Kopfhörer

Das Ziel des Versuchs war, extraaurale Kopfhörer mit herkömmlichen Bauarten hinsichtlich ihrer Eignung für die Wiedergabe dreidimensionalen Tons im Kontext von VR-Anwendungen zu vergleichen. Neben dem extraauralen sollten ein intraauraler und ein circumauraler Kopfhörer zum Einsatz kommen. Auf ein supraaurales Modell wurde verzichtet, da die Bauweise der circumauralen vergleichsweise ähnlich ist. Um zu verhindern, dass Qualitätsunterschiede einen signifikanten Einfluss auf die Versuchsergebnisse haben konnten, sollten ausschließlich hochwertige Studiokopfhörer oder solche, die den gleichen Standards entsprechen, gewählt werden.

Individuelle Kopfhörerentzerrungen wurden nicht gemacht, da gängige HMD-Systeme diese Option derzeit nicht anbieten und die Versuchsergebnisse für den Entwicklungsstand kommerzieller Produkte repräsentativ sein sollten. Dass es zu interindividuellen Unterschieden kommen würde, hat man daher bewusst in Kauf genommen.

Für die extraaurale Bauart kam der Prototyp der sog. *NearPhones* zum Einsatz, die 2016 vom Alumnus Nicolas Sünm im Rahmen seiner Bachelorarbeit an der Hochschule Düsseldorf entwickelt wurden [5] (Abb. 2). In den *NearPhones* sind *Visaton SL 87 ND* Breitbandlautsprecher als Dipolstrahler eingebaut. Die Resonanzfrequenz der Wandler liegt bei etwa 125 Hz und der Maximalpegel von 100 dBspl bei 100 Hz. Die offenen Lautsprechergehäuse wurden eigens für die *NearPhones* designt mit einem 3D-Drucker aus Polylactid (PLA) hergestellt.



Abbildung 2: Prototyp der NearPhones [4]

Bei einer Messung des Prototyps mit einem Kunstkopf offenbarte sich, dass sich Interferenzen zwischen den Lautsprechermembranen und den Seitenwänden des Kopfes bilden. Sie wurden bei der Entzerrung allerdings nicht weiter beachtet, da die *NearPhones* für eine Beschallungssimulation

entwickelt wurden, bei der die Unregelmäßigkeiten mithilfe eines DSPs eliminiert werden, der personalisierte Raumimpulsantwort (PRIR) und Kopfhörerentzerrungen (HpEQ) für jeden Nutzer berechnet. Stattdessen wurde im Übertragungsbereich zwischen 100 Hz und 17.000 Hz ein linearer Frequenzgang mit einer Toleranz von $\pm 1,5$ dB in 3 cm Entfernung auf den Lautsprecherachse angestrebt, aufgenommen mit einem Mikrofon. Um eine Überbetonung der Höhen auszugleichen, wurde die Zielfunktion nach mehreren Hörtests mit einem Highshelve (-3 dB bei 7,5 kHz) korrigiert.

Die Wahl des circumauralen Vergleichs-Kopfhörers fiel auf das geschlossene und diffusfeldentzerrte Modell *DT 770 PRO* von *beyerdynamic*. Der *ATH-E70* von der Marke *Audio-Technica* kam als intraauraler Vergleichs-Kopfhörer zum Einsatz. Da nicht zu erwarten war, dass alle Teilnehmer eigene Otoplastiken besitzen würden, hat man stattdessen austauschbare Schaumstoffaufsätze verwendet.

Stimulus

Für den Versuch war es besonders wichtig, dass im Stimulus alle Frequenzen im Übertragungsbereich (100 Hz bis 17000 Hz) vertreten sind, um spektrale Einflüsse der Kopfhörer messen zu können. Darüber hinaus sollte das Signal zwecks einer guten Lokalisierbarkeit impulshaltig sein. Die Lösung war ein Loop aus sich kontinuierlich wiederholenden 1/f-Rauschen Bursts. Die einzelnen Bursts waren 250 ms lang, ebenso die Pausen zwischen den Wiederholungen. Gefiltert wurde das Signal mit einer hohen Flankensteilheit mit einem Hochpass bei 100 Hz (18 dB/Okt.) und einem Tiefpass bei 17 kHz (30 dB/Okt.), um die unterschiedlichen Übertragungsbereiche der Kopfhörer auszugleichen und gleichzeitig möglichst viel von der Bandbreite zu nutzen.

Probanden

Von den 20 Probanden (17 männlich, 3 weiblich), deren Ergebnisse ausgewertet wurden, gaben 15 an, Studierende oder Beschäftigte der Hochschule Düsseldorf zu sein, darunter 14 im Fachbereich Medien. Das Alter der Probanden lag zwischen 20 und 59 Jahren (Mittelwert 32,4; Standardabweichung 11,1). Des Weiteren gaben 14 an, dass sie sich beruflich, innerhalb des Studiums oder der Berufsausbildung schwerpunktmäßig mit den Themen Tontechnik, Tonproduktion, Akustik und/oder AV-Medien befassen. 17 Probanden gaben außerdem an, schon einmal ein VR-Headset getragen zu haben. Es wurden nur volljährige Teilnehmer zugelassen, bei denen nach eigener Aussage kurz- oder langfristig keine erwähnenswerte Einschränkung des Hörvermögens vorlag.

Versuchsdurchführung

Vor Beginn eines Versuchsdurchlaufs sollte ein Teilnehmerbogen ausgefüllt werden, in dem, neben Informationen zur Person, wie Alter und Geschlecht, auch abgefragt wurde, ob man Beschäftigter oder Studierender der Hochschule Düsseldorf ist und ob man Vorerfahrungen mit VR-Headsets und verwandten Themen hatte.

Der Versuch wurde nach einer mündlichen Freigabe über einen entsprechenden Button im virtuellen Raum durch die Probanden gestartet. Daraufhin ertönte der Stimulus von einer der virtuellen Schallquellen. Um diese zu lokalisieren, durften sich die Teilnehmer zwar drehen, aus Sicherheitsgründen

allerdings nicht mehr als um 180° in eine Richtung. Außerdem sollten sie den markierten Bereich, auf dem sie standen, nicht verlassen. Nach dem Tätigen einer Eingabe über einen VR-Controller wurde die Schallquelle unabhängig davon, ob die Auswahl richtig oder falsch war, deaktiviert und eine andere wurde gestartet. Ein Teildurchgang wurde automatisch beendet, wenn der Stimulus über alle 36 Schallquellen einmal ertönt war. Die Probanden wurden angewiesen, die Aufgabe so langsam wie nötig und gleichzeitig so zügig wie möglich zu lösen. Der gesamte Vorgang wurde dann mit den anderen Kopfhörer-Modellen wiederholt.

Nach dem Versuch wurden die Teilnehmer gebeten, einen Fragebogen auszufüllen. Sie sollten mithilfe von Bewertungsskalen angeben, wie schwer sie die Aufgabe fanden, wie sie ihre eigene Trefferquote einschätzen würden und wie realitätsnah die auditive Räumlichkeit der virtuellen Umgebung auf sie gewirkt hat. Außerdem sollten sie die Kopfhörermodelle hinsichtlich ihres (gefühlten) Einflusses auf die Ortbarkeit virtueller Schallquellen und ihres Tragekomforts im Vergleich zueinander bewerten.

Ergebnisse

Während des Versuchs wurde gemessen, wie viele Fehler in den einzelnen Teildurchgängen gemacht wurden, wie groß dabei die Winkelabweichungen waren und wie viel Zeit die Probanden gebraucht haben, um nach den Positionswechsel des Stimulus ihre Eingaben zu tätigen. Für die gemessenen Größen wurden folgende Gesamtmittelwerte und Standardabweichungen berechnet:

Tabelle 1: Mittelwerte und Standardabweichungen der gemessenen Größen

Modell	NearPhones	DT 770 Pro	ATH-E70
Fehler			
Mittelwert	12,6	7,5	13,6
Standardabw.	8,5	8,7	11,2
Winkelabweichung			
Mittelwert	11,0°	8,2°	9,8°
Standardabw.	3,8°	5,1°	3,6°
Entscheidungsdauer			
Mittelwert	8,5 s	7,3 s	7,9 s
Standardabw.	5,1 s	3,1 s	5,0 s

Die berechneten Standardabweichungen deuten an, dass die individuellen Treffer- bzw. Fehlerquoten der Teilnehmer sehr unterschiedlich waren. Die Gesamtmittelwerte haben daher nur eine bedingte Aussagekraft. Mithilfe einer Box-Plot-Darstellung lässt sich die Fehlerverteilung unter Einbezug inter-individueller Unterschiede veranschaulichen (Abb. 3).

Um zu überprüfen, ob die gemessenen Unterschiede zwischen den Kopfhörer-Modellen statistisch signifikant sind, wurde als Methode die einfaktorische Varianzanalyse (ANOVA) gewählt, da sie sich für einen Vergleich der Mittelwerte von mehr als zwei Gruppen eignet. Mithilfe des Kolmogorow-Smirnow-Tests konnte für die Werte der Fehler und Entscheidungsdauern zum Signifikanzniveau $\alpha = 0,05$ eine Normalverteilung nachgewiesen werden, jedoch nicht für die Winkelabweichungen. Voraussetzung für die ANOVA sind zwar normalverteilte Daten, allerdings zeigt sich die Methode

besonders bei gleichen Gruppengrößen gegenüber einer Verletzung der Annahme robust [6]. Die zweite Voraussetzung für das Verfahren ist die Varianzhomogenität der Gruppen. Diese konnte für alle Parameter mithilfe des Levene-Tests nachgewiesen werden. Die Analyse ergab, dass die Unterschiede zwischen den verschiedenen Kopfhörer-Modellen für die gemessenen Größen keine statistische Signifikanz haben ($p > 0,05$). Es muss daher in allen drei Fällen die Nullhypothese beibehalten werden, dass zwischen den Mittelwerten der Gruppen keine signifikanten Unterschiede bestehen. Die zu Beginn aufgestellten Hypothesen, dass mit dem extraauralen Kopfhörer bei der Lokalisation virtueller Schallquellen die Fehlerquote (1.) und die Winkelabweichung (2.) geringer sei als bei den Vergleichsmodellen, konnten daher nicht bestätigt werden. Gleiches gilt für die Annahme, dass der extraaurale Kopfhörer eine schnellere Lokalisierung (3.) ermögliche.

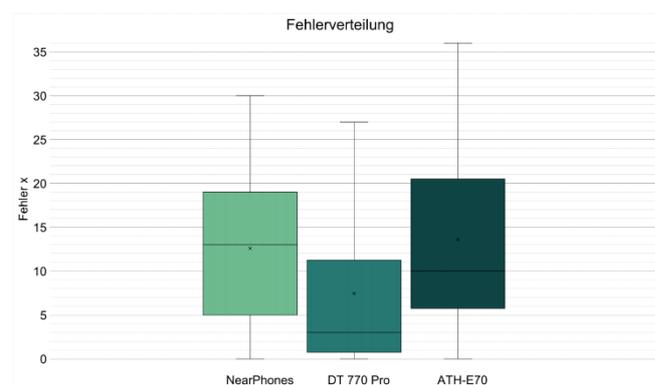


Abbildung 3: Box-Plot Darstellung der Fehlerverteilung

Die Teilnehmer sollten nach dem Versuch angeben, wie schwer sie die gestellte Aufgabe fanden, indem sie eine Bewertung auf einer Skala von 1 („sehr einfach“) bis 5 („sehr schwer“) abgaben. Die Durchschnittsbewertung lag bei 2,2 mit einer Standardabweichung von 0,9. 5 Punkte wurden von keinem Teilnehmer vergeben. Nach dem gleichen Prinzip sollten die Teilnehmer ihre eigene Trefferquote einschätzen und die auditive Räumlichkeit der VR-Szene bewerten. Für ersteres ergab die Auswertung einen Mittelwert 2,2 und für letzteres 2,1, die Standardabweichung betrug bei beiden 0,8.

Im zweiten Teil des Fragebogens sollten die drei Kopfhörermodelle hinsichtlich ihres Einflusses auf die Ortung der virtuellen Schallquellen mithilfe einer Bewertungsskala, ebenfalls von 1 („vergleichsweise sehr gut“) bis 5 („vergleichsweise sehr schlecht“), miteinander verglichen werden. Die Durchschnittswerte unterscheiden sich dabei mit 2,4 (SD = 1,0) für die *NearPhones*, 2,3 (SD = 1,1) für den *DT 770 PRO* und 2,4 (SD = 1,1) für den *ATH-E70* nur geringfügig. Bei der Gegenüberstellung der individuellen Bewertungen und Fehlerquoten konnte kein systematischer Zusammenhang beobachtet werden.

Tabelle 2: Bewertung des Einflusses auf die Ortung der virtuellen Schallquellen

Modell	NearPhones	DT 770 Pro	ATH-E70
Mittelwert	2,4	2,3	2,4
Standardabw.	1,0	1,1	1,1

Darüber hinaus sollten die Teilnehmer angeben, bei welcher Kopfhörer-Bauart ihres Empfindens nach der Tragekomfort

am höchsten war. 60 % entschieden sich dabei für das extraaurale Modell, 35 % für das circumaurale und nur 5 % (1 Teilnehmer) für das intraaurale (Abb. 4). Die Hypothese, dass der Tragekomfort beim extraauralen Kopfhörer höher ist als bei den Vergleichsmodellen, wurde damit bestätigt.

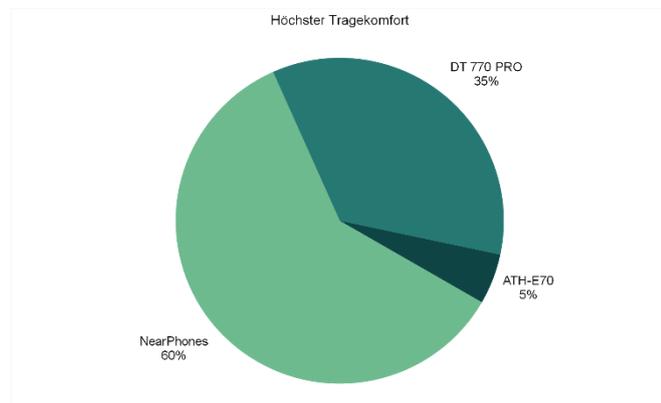


Abbildung 4: Bewertung des Tragekomforts

Diskussion

Bei der Auswertung der Versuchsergebnisse zeigten sich hinsichtlich der gemessenen Größen zwischen den verschiedenen Kopfhörer-Modellen zwar Unterschiede, allerdings konnte dafür mittels der gewählten Analyseverfahren keine statistische Signifikanz nachgewiesen werden.

Interessant zu beobachten war, wie stark sich die Ergebnisse der Teilnehmer interindividuell unterschieden haben. Grund dafür waren möglicherweise die unterschiedlichen Vorerfahrungen, vor allem aber die Tatsache, dass für die dynamische Binauralsynthese gemittelte HRTFs verwendet wurden, die zu manchen Teilnehmern besser gepasst haben als zu anderen. Unabhängig von der statistischen Aussagekraft der Ergebnisse, war es auch überraschend, dass der circumaurale Kopfhörer insgesamt besser abgeschnitten hat als der extraaurale. Eine Erklärung dafür könnte die Methode sein, mit der die *NearPhones* entzerrt wurden. Da für den Versuch keine individuellen Übertragungsfunktionen berechnet wurden, waren Nichtlinearitäten im Frequenzgang maßgeblich von physiologischen Eigenschaften der einzelnen Teilnehmer abhängig. Ein extraauraler Kopfhörer, der nicht dediziert für ein ganz spezielles Einsatzgebiet optimiert ist, wie die *NearPhones*, sondern für eine systemunabhängigere Nutzung ausgelegt ist, hätte vermutlich bessere Ergebnisse geliefert.

In Bezug auf die Anzahl der Fehler waren die interindividuellen Unterschiede für den intraauralen Kopfhörer besonders groß. Die auffällig hohe Varianz lässt sich eventuell dadurch erklären, dass die verwendeten Schaumstoffaufsätze keine perfekte Passform hatten, wodurch es bei manchen Teilnehmern möglicherweise zu Falschpositionierungen der Hörer innerhalb der Ohren und dadurch zu erheblichen Veränderungen des Frequenzgangs kam.

Die Auswertung der Fragebögen ergab, dass die Probanden allgemein gut mit der gestellten Aufgabe zurechtkamen und ihre eigene Performance größtenteils als gut einschätzten. Die Bewertung der akustischen Räumlichkeit war mit 2,1 Punkten im Durchschnitt zufriedenstellend. Unter den Kopfhörermodellen gab es hinsichtlich des Einflusses auf die Lokalisierung

keinen Favoriten, da alle ähnlich stark polarisiert haben. Dabei wurde kein Zusammenhang mit den Messwerten beobachtet. Der höchste Tragekomfort wurde von einer klaren Mehrheit dem extraauralen Kopfhörer zugesprochen. Ein noch eindeutigeres Ergebnis hätte möglicherweise ein noch leichteres Modell geliefert.

Fazit

Die Ergebnisse des Hörversuchs haben deutlich gemacht, dass für die korrekte Übertragung von Spatial Cues die richtige Kopfhörerentzerrung entscheidend wichtig ist und die Bauart allein nicht zwangsläufig einen Vorteil verschafft.

Des Weiteren zeigte sich, dass natürliche akustische Phänomene mit gängigen VR-Tools bereits gut nachgestellt werden können. Es scheint naheliegend, dass Hersteller den Bereich 3D-Audio künftig weiterentwickeln und auditive Elemente verstärkt nutzen werden, um den Immersionsgrad ihrer Anwendungen zu erhöhen. Dadurch würde sicherlich auch die Ton-Wiedergabequalität bei VR-Headsets eine noch größere Rolle spielen als bisher.

Literatur

- [1] Møller, H. (1992): *Fundamentals of Binaural Technology*. In: *Applied Acoustics* 36, S. 171-218.
- [2] Møller, H.; Hammershøi, D.; Jensen, C. B.; Sørensen, M. F. (1995): *Transfer Characteristics of Headphones Measured on Human Ears*. In: *JAES* 43 (4), S. 203-217.
- [3] Schärer, Z.; Lindau, A. (2009): *Evaluation of Equalization Methods for Binaural Signals*. In: *AES Convention* 126.
- [4] Kitanovic, J. (2020): *Evaluation der Ton-Wiedergabesysteme von Virtual Reality-Brillen mit besonderem Fokus auf den Einsatz extraauraler Kopfhörer* (Bachelorarbeit, Ton und Bild). Hochschule Düsseldorf.
- [5] Sünne, N. (2016): *Konzeption und Einrichtung eines Labors für Beschallungstechnik inklusive Beschallungssimulation* (Bachelorarbeit, Ton und Bild). Hochschule Düsseldorf.
- [6] Schmider, E.; Ziegler, M.; Danay, E.; Beyer, L.; Bühner, M. (2010): *Is It Really Robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption*. In: *Methodology* (2010). Göttingen: Hogrefe Publishing. S. 147-151.