

Untersuchung zur Abhängigkeit von Lokalisationsschärfe und Wiedergabeordnung bei Higher Order Ambisonics Lautsprecherwiedergabesystemen mittels Head-Tracking in VR

Piet Brodowski¹ | Oliver Gollan¹ | Marc Hartung¹ | Ulrich Reiter¹

¹ Technische Hochschule Köln

Zusammenfassung

Ambisonics beschreibt ein System, mit dem sich Klangfelder (Gerzon, 1973) aufnehmen und wiedergeben lassen. Während das System in seiner Grundform aus vier Lautsprechern besteht, profitieren hohe Frequenzen von einer höheren Dichte an Lautsprechern, um ein Klangfeld korrekt wiedergeben zu können. Die minimale Anzahl der Lautsprecher eines solchen Systems wird durch die Ordnungszahl gekennzeichnet (Fellgett, 1974). Sie ist damit synonym zur räumlichen Auflösung eines Klangfeldes. Bei Verwendung einer höheren Ordnungszahl als eins spricht man von Higher Order Ambisonics (HOA) (Daniel, 2001). Für den wissenschaftlichen Einsatz eines Systems ist es wichtig, die Grenzen eines solchen Aufbaus zu kennen, um zu wissen, in welchem Rahmen verlässliche Ergebnisse hinsichtlich der räumlichen Auflösung zu erwarten sind. In diesem Paper beschreiben wir eine neue, sehr schnelle Methode, mit der sich die Auflösung eines Systems testen lässt. Hierfür präsentieren wir Testpersonen eine Reihe von auditiven Stimuli und erfassen die dabei wahrgenommene Einfallsrichtung mithilfe eines Virtual-Reality-Headsets (VR-Headsets). Anhand der subjektiven Tests mit dem HOA-System der Technischen Hochschule Köln zeigt sich, dass eine signifikante Verbesserung der Lokalisationsgenauigkeit im Sprung von der ersten auf höhere Ordnungen erreicht wird. Zudem sind Schallquellen auf und oberhalb der Ohrhöhe präziser lokalisierbar als solche die darunterliegen. Dasselbe gilt auch für Sprachsignale gegenüber Rauschsignalen. Interessant ist insbesondere die hier verwendete neue und schnelle Methodik zur Bestimmung der Lokalisationsgenauigkeit mittels eines VR-Headsets.

1. Einleitung

Konventionelle Methoden zur Audiowiedergabe wie Mono oder Stereo sind dazu in der Lage, Schallereignisse in hoher Qualität zu reproduzieren. Jedoch mangelt es diesen Methoden daran, die Direktionalität einer Aufnahme – beispielsweise eines Konzertes – vollumfänglich wiederzugeben. Mit dem System Ambisonics ist eine Möglichkeit gegeben, eine solche Aufnahme als dreidimensionales Klangfeld zu speichern und mit einer Vielzahl von sphärisch angeordneten Lautsprechern zu reproduzieren. Der Name des Systems entspringt von der Intention, die Umgebung (Ambiance), in der eine

Aufnahme durchgeführt wurde, originalgetreu zu reproduzieren (Fellgett, 1974).

Für die räumliche Auflösung eines Klangfeldes ist unter anderem entscheidend, in welcher Ordnung das Wiedergabesignal vorliegt. Mit Zunahme der Ordnungszahl steigt die theoretisch erreichbare räumliche Auflösung, gleichzeitig aber auch die Anzahl der zur Wiedergabe benötigten Lautsprecher. Diverse Mindestanzahlen sind Tabelle 1 zu entnehmen. Dabei zeichnet sich Ambisonics insbesondere durch ein hohes Maß an Flexibilität aus. So lässt sich beispielsweise eine Produktion, die in niedrigerer Ordnung vorliegt, auf einem System höherer Ordnung wiedergeben. Dabei sind, obwohl die räumliche Auflösung des Systems in diesem Fall nicht ausgereizt werden kann, weiterhin sämtliche vorhandene Lautsprecher an der Wiedergabe eines Schallereignisses beteiligt. Darüber hinaus lässt sich unter Verwendung eines Headtrackers eine Ambisonics-Produktion auch auf Kopfhörern wiedergeben. Des Weiteren hebt sich Higher Order Ambisonics (HOA) gegenüber binauraler Kopfhörerwiedergabe dadurch ab, dass keine Gefahr einer im-Kopf-Lokalisation besteht, die häufig durch Binauralsynthese auf Basis nicht-individualisierter Head Related Transfer Functions (HRTFs)

Korrespondenz zum Artikel

Ulrich Reiter

E-Mail: ulrich.reiter@th-koeln.de

Bitte zitieren als

Brodowski, P., Gollan, O., Hartung, M., & Reiter, U. (2023). Untersuchung zur Abhängigkeit von Lokalisationsschärfe und Wiedergabeordnung bei Higher Order Ambisonics Lautsprecherwiedergabesystemen mittels Head-Tracking in VR. *LiveCom-LAB Conference Proceedings*, 1, 81-87. <https://doi.org/10.25716/thm-228>

verursacht wird (Møller et al., 1999). Damit wird HOA neben der reinen Audioszenenwiedergabe auch für Anwendungen in Augmented Reality (AR) und Virtual Reality (VR) interessant.

Tabelle 1
Anzahl benötigter Lautsprecher je Ordnungszahl (Fellgett, 1974)

Ordnung	Mindestanzahl benötigter Lautsprecher
0	1
1	4
2	9
3	16
n	$(n + 1)^2$

Wie in Tabelle 1 zu erkennen ist, steigt die Anzahl der notwendigen Lautsprecher schon mit den ersten höheren Ordnungen rapide. In der Vergangenheit wurden bereits subjektive Studien mit der Absicht durchgeführt, einen Gewinn an Lokalisationsschärfe mit höheren Ordnungen von Ambisonics zu bemessen. Beispielsweise haben Thresh et al. (2017) die räumliche Auflösung von Stimuli eines Ambisonics-Lautsprecherarrays bis zur fünften Ordnung untersucht. Dort wurden die Probanden allerdings angewiesen, ihre Blickrichtung während des Hörversuchs beizubehalten, sodass nicht die Auflösung des Systems, sondern die des menschlichen Gehörs im Vordergrund stand. Darüber hinaus wurden Daten erfasst, indem Probanden zwei identische Schallquellen pro Item präsentiert bekamen, von der die erste die Referenzschallquelle mit einer festgelegten Position darstellte und die zweite über einen Controller bewegt und mit der ersten zur Deckung gebracht werden sollte.

Valzolgher et al. (2020) erforschten in ihrer Arbeit den Einfluss von visueller Unterstützung bei der Lokalisation von Schallquellen. Dabei machten sie von einer VR-Umgebung Gebrauch, in der Probanden den Ursprung einer Schallquelle mit und ohne Gitterlinien verorten sollten und mithilfe eines Trackers in der Hand in Richtung der Quelle zeigen sollten.

Im Zuge einer Vermessung der räumlichen Auflösung des Lautsprecherarrays an der Technischen Hochschule Köln stellen wir ein neues VR-basiertes Verfahren vor, mit dem sich derartige Hörversuche schnell und präzise durchführen lassen. Der Fokus der Vermessung liegt auf einem Vergleich zwischen den ersten drei Ordnungen, die wir mit dem Lautsprecherarray reproduzieren können. Mithilfe des Verfahrens führen wir eine Studie durch, um unserer Forschungsfrage nachzugehen, mit welcher Genauigkeit Stimuli verschiedener Ordnungen in einem Ambisonics-Lautsprecherarray lokalisiert werden können. Im Vergleich zu den vorigen Werken ist hier Gegenstand der Untersuchung, welche Lokalisationsschärfe von Schallereignissen mit einem spezifischen Lautsprecherarray erreicht werden kann. Daher werden Probanden angewiesen, sich um die eigene Achse zu drehen und ihren Kopf frei zu bewegen. Dabei wird ihnen eine Reihe von unterschiedlich positionierten Stimuli präsentiert, die jeweils in der ersten, zweiten und dritten Ordnung vorliegen. Für die Datenerhebung machen wir uns das Tracking eines VR-

Headsets zunutze. Anders als Valzolgher et al. (2020) verwenden wir keinen getrackten Zeiger, sondern erfassen die Blickrichtung des VR-Headsets. Auf diese Art und Weise sind Probanden dazu befähigt, intuitiv und eigenständig ihre Vermutung einzugeben. Abschließend werten wir die subjektiv ermittelte Lokalisationsschärfe zwischen den untersuchten Ordnungen aus und stellen die Veränderungen graphisch dar, bevor wir unsere gewonnenen Erkenntnisse zur Diskussion stellen.

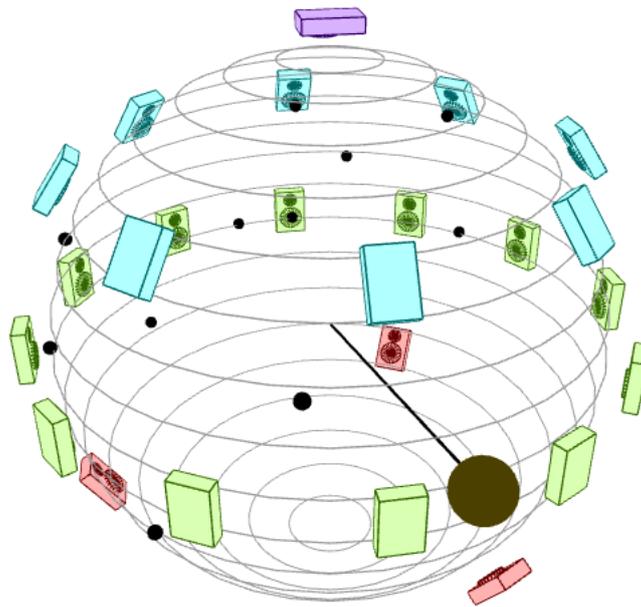
2. Methodik

Das Ziel dieses Versuches ist es, das Wiedergabesystem zu charakterisieren. Somit sollen für die Teilnehmer:innen die bestmöglichen Bedingungen geschaffen werden, die Einfallsrichtung so genau wie möglich zu bestimmen. Deshalb wurde festgelegt, dass der mit der Blickrichtung zu erfassende Stimulus fortlaufend wiederholt wird, bis die Teilnehmenden eine Eingabe tätigen. Des Weiteren wurden die Frequenzbereiche der Stimuli so gewählt, dass sich deren Einfallsrichtungen sowohl per *Interaural Time Difference* (ITD) als auch per *Interaural Level Difference* (ILD) bestimmen lassen. ITD und ILD sind die beiden Hauptmechanismen, anhand derer das menschliche Gehör die Einfallsrichtung eines Schallereignisses bestimmt (Yost & Dye, 1997). ITD beschreibt in diesem Fall den Laufzeitunterschied zwischen dem der Quelle zugewandten und dem der Quelle abgewandten Ohr. Als ILD beschreibt man den Pegelunterschied zwischen linkem und rechtem Ohr. Aufgrund physikalischer und anatomischer Limitationen sind diese Mechanismen nicht gleichermaßen effektiv bei der Lokalisation von Schallereignissen im gesamten Frequenzbereich des menschlichen Gehörs und ergänzen sich somit.

Abbildung 1
Ein Versuchsteilnehmer bei der Durchführung



Abbildung 2
Schematische Ansicht des Versuchsaufbaus



Anmerkung: Schwarze Punkte repräsentieren die verwendeten Positionen für Stimuli. Der schwarze Strahl stellt die Blickrichtung einer im Zentrum des Systems befindlichen Testperson dar. Lautsprecher in grün liegen auf Ohrhöhe bei 0° Elevation. Lautsprecher in rot liegen bei $-17,5^\circ$, während blau eingefärbte Lautsprecher bei 35° Elevation liegen. Der violette Lautsprecher liegt bei 90° Elevation.

Als Stimulus diente in dieser Studie ein Noise Burst im Frequenzbereich von 250 Hz bis 4000 Hz und einer Dauer von 200 ms gefolgt von einer ebenso langen Pause. Die Pausen dienen dazu Transienten zu schaffen, welche für die laufzeitbasierte Lokalisation notwendig sind. Zusätzlich wurde ein Sprachsignal als zweiter Stimulus gewählt, das 2 Sekunden dauert. Das Noise Burst Signal lässt sich aufgrund seiner Eigenschaften besonders gut lokalisieren, während das Sprachsignal einen praktischen Anwendungsfall darstellt.

Das an der Technischen Hochschule Köln zur Verfügung stehende Lautsprecherarray ist mit insgesamt 24 Lautsprechern dazu in der Lage, Ambisonics bis zur dritten Ordnung wiederzugeben. Die Lautsprecher sind auf einer gedachten Kugeloberfläche mit einem Durchmesser von ca. 4 m angeordnet. Der Testraum entspricht hinsichtlich Nachhallzeitverlauf mit einer mittleren Nachhallzeit von 221 ms zwischen 200 Hz und 4 kHz der ITU Recommendation BS.1116-3 (2015) und liegt auch in den anderen Frequenzbereichen innerhalb der vorgegebenen Toleranzen. Der Raum weist einen nahezu ebenmäßigen Abfall des Schallpegels auf.

Um überprüfen zu können, welchen Einfluss die Ordnung des Wiedergabesignals auf die Lokalisationsschärfe hat, lagen beide Signale in den Ordnungen 1 bis 3 vor. Durch die Platzierung dieser sechs Audiosignale auf einer Sphäre anhand von fünf Kategorien ergeben sich 30 unterschiedliche Stimuli, die den Proband:innen präsentiert wurden. Die Kategorisierung wurde vorgenommen, um die relative Position der Phantomschallquelle zu den realen Lautsprechern bei der Auswertung berücksichtigen zu können. Die 5 Kategorien lauten:

1. Der Einfallswinkel des Stimulus hat keine Elevation und der Azimut entspricht dem eines Lautsprechers
2. Der Einfallswinkel des Stimulus hat keine Elevation und der Azimut liegt zwischen zwei Lautsprechern
3. Elevation und Azimut sind so gewählt, dass sich der Einfallswinkel mittig zwischen drei Lautsprechern befindet
4. Der Einfallswinkel des Stimulus hat positive Elevation und der Azimut entspricht dem eines Lautsprechers
5. Der Einfallswinkel des Stimulus hat negative Elevation und der Azimut entspricht dem eines Lautsprechers

Wenn pro Kategorie ein Stimulus des Sprachsignals und ein Stimulus des Noise Bursts vertreten ist, ergeben sich 10 unterschiedliche Stimuli für jede Ordnung. Durch die symmetrische Anordnung der Lautsprecher ist es möglich, die Voraussetzungen der einzelnen Kategorien in mehreren Azimut-Winkeln auf der Sphäre zu erfüllen. Somit lassen sich beispielsweise Stimuli der ersten Kategorie auf den Azimut-Winkeln 0° , $\pm 90^\circ$ sowie 180° untersuchen. Um für die Kategorien jeweils zusätzliche Datenpunkte zu generieren, wurde jeder zweite Stimulus doppelt erfasst. Dies führt dazu, dass zu den 10 unterschiedlichen Stimuli noch 5 Wiederholungen pro Ordnung dazukommen, weshalb also 15 Stimuli je Ordnung präsentiert werden. Auf Grundlage des Lateinischen Quadrats (Campbell, G., & Geller, S., 1980) wird eine Pseudorandomisierung vorgenommen, um für jede Testperson eine unterschiedliche Reihenfolge der

Stimuli zu generieren. Durch diese Maßnahme sollen Reihenfolgeeffekte vermieden werden.

2.1 Technische Umsetzung

Für das Experiment wurde eine HTC Vive Pro inklusive eines Vive Controllers und zweier Basisstationen vom Typ SteamVR Base Station 2.0 eingesetzt, die zum Tracking der Brille und des Controllers im Raum dienten. Die Ansteuerung sowie Abfrage der Sensordaten von Headset und Controller erfolgte über die Schnittstelle SteamVR, die in die Videospiele-Engine Unity eingebunden wurde. Im Editor der besagten Engine erstellten wir eine simple VR-Umgebung, bestehend aus einer schwach texturierten und farblich einfach gestalteten Kugel, die deckungsgleich mit den Dimensionen der kugelförmigen Anordnung des Ambisonics-Lautsprecher-systems war. Zum einen diente sie dazu, der Testperson ein minimales visuelles Feedback zu geben, während diese innerhalb des realen Lautsprechersystems auf einem Drehhocker frei rotieren durfte, ohne jedoch visuelle Anhaltspunkte zur Ortung der Schallquellen zu bieten. Zum anderen wurde sie dazu genutzt, einen Schnittpunkt zwischen dem Vektor der Blickrichtung und der Kugeloberfläche – und damit die Einfallrichtung einer Schallquelle – zu bestimmen. Die dabei entstehenden Schnittpunkte wurden durch Umrechnung von kartesischen zu Kugelkoordinaten mit den tatsächlichen Koordinaten der Schallquellen vergleichbar gemacht, welche in der zum Abspielen der Stimuli verwendeten Digitalen Audio Workstation (DAW) über den Radius sowie Winkel von Azimut und Elevation definiert wurden.

2.2 Versuchsablauf

Die Durchführung des Versuches fand vom 25. Januar bis 28. Januar 2022 in den Räumlichkeiten der Technischen Hochschule Köln statt. Einladungen für die Teilnahme wurden im Vorfeld an die Studierenden versandt. Die Versuchsdauer für die insgesamt 45 Stimuli betrug mit kurzer Einführung für die Testpersonen jeweils etwa 20 Minuten.

Die Versuchsleiter gaben eine kurze Einführung in die Funktionsweise des Testaufbaus. Sie assistierten bei der Anpassung der VR-Brille und erklärten die Steuerung durch den Controller. Mit Beginn eines Versuchsdurchlaufs wurden den Teilnehmenden nun die insgesamt 45 Stimuli nacheinander in der festgelegten Reihenfolge dargeboten. Jeder der Stimuli wurde in einer Wiederholungsschleife abgespielt, bis eine Eingabe über die entsprechende Taste des Controllers erfolgte. Waren die Teilnehmenden sicher, die Einfallrichtung des Schallereignisses lokalisiert zu haben, bestätigten sie dies durch eine erneute Eingabe über eine separate Taste am Controller. Die Blickrichtung wurde damit erfasst und gespeichert. Im Anschluss begann die Wiedergabe des nächsten Stimulus. Mit Erfassung des letzten Stimulus endete der Versuch.

3. Auswertung

3.1 Deskription des Datensatzes

Die erhobenen Daten setzen sich aus den Ergebnissen von 33 subjektiven Tests zusammen, die von den Proband:innen in der Hochschule absolviert wurden. Mit je 45 Items pro Test beläuft sich die Zahl der insgesamt vorhandenen Datenpunkte auf $N = 1485$. In der Studie sind männliche Probanden mit einem Anteil von knapp über 80 % überrepräsentiert. Das durchschnittliche Alter der Teilnehmenden beträgt etwa 27 Jahre. Alle Teilnehmenden schätzen sich selbst als normalhörend ein.

Neben den erfassten Variablen wurden außerdem dichotome Unterscheidungen zwischen Signalen codiert. Dazu zählen Informationen wie Stimulus genau auf Lautsprecher oder nicht und Stimulus unterhalb Kopfhöhe oder nicht. Zudem wurde eine kombinierte Kenngröße als Addition von Azimut- und Elevationsdifferenz berechnet, die horizontale und vertikale Abweichungen zur Schallquelle zusammenfasst. Dabei wird mit den absoluten Werten gerechnet, wodurch sich Abweichungen in negativer und positiver Richtung nicht aufheben.

Bei der Auswertung der Studienergebnisse wurde zudem mit Daten gearbeitet, die um den Gruppenmittelwert zentriert wurden. Durch die Subtraktion des Mittelwertes innerhalb jedes Clusters von je 45 Items wird die Integrität der Vergleichbarkeit gewahrt, da die Trageposition der VR-Brille bei jeder Person einer minimalen Schwankung unterliegt. Durch das Prinzip des *Cluster Mean Centering* (Enders & Tofighi, 2007) wird eine entsprechende Offsetkorrektur vorgenommen, die es erlaubt, Effekte unabhängig von dieser Problematik zu betrachten.

3.2 Quantitative Analysen

Zunächst wird ein Überblick zur grundlegenden Korrelation der verwendeten Variablen aufgezeigt. Dazu sind in Tabelle 2 die wichtigsten Kenngrößen der zugehörigen ANOVA zur Varianzanalyse entnehmbar.

Dabei stellt die gepaarte Differenz von Azimut und Elevation zur ermittelten Blickrichtung die abhängige Variable (AV) dar. Es ist festzuhalten, dass sowohl die Ordnung des Ambisonics-Systems als auch die Frage, ob der Stimulus auf Ohrhöhe liegt oder nicht, eine höchstsignifikante Korrelation mit der AV aufweisen. Auch der Stimulustyp, getrennt nach Rausch- und Sprachsignal, korreliert sehr signifikant mit der Ausprägung der AV. Hingegen ist kein signifikanter Zusammenhang zwischen der Stimulusposition in Relation zum Lautsprecher festzustellen. Damit lässt sich erwartungsgemäß keine Abhängigkeit zwischen Lokalisationsschärfe und der Unterscheidung von Direkt- und Phantomschallquellen beobachten. Um einen Eindruck der horizontalen und vertikalen Differenzwerte zur ermittelten Blickrichtung zu erhalten, sind in Abbildung 3 der Azimut (links) und die Elevation (rechts) getrennt nach Ordnung abgebildet.

Tabelle 2
Varianzanalyse

AV: Azimut und Elevation	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Ordnung***	2	6079.44	3039.72	7.48	.0006
Stimulustyp**	1	3771.69	3771.69	9.28	.0024
Stimulus auf LS	1	324.57	324.57	0.80	.3716
Stimulus unter Ohrhöhe***	1	24425.67	24425.67	60.10	.0000
Residuals	1479	346695.59	234.41		

Anmerkung: AV = Abhängige Variable. Signifikanzwerte: * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$

Dabei kennzeichnet die horizontale Achse jeweils die in der DAW konfigurierte Position des Stimulus. Bei der Betrachtung des Azimuts entspricht ein negativer Wert einer Abweichung in die linke Richtung; folglich nach rechts bei einem positiven Wert. Negative Werte gleichen bei der Elevation einer Abweichung nach unten, während positive nach oben zeigen.

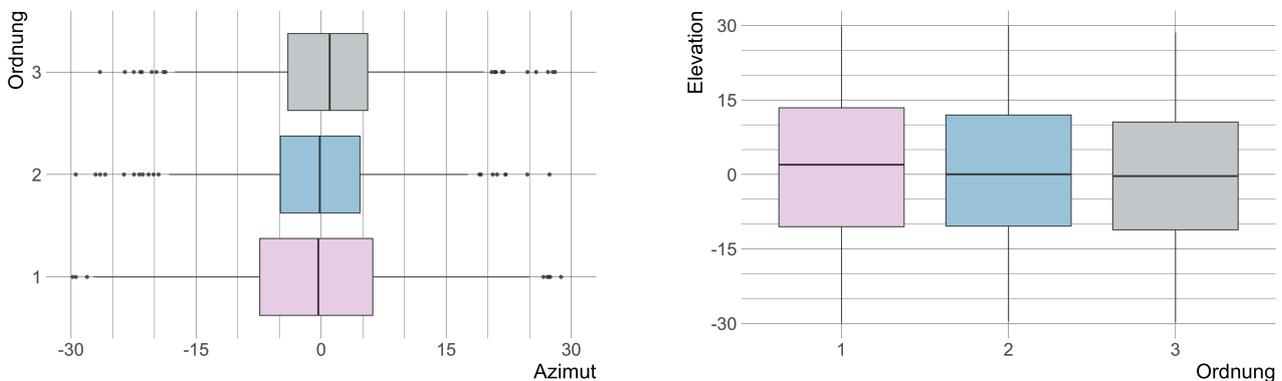
Es ist auffällig, dass vor allem die Azimutdifferenz bei Stimuli 1. Ordnung einen großen Unterschied zu jenen der 2. und 3. Ordnung aufzeigen. Die Streuung um die Schallquelle ist bei Signalen 1. Ordnung vergleichsweise breit. Auch statistisch ist dabei eine moderate negative Korrelation mit $r = -.048$ zu sehen, die eine sinkende Abweichung (also einen geringeren Lokalisationsfehler) bei steigender Ordnung impliziert. Jedoch

wirkt die Signifikanz mit $p = .063$ dieser These entgegen. Unabhängig davon ist die Beobachtung trotzdem belastbar, wenn die 2. und 3. Ordnung kombiniert betrachtet werden. In diesem Fall zeigt sich eine negative Korrelation mit $r = -.068$ bei $p = .001$, die einen signifikant geringeren Lokalisationsfehler mit dem Wechsel von der 1. Ordnung auf die höheren Ordnungen 2 und 3 assoziiert.

Ein gerichteter Zusammenhang zwischen steigender Ordnung und sinkender Abweichung zur tatsächlichen Position der Schallquelle ist bei Betrachtung der Elevation in Abbildung 3 (rechts) deutlich erkennbar. Auch hier zeigt sich mit $r = -.080$ bei $p = .002$ eine signifikante Korrelation, die gegenüber der horizontalen Lokalisationsschärfe noch deutlicher ausgeprägt ist.

Abbildung 3

Horizontale und vertikale Differenz zwischen Schallquelle und erfasster Blickrichtung in Grad



Gleichermaßen wird anhand eines direkten Vergleichs der gemittelten Azimut- und Elevationsdifferenzwerte deutlich, dass die Lokalisation auf der horizontalen Achse mit $\bar{X}_h = 9.4^\circ$ gegenüber der vertikalen Achse mit $\bar{X}_v = 16.3^\circ$ weniger stark um die Schallquelle gestreut ist. Im Durchschnitt fällt damit die vertikale Lokalisation mit einem Faktor von 1.7 unpräziser gegenüber der horizontalen aus.

Die Ergebnisse sind grafisch in Form einer Heatmap in Abbildung 4 zusammengefasst, die als gerasterte Dichtefunktion der ermittelten Lokalisationspunkte getrennt nach 1. Ordnung und der Kombination von 2. und 3. Ordnung skizziert ist. Der Plot bildet die Häufigkeit der gemessenen Lokalisationspunkte innerhalb der

begrenzten Fläche von $A=90 \times 90$ Einheiten ab. Zur vereinfachten Beschreibung wird die Dichte als Intervall von $[0; \max]$ definiert. Für die erste Ordnung entspricht der Maximalwert etwa 6 Beobachtungen pro Rasterpunkt, während es für die 2. und 3. Ordnung etwa 10 sind.

Zuletzt sollen die Auswirkungen ausgewählter Variablen in einem linearen Regressionsmodell in Tabelle 3 betrachtet werden. Hierbei zeigt sich, welche Vorhersage durch Änderung der unterschiedlichen Prädiktoren zur Auswirkung auf die mittlere Abweichung zwischen Stimulusposition und detektierte Blickrichtung formuliert werden kann.

Abbildung 4

Koordinaten der erfassten Blickrichtungen in Relation zur Schallquelle getrennt nach 1. Ordnung und der Kombination von 2. und 3. Ordnung

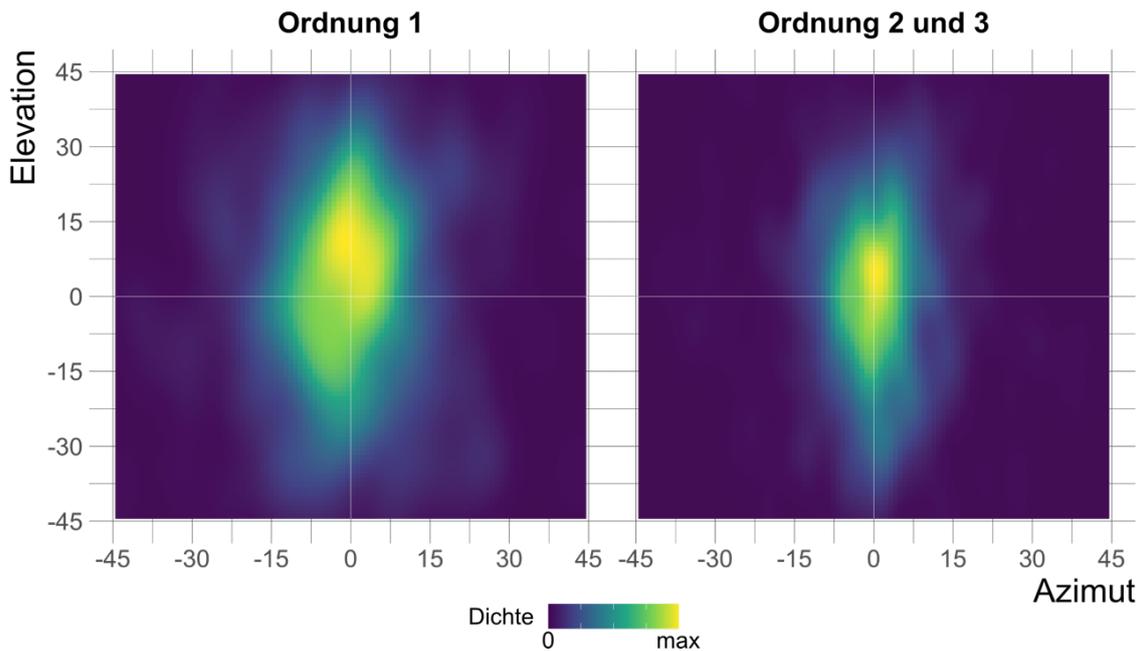


Tabelle 3
Lineares Regressionsmodell

AV: Azimut und Elevation	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	39.7981	1.7160	23.19	.0000***
Ordnung				
1 (ref)				
2	-4.4770	1.2874	-3.48	.0005***
3	-4.5412	1.2874	-3.53	.0004***
Stimulustyp				
Rauschen (ref)				
Sprache	-3.2596	1.0520	-3.10	.0020**
Stimulus unter Ohrhöhe				
Ja (ref)				
Nein	-10.9297	1.5439	-7.08	.0000***

Anmerkung: AV = Abhängige Variable. ref = Referenzkategorie. Signifikanzwerte: * = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$

Alle unabhängigen Variablen sind hierbei mindestens sehr signifikant. Laut linearem Regressionsmodell begünstigt der Wechsel von 1. Ordnung zu einer höheren Ordnung die Lokalisationsschärfe durchschnittlich um etwa 4.5° . Auch die Unterscheidung zwischen Rausch- und Sprachsignal hat einen mit einem Schätzwert von leicht über 3° nennenswerten Einfluss auf die Lokalisationsschärfe. Den am stärksten ausgeprägten Einfluss stellt der Prädiktor zur Relation von Stimulus- und Kopfposition in vertikaler Richtung dar. Hierbei wird unterschieden, ob sich die Schallquelle unterhalb oder auf Ohrhöhe und darüber befindet. Eine mögliche Erklärung für diesen stark gerichteten Effekt von etwa -11° ist die vergleichsweise niedrige Lautsprecherauflösung auf Höhe des unteren Rings des verwendeten Ambisonics-Systems mit lediglich drei Lautsprechern.

4. Fazit

Anhand quantitativer Analysen konnten wir zeigen, dass sowohl die Ordnung des Ambisonics-Systems als auch die Frage, ob Stimuli auf Ohrhöhe liegen oder nicht, eine höchstsignifikante Korrelation mit der Lokalisationsgenauigkeit aufweisen. Auch der Stimulustyp, getrennt nach Rausch- und Sprachsignal, korreliert sehr signifikant mit der Ausprägung der Präzision. Hingegen ist kein signifikanter Zusammenhang zwischen der Stimulusposition in Relation zum Lautsprecher festzustellen. Damit lässt sich keine Abhängigkeit zwischen Lokalisationsschärfe und der Unterscheidung von Direkt- und Phantomschallquellen wie bei herkömmlichen Lautsprecher-Setups beobachten.

Vor allem die Azimutdifferenz bei Stimuli 1. Ordnung zeigt einen großen Unterschied zu jenen der 2.

und 3. Ordnung. Konkret bedeutet dies für die Anwendung im VR- und AR-Umfeld, dass der Mehraufwand (mehr Lautsprecher, mehr DAW-Rechenleistung, etc.), der für den Übergang von der ersten zur zweiten Ordnung getrieben werden muss, für Wiedergabesetups wie das hier beschriebene gerechtfertigt erscheint. Ein weiterer Übergang zur dritten Ordnung hingegen erhöht die Lokalisationsschärfe – zumindest im von uns getesteten Setup – nicht signifikant.

Die präsentierten Ergebnisse belegen, dass die von uns neuentwickelte Testmethodik, die im Wesentlichen auf dem Einsatz des in der VR-Brille eingebauten Headtrackers basiert, schnelle und präzise Ergebnisse liefert, die mit den Ergebnissen anderer ähnlicher Studien zur Lokalisationsgenauigkeit von HOA-Systemen unterschiedlicher Ordnung vergleichbar sind.

Aufgrund des VR-Headsets, das während der Studie von den Proband:innen getragen wurde, lässt sich argumentieren, dass die Größe und Form dieses zusätzlichen Fremdkörpers im Kopfbereich Einfluss auf die Pegel- und Laufzeitunterschiede bei der auditiven Wahrnehmung nehmen könnten, dass also die probanden-eigenen HRTFs durch das Tragen der VR-Brille verändert wurden. Dieser Argumentation folgend könnte die Lokalisationsschärfe durch Training der Testpersonen bei aufgesetzter VR-Brille verbessert werden, so dass dann u.U. auch ein Übergang zu höheren Ordnungen zu besserer Diskrimination führen könnte. Es ist also nicht ausgeschlossen, dass sich der Aufwand einer Wiedergabe in höheren Ordnungen doch „lohnen“ könnte. Insbesondere ist festzuhalten, dass die hier erzielten Ergebnisse nur für die spezifisch an der TH Köln existierende Lautsprecheranordnung gelten.

Aus diesem Grund wäre es ebenfalls interessant, die hier vorgestellte Methodik auf unterschiedlichen HOA-Systemen zu testen und die Ergebnisse bzgl. der Lokalisationsschärfe bei unterschiedlichen Ordnungen zu vergleichen. Erst dann könnte eine allgemeingültige Aussage hinsichtlich der für eine bestimmte gewünschte Lokalisationsschärfe (Präzision) erforderlichen Ordnung gemacht werden.

Literaturverzeichnis

- Campbell, G., & Geller, S. (1980). Balanced latin squares. *Purdue University Department of Statistics Mimeoseries*, 80(26), 3-1
- Daniel, J. (2001). *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. PhD thesis, Université Paris 6
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121–138. <https://doi.org/10.1037/1082-989x.12.2.121>
- Fellgett, P. B. (1974). Ambisonic reproduction of directionality in surround-sound systems. *Nature*,

252(5484), 534–538.
<https://doi.org/10.1038/252534b0>

- Gerzon, M. A. (1973). Periphony: with-height sound reproduction. *J. Audio Eng. Soc.* vol. 21, no. 1, 2-10
- Møller, H., Hammershøi, D., Johnson, C.B., and Sørensen, M.F. (1999). Evaluation of Artificial Heads in Listening Tests. *J. Audio Eng. Soc.* vol. 47, no. 3, 83-100
- Recommendation ITU-R BS.1116-3. (2015). Methods for the subjective assessment of small impairments in audio systems.
- Thresh, L., Armstrong, C., & Kearney, G. (2017). A Direct Comparison of Localisation Performance When Using First, Third and Fifth Order Ambisonics For Real Loudspeaker And Virtual Loudspeaker Rendering. *Proc. AES 143th Convention, New York*.
- Valzolgher, C., Alzhaler, M., Gessa, E., Todeschini, M., Nieto, P., Verdelet, G., Salemme, R., Gaveau, V., Marx, M., Truy, E., Barone, P., Farnè, A., & Pavan, F. (2020). The impact of a visual spatial frame on real sound-source localization in virtual reality. *Current Research in Behavioral Sciences*, 1, 100003.
<https://doi.org/10.1016/j.crbeha.2020.100003>
- Yost, W. A., & Dye, R. H. (1997). Fundamentals of directional hearing. In *Seminars in Hearing*. vol. 18, No. 04, 321-344