Herausforderungen bei der Beurteilung der wahrgenommenen Qualität räumlicher Audiosignale

Judith Liebetrau[1], Thomas Sporer[2]

[1] *Fraunhofer IDMT, 98693 Ilmenau, Deutschland, Email: judith.liebetrau@idmt.fraunhofer.de*

[2] *Fraunhofer IDMT, 98693 Ilmenau, Deutschland, Email: thomas.sporer@idmt.fraunhofer.de*

Einleitung

Mittels räumlicher Audiowiedergabesysteme sollen Audio-Inhalte und akustischen Umgebungen richtungsgerecht und realitätsnah reproduziert sowie Immersion beim Hörer erzeugt werden. Unterschiedliche Wiedergabesysteme für dreidimensionales Audio wurden in den letzten Jahren entwickelt, erprobt und gegenübergestellt. Auch im Bereich der Standardisierung (SMPTE, ATSC, MPEG) ist ein Vergleich von unterschiedlichen Wiedergabeformaten und –systemen im Sinne der wahrgenommenen Qualität notwendig geworden. Dieser Beitrag diskutiert Probleme bei der Qualitätsbewertung räumlicher Wiedergabesysteme und zeigt verschiedene Alternativen zur Untersuchung auf.

Räumliche Audiowiedergabeverfahren

Das Ziel der räumlichen Audiowiedergabe ist die Erzeugung eines realitätsnahen, räumlichen Klangerlebnisses. Vielfältige Verfahren zur Kopfhörer- oder Lautsprecherwiedergabe wurden entwickelt. Einen umfassenden Überblick und eine detaillierte Beschreibungen einzelner Verfahren können in entsprechender Fachliteratur, wie z.B. [1] gefunden werden.

Die Binauralsynthese ist ein hörerzentrierter Ansatz, der auf einer korrekten Synthese der Ohrsignale beruht. Dieses Verfahren eignet sich besonders für Kopfhörerwiedergabe, ist aber nur mit Einschränkungen für die Lautsprecherwiedergabe nutzbar [2]. Für die Wiedergabe von räumlichen Audiosignalen über Lautsprecher werden kanalbasierte Ansätze oder aber Schallfeldsyntheseverfahren angewendet. Bei ersterem werden Phantomschallquellen zur Erzeugung des räumlichen Eindrucks genutzt. Die Anordnung der Lautsprecher in Mehrkanal-Tonsystemen ist oft standardisiert, vgl. [3, 4]. Im Fall von kanalbasiertem Audio muss bei der Produktion das Zielsetup bekannt sein, da entsprechende diskrete Lautsprechersignale vorproduziert werden.

Schallfeldreproduktionsverfahren, wie beispielsweise Ambisonics [5] und dessen Weiterentwicklung oder Wellenfeldsynthese (WFS) [7], zielen auf eine Synthese von Schallfeldern in einem gegebenen Raumvolumen ab [6]. Die Lautsprechersignale werden für jedes Wiedergabesetup berechnet. Bei Ambisonics entspricht dies den Werten für Schalldruck- und Schallschnelle für jede einzelne Lautsprecherposition. Objektbasierte Wiedergabeverfahren, wie beispielsweise WFS, beruhen auf Audiosignalen (Objekten) denen Metadaten zugeordnet sind. Die Metadaten beschreiben wie das Schallereignis durch den Hörer wahrgenommen werden soll und entsprechen den momentanen Eigenschaften des Audiosignals (Pegel und Frequenzgang, seine Koordinaten im Raum, den Phasenbezug zu anderen Kanälen, Breite, Lebensdauer oder Bewegung). Unter Einbeziehung der Eigenschaften des Wiedergaberaumes, insbesondere Anzahl und Ort der Lautsprecher, werden durch einen Signalprozessor (Renderer) die individuellen Audiosignale für jeden Lautsprecher berechnet. Die Summe aller Lautsignale bildet das gewünschte Schallfeld. Durch die angepassten Berechnungen des Signalprozessors ist die objektbasierte Wiedergabe skalierbar, solange dem Signalprozessor die akustischen Eigenschaften des Wiedergabeortes bekannt sind. Mischformen aus allen drei Paradigmen sind üblich, z.B. MPEG-H 3D Audio. Alle Verfahren haben gemein, dass sie eine Immersion des Zuhörers und sehr gute Klangqualität versprechen.

Standardisierte Methoden zur Bewertung der wahrgenommenen Audioqualität

Die beiden bekanntesten Standards zur Beurteilung von Audioqualität sind ITU-R BS.1116 [8] und BS.1534 [9]. Während ersterer Standard für die Untersuchung von kleinen wahrnehmbaren Unterschieden entwickelt wurde, sollte letztere Methode für die Evaluierung von moderaten Unterschieden eingesetzt werden. Beide Hörtestmethoden vergleichen die Qualität eines Systems unter Test (SUT) gegenüber der Qualität einer offenen Referenz. Jegliche wahrnehmbare Veränderung des SUT im Vergleich zur Referenz wird als Qualitätsminderung aufgefasst und entsprechend bewertet. Diese Verschlechterung wird bei einem Hörtest nach ITU-R BS.1116 anhand einer fünfstufigen „impairment scale“ durchgeführt, wie sie in Tabelle 1) abgebildet ist. Abbildung 1) zeigt eine „continuous quality scale“, die in fünf gleichgroße Bereiche eingeteilt und bei Hörtests nach ITU-R BS.1534 verwendet wird.

Tabelle 1: Fünfstufige Impairment scale zur Bewertung der wahrgenommenen Audioqualität nach ITU-R BS.1116.

Impairment	**Grade**
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

Die ITU-R BS.1116-Methodik ist eine Doppelblind-Hörtestmethode, mit zwei SUT: A und B. Eines der bei-

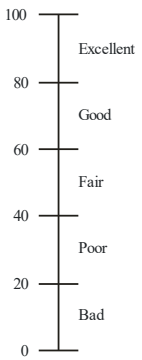

Abbildung 1: Kontinuierliche Qualitätskala zur Bewertung der wahrgenommenen Audioqualität nach ITU-R BS.1534.

den SUT ist immer die versteckte Referenz. Die Evaluierungsaufgabe umfasst zwei Schritte. Zunächst muss der Proband detektieren, ob A oder B die versteckte Referenz ist, d.h. welches Signal gleich der offenen Referenz ist. Im zweiten Schritt wird der Unterschied zwischen dem jeweils anderen Signal und der Referenz beurteilt. Bei einem ITU-R BS.1534-Test werden verschiedene SUT gleichzeitig zur offenen Referenz verglichen. Dabei ist ein SUT die versteckte Referenz und es gibt mindestens ein SUT mit einer besonders schlechten Qualität, der sogenannte Anker. Aus diesem Grund wird dieser Hörtest auch MUSHRA (multi stimulus with hidden reference and anchor) genannt. Auch hier findet eine mehrstufige Bewertung statt. Zunächst wird wieder die versteckte Referenz gesucht und danach die Audioqualität der verbliebenen SUT bewertet. Dabei muss dem Anker die schlechteste Qualitätsbewertung zugeordnet werden.
Beide Testmethoden resultieren in einem gemittelten Wert für die Gesamtqualität (basic audio quality). Die einzelnen Faktoren und deren Beitrag zur wahrgenommenen Qualität können durch diesen gemittelten Wert nicht bestimmt werden.

Probleme bei der Evaluierung von räumlichen Audiosinalen

Zwar gibt es einige standardisierte Testmethoden zur Beurteilung der wahrgenommenen Audioqualität, allerdings können diese nicht ohne weiteres auf die Evaluierung von räumlichen Audiosignalen angewendet werden. Wie oben beschrieben, wird bei diesen Methoden eine Gesamtaussage über die wahrgenommene Audioqualität ermittelt. Dabei stellt sich die Frage, in welchem Zusammenhang Audioqualität und räumliche Audioqualität stehen. Blauert definiert in [10] Audio- bzw. Soundqualität als „die Angemessenheit des Sounds im Kontext eines spezifischen technischen Ziels und/oder einer Aufgabe“. Qualität ist multidimensional und besteht aus unterschiedlichen Elementen [11]. Räumliche Audioqualität kann dementsprechend als Teil der Gesamtqualität verstanden werden. Aber auch die räumliche Qualität setzt sich aus vielen Faktoren, wie bspw. Lokalisierbarkeit, Räumlichkeit, Natürlichkeit, Klang oder Breitenausdehnung zusammen. Verschiedene Attribute [12] oder deskriptives Vokabular [13] werden genutzt, um Einzel- oder Gesamtfaktoren beschreibbar zu machen.
Durch die Vielschichtigkeit der räumlichen Audioqualität ist eine Bestimmung der basic audio quality, wie in den standardisierten Testmethoden üblich, nicht zielführend. Eine Untersuchung einzelner Aspekte der räumlichen Audioqualität erscheint sinnvoll. Nachfolgend werden drei Teilaspekte der räumlichen Audiowiedergabe diskutiert, die insbesondere eine Herausforderung für die Untersuchung von objektbasierten Wiedergabeverfahren darstellen: Lokalisation von Quellen, Klang und AV-Kohärenz. Die Auswahl ist damit begründet, dass die Lokalisationsgenauigkeit mit der Anzahl der Lautsprecher in einem System verbessert, die Größe des Sweet Spots vergrößert wird [14] aber Klangverfärbungen wahrscheinlicher werden. Die Audiowiedergabe mittels objektbasierten Ansätzen ermöglicht den Einsatz von sogenannten virtuellen Quellen, die auch durch den Zuhörerraum bewegt werden können. Dies ermöglicht neue künstlerische Gestaltungsmöglichkeiten und vergrößert unter Umständen die wahrgenommene Räumlichkeit sowie Natürlichkeit des Klangfelds. Die Grenzen der Lokalisationsgenauigkeit für statische Quellen ist relativ gut erforscht [15, 16]. Deutlich weniger Studien zur Lokalisiergenauigkeit von bewegten Quellen sind bekannt [17]. Ähnlich verhält es sich mit Untersuchungen bezüglich der zeitlichen und örtlichen Kohärenz zwischen auditiven und visuellen Stimuli.

Lokalisationsgenauigkeit

Bei objektorientierten Ansätzen beschreiben Metadaten, wie das Schallereignis durch den Hörer wahrgenommen werden soll. Dabei wird u.a. die Position des Audioobjekts, bezogen auf einen Referenzpunkt, angegeben. Um zu überprüfen, wie gut die Lokalisationsgenauigkeit in dem Wiedergabesystem ist, muss folglich die wahrgenommene Position mit der, in den Metadaten angegebenen, verglichen werden.
Für die Bestimmung der Lokalisationsgenauigkeit gibt es keine standardisierte Testmethodik. Je präziser der Testteilnehmer die wahrgenommene Position einer Audioquelle angeben kann, desto akkurater wird die Lokalisationsgenauigkeit gemessen. Diverse Methoden, wie beispielsweise verbale Beschreibung oder Zeigemethoden wurden in der Vergangenheit angewendet. Ein guter Überblick wird in [18] gegeben. Hier wird die Zeigemethode im Allgemeinen als geeignet für Lokalisationstests von räumlichen Audio vorgeschlagen. Auch hier gibt es wieder unterschiedliche Varianten, bei denen es zu großen Abweichungen in der Genauigkeit kommen kann. Bei Lichtzeigermethoden wird ein Lichtzeiger auf die wahrgenommene Schallrichtung eingestellt. Im einfachsten Fall zeigt der Testteilnehmer mit einem Laserpointer auf die entsprechende Position. Allerdings ist diese Variante mit Ungenauigkeiten aufgrund der willkürlichen Bewegung der Hand und Asymmetrie der Bewertungen zwischen Rechts- bzw. Linkshändern behaftet [20]. Um diese Probleme zu umgehen, wurde in einigen Untersuchungen, z.B. [21, 22], der Laserpointer über einen Trackball oder Joystick gesteuert. Ein anderer Ansatz ist die Benutzung

eines Rasters, das auf eine Leinwand vor den Lautsprechern aufgetragen ist. Jeder Sektor ist mit einer eindeutigen Nummerierung versehen, die der Proband zur Lokalisationsangabe nutzt. Hier kann zusätzlich zur Quellenposition auch die Quellenbreite ermittelt werden [23]. Auch wenn mit Zeigemethoden die Richtung der Schallwahrnehmung messbar gemacht wird, kann die Distanzwahrnehmung damit nicht abgebildet werden. In [24] wird eine Möglichkeit gezeigt, wie man neben der Richtung auch die Distanz von wahrgenommenen Schallquellen ermittelt. Es wurde eine grafische Oberfläche (GUI) entwickelt, bei der drei Einzelschallquellen in Richtung und Entfernung zur Sitzposition des Probanden arrangiert werden können (Abbildung 2).

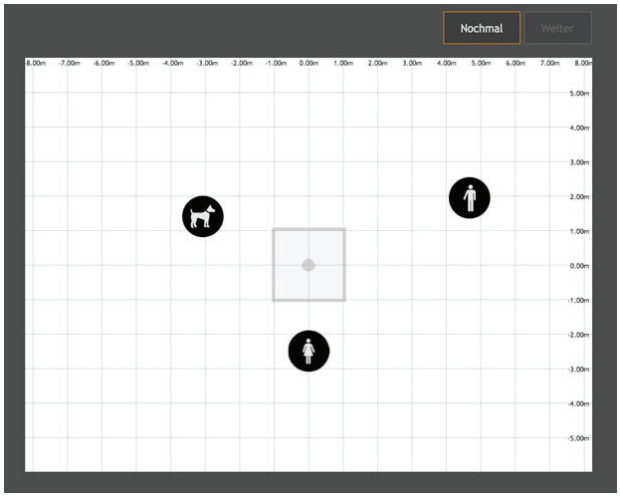

Abbildung 2: GUI zur Bewertung der wahrgenommenen Distanz und Richtung von drei verschiedenen Schallquellen.

Visuelle Eindrücke können die Lokalisation beeinflussen. Als Beispiel sei hier der Bauchrednereffekt genannt. Obwohl ein visueller Stimulus (Handpuppe) nicht örtlich mit einer Schallquelle (Redner) übereinstimmt, wird der Schall unter Umständen am Ort des visuellen Stimulus verortet [25]. Es ist wahrscheinlich, dass die Lautsprecheranordnung und die damit verbundene Erwartungshaltung des Testteilnehmers die Lokalisation prägt. Daher sollte eine visuelle Beeinflussung der Probanden vermieden werden. Dies kann entweder durch akustisch durchlässige Vorhänge vor den Lautsprechern oder aber durch Verbinden der Augen realisiert werden.
Idealerweise ist die Lokalisationsgenauigkeit auf allen Hörerpositionen innerhalb eines Wiedergabesystems gleich gut. Es empfiehlt sich daher die Lokalisationstests sowohl an der Idealposition (Sweet Spot) als auch außerhalb durchzuführen, um eine verallgemeinerte Aussage über die Lokalisationsgenauigkeit zu ermöglichen.

Klang

Die Bewertung des Klangs kann theoretisch mittels standardisierter Hörtestverfahren untersucht werden. Speziell bei dem Vergleich verschiedener objektbasierter Ansätze, ist die Definition der Referenz jedoch schwierig. Die Referenz, im eigentlichen Sinne, entspricht dem Schallereignis, wie es durch den Hörer wahrgenommen werden soll und in den Metadaten beschrieben ist. Aus pragmatischen Gründen wird oftmals das Verfahren mit der vermutlich besten Klangqualität in einem informellen Vortest ausgewählt und zur Referenz ernannt. Dennoch ist es denkbar, dass ein SUT in dem eigentlichen Hörtest eine höhere Klangqualität als die definierte Referenz aufweist. Diesem Problem kann auf unterschiedlicher Weise begegnet werden. Es gibt standardisierte Testmethoden, die ohne eine explizite Referenz auskommen, wie bspw. ACR [26] oder Paarvergleich [27]. Vorteilig ist, dass der Testteilnehmer eine eigene, innere Referenz bildet. Absolute Werte für die Klangqualität der SUT können somit ermittelt und anschließend eine Rangfolge aufgestellt werden. Der parallele Vergleich aller SUT ist zeiteffektiver. Um die Vorteile der MUSHRA-Methode zu erhalten aber Problematik der Referenz zu umgehen, wurde diese Methodik in einigen Studien modifiziert. Dabei wurde der offenen Referenz eine mittlere Audioqualität zugeordnet und die Skalenbeschriftung angepasst [28]. Dies ermöglicht eine "Besserbewertung"der Klangqualität von SUT bezogen auf die Referenz.

AV-Kohärenz

Räumliche Wiedergabesysteme wurden insbesondere für die Anwendung im Zusammenspiel mit Bildwiedergabe entwickelt. Mit objektorientierten Verfahren ergibt sich die Möglichkeit virtuelle Quellen durch den Zuhörerraum zu bewegen. Dadurch ergeben sich besondere Anforderungen an die Kohärenz von Audio und Video, die als Qualitätsmerkmal untersucht werden kann. Auch hier existieren diverse Testmethoden die zur Anwendung bereitstehen. Ein kurzer Überblick ist in [29] gegeben. Dort wird eine 3-AFC (alternative forced choice) Methode nach [30] angewendet, die jedoch nicht uneingeschränkt empfohlen werden kann, da einige Probanden Probleme mit der Testmethodik hatten. Wie beim Lokalisationstest sollte die AV-Kohärenz an unterschiedlichen Hörorten ermittelt werden.

Zusammenfassung

Die Beurteilung der räumlichen Audioqualität ist eine komplexe Aufgabe. Dies liegt u.a. daran, dass sich die räumliche Audioqualität aus vielen Faktoren zusammensetzt. Zusätzlich hat jedes Verfahren zur räumlichen Audiowiedergabe spezifische Vor- und Nachteile. Damit ist eine allgemeine Bewertung der räumlichen Qualität ohne Kenntnis der Systemmerkmale und der speziellen Aufgabenstellung nicht möglich. Die Bewertung einzelner, ausgewählter Qualitätsfaktoren wird als sinnvoll betrachtet. Jeder Faktor kann mit diversen Methoden untersucht werden. Eine allgemeingültige Empfehlung für die Anwendung einer spezifischen Testmethodik kann nicht gegeben werden. Vielmehr muss eine Fallentscheidung durchgeführt und eine der Untersuchungsaufgabe entsprechende Methode ausgewählt bzw. adaptiert werden. Bei der Beurteilung räumlicher Audiosignale sollte die jeweilige Evaluierung an verschiedenen Orten im Wiedergaberaum geschehen, um eine verallgemeinerte Aussage zu ermöglichen.

Literatur

[1] Weinzierl, S.: Handbuch der Audiotechnik. Springer-Verlag, Berlin & Heidelberg, 2008

[2] Klein, F.; Werner, S.: Perspektiven zur Anwendung der Binauralsynthese in der Medienproduktion. Medienproduktion - Ilmenau: Fachgebiet Kommunikationswissenschaft, TU Ilmenau, Bd. 5 (2014), S. 12-14

[3] Recommendation ITU-R BS.775-3: Multichannel stereophonic sound system with and without accompanying picture. 08/2012

[4] Recommendation ITU-R BS.2051: Advanced sound system for programme production. 02/2014

[5] Fellgett, P.: Ambisonics. Part one: general system description, Studio Sound, Vol. 17 no. 8, pp. 20-22, 40, 1975

[6] Daniel, J., Moreau, S., Nicol, R.: Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging. In Audio Engineering Society Convention 114. 2003

[7] Brandenburg, K., Brix, S., Sporer, T.: Wave field synthesis. 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2009 (pp. 1-4)

[8] Recommendation ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems. 02/2015

[9] Recommendation ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality levels of coding systems. 10/2015

[10] Blauert, J.: Product-sound assessment: An enigmatic issue from the point of view of engineering. Proc. Internoise 94 (1994), Vol.2, 857-862

[11] Blauert, J., Jekosch, U.: Sound-quality evaluation – a multi-layered problem. Acta Acustica united with Acustica 83(5) (1997), 747-753

[12] Bech, S., Zacharov, N.: Perceptual audio evaluation-Theory, method and application. John Wiley & Sons, (2007)

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

[14] Rebscher R., Theile G.: Enlarging the Listening Area by Increasing the Number of Loudspeakers, AES preprint No. 2932, 88th Convention Montreux, 1990

[15] Mills, A. W.: The Minimum Audible Angle. Harvard University, Harvard, 1958

[16] Perrott, D. R., Saberi,K.: Minimum audible angle thresholds for sources varying in both elevation and azimuth. Journal of the Acoustical Society of America, vol. 87, pp. 1728–1731, 1990

[17] Harris, J. D., Sergeant, R. L.: Monaural/binaural minimum audible angle for moving sound sources. J. Speech Hear. Res., vol. 14, pp. 618–629, 1971

[18] Majdak, P., Goupell, M. J., Laback, B.: 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. Attention, perception, & psychophysics, 72(2), 454-469, 2010

[19] Haber, L., Haber, R. N., Penningroth, S., Novak, K., Radgowski, H.: Comparison of nine methods of indicating the direction to objects: Data from blind adults. Perception, 22(1), 35-47, 1993

[20] Pinek, B., Brouchon, M.: Head turning versus manual pointing to auditory targets in normal subjects and in subjects with right parietal damage. Brain and cognition, 18(1), 1-11, 1992

[21] Seeber, B.: Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode, Technischen Universität München, Dissertationsschrift, 2002

[22] Melchior, F., Heusinger, U., Liebetrau, J.: Perceptual evaluation of a spatial audio algorithm based on wave field synthesis using a reduced number of loudspeakers. In Audio Engineering Society Convention 131. 2011

[23] Liebetrau, J., Sporer, T., Korn, T., Kunze, K., Mank, C., Marquard, D., Schnabel, M. A.: Localization in Spatial Audio - From Wave Field Synthesis to 22.2. In Audio Engineering Society Convention 123. 2007

[24] Sporer, T., Liebetrau, J., Werner, S., Kepplinger, S., Gabb, T., Sieder, T.: Localization of Audio Objects in Multichannel Reproduction Systems. In Audio Engineering Society 57th International Conference. 2015

[25] Seeber, B., Fastl, H.: On auditory-visual interaction in real and virtual environments. In Proc. ICA 2004, 18th Int. Congress on Acoustics, Japan, volume III, Int. Commission on Acoustics, pp. 2293–2296, 2004

[26] Recommendation ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures. 01/2012

[27] Recommendation ITU-R BS.1284-1. General methods for the subjective assessment of sound quality. 12/2003

[28] Sporer, T., Walther, A., Liebetrau, J., Bube, S., Fabris, C., Hohberger, T., Köhler, A.: Perceptual evaluation of algorithms for blind up-mix. In Audio Engineering Society Convention 121. 2006

[29] Sporer, T., Liebetrau, J., Goecke, D., Brandenburg, K.: Study on spatial coherence of moving audio-visual objects, in Proceedings of the 13th AES Brazil Conference. 2015

[30] Békésy, G.: Experiments in Hearing. Acoustical Society of America through the American Institute of Physics by arrangement with McGraw-Hill Book Company, 1960