



Relevanz von Next Generation Audio im Broadcast für Endnutzer

Untersuchung der Personalisierungsmöglichkeiten von
NGA-Broadcasts

Bachelorarbeit
im Studiengang Audiovisuelle Medien
zur Erlangung des akademischen Grades
Bachelor of Engineering (B. Eng.)

Vorgelegt von:	Marvin Schmid
Matrikelnummer:	38609
Abgabedatum:	29.03.2023
Abgabeort:	Hochschule der Medien Stuttgart
Erstprüfer:	Prof. Oliver Curdt
Zweitprüfer:	Prof. Dr. Frank Melchior

Ehrenwörtliche Erklärung

Hiermit versichere ich, Marvin Schmid, ehrenwörtlich, dass ich die vorliegende Bachelorarbeit mit dem Titel: „Relevanz von Next Generation Audio im Broadcast für Endnutzer“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

Horb am Neckar, den 28.03.2023

Marvin Schmid

Genderhinweis

Aus Gründen der besseren Lesbarkeit wird in dieser Bachelorarbeit das generische Maskulinum verwendet. Die in dieser Arbeit verwendeten Personenbezeichnungen beziehen sich – sofern nicht anders kenntlich gemacht – auf alle Geschlechter.

Kurzfassung

Next Generation Audio ist die umfassendste Revolution im Audiosegment seit Jahrzehnten und beinhaltet unter anderem weitläufige Neuerungen für erhöhte Personalisierung. Forschung in diesem Themengebiet beschränkt sich bislang hauptsächlich auf den technischen Aspekt, Interessen und Meinungen der Endnutzer wurden nur selten Gegenstand der Untersuchung. Die vorliegende Arbeit beschäftigt sich daher mit der Frage, wie perspektivische Zuschauer von Next Generation Audio Broadcasts deutscher Fernsehanstalten die erweiterten Personalisierungsmöglichkeiten beurteilen, wie die neuen Funktionalitäten in Zukunft angenommen werden könnten und wie allgemeingültig Beantwortungen auf diese Fragen abhängig von demografischen Merkmalen für die Gesamtbevölkerung getroffen werden können. Ziel ist es, als Informationsquelle für deutsche Broadcaster zu dienen, worauf der Fokus in der technischen Implementierung von Next Generation Audio gelegt werden könnte.

Hierzu wurde eine Umfrage in der deutschen Bevölkerung ausgetragen, wobei eine ausgesprochen positive Einstellung gegenüber drei der vier behandelten, gängigen Einsatzmöglichkeiten von Next Generation Audio Objekten festgestellt werden konnte. Die Bewertung der vierten Einsatzmöglichkeit fiel neutral aus. Abhängigkeiten der Bewertungen von demografischen Merkmalen fielen dabei größtenteils nicht oder nur mit kleinem Effekt signifikant aus. Basierend auf diesen Ergebnissen kann von einer äußerst zuträglichen Einstellung der deutschen Bevölkerung gegenüber Personalisierung allgemein und im Speziellen der erweiterten Personalisierung durch Next Generation Audio ausgegangen werden. Die beobachteten Einstellungen sind dabei weitestgehend für die gesamte Bevölkerung zutreffend, positive Evaluationen der Funktionen also nicht nur für vereinzelte Gruppen relevant. Bereitschaft, diese nächste Generation des Audiomaterials anzunehmen, scheint demnach vorhanden.

Abstract

Next Generation Audio is the most extensive revolution in the audio segment for decades and includes, but is not limited to, profound innovation for enhanced personalization. Research in this field has so far been largely constrained to the technical aspect, while interests and opinions of the consumers have been object of research seldomly. Therefore, the thesis at hand concerns itself with the question of how the perspective audience of Next Generation Audio broadcasts in Germany assesses the extended personalization options, how well these new functionalities may be embraced, and how universally answers to these questions may be formulated for the total population based on demographic traits. It is intended to serve as a source of information for German broadcasters, as to whereon the focus in the technical implementation of Next Generation Audio could be laid on.

For this, a survey among the German population was conducted. In this survey, three of the four discussed and well-established possible applications of Next Generation Audio objects were evaluated distinctively positive. The evaluation of the fourth possible application turned out neutral. Dependencies of these assessments based on demographic traits were observed to be mostly not significant or only significant with small effects. Therefore, on the basis of these results, personalization overall and especially the extended personalization by means of Next Generation Audio seem to be valued highly by the German population. The observed attitudes towards the discussed applications mostly apply to the entire population. Consequently, the positive evaluations are not only relevant for specific groups. Willingness to embrace this next generation of audio material seems to be existent.

Danksagung

An dieser Stelle möchte ich zunächst einigen Personen meinen Dank aussprechen, die auf verschiedene Arten Einfluss auf diese Arbeit genommen haben.

Zuallererst möchte ich Herrn Prof. Curdt sowie Herrn Prof. Dr. Melchior für die Übernahme von Betreuung und Prüfung meiner Bachelorarbeit sowie die Vermittlung vieler interessanter Inhalte während des gesamten Studiums danken. Weiterhin möchte ich im Speziellen noch einmal meine Wertschätzung für die Unterstützung bei der Akquirierung von Probanden für die quantitative Datenerhebung meiner Bachelorarbeit durch Herrn Prof. Curdt und Herrn Matthias „Mätt“ Adler ausdrücken.

Ein besonderer Dank gilt außerdem meinen Eltern und meinem Bruder sowie meinen Freunden, die dauerhaft unterstützend an meiner Seite standen, mir eine überaus wichtige moralische Stütze über die gesamte Bearbeitungszeit hinweg waren und ebenso eine elementare Rolle bei der Akquirierung von Probanden für das Kernstück meiner Forschungsarbeit übernommen haben.

Die Person, die ich mit meinen Launen und Sorgen rund um diese Arbeit wohl am meisten belastet haben dürfte, ist meine Freundin. Ich bin dankbar für den vielen guten Zuspruch, dein Durchhaltevermögen und die oft nötige Ablenkung, durch die ich meinen Kopf ab und an frei bekommen konnte.

Abschließend möchte ich, auch wenn sie bereits alle in den vorangegangenen Absätzen erwähnt wurden, nochmal ein riesiges Dankeschön an die aussprechen, die viel Zeit investiert haben und die es viele Nerven gekostet hat, diese Arbeit Korrektur zu lesen. Die genommene Zeit rechne ich euch immens hoch an.

Inhaltsverzeichnis

Ehrenwörtliche Erklärung.....	I
Kurzfassung.....	II
Abstract	III
Danksagung.....	IV
Abbildungsverzeichnis.....	V
Tabellenverzeichnis.....	VI
Abkürzungsverzeichnis.....	VII
1. Einleitung	1
1.1 Problemstellung.....	1
1.2 Zielsetzung	2
1.3 Aufbau	3
2. Geschichte des Fernsehtons in Deutschland	4
2.1 Mono	4
2.2 Stereo & Zweikanal	5
2.3 Surround.....	6
3. Next Generation Audio	9
3.1 Immersives Audio	10
3.2 Audiorepräsentationen.....	11
3.2.1 Kanalbasiertes Audio	12
3.2.2 Szenenbasiertes Audio	13
3.2.3 Objektbasiertes Audio.....	14
3.2.4 Hybride Ansätze.....	16
3.3 Codecs.....	17
3.3.1 AC-4	17
3.3.2 MPEG-H.....	18

3.3.3	DTS-UHD	18
3.4	Kompatibilität	19
4.	Funktionen von NGA-Objekten	21
4.1	Mehrsprachige Übertragung.....	21
4.2	Anpassung von Lautstärkelevels.....	22
4.3	Audiodeskription	22
4.4	Positionierbare Objekte	23
5.	Hypothesen	24
6.	Methodik	25
6.1	Aufbau des Fragebogens	27
6.2	Gütekriterien quantitativer Forschung	29
7.	Ergebnisse	31
7.1	Demografische Konstitution der Stichprobe	31
7.2	Allgemeine Bewertung (H ₁)	33
7.3	Abhängigkeit von Alter und Geschlecht (H ₂ & H ₃).....	35
7.4	Abhängigkeit von sensorischen Beeinträchtigungen (H ₄ & H ₅)	39
8.	Diskussion.....	41
8.1	Zusammenfassung der Ergebnisse.....	41
8.2	Einordnung und Deutung.....	44
8.3	Einschränkungen.....	48
9.	Fazit.....	49
	Nachwort.....	51
	Literaturverzeichnis	52
	Anhang	59

Abbildungsverzeichnis

Abbildung 1: Matrizierung auf Aufnahmeseite	7
Abbildung 2: Decodierung bei Dolby Surround	7
Abbildung 3: Kopfbezogenes Koordinatensystem.....	10
Abbildung 4: Kanalbasierter Workflow	12
Abbildung 5: Sphärische Harmonische nullter bis dritter Ordnung	14
Abbildung 6: Objektbasierter Workflow	15
Abbildung 7: Schematische Darstellung von S-ADM.....	20
Abbildung 8: Geschlechter- und Altersstruktur der Stichprobe	31
Abbildung 9: Geschlechterverteilung in den Altersgruppen.....	32
Abbildung 10: Bewertung nach Altersgruppen	36
Abbildung 11: Bewertung nach Geschlecht	38
Abbildung 12: Abhängigkeiten von sensorischen Beeinträchtigungen.....	40

Tabellenverzeichnis

Tabelle 1: Beispielhafte Datenraten kanalbasiert vs. hybrid	17
Tabelle 2: Cronbachs Alpha der Einstellungsskalen.....	30
Tabelle 3: Häufigkeiten sensorischer Beeinträchtigungen	32
Tabelle 4: t- (links) und Wilcoxon-Test (rechts) für "Mehrsprachige Übertragung"	33
Tabelle 5: t- (links) und Wilcoxon-Test (rechts) für "Anpassung von Lautstärkelevels"	34
Tabelle 6: t- (links) und Wilcoxon-Test (rechts) für "Audiodeskription"	35
Tabelle 7: t- (links) und Wilcoxon-Test (rechts) für "Positionierbare Objekte"	35

Abkürzungsverzeichnis

ADM	Audio Definition Model
AdvSS	Advanced Sound Systems
ANOVA	Varianzanalyse (eng. analysis of variance)
ARD	Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland
BBC	British Broadcasting Corporation
BR	Bayerischer Rundfunk
DTS	Digital Theater Systems
DVB	Digital Video Broadcasting
EBU	European Broadcasting Union
ESC	Eurovision Song Contest
FOA	First-order ambisonics
HOA	Higher-order ambisonics
HRTF	Head-related transfer function
IfD	Institut für Demoskopie Allensbach
IRT	Institut für Rundfunktechnik
ITU	International Telecommunication Union
KI	Künstliche Intelligenz
LFE	Low Frequency Effects
MPEG	Motion Picture Experts Group
NGA	Next Generation Audio
S-ADM	Serial Audio Definition Model
SDI	Serial Digital Interface
ZDF	Zweites Deutsches Fernsehen

1. Einleitung

„What is the most important development in audio technology for decades?“¹ wird in einem Dokument der European Broadcasting Union (EBU) zum Thema Next Generation Audio (NGA) die Frage aufgeworfen. Noch beinahe im selben Satz wird diese als rhetorisch zu bewertende Frage eigenständig beantwortet: „[NGA is] the biggest change in audio technology for 80 years“.² Begriffe wie Ultra High Definition, High Dynamic Range und High Frame Rate – Begrifflichkeiten, die ausschließlich den visuellen Teilbereich audiovisueller Erzeugnisse betreffen – scheinen nahezu allgegenwärtig und in aller Munde zu sein. Dies ist nicht zuletzt auch durch aktuelle Anwendungen im Kino, wie bei „Avatar: The Way of Water“, oder bei Broadcasts, wie bei Übertragungen der Formel 1, bedingt. Entwicklungen im zweiten Teilbereich, dem Audio-Bereich, erfreuen sich hingegen weniger allgemeiner Beachtung durch Konsumenten. Im Kino konnte mit Dolby Atmos mittlerweile ein Audioformat, das NGA zuzuordnen ist, Einzug finden – wenn auch nur bedingt, denn weniger als 60³ von über 1700⁴ Kinostätten in Deutschland haben derzeit mindestens ein Dolby Atmos Kino. In der deutschen Broadcastlandschaft hingegen ist NGA bisher noch nicht angekommen und dürfte den meisten Menschen noch kein Begriff sein. Grund hierfür ist, dass auch dieses einzig geläufige Format Dolby Atmos nur selten unter der Bezeichnung NGA genannt wird, sondern fast ausschließlich unter der eigenen Markenbezeichnung bekannt ist.

1.1 Problemstellung

Aus rein technischer Sicht wurde im Themenkomplex von NGA bereits vieles intensiv untersucht. Mittlerweile ist bekannt und in zahlreichen Testübertragungen auch bereits erprobt worden⁵, welche Personalisierungsmöglichkeiten den Konsumenten von Broadcasts mittels objektbasierter bzw. hybrider Audiorepräsentation bei zukünftigen NGA-Broadcasts zur Verfügung gestellt werden können. Die Endnutzer selbst sowie deren Wahrnehmung der einerseits bereits in klassisch kanalbasierter Form praktizierten und andererseits gänzlich

¹ EBU 2021, S. 6.

² Ebd., S. 7.

³ Vgl. Teufel 2023.

⁴ Vgl. Filmförderungsanstalt (FFA) 2023.

⁵ Vgl. Dolby 2018; Simon et al. 2018; Turnwald et al. 2018; de Jong et al. 2019; Mäsel et al. 2019; Bolt et al. 2021.

neuen Personalisierungsmöglichkeiten wurden demgegenüber bisher nur selten Untersuchungsgegenstand. Forscher des Research & Development Departments der British Broadcasting Corporation (BBC) trugen 2020 eine Studie zur Erforschung des von Endnutzern gewünschten Maßes an Personalisierung ihres Audioerlebnisses sowie der gewünschten optischen Repräsentation der bedienbaren Interfaces dieser Personalisierung aus.⁶ Eine Untersuchung dessen, welche der Funktionen den Zuschauern aus deren eigener Sicht einen realen Mehrwert in ihrer Wahrnehmung des Fernsehprogramms bieten könnte, fand bisher allerdings noch nicht statt.

Diese Bachelorarbeit forscht daher direkt am Endnutzer und widmet sich der Frage, wie die letztendlichen Konsumenten zukünftiger NGA-Broadcasts deutscher Fernsehsender die gängigsten Personalisierungsmöglichkeiten, die durch objektbasierte Audiorepräsentation bei NGA-Broadcasts zur Verfügung gestellt werden können, bewerten und in Zukunft annehmen könnten. Weiterhin soll evaluiert werden, ob über die beobachteten Einstellungen diesen Funktionen gegenüber eine allgemeingültige Aussage unabhängig von demografischen Merkmalen für alle Zuschauer getroffen werden kann bzw. welchen Einfluss die demografischen Merkmale Alter, Geschlecht und das Vorhandensein von sensorischen Beeinträchtigungen auf diese Bewertung haben. Hierzu wurde eine quantitative Umfrage in Deutschland durchgeführt, die Personalisierungsmöglichkeiten durch NGA-Objekte präsentierte und anschließend zu den einzelnen Funktionen befragte.

1.2 Zielsetzung

NGA ist bisher in der deutschen Fernsehlandschaft noch nicht angekommen. Anders als in Südkorea, wo NGA bereits seit 2017 ein fester Bestandteil im Broadcast ist und dauerhafte Anwendung im regelmäßigen Sendungsablauf findet⁷, sowie einigen weiteren Ländern wie beispielsweise Brasilien, wo mit MPEG-H ein Übertragungscodec gewählt und konkrete Pläne zur Einführung von NGA im kommenden Jahr 2024 bestehen⁸, wurde für zukünftige deutsche Broadcasts bisher weder ein Codec gewählt⁹, noch existieren konkrete, öffentlich zugängliche Zeitpläne, wann NGA bei deutschen Fernsehsendern Einzug finden soll. Ziel dieser Arbeit ist

⁶ Vgl. Pardoe et al. 2020.

⁷ Vgl. Fraunhofer Institute for Integrated Circuits (IIS) 2020.

⁸ Vgl. Fraunhofer Institute for Integrated Circuits (IIS) 2022.

⁹ Vgl. VRTONUNG 2022.

daher, einen Anhaltspunkt für deutsche Fernsehanstalten zu bieten, worauf der Fokus in der Bereitstellung von Personalisierungsmöglichkeiten bei NGA-Broadcasts gelegt werden könnte, noch bevor NGA Anwendung bei deutschen Broadcastern findet. Auf Grundlage der erhobenen empirischen Daten aus der Umfrage soll demnach aufgezeigt werden, welche Einsatzmöglichkeiten von NGA-Objekten besondere Relevanz für das Zielpublikum im bundesdeutschen Gebiet haben – direkt aus der Sicht ebendieses Publikums.

1.3 Aufbau

Diese Bachelorarbeit ist nachfolgend unterteilt in drei Abschnitte: den theoretischen Teil, den empirischen Teil sowie das abschließende Fazit. Der theoretische Teil befasst sich dabei zunächst in Kapitel 2 mit der Historie technischer Entwicklungen im Bereich des Fernsehens in Deutschland. Es werden drei zeitliche Abschnitte behandelt: Mono-, Stereo- und Zweikanal- sowie Surround-Übertragung. Dabei wird explizit auf die in diesen Abschnitten jeweils verfügbaren Personalisierungsmöglichkeiten eingegangen. Anschließend behandelt und erläutert Kapitel 3 den Begriff Next Generation Audio sowie die zugrundeliegenden Herangehensweisen bezüglich Audiorepräsentation für immersives Audiomaterial: kanal-, szenen- und objektbasiertes Audio. Dabei wird herausgearbeitet, wie und wodurch NGA im Vergleich zu bisherigen Broadcasts mehr Potenzial für erhöhte Personalisierung bietet. Zuletzt werden die verfügbaren NGA-Codierungen AC-4, MPEG-H und DTS-UHD, durch welche NGA-Inhalte für Broadcasts transportiert werden können, präsentiert sowie das einheitliche Metadatenmodell Audio Definition Model (ADM) für Interoperabilität zwischen diesen Codierungen vorgestellt. Als Bindeglied zwischen theoretischem Rahmen und empirischem Teil der Arbeit werden in Kapitel 4 die im empirischen Teil untersuchten Funktionen von NGA-Objekten anhand bereits erfolgter Testübertragungen definiert. Mit Kapitel 5 beginnt der empirische Teil der Arbeit. Hier werden zunächst die Forschungshypothesen formuliert, bevor in Kapitel 6 die zugrundeliegende Methodik der Forschung erläutert wird. Hierbei wird auf den Aufbau des erstellten Fragebogens sowie die Gütekriterien der Forschung eingegangen. Kapitel 7 beinhaltet anschließend die Ergebnisse der Datenerhebung unterteilt nach den formulierten Hypothesen aus Kapitel 5. Der empirische Teil dieser Arbeit endet mit Kapitel 8, wobei die Ergebnisse aus Kapitel 7 diskutiert und eingeordnet werden. Das abschließende Fazit fasst die wichtigsten Erkenntnisse zusammen und formuliert Antworten auf die eingangs in Kapitel 1.1 formulierten Forschungsfragen.

2. Geschichte des Fernsehens in Deutschland

Dieses Kapitel skizziert nachfolgend die technische Entwicklung des Fernsehens seit Beginn der Fernsehübertragung in Deutschland. Dabei wird in die drei Zeitabschnitte Mono, Stereo und Zweikanal sowie Surround unterteilt und auf die Personalisierungsmöglichkeiten, die den Zuschauern aufgrund der technischen Gegebenheiten innerhalb dieser Abschnitte zur Verfügung gestellt werden konnten bzw. können, eingegangen.

2.1 Mono

Als Startpunkt fernsehtechnischer Übertragung in Deutschland kann ein Bildübertragungsversuch, der mithilfe der von Paul Nipkow erfundenen mechanischen Lochscheibe am 11. Februar 1928 durchgeführt werden konnte, gesehen werden. Hierbei gelangen erstmals Bildübertragungen aus dem Labor der Telehor AG zum Telegrafentechnischen Reichsamt, wobei lediglich die Übertragung von Bildern erprobt, aber noch kein Ton übertragen wurde.¹⁰ Erst zwei Jahre später, im Jahr 1930, wurde schließlich auch ein erster Ton-Sender im Ultrakurzwellenbereich getestet.¹¹ Auf diesen ersten Test mit Ton folgten weitere Testläufe, bis der Fernsehsender „Paul Nipkow“ als erster in Deutschland am 22. März 1935 mit regelmäßigem Programm im Versuchsbetrieb startete, bevor am 15. Januar 1936 schließlich die offizielle und dauerhafte Inbetriebnahme als erster deutscher Fernsehsender folgte.¹²

Dieser Fernsehsender Paul Nipkow und nach dem Zweiten Weltkrieg Sender wie die Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland (ARD) und das Zweite Deutsche Fernsehen (ZDF) begleiteten Großereignisse wie die Olympischen Sommerspiele 1936, die erste Tagesschau-Übertragung 1952, die Krönung der britischen Monarchin Queen Elizabeth II. 1953, die erste live ausgestrahlte Fußball-Weltmeisterschaft 1954, den ersten Eurovision Song Contest (ESC) 1956, die Einführung des Farbfernsehens 1967 und die Mondlandung 1969 neben vieler weiterer bedeutender Ereignisse bis in die frühen 1980er Jahre ausschließlich monofonem Ton.¹³ „Monofon“ oder auch nur

¹⁰ Vgl. Bundesarchiv 2023.

¹¹ Vgl. Riegler 2019.

¹² Vgl. ebd.

¹³ Vgl. Slavik 2013, S. 973.

„mono“ bedeutet hierbei, dass exakt ein Kanal übertragen und wiedergegeben wird. Auf Produktions- bzw. Senderseite kann hierbei eine Abmischung aus vielen Quellen (Mikrofonen, Tonbändern etc.) stattfinden, die verschiedenen Quellen werden dabei allerdings zu nur einem monofonen Einkanalsignal zusammengemischt, das zu den Konsumenten übertragen wird. Auf deren Wiedergabeseite kann ebenso nicht nur über einen, sondern mehrere Lautsprecher wiedergegeben werden. Dabei erhält allerdings jeder Lautsprecher dasselbe Signal.¹⁴

In der Zeit der einkanaligen Übertragung, in der lediglich eine monofone Abmischung von den Fernsehsendern übertragen wurde, gab es daher keine Personalisierungsmöglichkeiten des Tons für die Zuschauer der übertragenen Programme, die über die Anpassung der Wiedergabelautstärke hinausging.

2.2 Stereo & Zweikanal

Während monofone Übertragung im deutschen Fernsehen bis in die 1980er-Jahre alternativlos war, wurde stereofone Übertragung für den Hörfunk bereits auf der Funkausstellung 1963 vorgestellt¹⁵ und fand in den folgenden Jahren zunehmende Anwendung. Ein früherer Ansatz, wie zweikanaliger, stereofoner Ton bei Fernsehübertragungen ermöglicht werden kann, war daher die synchrone Übertragung mit einem Radiosender. Beispielsweise übertrug das ZDF im April 1972 eine Tschaikowsky-Oper, für die der Bayerische Rundfunk (BR) synchron eine stereofone Fassung über den Hörfunk ergänzte.¹⁶ Fernsehzuschauer, die ein stereofähiges Radiogerät besaßen, erhielten damit die Option, den monofonen Fernsehton abzuschalten und an dessen Stelle den synchronen, stereofonen Radioton zu hören.

Um Stereoton auch direkt in den Bereich der Fernsehübertragungen zu bringen, begann das Institut für Rundfunktechnik (IRT) bereits ab den 1960er Jahren mit Forschungen im Bereich des Zweikanaltons.¹⁷ Diesen nutzte das ZDF erstmalig 1981, wodurch erstmals neben Mono- auch Stereo- und Zweikanalsendungen ermöglicht wurden.¹⁸ Bei der Stereowiedergabe werden, anders als bei monofoner Wiedergabe, zwei Kanäle genutzt und somit zwei

¹⁴ Vgl. Schätzlein 2005, S. 187.

¹⁵ Vgl. WDR 1963.

¹⁶ Vgl. Schätzlein 2005, S. 190.

¹⁷ Vgl. ebd., S. 187.

¹⁸ Vgl. Slavik 2013, S. 973.

verschiedene Signale ausgegeben – eines durch den linken und eines durch den rechten Lautsprecher. Hierbei können monaurale Quellen durch Panning in unterschiedlicher Intensität bzw. Lautstärke (Intensitätspanning) und/oder zueinander verzögert (Laufzeitpanning) auf beide Lautsprecher gegeben werden.¹⁹ Durch die resultierenden Intensitäts- und Laufzeitunterschiede zwischen den Lautsprechern zur Abhörposition und den beiden Ohren des Zuhörers ergeben sich interaurale Pegel- und Laufzeitdifferenzen, welche eine Lokalisierung der reproduzierten Schallereignisse in Horizontalebene ermöglichen.²⁰ Schallereignisse können hierbei an Positionen entlang der gesamten Basisbreite zwischen den beiden Lautsprechern in Form sogenannter Phantomschallquellen wahrgenommen werden. Phantomschallquellen sind dabei definiert als fiktive Schallquellen an Orten, an denen keine realen Schallwandler positioniert sind und die durch Überlagerung der Schallfelder zweier benachbarter Lautsprecher entstehen.²¹ Geschieht das beschriebene Panning dynamisch, können optische Bewegungsabläufe auf dem Bildschirm zusätzlich auch akustisch dargestellt werden. Ergebnisse dieser Eigenschaften von Stereo sind im Vergleich zu Mono „mehr Durchsichtigkeit, Räumlichkeit und Klangfülle“²².

Die beiden Sendekanäle des Zweikanaltons konnten neben der Anwendung für die stereofone Übertragung auch als zwei voneinander unabhängige Monokanäle behandelt werden, die auf Wiedergabeseite individuell ausgewählt werden konnten. Hierdurch war es Fernsehsendern beispielsweise möglich, einen Film in deutscher Synchronisation auf Kanal 1 und in englischer Originalsprache auf Kanal 2 simultan zu übertragen. Alternativ dazu wurde die Bereitstellung einer Audiodeskription – eine akustische Beschreibung dessen, was auf dem Bildschirm passiert – durch die Verwendung der beiden Audiokanäle als separate Monokanäle technisch ebenso erstmalig möglich.²³ Somit ging mit der Einführung des Zweikanaltons auch erstmalig die Möglichkeit der Personalisierung durch den Zuschauer einher.

2.3 Surround

Durch die üblich gewordene zweikanalige Übertragung des Fernsehtons war ebenso die Grundlage für den nächsten Entwicklungsschritt gegeben: Surround-Sound. Hierbei wurde

¹⁹ Vgl. Friesecke 2014, S. 693–694.

²⁰ Vgl. ebd., S. 134–135.

²¹ Vgl. Theile et al. 2013, S. 221–222.

²² Ebd., S. 218.

²³ Vgl. Schätzlein 2005, S. 188.

das Wiedergabesystem mit den üblichen Stereolautsprechern links und rechts vor dem Zuhörer um einen Centerlautsprecher mittig vor sowie weitere, sogenannte Surroundlautsprecher um den Zuhörer herum erweitert, wodurch die Erzeugung und Lokalisation von Phantomschallquellen nicht mehr nur vor dem Zuhörer, sondern ebenso um den Zuhörer herum ermöglicht werden konnte. Ab 1993 wurden somit erste Sendungen im deutschen Fernsehen im analogen Dolby Surround Format übertragen.²⁴ Dies gelang durch Matrizierung der vier Kanäle left (L), center (C), right (R) und surround (S) auf Aufnahme-seite zu den zu übertragenden Ausgangssignalen left-total (L_t) und right-total (R_t), wie sie in Abbildung 1 dargestellt ist.

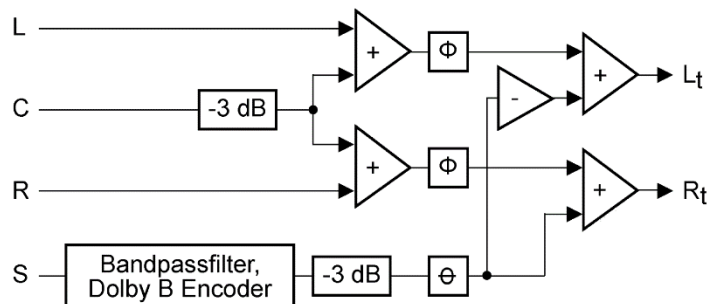


Abbildung 1: Matrizierung auf Aufnahme-seite²⁵

Die Nutzsignale auf der Wiedergabeseite entstehen bei Dolby Surround mittels Decodierung von L_t und R_t durch einen passiven Decoder, dargestellt in Abbildung 2. L_t und R_t werden wie übertragen und empfangen auf den linken bzw. rechten Lautsprecher gegeben, während das Signal für den Centerkanal durch gleichphasige und das des Surroundkanals durch gegenphasige Summierung von L_t und R_t gewonnen wird.²⁶

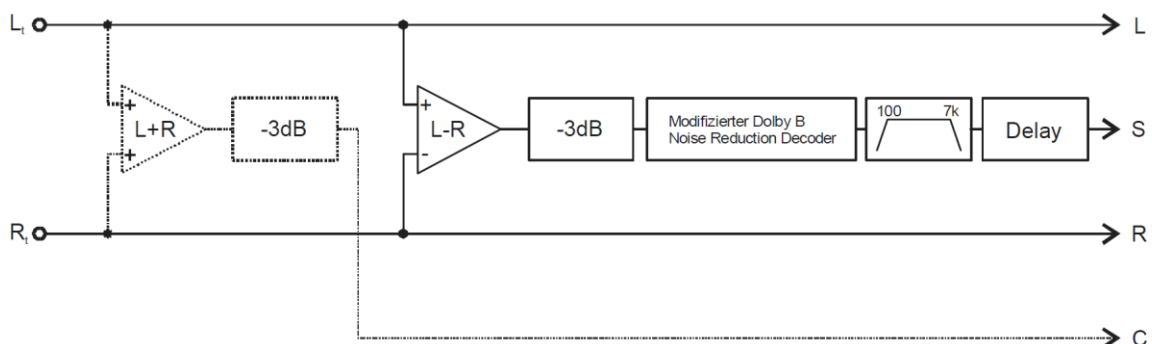


Abbildung 2: Decodierung bei Dolby Surround²⁷

²⁴ Vgl. Schätzlein 2005, S. 193.

²⁵ Slavik 2013, S. 975.

²⁶ Vgl. ebd., S. 974.

²⁷ Friesecke 2014, S. 793.

Über die Jahre fanden Weiterentwicklungen von Dolby Surround wie Dolby Pro Logic, Dolby Pro Logic II und weitere Iterationen Anwendung in der analogen Übertragung deutscher Broadcasts. Für deren komplexere, aktive Decodierung sei auf Frieseckes Audio-Enzyklopädie verwiesen.²⁸ Bei diesen analogen Surroundformaten war dementsprechend erneut keine Personalisierung auf der Wiedergabeseite durch den Endnutzer möglich, da die beiden vorhandenen Übertragungskanäle für die beiden Signale L_t und R_t genutzt wurden. Somit entfielen die Personalisierungsmöglichkeiten, die die Zweikanalübertragung ermöglicht hatte.

Im Jahr 1999 war ProSieben schließlich der erste deutsche Fernsehsender, der in einem digitalen Surroundformat (Dolby Digital) übertrug.²⁹ Bei digitalen Formaten wie Dolby Digital liegt in der Regel pro Lautsprecherkanal ein eigenes Signal vor, es erfolgt keine Signalgewinnung durch Matrizierung weniger Signale mehr.

Der für Deutschland relevante Digital Video Broadcasting (DVB)-Standard enthält heutzutage zahlreiche Audiocodierungsverfahren. Da DVB in allen Übertragungsformen (Satellit, Kabel, Terrestrik und IP) digital arbeitet, spielen analoge Übertragungsverfahren wie Dolby Surround und dessen Weiterentwicklungen hierbei keine Rolle mehr. Schließlich wurde analoges Fernsehen Mitte 2019 auch final abgestellt.³⁰ Bei diesen digitalen Übertragungsformaten kann Personalisierung ermöglicht werden, indem mehrere gesamte Mischungen zur Verfügung gestellt werden, die am Endgerät ausgewählt werden können. Die meisten Workflows basieren allerdings weiterhin auf dem Serial Digital Interface (SDI), wodurch bereits drei 5.1-Surroundmischungen, die jeweils sechs Audiokanäle besetzen, die 16 verfügbaren Audiokanäle bei SDI³¹ übersteigen. Dieser Ansatz wird beispielsweise für die bei den öffentlich-rechtlichen Sendern von ARD und ZDF seit Juni 2022 verfügbare, alternative Audiospur „Klare Sprache“ verfolgt. In Echtzeit wird basierend auf künstlicher Intelligenz (KI) aus dem Originalton die Sprache von etwaigen Hintergrundgeräuschen extrahiert und in ihrer Lautstärke angehoben, während die Lautstärke der Hintergrundgeräusche dazu analog leicht abgesenkt wird. Hierdurch soll eine bessere Sprachverständlichkeit, vor allem auch für sensorisch beeinträchtigte Menschen erzielt werden.³²

²⁸ Vgl. Friesecke 2014, S. 791–796.

²⁹ Vgl. Schätzlein 2005, S. 193.

³⁰ Vgl. Verbraucherzentrale 2022.

³¹ Vgl. Bolt et al. 2021, S. 6.

³² Vgl. ZDF 2022; ARD 2023b.

3. Next Generation Audio

Im aktuellen DVB-Standard enthalten sind unter anderem auch die Codecs AC-4, MPEG-H und DTS-UHD.³³ Mithilfe dieser sogenannten NGA-Codecs sollen in Zukunft die aktuellen Sendetonstandards um neue Funktionalität bereichert werden. Dieses Kapitel beschäftigt sich intensiv mit dieser nächsten Generation des (Fernseh-)Tons und arbeitet dabei heraus, wie NGA technisch ein höheres Maß an Personalisierung für Endnutzer ermöglicht.

In der Empfehlung BS.2051-3 des Untersektors Funkkommunikation der International Telecommunications Union (ITU) wird für NGA, auch Advanced Sound Systems (AdvSS) genannt, folgende Definition gegeben:

An advanced sound system is a system with a reproduction configuration beyond those specified in Recommendation ITU-R BS.775 or a system with any reproduction configuration that can support channel-based, object-based or scene-based input signals or their combination with metadata. An advanced sound system uses audio data in combination with an appropriate set of metadata to specify a sound scene to be delivered/broadcasted.³⁴

Die angesprochene ITU-R BS.775 umfasst hierbei Lautsprecherkonfigurationen mit optionalem Low Frequency Effects (LFE) Kanal von Mono (1/0) bis Dreikanal-Stereo mit zwei Surroundlautsprechern (3/2)³⁵, geläufiger bekannt als 5.0 bzw. 5.1 Surround. NGA bzw. AdvSS beinhaltet somit alle weitergehenden Konfigurationen über Surround mit fünf Lautsprechern hinaus. Weiterhin umfasst es Systeme unabhängig ihres Wiedergabesystems, die kanalbasiertes, objektbasiertes oder szenenbasiertes Audio sowie Kombinationen dieser Ansätze unterstützen.

Die EBU ergänzt:

NGA delivers audio experiences that are more accessible, personalized, and interactive – independent of how they are consumed, be it on headphones, soundbars or multi-speaker setups. NGA frees producers from the need to create multiple audio mixes for different reproduction systems by allowing them to deliver a single, multi-purpose audio master instead.³⁶

Weitere Aspekte von NGA sind demnach mehr Barrierefreiheit, Personalisierung und Interaktivität im Vergleich zu bisherigen Ansätzen. Unabhängig vom jeweiligen Wiedergabesystem soll mit nur einem erstell- und übertragbaren Audio-Master mit allen

³³ Vgl. ETSI 2022, S. 151.

³⁴ ITU-R 2022a, S. 1.

³⁵ Vgl. ITU-R 2022c, S. 5.

³⁶ EBU 2019, S. 1.

Wiedergabesystemen von Kopfhörern bis zu vielkanaligen Lautsprechersystemen umgegangen werden können.

Neben den von ITU und EBU bereits genannten Aspekten ergänzt die BBC mit dem Stichwort „immersive audio“ einen weiteren zentralen Faktor von NGA gegenüber herkömmlichen Audioerlebnissen³⁷, womit sich das nächste Unterkapitel befasst.

3.1 Immersives Audio

Die bisher in Kapitel 2 erläuterten Wiedergabeformate (Mono, Stereo, Surround) sind allesamt auf zweidimensionale Wiedergabe in horizontaler Ebene beschränkt. Wie im vorangegangenen Überblick beschrieben, ist immersives Audio, auch 3D-Audio genannt, ein zentraler Aspekt und eine bedeutende Veränderung von NGA gegenüber diesen traditionellen Wiedergabeformaten im Broadcast. Die übliche horizontale Wiedergabeebene wird bei diesen immersiven Wiedergabeformen um die Höhendimension erweitert³⁸, wodurch räumliches Hören in allen drei Raumebenen, wie in Abbildung 3 dargestellt, ermöglicht wird. Ziel hiervon ist ein vollumfänglich einhüllendes Audioerlebnis mit Schallereignissen aus allen Richtungen um den Zuhörer herum.

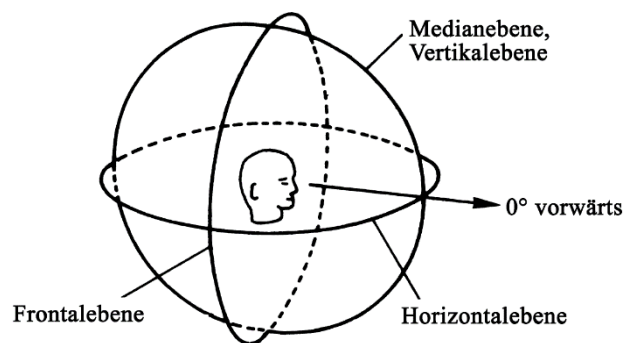


Abbildung 3: Kopfbezogenes Koordinatensystem³⁹

Entscheidend bei der Wahrnehmung von immersivem Audio ist zusätzlich zur Lokalisation in Horizontalebene die Lokalisation von Schallereignissen in der Medianebene. Hierbei geht es maßgeblich darum, die Höhe von Schallereignissen korrekt wahrzunehmen sowie korrekte Lokalisation vor oder hinter dem Zuhörer zu erreichen.⁴⁰ Anders als in der Horizontalebene, gibt es in der Medianebene höchstens minimale interaurale Laufzeit- und Pegeldifferenzen.

³⁷ Vgl. BBC 2020.

³⁸ Vgl. Bleidt et al. 2017, S. 203; Bolt et al. 2021, S. 3.

³⁹ Dickreiter und Goeres-Petri 2013, S. 128.

⁴⁰ Vgl. Friesecke 2014, S. 135.

Für die Wahrnehmung in Medianebene spielen daher vor allem richtungsbestimmende Frequenzbänder sowie die komplexe Filterwirkung von Kopf, Ohren und Rumpf, die durch kopfbezogene Außenohrübertragungsfunktionen (head-related transfer function, HRTF) beschrieben wird, eine wichtige Rolle.⁴¹

Dreidimensionale, immersive Wiedergabe, wie sie bei NGA zum Einsatz kommt, kann beispielsweise durch die Erweiterung von Surroundsystemen wie 5.1 oder 7.1 um sogenannte Höhen- bzw. Überkopflautsprecher stattfinden, um Lautsprecherkonfigurationen wie zum Beispiel 5.1.2, 5.1.4 oder 7.1.4 zu erhalten.⁴² Weiterhin ist 3D-Audio durch binaurale Kopfhörerwiedergabe und 3D-fähige Soundbars realisierbar. Bei den auf Lautsprechern basierenden, immersiven Wiedergabeverfahren geschieht die Lokalisation von Schallereignissen durch die Überlagerung der Schallfelder dreier benachbarter Lautsprecher, sogenannter Lautsprechertripel, wodurch per Summenlokalisation Phantomschallquellen entstehen.⁴³ Beim Sonderfall der 3D-Soundbars, die aus nur einer Schallquelle meist direkt unter dem Fernseher sowie häufig einem separaten Subwoofer für den LFE-Kanal bestehen, wird beim Zuhörer durch gezielte Reflexionen an Wand- und Deckenflächen oder softwarebasierte Höhenvirtualisation, bei der Filterungen auf Basis der richtungsbestimmenden Frequenzen vorgenommen werden, der Eindruck von weiteren, im Raum verteilten Schallquellen erzeugt.⁴⁴ Für binaurale (beidohrige) Wiedergabe werden die beiden Ohrsignale durch Binauralsynthese mittels mathematischer Faltung eines nachhallfreien Quellsignals mit kopfbezogenen Raumimpulsantworten erzeugt. Die hierbei entstehenden Ohrsignale enthalten alle für die dreidimensionale Lokalisation erforderlichen Informationen, also die Berücksichtigung richtungsbestimmender Frequenzbänder sowie die durch HRTFs beschriebene Filterwirkung, und müssen anschließend nur noch per Kopfhörer direkt in die Gehörgänge gegeben werden.⁴⁵

3.2 Audiorepräsentationen

Im vorangegangenen Zitat aus der ITU-R Empfehlung wurden drei Audiorepräsentationen, die bei NGA verwendet werden, genannt: Kanal-, szenen- und objektbasiertes Audio. Mittels

⁴¹ Vgl. Weinzierl 2008, S. 586; Friesecke 2014, S. 136.

⁴² Vgl. Bleidt et al. 2017, S. 203.

⁴³ Vgl. Slavik und Weinzierl 2008, S. 657.

⁴⁴ Vgl. Dolby 2023.

⁴⁵ Vgl. Slavik und Weinzierl 2008, S. 671.

dieser verschiedenen Ansätze kann zum Beispiel auch immersives Audiomaterial produziert und wiedergegeben werden. Die folgenden Unterkapitel behandeln diese drei verschiedenen Audiorepräsentationen und stellen sie kurz vor.

3.2.1 Kanalbasiertes Audio

Kanalbasierte Audiorepräsentation ist die traditionelle und vertrauteste Herangehensweise bei Audioproduktion und -distribution. Bei rein kanalbasiertem Workflow, wie in Abbildung 4 dargestellt, wird bereits auf Produktionsseite die spätere Wiedergabekonfiguration vordefiniert. In Abbildung 4 ist dies Stereo. Auf Wiedergabeseite werden die einzelnen Kanalsignale dann nur noch auf die entsprechenden Lautsprecher im definierten Wiedergabesystem geroutet.⁴⁶

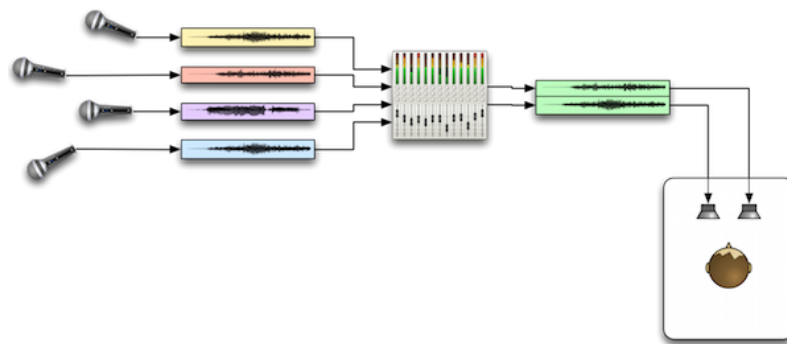


Abbildung 4: Kanalbasierter Workflow⁴⁷

Fertige Mischungen, die auf eine bestimmte Lautsprecherkonfiguration ausgelegt sind, können lediglich durch Upmixing (z.B. Upmix einer 2.0-Stereo-Mischung auf ein 5.1-Surroundsystem) oder Downmixing (z.B. Downmix einer 5.1-Surround-Mischung auf ein 2.0-Stereosystem) auf andere Wiedergabeformate angepasst werden. Empfehlungen, wie solche Up- und Downmixverfahren für alle Konfigurationen zwischen Mono und 5.0/5.1 ablaufen sollen, sind in ITU-R BS.775 zu finden.⁴⁸

Kanalbasierte Audiorepräsentation allein erfüllt dementsprechend nicht die in Kapitel 3 genannten Anforderungen an NGA. Die Audiomischung ist nicht vom Endnutzer anpassbar, sondern ein festes, in sich geschlossenes Konstrukt. Demnach ergibt sich keine zusätzliche Barrierefreiheit, keine Interaktivität mit dem vorliegenden Audiomaterial und keine weitergehende Personalisierung außer durch die Übertragung mehrerer kanalbasierter

⁴⁶ Vgl. ITU-R 2022a, S. 2–3.

⁴⁷ Centre for Vision, Speech and Signal Processing (CVSSP) 2018.

⁴⁸ Vgl. ITU-R 2022c, S. 8–9.

Gesamtmischungen, wie in Kapitel 2.3 beschrieben. Außerdem kann beim kanalbasierten Ansatz keine flexible Anpassung auf Wiedergabeseite stattfinden, sollten Lautsprecher nicht entsprechend den Vorgaben für bestimmte Konfigurationen aufgestellt sein (z.B. kein optimales 60°-Stereodreieck).

3.2.2 Szenenbasiertes Audio

Die szenenbasierte Audiorepräsentation löst sich von der starren Bindung von einzelnen Kanälen an diskrete Lautsprecher in einem Wiedergabesystem. Sie basiert auf Ambisonics – einem Format, das anstelle von diskreten Lautsprechersignalen in den Kanälen eine lautsprecherunabhängige Repräsentation eines Schallfeldes enthält.⁴⁹ Ambisonics in erster Ordnung geht in seinen Anfängen zurück auf Michael A. Gerzon und seinen im Journal of the Audio Engineering Society erschienenen Artikel aus dem Jahr 1973.⁵⁰

Die Aufnahme dreidimensionaler First-Order-Ambisonics (FOA) kann durch eine Weiterentwicklung des koinzidenten MS-Aufnahmeverfahrens bewerkstelligt werden. Hierbei wird die Aufnahmekonfiguration aus einem omnidirektionalen Mikrofon (Mitte, W) und einem seitlich (links/rechts) orientierten Achtermikrofon (Seite, X) um zwei weitere Mikrofone mit der Richtcharakteristik einer Acht erweitert, die auf vorne/hinten (Y) respektive oben/unten (Z) gerichtet sind. Alternativ kann ebenso eine tetraedrische Anordnung aus vier Nierenmikrofonen (A-Format) so encodiert werden, dass W, X, Y und Z (B-Format) gewonnen werden.⁵¹ FOA ist in der räumlichen Auflösung sowie in den Ausmaßen des optimalen Abhörbereichs, auch sweet spot genannt, eingeschränkt, da die Signale stark korrelieren. Um erhöhte Dekorrelation zwischen benachbarten Wiedergabesignalen zu erreichen und somit eine genauere Synthese des Schallfelds mit schärferer Richtungsabbildung sowie einen größeren Abhörbereich zu schaffen, kann FOA durch weitere Komponenten erweitert werden, um Higher-Order Ambisonics (HOA) zu erhalten.⁵² NGA nutzt diese Ambisonics-Ausprägungen höherer Ordnung. Abbildung 5 zeigt die Richtcharakteristiken aller Komponenten nullter bis dritter Ordnung, sogenannte sphärische Harmonische. Jede Zeile in der Abbildung entspricht hierbei einer Ambisonics-Ordnung, wobei alle Komponenten der niedrigeren Ordnungen auch in den höheren Ordnungen enthalten sind. Daraus ergeben sich

⁴⁹ Vgl. Olivieri et al., S. 5.

⁵⁰ Vgl. Gerzon 1973.

⁵¹ Vgl. Zotter und Frank 2019, S. 9–11.

⁵² Vgl. Slavik und Weinzierl 2008, S. 661–662; Zotter und Frank 2019, S. 19.

für Ambisonics dritter Ordnung beispielsweise 16 Komponenten, also 16 Signalkanäle. Allgemein gefasst spricht man von $(N+1)^2$ Komponenten, wobei N der Ordnungszahl entspricht.⁵³

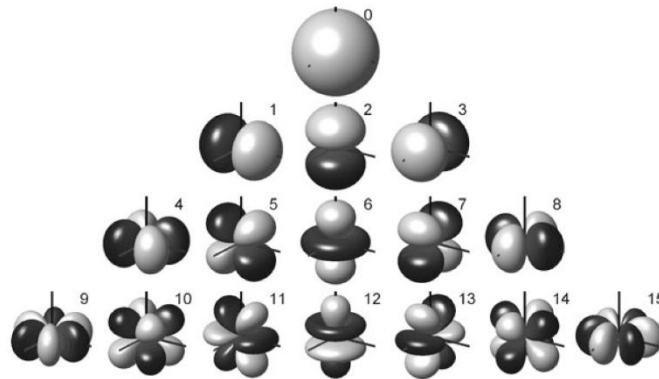


Abbildung 5: Sphärische Harmonische nullter bis dritter Ordnung⁵⁴

Anders als bei kanalbasierter Audiorepräsentation können die übertragenen Kanäle nicht schlicht auf Lautsprecher im Wiedergabesystem gegeben werden. Die auf Aufnahmeseite encodierten Signale werden bei szenenbasierter Audiorepräsentation hingegen zunächst auf der Wiedergabeseite decodiert und anschließend frei auf das vorhandene Wiedergabesystem gerendert. Erst im Renderer entstehen die schlussendlichen Lautsprechersignale basierend auf Anzahl und Positionen der Lautsprecher im Raum.⁵⁵

Szenenbasierte Audiorepräsentation erfüllt die Anforderung an den einen vom Wiedergabesystem unabhängigen Audio-Master durch das freie Rendering auf das jeweils vorliegende Wiedergabesystem besser als die kanalbasierte Audiorepräsentation. Auch Interaktion mit dem Audiomaterial ist möglich, da zum Beispiel eine Rotation des in den Ambisonics-Signalen enthaltenen Schallfeldes bei Wiedergabe über Kopfhörer per Headtracking möglich ist.⁵⁶ Durch ausschließlich szenenbasierte Audiorepräsentation ist es allerdings ebenso wie bei kanalbasierter Repräsentation nicht möglich, erhöhte Personalisierung und erhöhte Barrierefreiheit zur Verfügung zu stellen.

3.2.3 Objektbasiertes Audio

Objektbasierte Audiorepräsentation ist charakterisiert durch sogenannte Audioobjekte anstelle von einzelnen Audiokanälen. Audioobjekte bestehen dabei aus einem Audiosignal

⁵³ Vgl. Bleidt et al. 2017, S. 204.

⁵⁴ Zotter und Frank 2019, S. 68.

⁵⁵ Vgl. Bleidt et al. 2017, S. 204.

⁵⁶ Vgl. Zotter und Frank 2019, S. 8–9.

sowie zugehörigen Informationen über beispielsweise Position und Gain des Signals, die in Form von Metadaten vorliegen. Diese Metadaten werden durch einen Renderer im Endgerät interpretiert, der Audiosignale für die jeweiligen Schallwandler der vorhandenen Wiedergabekonfiguration generiert. Anders als bei der kanalbasierten Audiorepräsentation ist dabei nicht festgelegt, in welcher Konfiguration wiedergegeben werden muss bzw. soll, der Renderer kann auf Grundlage des vorliegenden Wiedergabesystems passende Audiosignale frei generieren.⁵⁷ Ein solcher objektbasierter Workflow ist in Abbildung 6 dargestellt.

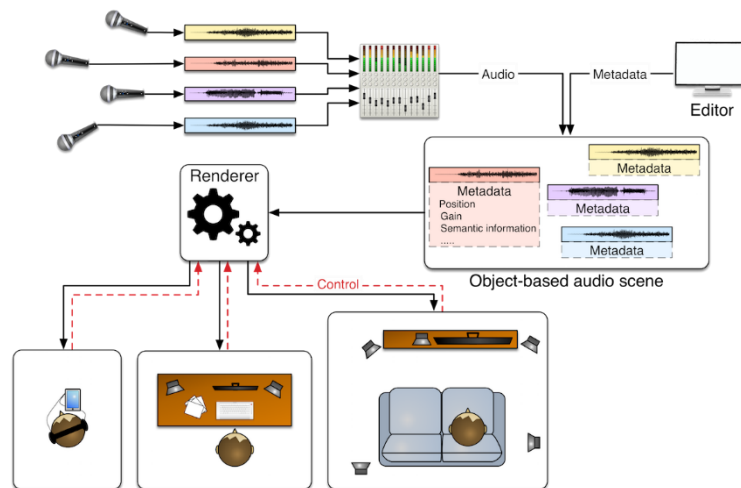


Abbildung 6: Objektbasierter Workflow⁵⁸

Objektbasierte Audiorepräsentation kann erhöhte Personalisierung für den Endnutzer ermöglichen (in Abbildung 6 mit „Control“ angedeutet). Broadcaster können durch den Prozess des Authorings der Metadaten festlegen, welche Art(en) der Personalisierung sie ermöglichen wollen und wie umfangreich der Einfluss des Endnutzers sein kann.⁵⁹ Personalisierungsmöglichkeiten, die bisher im deutschen Fernsehen durch die Wahl aus verschiedenen kanalbasierten Gesamtmischungen angeboten wurden und werden, können mittels objektbasierter Repräsentation alle gleichzeitig mit nur einem Audio-Master realisiert werden. Es kann beispielsweise die Wahl aus verschiedenen Wiedergabesprachen sowie eine zuschaltbare Audiodeskription (vgl. Kapitel 2.2, Stichwort Zweikanalton) und auch die Anpassung der Dialoglautstärke (vgl. Kapitel 2.3, Stichwort Klare Sprache) angeboten werden. Durch objektbasierte Repräsentation können demnach auch die Anforderungen an NGA bezüglich Personalisierung und Barrierefreiheit erfüllt werden.

⁵⁷ Vgl. Grewe et al. 2018, S. 92; Hestermann et al. 2018, S. 96–97.

⁵⁸ Centre for Vision, Speech and Signal Processing (CVSSP) 2018.

⁵⁹ Vgl. Pardoe et al. 2020, S. 1.

3.2.4 Hybride Ansätze

Die drei vorgestellten Formen der Audiorepräsentation werden in der Praxis nicht ausschließlich getrennt voneinander behandelt. Oft bietet es sich an, verschiedene Ansätze zu kombinieren, um einen optimalen Kompromiss aus Komplexität, Qualität und erwünschtem Grad an Personalisierung zu erreichen. Um erweiterte Personalisierungsmöglichkeiten anzubieten, ist bei diesen hybriden Ansätzen die objektbasierte Repräsentation immer zumindest anteilig enthalten.

Verwendung findet häufig die Kombination aus kanalbasierter und objektbasierter Repräsentation in Form einer kanalbasierten, mehrkanaligen Grundmischung mit zusätzlichen Audioobjekten, wie beispielsweise verschiedenen Kommentatoren Sprachen bei Sportübertragungen. Praktische, vergangene Testanwendungsfälle von NGA-Broadcasts mit dieser Art des hybriden Ansatzes sind bei Dolby, de Jong et al. und Bolt et al. zu finden.⁶⁰

Ein weiterer hybrider Ansatz ist die Kombination aus szenenbasierter und objektbasierter Repräsentation. Auch hierbei wird mittels Audioobjekten die Personalisierung ermöglicht, HOA liefert die wiedergabesystemagnostische Grundmischung. Bei der Übertragung der Europameisterschaften der Leichtathletik 2018 wurde neben der Kombination aus kanalbasierter und objektbasierter Repräsentation auch dieser Ansatz getestet.⁶¹ Olivieri et al. schlagen die Verwendung dieses hybriden Ansatzes für eine erfolgreiche Implementierung von NGA unter Angabe einiger Argumente vor. So sei flexibles Rendering durch Verwendung eines szenenbasierten Betts verbunden mit einer weiterhin möglichst geringen Komplexität und einer möglichst geringen und klar voraussehbaren benötigten Bandbreite möglich, während trotzdem die Personalisierungsmöglichkeiten von objektbasierter Audiorepräsentation genutzt werden können.⁶²

Allgemein kann durch ein hybrides System die benötigte Datenrate für die Übertragung mehrerer Präsentationen eines Broadcast deutlich gesenkt werden. Tabelle 1 zeigt eine beispielhafte Einsatzmöglichkeit mit vier verschiedenen Präsentationen: 5.1 mit englischem Kommentar, 5.1 mit spanischem Kommentar, 5.1 mit chinesischem Kommentar und 5.1 ohne Kommentar.

⁶⁰ Vgl. Dolby 2018; de Jong et al. 2019; Bolt et al. 2021.

⁶¹ Vgl. de Jong et al. 2019.

⁶² Vgl. Olivieri et al., S. 20–21.

Conventional approach		Presentation approach	
Stream	Data rates/kbps	Stream	Data rates/kbps
English 5.1	144	English Commentary	48
Spanish 5.1	144	Spanish Commentary	48
Chinese 5.1	144	Chinese Commentary	48
Commentary-free 5.1	144	Commentary-free 5.1	144
TOTAL	576	TOTAL	288
		Savings	50%

Tabelle 1: Beispielhafte Datenraten kanalbasiert vs. hybrid⁶³

Während bei rein kanalbasierter Audiorepräsentation vier vollständige, einzelne 5.1-Surround-Mischungen übertragen werden müssten, reicht im hybriden Ansatz aus kanalbasierter und objektbasierter Repräsentation die Übertragung eines 5.1-Kanalbetts mit drei zusätzlichen Kommentatoren Sprachen in Form von Audioobjekten aus. Hierbei sei angemerkt, dass bei Workflows, die auf SDI basieren, eine kanalbasierte Produktion wie in der linken Hälfte von Tabelle 1 nicht möglich wäre, da jeder einzelne 5.1-Mix sechs der insgesamt verfügbaren 16 Audiokanäle verbrauchen würde.

3.3 Codecs

Wie werden nun aber die verschiedenen Ansätze sowie deren Kombinationen für NGA encodiert und übertragen? Zum Einsatz kommen die bereits in Kapitel 3 erwähnten NGA-Codecs AC-4, MPEG-H und DTS-UHD, die alle drei im aktuellen DVB-Standard enthalten sind. Da für Deutschland, wie bereits in der Einleitung dieser Arbeit erwähnt, noch nicht final festgelegt wurde, welcher dieser Codecs in Zukunft verwendet werden soll, werden die folgenden Unterkapitel alle drei Codecs kurz vorstellen.

3.3.1 AC-4

Die Dolby Laboratories mit Sitz in San Francisco liefern bereits seit einigen Jahrzehnten für die Fernsehtechnik relevante Systeme und Codierungen. Von analogen Surroundformaten wie Dolby Surround und Dolby Pro Logic (II) bis zu digitalen Formaten wie Dolby Digital (codiert in AC-3) und Dolby Digital Plus (codiert in Enhanced AC-3) fanden und finden viele Entwicklungen Einsatz bei Fernsehübertragungen (siehe dazu auch Kapitel 2). Mit AC-4 wird nun auch ein Codec für NGA-Anwendungen angeboten. AC-4 ist eine verlustbehaftete

⁶³ Dolby 2021, S. 18.

Audiocodierung, definiert in ETSI TS 103 190⁶⁴, die laut eigenen Angaben seitens der Dolby Laboratories über eine im Durchschnitt um etwa 50% effizientere Audiocodierung im Vergleich zur Vorgängercodierung Enhanced AC-3 verfügt.⁶⁵ Unterstützt wird immersives Audio in kanalbasierter und objektbasierter Repräsentation sowie Kombinationen beider Ansätze.⁶⁶ HOA werden hingegen nicht unterstützt. Mittels objektbasierten Ansatzes erfüllt AC-4 die Anforderungen an NGA bezüglich Personalisierung und Barrierefreiheit. Als objektbasiertes bzw. hybrides System kann wiedergabeseitig frei auf das jeweils vorliegende Wiedergabesystem gerendert werden, wodurch der geforderte Aspekt der Unabhängigkeit vom Wiedergabesystem ebenso gegeben ist.

3.3.2 MPEG-H

MPEG-H, definiert in ISO/IEC 23008-3:2015⁶⁷, ist der NGA-Codec der Motion Picture Experts Group (MPEG), die in der Vergangenheit, wie auch Dolby, bereits Audiocodierungen für Broadcastanwendungen geliefert haben. MPEG-H unterstützt wie auch Dolby AC-4 kanalbasiertes und objektbasiertes Audio, allerdings – anders als AC-4 – auch szenenbasiertes Audio, sowie jegliche Kombinationen dieser Repräsentationen.⁶⁸ Für Broadcastanwendungen wird dabei das Low Complexity Level 3 angewandt, das mit maximal 48 kHz Abtastrate arbeitet und bis zu 32 einzelne Audioelemente (Kanäle, Objekte, Ambisonics) verarbeiten kann, wobei maximal 16 Audioobjekte gleichzeitig gerendert werden können.⁶⁹ Damit erfüllt MPEG-H auf Basis von objektbasierter Audiorepräsentation die Anforderungen an Personalisierung und Barrierefreiheit sowie durch objekt- und szenenbasierte Repräsentation den Anspruch an systemagnostische Wiedergabe.

3.3.3 DTS-UHD

Der dritte NGA-Codec, DTS-UHD, ist definiert in ETSI TS 103 491⁷⁰ und kommt vom kalifornischen Unternehmen Digital Theater Systems (DTS) Incorporated. Wie Dolby und MPEG, hat auch DTS in der Vergangenheit bereits mehrere Formate und Codierungen für

⁶⁴ Vgl. ETSI 2018.

⁶⁵ Vgl. Dolby 2021, S. 13.

⁶⁶ Vgl. ebd., S. 5.

⁶⁷ Vgl. ISO/IEC 23008-3:2015 2015.

⁶⁸ Vgl. Grewe et al. 2021, S. 1.

⁶⁹ Vgl. Bleidt et al. 2017, S. 204; Turnwald et al. 2018, S. 90.

⁷⁰ Vgl. ETSI 2019.

Fernsehanwendungen geliefert und stellt mit dem NGA-Codec DTS-UHD die dritte Entwicklungsstufe nach den weiterhin verwendeten Codecs DTS und DTS-HD. Mittels DTS-UHD kann wie auch mit MPEG-H kanal-, objekt- und szenenbasiertes Audio codiert und übertragen werden.⁷¹ Dementsprechend erfüllt auch der Codec von DTS die in Kapitel 3 präsentierten Anforderungen an NGA-Produktionsformate bezüglich immersivem Audio und Systemagnostizität sowie mittels objektbasierter Repräsentation die endnutzerseitige Personalisierung und Barrierefreiheit.

3.4 Kompatibilität

Die drei behandelten NGA-Audiocodierungen nutzen jeweils eigene Metadatenstrukturen zur Beschreibung der enthaltenen Audioszenen. Hierdurch ergibt sich die Problematik, dass bei Konvertierung zwischen den verschiedenen Formaten Informationen verändert werden oder verloren gehen könnten. Ergebnis wäre ein fehlerhaftes Rendering und damit eine fehlerhafte Reproduktion der Audioszene auf das vorliegende Wiedergabesystem.

Aufgrund dieser Problematik wurde das ADM entwickelt. Das ADM, spezifiziert in ITU-R BS. 2076⁷², ist ein codec-agnostisches, universelles und offenes Metadatenmodell für die Produktion, den Austausch und die Archivierung von NGA-Inhalten in dateibasierten Workflows.⁷³ Mit dem ADM können alle drei in NGA-Workflows verwendeten Audiorepräsentationen (kanal-, objekt- und szenenbasiert) beschrieben werden. Es ist allerdings kein eigenes Audioübertragungsformat, sondern enthält lediglich eine einheitliche Beschreibung dessen, was im jeweiligen Übertragungsformat enthalten ist.⁷⁴ Diese Beschreibung der Metadaten wird in der menschenlesbaren Extensible Markup Language (XML) vorgenommen, kann anschließend aber auch in andere Sprachen wie beispielsweise JavaScript Object Notation (JSON) überführt werden und wird für die Übertragung Wave-Dateiformaten wie Broadcast Wave 64 (BW64) oder dem Broadcast Wave Format (BWF), die in BS.2088 respektive BS.1352 spezifiziert sind⁷⁵, angehängt.⁷⁶ Mithilfe der einheitlichen

⁷¹ Vgl. ETSI 2019, S. 24.

⁷² Vgl. ITU-R 2019b.

⁷³ Vgl. Fraunhofer Institute for Integrated Circuits (IIS) 2023.

⁷⁴ Vgl. ITU-R 2019b, S. 6.

⁷⁵ Vgl. ITU-R 2019c, ITU-R 2007.

⁷⁶ Vgl. ITU-R 2019b, S. 6–7.

Metadatenstruktur des ADMs können vorhandene NGA-Inhalte fehlerfrei zwischen den drei in Kapitel 3.3 vorgestellten Codierungen konvertiert und weiterverarbeitet werden.

Wie eingangs dieses Unterkapitels beschrieben, ist das ADM ein Metadatenmodell für dateibasierte Workflows. Für lineare Workflows wie Live- und Echtzeitproduktionen, also auch Broadcasts, bei denen Audiomaterial in einzelnen Frames anstelle von geschlossenen Dateien übertragen werden muss, ist das ADM daher nicht geeignet. Für solche Anwendungsfälle gibt es daher eine serielle Repräsentation des ADM (Serial Audio Definition Model, S-ADM), die in BS.2125 spezifiziert ist.⁷⁷ S-ADM kann dieselben Metadaten wie ADM beschreiben, segmentiert diese allerdings für serielle, also nacheinander abfolgende Übertragung. Abbildung 7 zeigt eine schematische Darstellung dieser Segmentierung der ADM-Metadaten in einzelne Metadaten-Frames.

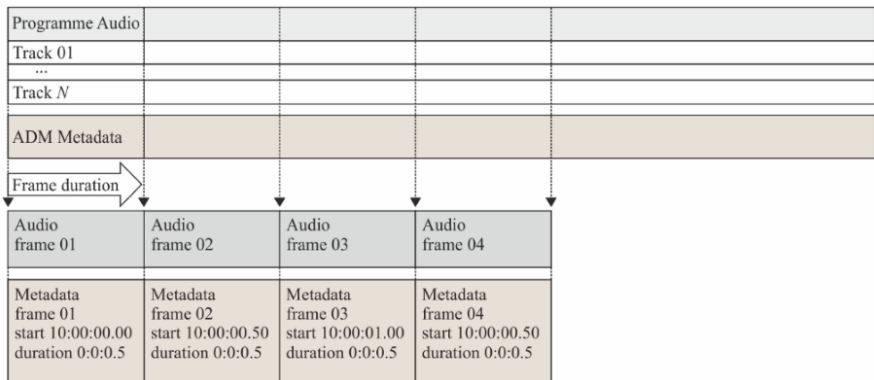


Abbildung 7: Schematische Darstellung von S-ADM⁷⁸

Für korrektes Rendering von Audiomaterial mit ADM-Metadaten wurde außerdem ein eigener, ebenfalls offen verfügbarer Renderer entwickelt, der in BS.2127 spezifiziert ist.⁷⁹ Basierend auf den auf Produktionsseite erstellten Metadaten sowie dem Endnutzer-Input am Endgerät erzeugt dieser Renderer die passenden Signale für die Schallwandler des jeweiligen Wiedergabesystems.

⁷⁷ Vgl. ITU-R 2022b.

⁷⁸ Ebd., S. 4.

⁷⁹ Vgl. ITU-R 2019a.

4. Funktionen von NGA-Objekten

Nachdem das vorausgehende Kapitel herausgearbeitet hat, welche Neuerungen NGA mit sich bringt und weiterhin erläutern konnte, dass bei rein objektbasierter bzw. einer hybriden Herangehensweise aus objektbasierter sowie entweder kanal- oder szenenbasierter Audiorepräsentation weitgehende Personalisierungsmöglichkeiten für Endnutzer ermöglicht werden können, behandelt dieses Kapitel nun die spezifischen, im empirischen Teil der Arbeit untersuchten Einsatzmöglichkeiten, die Endnutzern bei NGA-Broadcasts zur Verfügung gestellt werden können.

4.1 Mehrsprachige Übertragung

Mehrsprachige Broadcastübertragung ist bereits seit dem in Kapitel 2.2 beschriebenen Zweikanalton der 1980er-Jahre und auch eine bis heute weiterhin praktizierte Form der Audioübertragung deutscher Fernsehsender. Prominentester Vertreter mehrsprachiger Übertragung in Deutschland ist der deutsch-französische Kultursender Arte, der in Deutschland laut eigenen Angaben bei etwa 70% seiner Angebote neben der deutschsprachigen auch eine französischsprachige Audiotonspur anbietet.⁸⁰ Während hierfür bei klassisch kanalbasierter Repräsentation in SDI-Workflows beispielsweise 12 SDI-Kanäle für zwei 5.1-Surroundmischungen (deutsch und französisch) nötig wären, könnte bei einem hybriden NGA-Workflow sogar ein immersives Audioerlebnis durch ein 5.1.4-Kanalbett mit den beiden Sprachen in Form von Audioobjekten mit der gleichen Anzahl an SDI-Kanälen übertragen werden. Zwei rein kanalbasierte, immersive Mischungen im 5.1.4-Format wären bei SDI-Workflows hingegen überhaupt nicht möglich, da 20 Kanäle benötigt würden, SDI allerdings über nur 16 eingebettete Audiokanäle verfügt. Da SDI auf absehbare Zeit weiterhin Teil von Broadcastproduktionen bleiben wird⁸¹ und rein IP-basierte Workflows, bei denen theoretisch mit unbegrenzt vielen Kanälen gearbeitet werden könnte⁸², erst sukzessive adoptiert werden, bieten NGA-Produktionen für Broadcaster die Möglichkeit, ihren Zuschauern einen höheren Personalisierungsgrad selbst bei SDI-basierten Workflows

⁸⁰ Vgl. ARTE 2023.

⁸¹ Vgl. Bolt et al. 2021, S. 14.

⁸² Vgl. ebd., S. 5.

anbieten zu können. Mehrere Sprachen auf Basis von objektbasierter Audiorepräsentation zu realisieren, fand bereits vielfach Anwendung bei Testübertragungen großer Sportevents.⁸³

4.2 Anpassung von Lautstärkelevels

Bisher funktioniert die Anpassung der Lautstärke einzelner Elemente einer Gesamtmischung, ohne dabei Einfluss auf die restliche Audiomischung zu nehmen, im deutschen Fernsehen lediglich für die Dialoglautstärke bei den öffentlich-rechtlichen Sendern durch das bereits in Kapitel 2.3 erwähnte KI-basierte „Klare Sprache“-Verfahren. Hierbei wird aus fertigem, kanalbasierten Audiomaterial die Sprache herausgefiltert und in der Lautstärke verstärkt, wohingegen die restlichen Anteile der Gesamtmischung in der Lautstärke leicht abgesenkt werden. Das Ergebnis dieses Echtzeit-Prozesses wird dem Zuschauer als separate Audiospur ebenso wie die Standardabmischung als rein kanalbasierte Repräsentation des Audiomaterials angeboten. Sie kann als Alternative zum Standardton ausgewählt werden, ist aber nicht weitergehend veränder- bzw. personalisierbar. Die privaten deutschen Fernsehsender bieten ein solches Verfahren aktuell nicht an. Mit Umstellung auf NGA wäre es Fernsehsendern möglich, den Dialog eines Broadcast als Audioobjekt separat zu einem kanal- oder szenenbasierten Audiobett anzubieten und dem Zuschauer im Metadaten-Authoring verschiedene Grade an Personalisierung der Dialoglautstärke zu ermöglichen. Beispielsweise wäre eine Präsentation mit mehreren Abstufungen als vordefinierte Presets denkbar: Standard, Dialog+ (Dialog z.B. um +6dB verstärkt) und Dialog++ (Dialog z.B. um +9dB verstärkt). Alternativ kann dem Zuschauer ein vordefinierter Bereich angeboten werden, in dem der Dialog gegenüber der restlichen Mischung anpassbar ist (z.B. von -6dB bis +12 dB). Beispiele für bisherige Einsätze von objektbasierter Erhöhung der Sprachverständlichkeit finden sich bei Turnwald et al. (ESC 2018) und Bolt et al. (Fußball-Europameisterschaft 2020).⁸⁴

4.3 Audiodeskription

Auch die Audiodeskription fand bei deutschen Fernsehsendern frühzeitig Anwendung. Ebenso wie die mehrsprachige Übertragung wurden Broadcasts mit einer optionalen Audiodeskription durch die Verwendung der beiden Kanäle der Zweikanalübertragung des

⁸³ Vgl. Turnwald et al. 2018, S. 90; Bolt et al. 2021, S. 7, 2021, S. 8–9, 2021, S. 10.

⁸⁴ Vgl. Turnwald et al. 2018, S. 90; Bolt et al. 2021, S. 7.

IRT als separate Monokanäle technisch erstmals in den 1980er-Jahren möglich (siehe Kapitel 2.2), Anwendung fand sie allerdings erstmalig beim ZDF im Jahr 1993 und regelmäßig ab 1997 bei Arte und dem BR, woraufhin weitere öffentlich-rechtliche Sender sukzessive folgten.⁸⁵ Wie auch bei den beiden bisher behandelten Funktionen, kann Audiodeskription bei kanalbasiertem Audiomaterial nur dann angeboten werden, wenn eine vollständige eigene Mischung zusätzlich zu etwaigen anderen Stereo- bzw. Surroundmischungen übertragen wird. Bei NGA kann die Audiodeskription hingegen als eigenes, separates Audioobjekt zusätzlich zum kanal- oder szenenbasierten Audiobett, wie auch eine oder mehrere Wiedergabesprachen (vgl. Kapitel 4.1), übertragen und durch den Zuschauer hinzugeschaltet werden. Die optionale, durch objektbasierte Audiorepräsentation zur Verfügung gestellte Audiodeskription verwendete beispielsweise das Schweizer Radio und Fernsehen (SRF) bei einer Testübertragung 2019.⁸⁶

4.4 Positionierbare Objekte

Die vierte zu untersuchende Einsatzmöglichkeit von NGA-Objekten ist eine gänzlich neue Funktion, die beim herkömmlichen, kanalbasierten Ansatz nicht möglich ist. Hierbei handelt es sich um die Positionierung von Audioobjekten im Raum. Im Metadaten-Authoring kann auf Produktionsseite nicht nur der Gain eines Audioobjekts vordefiniert werden, sondern auch die Position des Objekts im Raum. Dem Endnutzer kann dementsprechend nicht nur die Option der Lautstärkeanpassung eines Audioobjekts gegeben werden, sondern auch die Option, die Position eines Objekts zu beeinflussen. Beispielsweise könnte ermöglicht werden, eine Audiodeskription von standardmäßig zentral vorne in der Mitte durch mehrere Regler (links-rechts, vorne-hinten, oben-unten) auf eine beliebige gewünschte Position im Raum zu verschieben, um sie näher in die Richtung des Zuhörers zu bewegen, der sie aufgrund einer sensorischen Beeinträchtigung benötigt. Turnwald et al. beschreiben das Authoring dieser neuartigen Funktion im Rahmen der Live-Übertragung des ESC 2018, bei dem den Zuschauern das Panning des Kommentars in zwei Ebenen (links-rechts, oben-unten) ermöglicht wurde.⁸⁷

⁸⁵ Vgl. ARD 2023a.

⁸⁶ Vgl. Bolt et al. 2021, S. 10.

⁸⁷ Vgl. Turnwald et al. 2018, S. 90.

5. Hypothesen

Die allgemeine Bewertung dieser einzelnen Einsatzmöglichkeiten wurde anhand eines Hypothesenpaars analysiert. Dieses Hypothesenpaar wurde dabei für alle vier genannten Einsatzmöglichkeiten überprüft. Für die in Kapitel 6 näher beschriebenen Signifikanztests ergab sich der Erwartungswert μ_0 , gegen den getestet wurde, aus der in der Umfrage verwendeten siebenstufigen Likert-Skala, bei der der Wert 4 eine neutrale Bewertung der jeweiligen Funktion darstellt ($\mu_0 = 4$). Es wurde das folgende Hypothesenpaar formuliert:

H₀: Die Funktion wird neutral bewertet ($\mu = \mu_0$).

H₁: Die Funktion wird nicht neutral bewertet ($\mu \neq \mu_0$).

Weiterhin sollte anhand zweier weiterer Hypothesenpaare überprüft werden, ob eine Abhängigkeit der Bewertung der Funktionen nach Altersgruppen oder dem Geschlecht der Probanden ausgemacht werden kann. Hierzu wurden folgende Hypothesenpaare formuliert:

H₀: Die Bewertung der Funktionen ist unabhängig vom Alter.

H₂: Die Bewertung der Funktionen ist abhängig vom Alter.

H₀: Die Bewertung der Funktionen ist unabhängig vom Geschlecht.

H₃: Die Bewertung der Funktionen ist abhängig vom Geschlecht.

Zuletzt sollte noch überprüft werden, ob das Vorhandensein einer Beeinträchtigung des Hör- oder des Sehapparates einen signifikanten Einfluss auf die Bewertung derjenigen NGA-Funktionen hat, die Einfluss auf diese Beeinträchtigungen nehmen können:

H₀: Die Bewertung der Funktion „Anpassung der Lautstärkelevels“ wird durch das Vorhandensein einer Beeinträchtigung des Hörapparats nicht beeinflusst.

H₄: Die Bewertung der Funktion „Anpassung der Lautstärkelevels“ wird durch das Vorhandensein einer Beeinträchtigung des Hörapparats beeinflusst.

H₀: Die Bewertung der Funktion „Audiodeskription“ wird durch das Vorhandensein einer Beeinträchtigung des Sehapparats nicht beeinflusst.

H₅: Die Bewertung der Funktion „Audiodeskription“ wird durch das Vorhandensein einer Beeinträchtigung des Sehapparats beeinflusst.

6. Methodik

In der Einleitung dieser Arbeit wurde bereits erläutert, wie minimal die bisherige Forschung der konkreten Anwendung von NGA direkt am Konsumenten stattgefunden hat. Diese Arbeit soll daher Teil der Forschung direkt an den potenziellen Endnutzern von zukünftigen NGA-Broadcastangeboten sein. Um die Forschungsfragen der vorliegenden Arbeit zu beantworten, wurde dementsprechend eine quantitative Umfrage innerhalb der deutschen Bevölkerung durchgeführt, um die Sicht deutscher Broadcast-Konsumenten auf die technisch realisierbaren Personalisierungsmöglichkeiten bei NGA-Broadcasts sowie deren Relevanz für ihr Broadcasterlebnis zu überprüfen. Die Datenerhebung wurde mithilfe des Umfrageportals UmfrageOnline.com über einen Zeitraum von zwei Wochen in der Zeitspanne vom 24. Januar bis zum 7. Februar 2023 vollzogen. Teilnahmeberechtigt waren dabei alle Menschen mit aktuellem Wohnsitz in der Bundesrepublik Deutschland ab Vollendung ihres 16. Lebensjahres. Die Altersbegrenzung ergab sich hierbei durch die gesetzliche Regelung für Onlinebefragungen, aufgrund welcher für internetbasierte Befragungen bei Personen unter 16 Jahren die Einwilligungserklärung eines Sorgeberechtigten vonnöten ist.⁸⁸ Da alle Einwohner der Bundesrepublik Deutschland potenzielle oder bereits aktuelle Konsumenten von Broadcasts deutscher Fernsehsender sind, ergaben sich keine weitergehenden Einschränkungen der Umfragepopulation. Je nach Beantwortung der Probanden enthielt die Umfrage zwischen 19 und 22 Fragen. Die verschiedenen Umfänge des Fragebogens ergaben sich dadurch, dass einige Fragen durch Fragebogenlogik entfernt wurden, wenn eine bestimmte Antwort bei einer vorausgehenden Frage gewählt wurde. Verteilt wurde der Link zur Umfrage im privaten Umfeld des Forschenden, über Mailverteiler des Erstprüfers, Herrn Professor Curdt, und des technischen Mitarbeiters im TV-Bereich der Hochschule der Medien, Matthias Adler, sowie in deutschsprachigen Facebook-Gruppen zur Akquirierung von Umfrageteilnehmern und die Umfrageverteilungsplattform SurveyCircle.

Insgesamt nahmen 287 Personen an der Befragung teil. In die statistische Auswertung flossen allerdings ausschließlich vollständige Beantwortungen des Fragebogens ein. Einige Probanden schlossen die Befragung nicht ab, weswegen durch dieses Kriterium 56 Beantwortungen ausgeschlossen wurden. Weiterhin wurden manuell solche Beantwortungen aussortiert, bei denen die Einwilligungserklärung durch den jeweiligen Probanden abgelehnt

⁸⁸ Vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM) et al. 2021, S. 1.

wurde. Diese wurden in der automatischen Erkennung durch UmfrageOnline.com zunächst ebenso als vollständig bearbeitet gekennzeichnet. Damit wurden fünf weitere Beantwortungen ausgeschlossen. Das letzte Kriterium stellte die Beantwortung mindestens einer der Bewertungsfragen aus dem Hauptteil, wodurch zwei weitere Beantwortungen aussortiert wurden, bei denen alle Bewertungsfragen übersprungen wurden. Nach Berücksichtigung aller Ein- und Ausschlusskriterien ergab sich daher schlussendlich eine Stichprobengröße von $n = 224$ Personen für die Auswertung. Der ursprünglich erwünschte Stichprobenumfang von $n \geq 385$, der sich aus der Formel für die Schätzung von Anteilswerten mit kleinem Auswahlssatz⁸⁹ mit 95%-Konfidenzniveau ($Z_{\alpha} = 1,96$) und einer maximalen Fehlermarge von 5% ($e = 0,05$) ergibt, konnte demnach nicht erreicht werden. Daher muss für die im späteren Verlauf präsentierten Ergebnisse von einer maximalen Fehlermarge von 7% ($e = 0,07$) ausgegangen werden. Die hierfür erforderliche Stichprobengröße von $n \geq 196$ konnte bei gleichem 95%-Konfidenzniveau erreicht werden.

Als Statistiksoftware für die Auswertung wurde SPSS der International Business Machines Corporation verwendet. Die erhobenen und ausgewerteten Daten der Einstellungsvariablen aus dem Hauptteil der Befragung wiesen dabei keine Normalverteilung auf.⁹⁰ Nach Rasch und Guiard sind parametrische Testmethoden für große Stichproben ausreichend robust gegenüber Verletzungen der Normalverteilung⁹¹, die in Kapitel 7 präsentierten Ergebnisse der angewandten parametrischen Verfahren zu den in Kapitel 5 formulierten Forschungshypothesen wurden dennoch zusätzlich durch die jeweils korrespondierenden, nicht-parametrischen Testverfahren, die keine Normalverteilung voraussetzen, überprüft. Zur Testung der Hypothese H_1 auf Signifikanz wurden parametrische Einstichproben t-Tests sowie zur Absicherung derer Ergebnisse nicht-parametrische Einstichproben Wilcoxon-Tests angewandt. Die Überprüfung der Hypothese H_2 wurde durch einfaktorielle Varianzanalysen (ANOVA) sowie nicht-parametrische Kruskal-Wallis-Tests für mehr als zwei unabhängige Stichproben durchgeführt. Zur Testung der Signifikanzen der Hypothesen H_3 bis H_5 wurden zuletzt Zweistichproben t-Tests sowie nicht-parametrische Mann-Whitney-U-Tests zum Vergleich von zwei unabhängigen Stichproben verwendet.

⁸⁹ Vgl. Häder und Häder 2014, S. 288.

⁹⁰ Für die Tests auf Normalverteilung siehe Anhang 1.

⁹¹ Vgl. Rasch und Guiard 2004.

6.1 Aufbau des Fragebogens

Der erstellte Fragebogen bestand in seiner groben Struktur aus vier Teilen: Einleitung mit Begrüßungstext und Einwilligungserklärung, Beginn der Befragung mit einigen Einstiegsfragen, Hauptteil mit den inhaltlichen Fragen zu Personalisierungsmöglichkeiten bei NGA-Broadcasts und Schluss mit demografischen Fragen.

In der Einleitung wurden die Probanden zunächst begrüßt und anschließend das Thema der Befragung genannt sowie erläutert, zu welchem Zweck die Befragung durchgeführt wird. Weiterhin wurden die Probanden über ihre Aufgabe bei der Befragung aufgeklärt und auf die Notwendigkeit von Kopfhörern hingewiesen, um die in der Umfrage enthaltenen Videos bestmöglich wiedergeben zu können. Für die anschließende Einwilligungserklärung wurde die empfohlene, vorgefertigte Einwilligungserklärung der Hochschule der Medien abgewandelt übernommen und um die Bedingungen der Alters- sowie der Wohnortbegrenzung ergänzt.⁹² Bei Ablehnung der Einwilligungserklärung erfolgte das sofortige Ende der Befragung.

Zu Beginn der eigentlichen Befragung wurden einige einfache, schnell beantwortbare Einstiegsfragen angebracht, die für die Beantwortung der Forschungsfragen eine nur sekundäre Rolle spielten. Hiermit sollte den Probanden eine niederschwellige Einführung in die Befragung ermöglicht werden. Es wurde dabei erfragt, wie oft Probanden Angebote deutscher Fernsehsender konsumieren sowie über welche Wiedergabekonfigurationen sie diese Angebote an ihren stationären (Fernseher, Desktop-PC) bzw. mobilen (Smartphone, Tablet, Notebook) Endgeräten konsumieren. Falls bei der ersten Frage die Antwortmöglichkeit „Gar nicht“ gewählt wurde, wurden die beiden weiteren Fragen automatisiert ausgeblendet. Die Erhebung der Antworten auf diese Fragen erfolgte durch Multiple Choice, wobei die erste Frage nur per Einzelantwort ausgefüllt, die beiden anderen Fragen ebenso durch die Auswahl mehrerer Antwortmöglichkeiten beantwortet werden konnten.

Im Anschluss an die Einstiegsfragen folgte der Hauptteil, in dem die vier in Kapitel 4 erläuterten Einsatzmöglichkeiten und Funktionen von Audioobjekten zur Personalisierung von NGA-Broadcasts durch kurze Videos vorgestellt und anschließend dazu

⁹² Vgl. Hochschule der Medien (HdM) 2022, S. 4–5.

Bewertungsfragen gestellt wurden. Entsprechend den vier vorgestellten Einsatzmöglichkeiten wurde dieser Teil der Befragung in vier Blöcke unterteilt. Jeder Block enthielt ein Video, das die jeweilige Einsatzmöglichkeit präsentierte, sowie drei bis vier Bewertungsfragen mit siebenstufigen, endpunktbenannten Likert-Skalen. Diese reichten von 1, wie „Überhaupt nicht sinnvoll“, bis 7, wie „Ausgesprochen sinnvoll“, und wurden gewählt, da sie als intervallskaliert betrachtet werden dürfen⁹³, wodurch mehr mathematische Operationen im Umgang mit den Ergebnissen verwendet werden durften. Die ersten drei Fragen jedes Blocks erfragten die Bewertung der jeweiligen Personalisierungsmöglichkeit in drei Kategorien bzw. Formaten: Filme und Serien, Nachrichtensendungen und Sportübertragungen. Diese drei Kategorien wurden gewählt, um eine Bewertung der Personalisierungsmöglichkeiten in drei der beliebtesten Formatkategorien der deutschen Zuschauerschaft basierend auf einer Umfrage des Instituts für Demoskopie Allensbach (IfD) aus dem Jahr 2022 zu erfragen.⁹⁴ Aus diesen drei Items wurden für jede behandelte Funktion durch Mittelwertberechnungen der entsprechenden drei Items eigene latente Variablen erstellt, an welchen das jeweilige Merkmal, also die Bewertung der einzelnen Funktion bzw. die Einstellung der deutschen Bevölkerung zu den Personalisierungsmöglichkeiten, untersucht werden sollte. Für die beiden Funktionen „Mehrsprachige Übertragung“ und „Positionierbare Objekte“ gab es jeweils eine weitere Frage, die besondere Einsatzmöglichkeiten basierend auf Ansätzen aus der Literatur beinhalteten. Bei „Mehrsprachige Übertragung“ handelte es sich hierbei um den Sonderfall von mehreren Kommentatoren bei Sportübertragungen abhängig von ihrer Teamzugehörigkeit⁹⁵, bei „Positionierbare Objekte“ um den Sonderfall positionierbarer bzw. beweglicher Audiodeskription.⁹⁶ Diese Sonderfälle gingen in die allgemeinen Skalen für die Merkmalsuntersuchungen nicht ein und werden lediglich im Diskussionsteil in Kapitel 8 erneut aufgegriffen. Anschließend wurde abgefragt, ob der Einsatz dieser Funktionen dazu führen könnte, dass der Proband mehr Angebote deutscher Fernsehsender schauen würde. Falls diese Frage mit „Ja“ beantwortet wurde, erschien – den Hauptteil abschließend – die

⁹³ Vgl. Keller 2013.

⁹⁴ Vgl. Institut für Demoskopie Allensbach (IfD) 2022a.; Hier wird in viele Unterkategorien unterteilt, wobei fünf der ersten zehn zur Überkategorie „Filme und Serien“ gezählt werden können. Daher wurde diese Kategorie formuliert. „Sportübertragungen“ wurde als dritte Kategorie anstelle von „Reisemagazinen und Sendungen über fremde Länder und Kulturen“ gewählt, da diesen in der bisherigen Forschung und bei bisherigen Testübertragungen Hauptaugenmerk zukam.

⁹⁵ Vgl. Dolby 2018, S. 8; EBU 2021, S. 11.

⁹⁶ Vgl. Turnwald et al. 2018, S. 90; Pardoe et al. 2020, S. 5; EBU 2021, S. 7.

Frage, welche der Funktionen dies bewirken könnten. Diese Fragenfolge ist ebenso wie die Fragen aus dem Abschnitt mit den Einstiegsfragen nicht für die Untersuchung der Hypothesen relevant und wird daher erst in der Diskussion erneut aufgegriffen.

Der abschließende Teil des Fragebogens enthielt zuletzt demografische Fragen zum Geschlecht und dem Alter der Probanden sowie die Frage, ob eine sensorische Beeinträchtigung vorliegt, die durch eine der Funktionen beeinflusst werden kann. Diese Angaben wurden aufgrund der Hypothesen H₂ bis H₅ erhoben.

Die vollständigen Fragebogen in deutscher wie englischer Sprache sowie die Links zu den dazugehörigen Videos sind im Anhang dieser Arbeit enthalten (siehe Anhang 4).

6.2 Gütekriterien quantitativer Forschung

Als quantitative Studie im Rahmen einer Umfrage obliegt diese Forschung den drei Gütekriterien quantitativer Forschung: Objektivität, Reliabilität und Validität. Objektivität bedeutet dabei „im Kontext quantitativer Forschung [...], in wie weit ein Messwert unabhängig ist von der Person, die eine Messung durchführt.“⁹⁷ Krebs und Menold unterteilen diesen Begriff der Objektivität in die Unterkategorien der Durchführungs-, Auswertungs- und Interpretationsobjektivitäten.⁹⁸ Die Durchführungsobjektivität wurde bei der zugrundeliegenden Datenerhebung, wie es ebendort beschrieben wird⁹⁹, durch die Standardisierung des Fragebogens, also durch das Zurverfügungstellen lediglich eines einheitlichen Fragebogens an alle Probanden, weitgehend gesichert. Krebs und Menold führen überdies an, dass für die Sicherstellung der Auswertungsobjektivität eine detaillierte Dokumentation der Datenaufbereitung erfolgen sowie die Modifikation der Daten für die statistischen Analysen und der Umgang mit fehlenden Werten erläutert werden muss.¹⁰⁰ Die Datenaufbereitung wurde in Kapitel 6 behandelt, welches die Aufbereitung der Rohdaten unter Berücksichtigung einiger Ein- und Ausschlusskriterien erläutert. In Kapitel 6.1 wurde zudem die Modifikation der Daten in Form der Zusammenfassung jeweils dreier Items durch Mittelwertberechnungen erläutert. Fehlende Werte im Datensatz wurden in der Variablenansicht von SPSS als fehlend definiert und mit dem Wert „-99“ beschriftet. SPSS

⁹⁷ Flick 2014, S. 412.

⁹⁸ Vgl. Krebs und Menold 2014, S. 426.

⁹⁹ Vgl. ebd.

¹⁰⁰ Vgl. ebd., S. 427.

ignoriert diese Werte bei analytischen Verfahren. Zuletzt schränken Krebs und Menold die Interpretationsobjektivität ein, da sie für die Interpretation sozialwissenschaftlicher Forschungsergebnisse nicht existieren kann, weil sie zwangsläufig subjektiven Bewertungen unterliegt.¹⁰¹

Die Reliabilität ist in der quantitativen Forschung definiert als „das Ausmaß, in dem wiederholte Messungen eines Einstellungsobjekts zu gleichen Werten führen [, ...] bezieht sich also auf die Replizierbarkeit von Messungen und beschreibt die Genauigkeit (Präzision, Zuverlässigkeit), mit der ein (Einstellungs-)Objekt erfasst werden kann.“¹⁰² Zur Sicherstellung der internen Reliabilität der erfassten Daten wurde der Reliabilitätskoeffizient Cronbachs Alpha für die Einstellungsskalen der Personalisierungsmöglichkeiten berechnet. Die Ergebnisse sind in Tabelle 2 zu sehen.

<i>Mehrsprachige Übertragung</i>	<i>Anpassung von Lautstärkelevels</i>	<i>Audiodeskription</i>	<i>Positionierbare Objekte</i>
0,825	0,734	0,852	0,845

Tabelle 2: Cronbachs Alpha der Einstellungsskalen

Mit Werten von Cronbachs Alpha > 0,70 für alle vier bzw. > 0,80 für drei der Einstellungsskalen liegen die Werte durchgängig im akzeptablen bis wünschenswerten Bereich.¹⁰³

Das dritte Gütekriterium der Validität beschreibt das „Ausmaß, in dem ein Test das misst, was er zu messen vorgibt.“¹⁰⁴ Sie wurde für die durchgeführte Umfrage anhand der begründet ausgewählten Items für die einzelnen Personalisierungsmöglichkeiten, die die gesamten Einstellungsskalen ergeben, sichergestellt. Durch diese Items wird explizit die Einstellung der Probanden zu den Personalisierungsmöglichkeiten und kein anderer Aspekt erfragt.

¹⁰¹ Vgl. Krebs und Menold 2014, S. 426.

¹⁰² Ebd., S. 427–428, Hervorhebung im Original.

¹⁰³ Vgl. ebd., S. 430.

¹⁰⁴ Bühner 2006, S. 36.

7. Ergebnisse

In diesem Kapitel werden nachfolgend die Ergebnisse der empirischen Datenerhebung präsentiert. Zunächst wird hierbei die demografische Konstitution der Stichprobe in Kapitel 7.1 vorgestellt. Anschließend werden in den Kapiteln 7.2 bis 7.4 die relevanten Ergebnisse für die Hypothesen H_1 bis H_5 vorgestellt.

7.1 Demografische Konstitution der Stichprobe

Wie bereits in Kapitel 6.1 beschrieben, wurden am Ende des Fragebogens einige demografische Merkmale der Probanden erfragt. Hierbei wurde das Alter sowie das Geschlecht erhoben und zusätzlich erfragt, ob bei den Probanden eine sensorische Beeinträchtigung des Hör- und/oder des Sehapparats vorliegt. Um einen Überblick über die allgemeine demografische Konstitution der Stichprobe zu geben, werden die Ergebnisse dieser Fragen nachfolgend kurz vorgestellt und veranschaulicht.

Abbildung 8 zeigt die Geschlechter- sowie die Altersstruktur innerhalb der erhobenen Stichprobe. Diese Verteilungen sind für die Betrachtung der Repräsentativität der Stichprobe für die Gesamtbevölkerung Deutschlands und speziell für die Hypothesen H_2 und H_3 von Bedeutung. Die dargestellten relativen Werte der Geschlechterverteilung ergeben sich aus der bereits genannten Stichprobengröße von $n = 224$ sowie den Häufigkeiten: 96 Frauen, 126 Männer sowie 2 Probanden, die zu ihrem Geschlecht keine Angabe machen wollten. Die Option „divers“ wählte kein Proband.

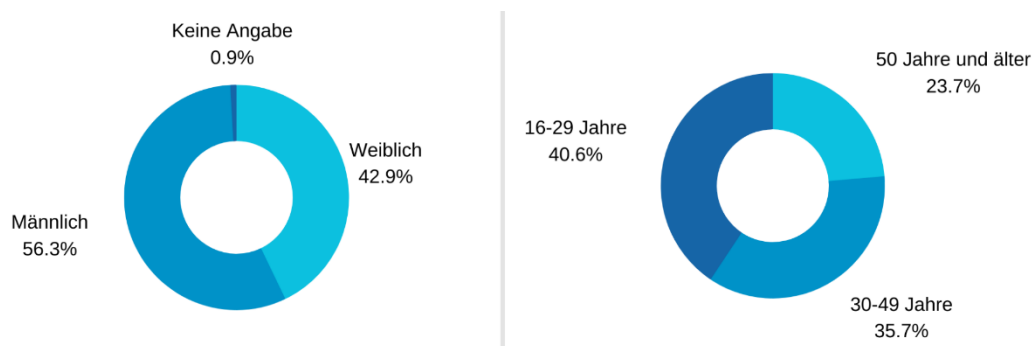


Abbildung 8: Geschlechter- und Altersstruktur der Stichprobe

Die ursprünglich erhobenen Altersgruppen (16-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70 und älter) wurden für die Auswertung in drei umfassendere, etwa gleich große Gruppen unterteilt. Näherungsweise folgt dies einer Einteilung, bei denen Menschen, die noch reinen Mono-Fernsehton kennengelernt haben, in die älteste Gruppe aufgenommen wurden. Die, die bereits

ab sehr jungem Alter mit Stereoton im Fernsehen konfrontiert wurden, gelangten in die mittlere Gruppe und Menschen, die zumindest technisch gesehen immer Zugriff auf Surround-Fernsehton hatten, wurden Teil der jüngsten Gruppe. Somit umfasste die jüngste Gruppe der 16- bis 29-Jährigen 91 Probanden, die mittlere Gruppe der 30- bis 49-Jährigen 80 Probanden und die älteste Gruppe der 50-Jährigen und älter 53 Probanden.¹⁰⁵ Die Geschlechterverteilung innerhalb dieser definierten Gruppen ist in Abbildung 9 dargestellt. Die Verteilungen in der jüngsten und ältesten Gruppe sind gleichmäßig, lediglich in der mittleren Gruppe überwiegt der Anteil der männlichen Probanden deutlich.

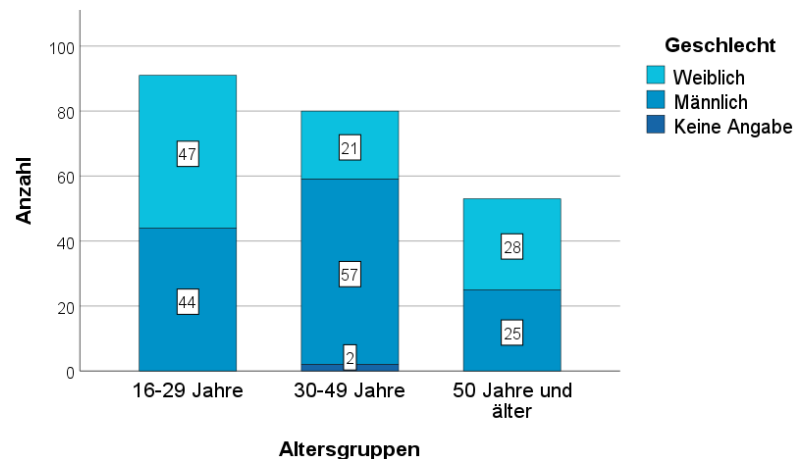


Abbildung 9: Geschlechterverteilung in den Altersgruppen

Die Beschreibung der Konstitution der Stichprobe abschließend zeigt Tabelle 2 die Häufigkeiten sensorischer Beeinträchtigungen innerhalb der Stichprobe. Diese Häufigkeiten sind speziell für die Untersuchung von H₄ und H₅ von Bedeutung. Da diese Frage als Multiple Choice ausgelegt war und eine Person angegeben hat, sowohl eine Beeinträchtigung des Hör- als auch des Sehapparats zu haben, wird diese zweimal in der Zählung von N aufgenommen. Somit ergibt sich die Gesamtanzahl von 225 bei weiterhin 224 Probanden.

	Antworten		
	N	Prozent	Prozent der Fälle
Hörbeeinträchtigung	7	3,1%	3,1%
Sehbeeinträchtigung	5	2,2%	2,2%
Keine (der genannten)	197	87,6%	87,9%
Keine Angabe	16	7,1%	7,1%
Gesamt	225	100,0%	100,4%

Tabelle 3: Häufigkeiten sensorischer Beeinträchtigungen

¹⁰⁵ Für die ursprünglich erhobenen Altersgruppen siehe Anhang 2.

7.2 Allgemeine Bewertung (H₁)

Hypothese H₁ dient der Überprüfung der allgemeinen Bewertung der vier Personalisierungsmöglichkeiten aus Sicht der deutschen Bevölkerung und demnach den potenziellen Endnutzern von NGA-Broadcasts deutscher Fernsehanstalten. Aus der erhobenen Stichprobe ergab sich für die Einstellungsskala der Personalisierungsmöglichkeit „Mehrsprachige Übertragung“ ein Median von $Mdn = 5,3$, das arithmetische Mittel lag bei $M = 5,2$ ($SD = 1,5$). Übertragen auf die Grundpopulation ergab sich für den Mittelwert das Konfidenzintervall 95% CI [5,0; 5,4].

Die linke Hälfte von Tabelle 4 enthält die Ergebnisse des angewandten, parametrischen Einstichproben t-Tests. Der beobachtete Mittelwert der Stichprobe ($M = 5,2$; $SD = 1,5$) lag für „Mehrsprachige Übertragung“ demnach signifikant höher als der hypothetische Vergleichswert von 4 ($t(223) = 12,03$; $p < 0,001$). Die Effektstärke beläuft sich auf $d = \frac{\bar{x} - \bar{x}_0}{s} = \frac{5,2 - 4}{1,5} = 0,8$ und entspricht damit nach Cohen einem großen Effekt.¹⁰⁶ Die rechte Hälfte von Tabelle 4 zeigt die Ergebnisse des zur Kontrolle des parametrischen Tests angewandten Einstichproben Wilcoxon-Tests. Verglichen mit dem hypothetischen Referenzmedian von 4 hatte die Stichprobe einen signifikant höheren Median von $Mdn = 5,3$ ($Z = 9,16$; $p < 0,001$). Die Effektstärke beträgt $r = \frac{Z}{\sqrt{N}} = \frac{9,16}{\sqrt{224}} = 0,61$, was nach Cohen einem großen Effekt entspricht.¹⁰⁷

Beide Tests ergeben eine signifikant positive Einstellung der Personalisierungsmöglichkeit „Mehrsprachige Übertragung“ durch die Probanden. Die Effektgröße ist dabei für beide Tests gleichermaßen als groß zu bewerten. Die Nullhypothese wird für diese Funktion verworfen und die Hypothese H₁ akzeptiert.

Testwert = 4				Teststatistik	18164,500
Mittlere Differenz	T	df	Signifikanz	Standardfehler	841,958
			Zweiseitiges p	Standardisierte Teststatistik	9,157
1,20313	12,027	223	0,000	Asymptotische Sig. (zweiseitiger Test)	0,000

Tabelle 4: t- (links) und Wilcoxon-Test (rechts) für „Mehrsprachige Übertragung“

Für die Einstellungsskala der Personalisierungsmöglichkeit „Anpassung von Lautstärkelevels“ ergab sich ein Median von $Mdn = 5,6$ sowie ein Mittelwert von $M = 5,42$ (SD

¹⁰⁶ Vgl. Cohen 1992, S. 157.

¹⁰⁷ Vgl. ebd.

= 1,23). Das Konfidenzintervall des Mittelwerts für die Gesamtpopulation umfasste 95% CI [5,26; 5,58].

Tabelle 5 zeigt die Ergebnisse des durchgeführten Einstichproben t-Tests für „Anpassung von Lautstärkelevels“. Es konnte ein signifikant höherer Mittelwert ($M = 5,42$; $SD = 1,23$) gegenüber dem hypothetischen Referenzwert von 4 beobachtet werden ($t(223) = 17,32$; $p < 0,001$). Der Effekt ist mit einer Effektstärke von $d = 1,15$ nach Cohen groß. Tabelle 5 enthält in der rechten Hälfte ebenso die Ergebnisse des Einstichproben Wilcoxon-Tests zur Überprüfung der Ergebnisse des parametrischen Tests. Verglichen mit dem hypothetischen Referenzmedian von 4 hatte die Stichprobe einen signifikant höheren Median von $Mdn = 5,6$ ($Z = 10,99$; $p < 0,001$). Die Effektstärke beträgt $r = 0,73$, was einem großen Effekt nach Cohen entspricht.

Beide Tests ergaben eine signifikant positivere Einstellung mit großer Effektstärke. Die Nullhypothese wird für die Personalisierungsmöglichkeit „Anpassung von Lautstärkelevels“ verworfen und die Hypothese H_1 akzeptiert.

Testwert = 4				Teststatistik	20392,500
Mittlere Differenz	T	df	Signifikanz	Standardfehler	866,983
			Zweiseitiges p	Standardisierte Teststatistik	10,986
1,41815	17,324	223	0,000	Asymptotische Sig. (zweiseitiger Test)	0,000

Tabelle 5: t- (links) und Wilcoxon-Test (rechts) für "Anpassung von Lautstärkelevels"

Der Median für die Personalisierungsmöglichkeit „Audiodeskription“ betrug $Mdn = 6$, das arithmetische Mittel lag bei $M = 5,78$ ($SD = 1,35$). Das Konfidenzintervall für den Mittelwert der Population belief sich auf 95% CI [5,6; 5,95].

Das Ergebnis des Einstichproben t-Tests, dargestellt in Tabelle 6, ergab für „Audiodeskription“ einen signifikant höheren Mittelwert ($M = 5,78$; $SD = 1,35$) verglichen mit dem hypothetischen Referenzwert von 4 ($t(223) = 19,7$; $p < 0,001$). Die Effektstärke beträgt $d = 1,32$, es liegt ein großer Effekt vor. Das Ergebnis des zur Überprüfung des parametrischen Tests durchgeführten Einstichproben Wilcoxon-Test bestätigte einen signifikant höheren Median der Stichprobe von $Mdn = 6$ im Vergleich zum hypothetischen Referenzmedian von 4 ($Z = 11,44$; $p < 0,001$). Der Effekt ist groß bei einer Effektstärke von $r = 0,76$.

Auch für die Personalisierungsmöglichkeit „Audiodeskription“ liegen signifikante Ergebnisse beider Testverfahren mit großen Effektstärken vor. Die Nullhypothese wird verworfen und die Hypothese H_1 akzeptiert.

Testwert = 4				Teststatistik	21257,000
Mittlere Differenz	T	df	Signifikanz	Standardfehler	880,723
			Zweiseitiges p	Standardisierte Teststatistik	11,438
1,77530	19,699	223	0,000	Asymptotische Sig. (zweiseitiger Test)	0,000

Tabelle 6: t- (links) und Wilcoxon-Test (rechts) für "Audiodeskription"

Die Personalisierungsmöglichkeit „Positionierbare Objekte“ wurde von einem Probanden übersprungen, weshalb sich die Anzahl der bearbeiteten Fälle auf 223 beschränkt. Der Median der Stichprobe betrug $Mdn = 4$, der Mittelwert lag bei $M = 4,07$ ($SD = 1,67$). Das ermittelte Konfidenzintervall für den Mittelwert in der Population war 95% CI [3,85; 4,29].

Der Einstichproben t-Test ergab ein nicht signifikantes Ergebnis für den beobachteten Mittelwert ($M = 4,07$; $SD = 1,67$) verglichen mit dem hypothetischen Referenzwert von 4 ($t(222) = 0,64$; $p = 0,52$). Auch der Einstichproben Wilcoxon-Test fiel nicht signifikant für den beobachteten Median von $Mdn = 4$ gegenüber dem hypothetischen Referenzmedian von 4 aus ($Z = 0,44$; $p = 0,66$).

Auf Grundlage der Ergebnisse aus den angewandten Testverfahren kann die Nullhypothese für die Personalisierungsmöglichkeit „Positionierbare Objekte“ nicht abgelehnt werden und wird beibehalten.

Testwert = 4				Teststatistik	9699,500
Mittlere Differenz	T	df	Signifikanz	Standardfehler	774,661
			Zweiseitiges p	Standardisierte Teststatistik	0,438
0,07175	0,642	222	0,522	Asymptotische Sig. (zweiseitiger Test)	0,662

Tabelle 7: t- (links) und Wilcoxon-Test (rechts) für "Positionierbare Objekte"

7.3 Abhängigkeit von Alter und Geschlecht (H_2 & H_3)

Dieses Unterkapitel befasst sich mit den Bewertungen der Personalisierungsmöglichkeiten in Abhängigkeit von den demografischen Merkmalen Alter und Geschlecht. Hypothese H_2 behandelt hierbei zunächst die Abhängigkeit der Bewertungen vom Alter, anschließend befasst sich Hypothese H_3 mit der Abhängigkeit der Bewertungen vom Geschlecht.

Abbildung 10 stellt die Bewertungen unterteilt nach den drei in Kapitel 7.1 definierten Altersgruppen in Form von Boxplots grafisch dar.

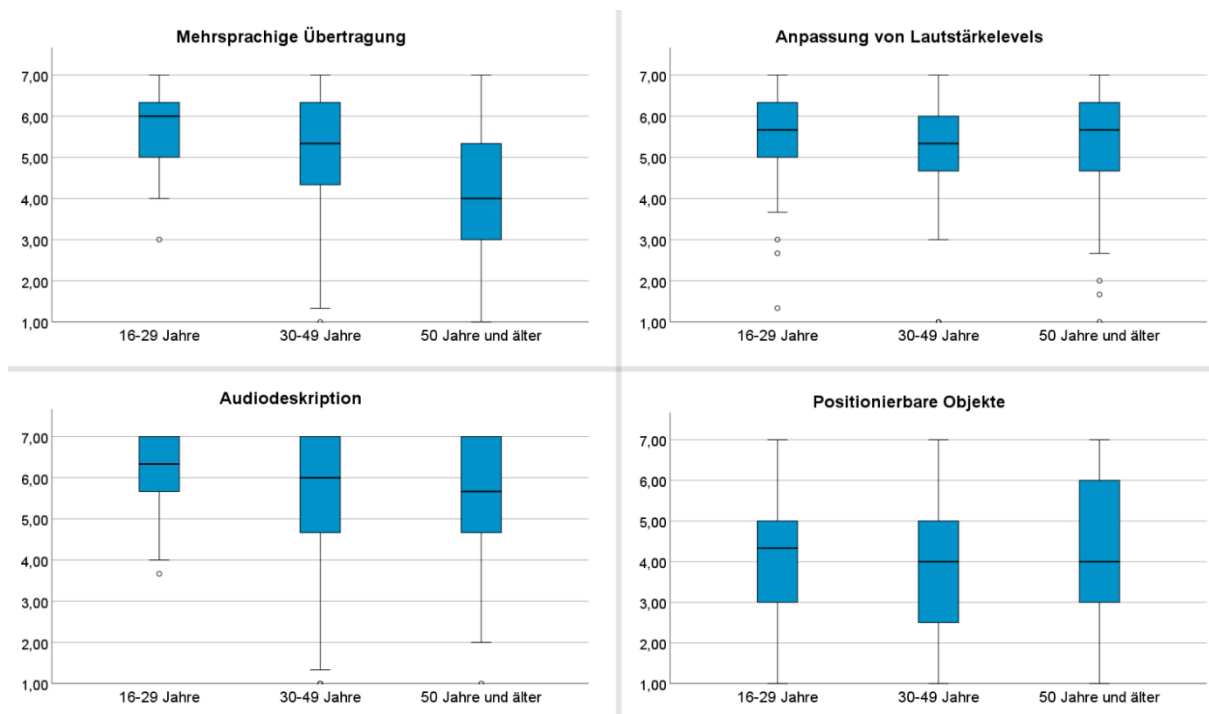


Abbildung 10: Bewertung nach Altersgruppen

Für „Mehrsprachige Übertragung“ ergab sich ein Median von $Mdn = 6$ sowie ein Mittelwert von $M = 5,75$ ($SD = 0,94$) für die Gruppe der 16- bis 29-Jährigen, ein Median von $Mdn = 5,3$ sowie ein Mittelwert von $M = 5,26$ ($SD = 1,46$) für die 30- bis 49-Jährigen und ein Median von $Mdn = 4$ sowie ein Mittelwert von $M = 4,17$ ($SD = 1,79$) für die Menschen im Alter von 50 Jahren und älter. Zur Überprüfung, ob hierbei statistisch signifikante Unterschiede zwischen den Altersgruppen vorliegen, wurde eine einfaktorielle ANOVA berechnet. Gemäß Levene-Test war keine Varianzhomogenität gegeben ($p < 0,001$), weswegen ausschließlich die Ergebnisse der robusteren Welch-ANOVA, die keine Varianzhomogenität voraussetzt, berücksichtigt wurden. Es konnte demnach ein statistisch signifikanter Unterschied der Bewertungen zwischen den Altersgruppen beobachtet werden ($F(2, 110,92) = 18,77$; $p < 0,001$). Eine Post-hoc-Analyse mittels Games-Howell-Test attestierte zwischen allen drei Altersgruppen signifikante Unterschiede: zwischen der jüngsten und der mittleren Gruppe ($p = 0,03$), zwischen der jüngsten und der ältesten Gruppe ($p < 0,001$) sowie zwischen der mittleren und der ältesten Gruppe ($p = 0,001$). Der zur Kontrolle durchgeführte Kruskal-Wallis-Test berichtete ebenso signifikante Unterschiede der Bewertung dieser Personalisierungsmöglichkeit in Abhängigkeit vom Alter ($H(2) = 29,49$; $p < 0,001$). Anschließend durchgeführte Post-hoc-Tests mit Bonferroni-Korrektur zeigten, dass sich die jüngste und die älteste Gruppe signifikant mit mittlerem Effekt ($p < 0,001$; $r = 0,36$) sowie die mittlere und die älteste Gruppe signifikant mit kleinem Effekt ($p = 0,001$; $r = 0,24$) unterscheiden. Der Unterschied zwischen der jüngsten und

der mittleren Gruppe fiel hingegen nicht signifikant aus ($p = 0,157$). Für „Mehrsprachige Übertragung“ wird die Nullhypothese abgelehnt und H_2 angenommen.

Bei der Personalisierungsmöglichkeit „Anpassung von Lautstärkelevels“ lagen Median und Mittelwert der Gruppe der 16- bis 29-Jährigen bei $Mdn = 5,6$ und $M = 5,59$ ($SD = 1,12$), bei der Gruppe der 30- bis 49-Jährigen bei $Mdn = 5,3$ und $M = 5,29$ ($SD = 1,21$) und bei der Gruppe der 50-Jährigen und älter bei $Mdn = 5,6$ und $M = 5,33$ ($SD = 1,4$). Der Levene-Test bestätigte Varianzhomogenität ($p = 0,293$), weshalb die Ergebnisse der normalen einfaktoriellen ANOVA verwendet werden konnten. Die ANOVA fiel hierbei nicht signifikant aus ($F(2, 221) = 1,46$; $p = 0,234$). Der Kruskal-Wallis-Test bestätigte dieses nicht signifikante Ergebnis der ANOVA ($H(2) = 2,88$; $p = 0,24$). Da kein signifikanter Unterschied der Bewertung der Funktion abhängig vom Alter attestiert werden konnte, wird die Nullhypothese beibehalten.

Die Mediane und Mittelwerte der Personalisierungsmöglichkeit „Audiodeskription“ lagen bei $Mdn = 6,3$ und $M = 6,21$ ($SD = 0,88$) bei der jüngsten Gruppe der 16- bis 29-Jährigen, für die mittlere Gruppe der 30- bis 49-Jährigen bei $Mdn = 6$ und $M = 5,49$ ($SD = 1,5$) sowie bei der ältesten Gruppe der 50-Jährigen und älter bei $Mdn = 5,6$ und $M = 5,47$ ($SD = 1,58$). Zwischen diesen Gruppen bestand erneut keine Varianzhomogenität (Levene-Test $p < 0,001$), weswegen die Ergebnisse der Welch-ANOVA betrachtet wurden. Diese bezeugte signifikante Unterschiede in der Bewertung abhängig vom Alter ($F(2, 111,06) = 10,13$; $p < 0,001$). Die Post-hoc-Analyse durch einen Games-Howell-Test ergab signifikante Unterschiede zwischen der jüngsten und der mittleren ($p < 0,001$) sowie zwischen der jüngsten und der ältesten Gruppe ($p = 0,007$). Zwischen der mittleren und der ältesten Gruppe konnte hingegen kein statistisch signifikantes Ergebnis festgestellt werden ($p = 0,996$). Der zur Kontrolle ergänzte Kruskal-Wallis-Test stellte ebenfalls signifikante Unterschiede in der Bewertung der Funktion abhängig vom Alter fest ($H(2) = 12,09$; $p = 0,002$). Post-hoc-Tests mit Bonferroni-Korrektur attestierten signifikante Unterschiede zwischen den Gruppen „16-29 Jahre“ und „30-49 Jahre“ mit kleinem Effekt ($p = 0,005$; $r = 0,21$) sowie zwischen den Gruppen „16-29 Jahre“ und „50 Jahre und älter“ mit ebenso kleinem Effekt ($p = 0,024$; $r = 0,18$). Die Nullhypothese wird verworfen und H_2 angenommen.

Für die Personalisierungsmöglichkeit „Positionierbare Objekte“ lagen Median und Mittelwert der jüngsten Gruppe bei $Mdn = 4,3$ sowie $M = 4,15$ ($SD = 1,58$), die der mittleren Gruppe der 30- bis 49-Jährigen bei $Mdn = 4,3$ sowie $M = 3,84$ ($SD = 1,67$) und die der ältesten Gruppe der

50 Jahre alten und älteren bei $Mdn = 4$ sowie $M = 4,29$ ($SD = 1,81$). Es bestand Varianzhomogenität (Levene-Test $p = 0,291$). Per einfaktorieller ANOVA konnten keine signifikanten Unterschiede festgestellt werden ($F(2, 220) = 1,36$; $p = 0,259$). Auch durch den Kruskal-Wallis-Test konnte kein signifikantes Ergebnis beobachtet werden ($H(2) = 1,77$; $p = 0,41$). Demnach wird für diese Personalisierungsfunktion die Nullhypothese beibehalten.

Die nachfolgend behandelte Hypothese H_3 dient der Überprüfung der Abhängigkeit der Bewertungen vom Geschlecht. Die beiden Probanden, die zu ihrem Geschlecht keine Angabe machen wollten, wurden für die Auswertung ausgeschlossen, die Unterteilung erfolgte demnach nur nach weiblichem und männlichem Geschlecht, wie in Abbildung 11 dargestellt.

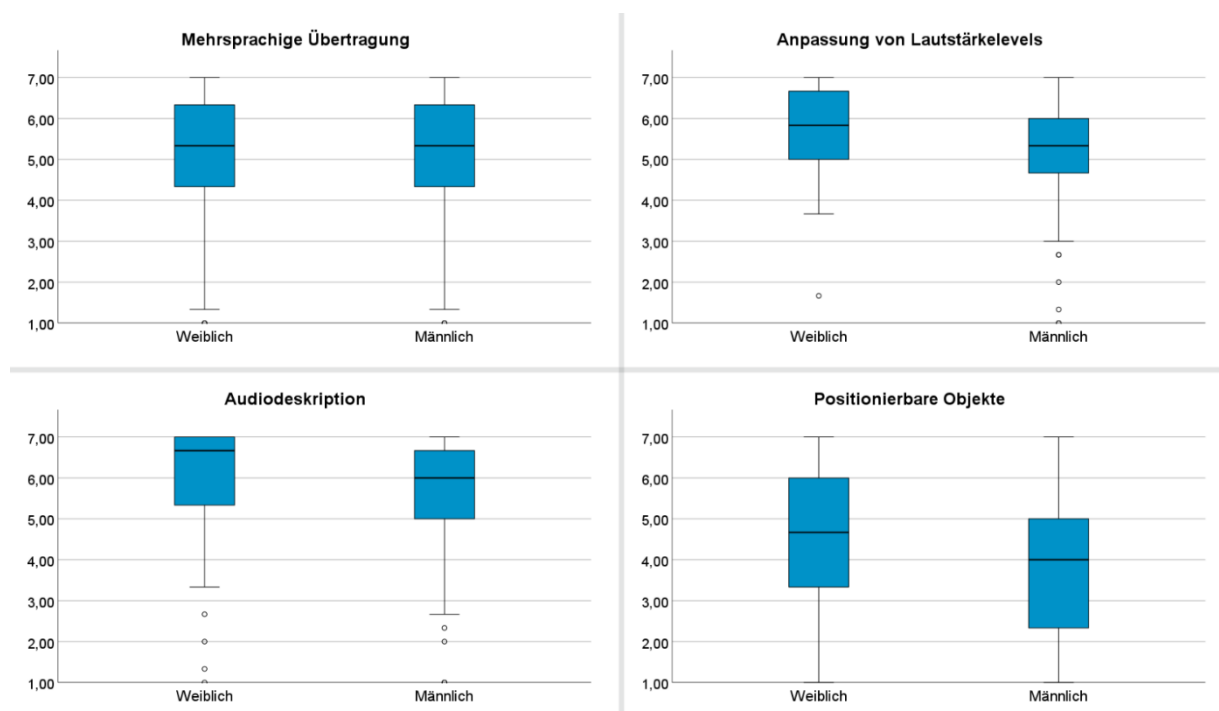


Abbildung 11: Bewertung nach Geschlecht

Die Bewertung der Personalisierungsmöglichkeit der mehrsprachigen Übertragung ergab für die Frauen einen Median von $Mdn = 5,3$ sowie ein arithmetisches Mittel von $M = 5,23$ ($SD = 1,45$). Für die Männer ergab sich ein Median von ebenso $Mdn = 5,3$, der Mittelwert betrug $M = 5,22$ ($SD = 1,5$). Ein t-Test für zwei unabhängige Stichproben konnte keinen signifikanten Unterschied zwischen den beiden Gruppen feststellen ($t(220) = 0,04$; $p = 0,965$). Dieses Ergebnis konnte durch einen nicht-parametrischen Mann-Whitney-U-Test bestätigt werden ($U = 6134$; $Z = 0,18$; $p = 0,855$). Für „Mehrsprachige Übertragung“ kann die zu H_3 zugehörige Nullhypothese daher nicht verworfen werden.

Die Personalisierungsmöglichkeit „Anpassung von Lautstärkelevels“ bewerteten Frauen bei einem Median von $Mdn = 5,8\bar{3}$ sowie einem Mittelwert von $M = 5,7$ ($SD = 1,07$). Der Median bei den Männern lag bei $Mdn = 5,3$, das arithmetische Mittel bei $M = 5,24$ ($SD = 1,25$). Ein t-Test suggerierte hierfür einen signifikanten Unterschied zwischen beiden Geschlechtern ($t(220) = 2,86$; $p = 0,005$) bei kleiner Effektstärke von $d = 0,39$. Dieses Ergebnis konnte durch einen Mann-Whitney-U-Test, der ebenfalls signifikant bei ebenso kleiner Effektstärke ausfiel ($U = 4800$; $Z = -2,65$; $p = 0,008$; $r = 0,18$), bestätigt werden. Die Nullhypothese wird demnach für diese Personalisierungsmöglichkeit abgelehnt und H_3 angenommen.

Frauen bewerteten „Audiodeskription“ bei einem Median von $Mdn = 6,6$ sowie einem Mittelwert von $M = 5,99$ ($SD = 1,37$). Die Bewertung seitens der Männer lag bei dieser Funktion bei einem Median von $Mdn = 6$ sowie einem Mittelwert von $M = 5,62$ ($SD = 1,31$). Der angewandte t-Test bezeichnete diesen Unterschied als signifikant mit kleinem Effekt ($t(220) = 2,03$; $p = 0,044$; $d = 0,27$). Der kontrollierende Mann-Whitney-U-Test ergab ebenso einen signifikanten Unterschied zwischen den Geschlechtern bei kleiner Effektgröße ($U = 4764$; $Z = -2,78$; $p = 0,006$; $r = 0,19$). Für die Personalisierungsmöglichkeit „Audiodeskription“ wird dementsprechend die Nullhypothese verworfen und H_3 angenommen.

Die Bewertung der letzten Funktion der positionierbaren Objekte fiel bei den Frauen mit einem Median von $Mdn = 4,6$ sowie einem arithmetischen Mittel von $M = 4,58$ ($SD = 1,64$) aus. Die Männer bewerteten diese Funktion bei einem Median von $Mdn = 4$ sowie einem Mittelwert von $M = 3,72$ ($SD = 1,59$). Dieser Unterschied konnte durch die Berechnung des t-Tests als signifikant mit mittlerer Effektstärke erkannt werden ($t(219) = 3,91$; $p < 0,001$; $d = 0,53$). Der nicht-parametrische Kontroll-Mann-Whitney-U-Test fiel ebenfalls signifikant aus ($U = 4356,5$; $Z = -3,47$; $p < 0,001$). Die Effektstärke entsprach mit $r = 0,23$ anders als beim t-Test nur einem kleinen Effekt. Die Nullhypothese wird für „Positionierbare Objekte“ ebenfalls abgelehnt und H_3 angenommen.

7.4 Abhängigkeit von sensorischen Beeinträchtigungen (H_4 & H_5)

Dieses Unterkapitel befasst sich, den Ergebnisteil dieser Arbeit abschließend, mit der Abhängigkeit der Bewertung ausgewählter Funktionen von dem Vorhandensein von sensorischen Beeinträchtigungen, die diese Bewertungen beeinflussen könnten. Die zugehörigen Ergebnisse sind in Abbildung 12 grafisch dargestellt.

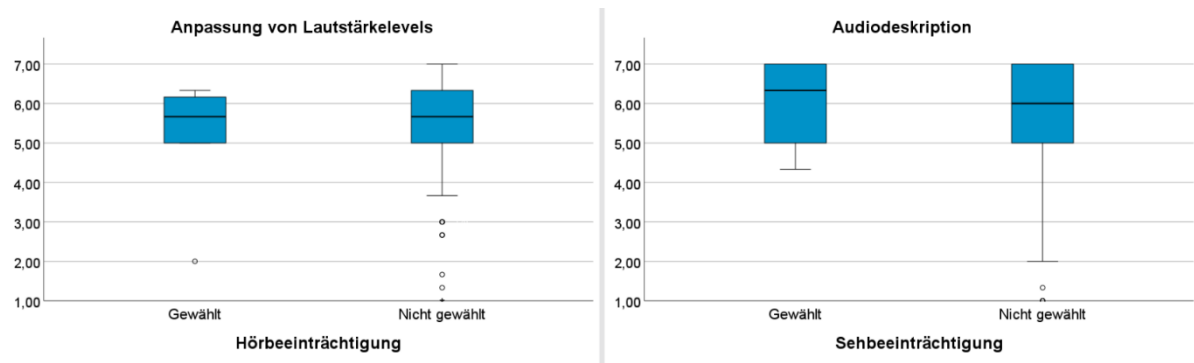


Abbildung 12: Abhängigkeiten von sensorischen Beeinträchtigungen

Bei der Hypothese H_4 handelt es sich zunächst um die Untersuchung der Abhängigkeit der Bewertung der Funktion „Anpassung von Lautstärkelevels“ von einer vorliegenden Beeinträchtigung des Hörapparats. Ein berechneter t-Test konnte keinen signifikanten Unterschied zwischen den Bewertungen der Personen mit ($M = 5,19$; $SD = 1,51$) und der Personen ohne Hörbeeinträchtigung ($M = 5,43$; $SD = 1,22$) feststellen ($t(222) = 0,5$; $p = 0,618$). Dieses Ergebnis konnte durch einen nicht-parametrischen Mann-Whitney-U-Test, der ebenso keinen signifikanten Unterschied zwischen den Bewertungen der Personen mit Hörbeeinträchtigung ($Mdn = 5,6$) und denjenigen ohne ($Mdn = 5,6$) ausmachen konnte, bestätigt werden ($U = 737,5$, $Z = -0,13$; $p = 0,896$). Demnach wird die Nullhypothese nicht verworfen und H_4 nicht angenommen.

H_5 behandelt zuletzt die Abhängigkeit der Bewertung der Personalisierungsmöglichkeit „Audiodeskription“ vom Vorliegen einer Beeinträchtigung des Sehapparats. Der berechnete t-Test konnte auch hier keinen signifikanten Unterschied der Bewertungen zwischen den Personen mit ($M = 5,93$; $SD = 1,21$) und den Personen ohne Beeinträchtigung des Sehapparats ($M = 5,77$; $SD = 1,35$) ausmachen ($t(222) = -0,26$; $p = 0,792$). Der kontrollierende Mann-Whitney-U-Test bestätigte dieses nicht signifikante Ergebnis zwischen den beiden Gruppen mit ($Mdn = 6,3$) und ohne Beeinträchtigung ($Mdn = 6$) des Sehsinns ($U = 573,5$; $Z = 0,19$; $p = 0,852$). Demnach kann auch die Nullhypothese von H_5 nicht verworfen werden und wird beibehalten.

8. Diskussion

Dieser Teil der Arbeit soll anschließend zunächst die Forschungsergebnisse aus Kapitel 7 zusammenfassen, bevor kurz die Ergebnisse aus den Fragebogenabschnitten thematisiert und vorgestellt werden, die für die Ergründung der Forschungshypothesen und -fragen keine bzw. nur eine sekundäre Rolle spielen. Daraufgehend werden die Ergebnisse interpretiert und eingeordnet, bevor abschließend noch einige Einschränkungen dieser Forschung erwähnt werden sollen.

8.1 Zusammenfassung der Ergebnisse

Ziel dieser Forschung war es, zu ergründen, wie die Endnutzer von zukünftigen NGA-Broadcasts die gängigsten Personalisierungsmöglichkeiten, die durch den in NGA enthaltenen objektbasierten Ansatz von Audiorepräsentation technisch realisierbar sind, bewerten und bei Einführung im regelmäßigen Programm deutscher Fernsehanstalten annehmen könnten. Hierzu wurde eine quantitative Umfrage innerhalb der deutschen Bevölkerung ausgetragen, wobei eine Stichprobengröße von $n = 224$ erreicht werden konnte. Im Vorfeld wurde erwartet, dass die zusätzliche Personalisierung und Individualisierung positiv von den Endnutzern evaluiert würde. Diese Erwartungen konnten durch weitestgehend positive Bewertungen der Personalisierungsmöglichkeiten bestätigt werden. So konnte die zu H_1 zugehörige Nullhypothese, die stellvertretend für eine neutrale Einstellung gegenüber der jeweiligen Personalisierungsmöglichkeit stand, für drei der vier Fälle, genauer gesagt „Mehrsprachige Übertragung“, „Anpassung von Lautstärkelevels“ und „Audiodeskription“, anhand signifikant positiver Bewertungen mit jeweils großen Effektstärken verworfen werden. Lediglich für „Positionierbare Objekte“ konnte keine statistisch signifikant positive Bewertung festgestellt werden, allerdings fiel die Bewertung ebenso nicht signifikant negativ aus. Demnach wurde die Nullhypothese beibehalten, also eine neutrale Einstellung gegenüber positionierbaren Objekten attestiert. Die statistischen Signifikanzen wurden hierbei mittels Einstichproben t-Tests ermittelt. Die resultierenden Ergebnisse aus diesem parametrischen Testverfahren wurden anschließend durch nicht-parametrische Einstichproben Wilcoxon-Tests verifiziert.

Weiterhin sollte überprüft werden, ob diese Bewertungen allgemeingültig für die deutsche Zuschauerschaft sind oder ob bedeutende Unterschiede anhand demografischer Merkmale

wie Alter, Geschlecht oder dem Vorhandensein von sensorischen Beeinträchtigungen auszumachen sind. Um auf Abhängigkeit vom Alter zu testen, wurden als Testverfahren die parametrische einfaktorielle ANOVA sowie der nicht-parametrische Kruskal-Wallis-Test eingesetzt. Bei signifikanten Ergebnissen kam es weiterhin zu Post-hoc-Analysen, um zu überprüfen, zwischen welchen Altersgruppen signifikante Ergebnisse vorliegen. Für „Mehrsprachige Übertragung“ konnte eine Abhängigkeit vom Alter festgestellt werden. Demnach fiel die Bewertung von Menschen zwischen 16 und 29 Jahren sowie die der 30- bis 49-Jährigen bei beiden Testverfahren signifikant positiver als die der 50-Jährigen und Älteren mit mittlerer respektive kleiner Effektgröße aus. Das parametrische Testverfahren attestierte ebenso einen signifikanten Unterschied zwischen der jüngsten und der mittleren Altersgruppe, der zur Kontrolle eingesetzte Kruskal-Wallis-Test fiel hierfür allerdings nicht signifikant aus. Demnach kann lediglich ein höchstens schwach signifikanter Unterschied zugestanden werden. Für „Anpassung von Lautstärkelevels“ konnte keine signifikante Abhängigkeit der Bewertung vom Alter festgestellt werden. Bei „Audiodeskription“ konnte wiederum eine signifikante Abhängigkeit erkannt werden, wobei die Bewertung der jüngsten Gruppe gegenüber der mittleren sowie ältesten Gruppe jeweils signifikant positiver mit kleiner Effektstärke ausfiel. Für „Positionierbare Objekte“ fielen beide Testverfahren zuletzt erneut nicht signifikant aus, es konnte also keine Abhängigkeit vom Alter festgestellt werden. Die zu H_2 zugehörige Nullhypothese wurde demnach für „Mehrsprachige Übertragung“ und „Audiodeskription“ abgelehnt und blieb für „Anpassung von Lautstärkelevels“ und „Positionierbare Objekte“ angenommen.

Die Überprüfung der Abhängigkeit vom Geschlecht fand mittels parametrischen t-Tests sowie zur Kontrolle aufgebrauchten nicht-parametrischen Mann-Whitney-U-Tests statt. Bei „Mehrsprachige Übertragung“ konnte durch beide Testverfahren kein signifikanter Unterschied zwischen Frauen und Männern festgestellt werden. Für „Anpassung von Lautstärkelevels“ sowie „Audiodeskription“ konnte hingegen eine signifikant positivere Bewertung mit kleiner Effektstärke durch Frauen identifiziert werden. Zuletzt konnte ebenso eine signifikant positivere Bewertung von „Positionierbare Objekte“ durch Frauen gegenüber Männern bei kleinem (t-Test) bis mittleren Effekt (Mann-Whitney-U-Test) beobachtet werden. Die zu H_3 zugehörige Nullhypothese wurde daher in drei von vier Fällen, aber mit überwiegend kleinen Effektstärken verworfen.

Letzte zur Überprüfung ausstehende demografische Merkmale waren sensorische Beeinträchtigungen des Hör- sowie des Sehsinns. Hierzu wurden ebenfalls t-Tests sowie Mann-Whitney-U-Tests ausgeführt, um mögliche Abhängigkeiten zu überprüfen. Entgegen den Erwartungen ergab hierbei weder das Vorhandensein einer Beeinträchtigung des Hörapparats eine signifikant andere Bewertung von „Anpassung von Lautstärkelevels“ noch das Vorhandensein einer Beeinträchtigung des Sehapparats eine signifikant andere Bewertung der Funktion „Audiodeskription“ als bei Menschen ohne solche Beeinträchtigungen.

Weitere, nur als sekundär zu bewertende Ergebnisse, die in Kapitel 7 nicht tiefergehend analysiert wurden, umfassten die genannten Sonderfälle der Funktionen „Mehrsprachige Übertragung“ und „Positionierbare Objekte“ sowie die Fragen, wie oft Menschen in Deutschland Angebote deutscher Fernsehsender schauen, welche Systeme für die Wiedergabe des Tons an stationären (TV, Desktop-PC) und mobilen Geräten (Smartphones, Tablets, Notebooks) verwendet werden und ob der Einsatz der Personalisierungsmöglichkeiten bewirken könnte, dass mehr Fernsehen geschaut wird. Bei Beantwortung mit „Ja“ wurde außerdem noch erhoben, welche der Funktionen dazu führen könnten. Der Sonderfall mehrerer verschiedener Kommentatoren nach Teamzugehörigkeit bei Sportübertragungen wurde neutral mit einem Median von $Mdn = 4$ bewertet, das arithmetische Mittel lag mit $M = 3,86$ ($SD = 1,81$) leicht unter neutraler Bewertung. Demnach scheint eine Umsetzung dieser Funktion nur begrenzt empfehlenswert. Dementgegen wurde der zweite Sonderfall von positionierbarer Audiodeskription im Median mit $Mdn = 6$ sowie einem Mittelwert von $M = 5,21$ ($SD = 1,6$) bewertet, also bedeutend positiv. Bei Einsatz von Audiodeskription bei objektbasierter bzw. hybrider Herangehensweise scheint die Ermöglichung der freien Positionierung im Metadaten-Authoring daher empfehlenswert. Es konnte weiterhin ermittelt werden, dass Menschen im Median ein bis drei Mal pro Woche Angebote deutscher Fernsehsender konsumieren. Bei stationären Geräten finden die in den Geräten verbauten Lautsprecher dabei die höchste Anwendung mit 99 Angaben der Befragten. Stereoanlagen mit 47, Surroundanlagen mit 29 und Soundbars mit 28 Angaben folgten im Ranking darauf. 13 Personen gaben an, Inhalte per Kopfhörer zu hören, immersive Lautsprecherkonfigurationen nutzen 9 der Befragten Personen, während 3 Personen angaben, Bluetooth-Lautsprecher für die Wiedergabe zu verwenden. Demnach scheinen die Voraussetzungen an immersive Wiedergabe bei den meisten Personen noch nicht gegeben zu sein, da lediglich die wenigen

immersiven Lautsprecherkonfigurationen sowie Kopfhörer und evtl. einige wenige Soundbars immersive Wiedergabe ermöglichen. Bei Mobilgeräten scheinen Voraussetzungen an immersive Wiedergabe besser erfüllt, da hier zusammengefasst aus kabelgebundenen und drahtlosen insgesamt 129 Angaben zu Kopfhörerverwendung gemacht wurden. Verbaute Lautsprecher lagen mit 108, Bluetooth-Lautsprecher mit 43 und Stereoanlagen mit 2 Angaben dahinter. Dementsprechend scheinen hier mehr Menschen auf immersive Wiedergabe vorbereitet zu sein, vor allem in Form von binauraler Wiedergabe. Bezüglich der Frage, ob die Einführung und Anwendung der genannten Personalisierungsmöglichkeiten bewirken könnte, dass die Befragten mehr Angebote der Fernsehanstalten ansehen, antworteten etwa zwei Drittel mit „Nein“ und ein Drittel mit „Ja“. Von den 76 Personen, die mit „Ja“ antworteten, gaben bei Möglichkeit der Mehrfachnennung 56 Personen an, dass „Mehrsprachige Übertragung“ dies bewirken könnte, „Anpassung von Lautstärkelevels“ kam auf 54, „Audiodeskription“ auf 20 und „Positionierbare Objekte“ auf 29 Nennungen.¹⁰⁸

8.2 Einordnung und Deutung

Mehrsprachige Übertragung: Die allgemein signifikant positive Bewertung von mehrsprachiger Übertragung spiegelt den bereits in einem 2010 veröffentlichten Artikel der Süddeutschen Zeitung beschriebenen Trend wider, dass das Interesse an Originalfassungen von Filmen und Serien anstelle von deutschen Synchronisationen zunehmend steigt.¹⁰⁹ Obgleich das Marktforschungsinstitut GWI für Deutschland weiterhin eine Präferenz von Synchronisation im Vergleich zu Originalfassungen (mit Untertiteln) feststellen konnte¹¹⁰, scheint die Option der eigenen Entscheidung der Wiedergabesprache dennoch hoch angesehen. Da in die Einstellungsskala ebenso die Bewertungen für Nachrichtensendungen und Sportübertragungen mit verschiedenen Wiedergabesprachen eingingen, kann davon ausgegangen werden, dass zusätzlich der Aspekt erhöhter Zugänglichkeit auch für Menschen, die kein Deutsch sprechen, ein treibendes Motiv der positiven Bewertung war. Dennoch soll an dieser Stelle erwähnt sein, dass die Bewertung der mehrsprachigen Übertragung erwartungsgemäß am positivsten für Filme und Serien ausfiel ($M = 5,69$) und zu Nachrichtensendungen ($M = 5,29$) sowie noch weiter zu Sportübertragungen ($M = 4,64$) abfiel.

¹⁰⁸ Für grafische Darstellungen aller sekundärer Ergebnisse siehe Anhang 3.

¹⁰⁹ Vgl. Schmieder 2010.

¹¹⁰ Vgl. Hedges 2022.

Die signifikant bessere Bewertung mehrsprachiger Broadcasts durch die jüngeren Altersgruppen gegenüber der ältesten Altersgruppe korreliert mit besseren Fremdsprachenkenntnissen der jüngeren Generationen. So gaben beispielsweise nur 23% aller in einer Studie des IfD befragten 14-19-Jährigen an, geringe oder keine Kenntnisse der englischen Sprache zu haben, während dieser Anteil für jede weitere Altersgruppe bis zu den 70-Jährigen und Älteren auf etwa 80% anstieg.¹¹¹ Vergleichbare Statistiken für andere Fremdsprachen konnten nicht ausfindig gemacht werden, allerdings ist die Prävalenz von englischsprachigen Inhalten – vor allem bei Filmen und Serien sowie bei Sportübertragungen – ungleich höher als jene anderssprachiger Inhalte. Zuletzt kann aufgrund der statistisch nicht signifikanten Abhängigkeit vom Geschlecht davon ausgegangen werden, dass die Funktion bei flächendeckender Einführung gleichermaßen positiv sowohl bei Frauen als auch bei Männern angenommen würde.

Anpassung von Lautstärkelevels: Die signifikant positive Bewertung von anpassbaren Lautstärken einzelner Elemente im Gesamtmix folgt den qualitativ erhobenen Ergebnissen einer von der BBC ausgetragenen Studie mit einer kleinen Stichprobengröße von n = 11. Auch hier fand diese Funktionalität hohen Zuspruch und wurde als eine zentrale Verbesserung gegenüber herkömmlichen Broadcasts bezeichnet.¹¹² Die präsentierten Ergebnisse dieser Studie suggerieren eine ebenbürtig lobende Einstellung der deutschen Bevölkerung. Die ausbleibende signifikante Abhängigkeit vom Alter sowie die zwar signifikante, aber nur mit einem kleinen Effekt einhergehende positivere Bewertung durch Frauen gegenüber Männern und die nicht nachweisbare Abhängigkeit von einer vorliegenden Hörbeeinträchtigung deutet auf eine durchweg positive Einstellung gegenüber der Funktion weitgehend unabhängig von demografischen Merkmalen und damit eine sinnvolle Funktion für alle hin. Dies folgt dem Eindruck, der auf Grundlage diverser Leserbriefe an die Süddeutsche Zeitung entsteht, in denen Beschwerden über die Sprachunverständlichkeit aktueller, traditionell kanalbasiert zur Verfügung gestellter Programme von Frauen, Männern und sogar ganzen Familien aller Altersklassen und mit und ohne Hörbeeinträchtigungen und -hilfen geäußert werden.¹¹³

Audiodeskription: Die im Vergleich zwischen den behandelten Personalisierungsmöglichkeiten positivste Bewertung wurde der Audiodeskription zuteil. Verbunden mit dem im vorherigen

¹¹¹ Vgl. Institut für Demoskopie Allensbach (IfD) 2022b.

¹¹² Vgl. Pardoe et al. 2020, S. 5–6.

¹¹³ Vgl. Musik zu laut, Dialoge genuschelt 2018.

Kapitel 8.1 vorgestellten sekundären Ergebnis, dass diese Funktion den wenigsten Einfluss auf ein mögliches Mehrschauen der Konsumenten hat, scheint möglich, dass die Bewertung mit einer gewissen sozialen Erwünschtheit bzw. hoher Empathie für die Menschen, die für ihr Erleben von Broadcasts auf Audiodeskriptionen angewiesen sind, einhergehen könnte. Da die Befragung anonym und online durchgeführt wurde und somit keine weiteren Personen von außen Einfluss nehmen konnten, sollte von Empathie gesprochen werden. Die signifikant positivere Bewertung durch die jüngste gegenüber den anderen beiden Altersgruppen wie auch die ebenso signifikant positivere Bewertung durch Frauen gegenüber Männern gingen allesamt mit einem jeweils kleinen Effekt einher. Aufgrund der dennoch durchweg über alle Gruppen sehr positiven Evaluierungen kann weiterhin von einer weitestgehend gleichermaßen positiven Einstellung der Audiodeskription gegenüber gesprochen werden. Die erwartete, aber ausgebliebene positivere Bewertung durch Menschen mit Sehbeeinträchtigungen, die möglicherweise auf Audiodeskriptionen für durchgängiges Verständnis der gezeigten Programme angewiesen sind, könnte ebenso auf die möglicherweise durch große Empathie unterstützte Beantwortung zurückzuführen sein.

Positionierbare Objekte: Ähnlich wie die wenigen Befragten aus der BBC-Studie von Pardoe et al. scheint auch die deutsche Bevölkerung dieser Personalisierungsmöglichkeit gegenüberzustehen. Dort wurde sie sinngemäß als technisches Novum bzw. als „Spielerei“ ohne echten Mehrwert bezeichnet.¹¹⁴ Die neutrale Bewertung innerhalb dieser vorliegenden Studie scheint damit einherzugehen. Lediglich der Sonderfall positionierbarer Audiodeskriptionen kam bei Pardoe et al. einiger Zuspruch zu¹¹⁵, der sich ebenso in den sekundären Ergebnissen dieser Befragung widerspiegelt. Dabei ist diese Neutralität unabhängig vom Alter für alle Altersklassen sehr ähnlich. Frauen bewerten die Funktion allerdings signifikant positiver als Männer. Da hierbei nicht ausschließlich ein kleiner, sondern je nach Testverfahren ein kleiner bis mittlerer Effekt vorlag, soll auf diesen Unterschied explizit hingewiesen werden. Mögliche Erklärungen für diesen Unterschied oder Hinweise, worin Gründe für diesen signifikant ausfallenden Unterschied zwischen Männern und Frauen liegen könnten, konnten im Rahmen der durchgeführten Literaturstudie allerdings nicht ausgemacht werden.

¹¹⁴ Vgl. Pardoe et al. 2020, S. 5.

¹¹⁵ Vgl. ebd.

Die Bewertung der einzigen der vier Personalisierungsmöglichkeiten, die vollständig neu ist, bei aktuell praktizierten, kanalbasierten Workflows also nicht ermöglicht werden kann und erst durch den mindestens anteilig objektbasierten Ansatz technisch realisierbar wird, fiel demnach relativ zu den anderen drei Funktionen am negativsten aus. Dennoch bietet der objektbasierte Ansatz auch in Relation zu den drei wesentlich positiv evaluierten Funktionen immense Vorteile, da alle behandelten Funktionen auch bei SDI-basierten Workflows mit nur 16 embedded audio Kanälen vollständig simultan und mit erheblichen Einsparungen der benötigten Datenraten gegenüber ausschließlich kanalbasierten Workflows ermöglicht werden können. Ebenso kann ein insgesamt höheres Maß an Personalisierung realisiert werden, da nicht nur zwischen verschiedenen festen Gesamtmischungen gewählt, sondern innerhalb der im Authoring bestimmten Begrenzungen weitergehende Personalisierungen definiert werden können. So kann beispielsweise nicht nur eine andere Wiedergabesprache gewählt, sondern gleichzeitig ermöglicht werden, die Lautstärke dieses Sprachobjekts zusätzlich individuell anzupassen. Für eine annähernd gleiche, aber dennoch weniger tiefgreifende Funktionalität bräuchte es bei kanalbasierter Audiorepräsentation mit aktuellen Methoden zwei Gesamtmischungen in den zwei verschiedenen Sprachen sowie zwei weitere Gesamtmischungen nach dem „Klare Sprache“-Verfahren der öffentlich-rechtlichen Rundfunkanstalten, um eine Anpassung der Dialoglautstärke zu ermöglichen. Der Umfang dieser Lautstärkeanpassung wäre entgegen dem objektbasierten Ansatz aber vordefiniert, also nicht weiter individuell anpassbar und aufgrund der begrenzten Anzahl an SDI-Audiokanälen oft auch nur mit Stereomischungen realisierbar, da vier Surroundmischungen die Kapazitäten von SDI überschreiten würden. Möglicherweise notwendige Abwägungen, welche der Funktionen angeboten werden können oder sollen, werden durch NGA weniger notwendig oder gar obsolet. Dies gilt speziell auch unter dem Gesichtspunkt, dass kommende immersive Audioinhalte, die bis hin zum 7.1.4-Format übertragen werden könnten, bedeutend mehr einzelne Audiokanäle benötigen als die bisherigen 5.0- bzw. 5.1-Surroundmischungen, wodurch eine effizientere Nutzung der verfügbaren Kapazitäten unumgänglich ist. Dies gilt erst recht, wenn Personalisierung in Zukunft nicht mehr nur den Standardton und maximal zwei zusätzliche Optionen (bspw. Audiodeskription und „Klare Sprache“), sondern noch weitere der behandelten Funktionen gleichzeitig umfassen soll.

8.3 Einschränkungen

An dieser Stelle sollen einige Umstände erwähnt werden, anhand derer die Repräsentativität dieser Studie eingeschränkt ist. Zunächst sei darauf hingewiesen, dass die erhobene Stichprobe keiner Zufallsstichprobe aus der Grundpopulation der deutschen Bevölkerung entspricht, sondern auf einer Basis von Verfügbarkeit und Erreichbarkeit fußt. Demnach ergaben sich, wie in Kapitel 7.1 dargestellt, abweichende Verteilungen demografischer Merkmale innerhalb der Stichprobe gegenüber der Grundpopulation. So sind beispielsweise Männer mit 56% insgesamt leicht überrepräsentiert gegenüber der durch das Statistische Bundesamt hochgerechneten Geschlechterverteilung im Jahr 2023, die in der Grundpopulation bei etwa 50,7% Frauen und 49,3% Männern liegt.¹¹⁶ Dieser Umstand entspringt dabei vor allem der starken Überrepräsentation in der Gruppe der 30- bis 49-Jährigen, in der jüngsten sowie der ältesten Altersgruppe ist hingegen eine weitestgehende Annäherung an die Verteilung in der Grundpopulation gegeben. Weiterhin sind junge Menschen den älteren gegenüber überrepräsentiert, was dem Umstand geschuldet ist, dass überwiegend Menschen aus dem persönlichen Umfeld des Forschenden als Probanden gewonnen werden konnten. Diese Umstände konnten durch die überwiegend nicht signifikant ausfallenden Abhängigkeiten von Geschlecht und Alter weitgehend abgefangen werden, allerdings gibt es ebenso einige wenige Einzelfälle, näher thematisiert im vorangegangenen Kapitel 8.2, bei denen ein Effekt dieser demografischen Merkmale festgestellt werden konnte. Zuletzt sollen die nicht signifikant ausgefallenen Abhängigkeiten von sensorischen Beeinträchtigungen bei den entsprechenden Funktionen keinen Anspruch an allgemeingültige Repräsentativität stellen, sondern lediglich als erster Eindruck dienen. Grund hierfür sind die sehr geringen Stichprobengrößen von $n = 7$ für Menschen mit Hörbeeinträchtigungen respektive $n = 5$ für Menschen mit Sehbeeinträchtigungen innerhalb der Gesamtstichprobe.

¹¹⁶ Vgl. Statistisches Bundesamt (Destatis) 2021, S. 5–6.

9. Fazit

Die vorliegende Studie ging der Frage nach, wie die perspektivischen Konsumenten von zukünftigen NGA-Broadcasts deutscher Fernsehsender die gängigsten Einsatzmöglichkeiten von NGA-Objekten evaluieren und wie der Einsatz dieser Funktionen in Zukunft angenommen werden könnte. Dabei sollte weiterhin abgewogen werden, wie allgemeingefasst eine schlussendliche Antwort auf diese Frage getroffen werden kann bzw. welche Abhängigkeiten von diversen demografischen Merkmalen beobachtbar sind. Zur Ergründung dieser Frage wurde eine quantitative Umfrage innerhalb der deutschen Bevölkerung durchgeführt, die die gängigsten, bereits in zahlreichen Testübertragungen erprobten Personalisierungsmöglichkeiten thematisierte.

Durch die an den Ergebnissen der Erhebung durchgeführten statistischen Hypothesentests konnten stark signifikant positive Einstellungen der deutschen Bevölkerung gegenüber drei der vier behandelten Personalisierungsmöglichkeiten beobachtet werden: Mehrsprachige Übertragung, Anpassung von Lautstärkelevels und Audiodeskription. Die vierte Funktion der positionierbaren Audioobjekte wurde demgegenüber nur neutral bewertet. Abhängigkeiten von den demografischen Merkmalen Alter, Geschlecht und der Existenz von sensorischen Beeinträchtigungen fielen dabei zumeist nicht signifikant aus. In den Fällen, bei denen eine statistische Signifikanz nachgewiesen werden konnte, handelte es sich beständig um kleine Effekte. Lediglich für „Mehrsprachige Übertragung“ ergab sich ein mittlerer Effekt zwischen der jüngsten und der ältesten Altersgruppe und bei „Positionierbare Objekte“ handelte es sich in Abhängigkeit vom Geschlecht je nach Testverfahren nicht ausschließlich um einen kleinen, sondern einen kleinen bis mittleren Effekt. Hierbei bewerteten Frauen besser als Männer.

Aus diesen Ergebnissen konnte geschlossen werden, dass die deutsche Bevölkerung Personalisierung und Individualisierung ihrer Broadcast-Audioerlebnisse ausgesprochen positiv gegenüberzustehen scheint und diese befürwortet. Die positive Bewertung mehrsprachiger Broadcasts konnte auf das zunehmende Interesse, Inhalte in ihrer originalen Fassung schauen zu wollen sowie die erhöhte Zugänglichkeit für Menschen, die kein Deutsch sprechen, zurückgeführt werden. Vermehrte Beschwerden über problematische Sprachverständlichkeit konnten als Erklärung für die ebenso positive Bewertung der individuellen Anpassung von Lautstärken einzelner Elemente ausgemacht werden. Da Audiodeskriptionen als seltenster Grund für eine höhere persönliche Einschaltquote genannt

wurden, konnte darauf geschlossen werden, dass womöglich nicht der persönliche Nutzen für die ausgesprochen positive Bewertung verantwortlich ist, sondern eine hohe Empathie den Menschen gegenüber, die Audiodeskriptionen für ihr persönliches Erlebnis des Programms benötigen. Das schwache Abschließen positionierbarer Audioobjekte konnte zuletzt darauf zurückgeführt werden, dass diese Funktionalität vermutlich als Banalität ohne wirklichen Mehrwert angesehen wird. Bezogen auf die demografischen Unterschiede konnte ermittelt werden, dass mehrsprachige Programme ein jüngeres Publikum begründet durch bessere Fremdsprachenkenntnisse mehr als ältere Menschen ansprechen könnten. Die ansonsten nicht oder nur mit kleinem Effekt signifikant ausfallenden Abhängigkeiten von den erhobenen demografischen Merkmalen lassen ansonsten eine weitestgehend allgemeingültige Annahme der beobachteten Einstellungen über die gesamte Bevölkerung hinweg zu.

Deutschen Rundfunkanstalten ist auf Grundlage der Erkenntnisse dieser Studie zu empfehlen, die auch mittels kanalbasierter Audiorepräsentation bereits technisch umsetzbaren und teilweise bereits praktizierten Personalisierungsmöglichkeiten beizubehalten, dabei allerdings von den Vorzügen hybrider Ansätze, speziell der objektbasierten Repräsentation der übertragenen Audioinhalte, Gebrauch zu machen. Gemeint sei damit, Nutzen aus den geringeren Datenraten sowie den weniger notwendigen Audiokanälen zu ziehen und das Potenzial erhöhter Personalisierung und Barrierefreiheit auszuschöpfen, während gleichzeitig auf immersives Audio umgestellt werden kann. NGA vereint all diese Vorzüge, die den Ergebnissen der Umfrage entsprechend bei einem beachtlichen Anteil von etwa einem Drittel der Menschen eine erhöhte Einschaltquote bedingen könnte. Vom zusätzlichen Aufwand beim Authoring, um Objekte positionierbar zu machen, kann dabei eher abgeraten werden. Allerdings scheinen flächendeckende oder zumindest bedeutend verstärkte, mehrsprachige Angebote sowie bedeutend häufiger angebotene Audiodeskriptionen und die Option, einzelne Elemente in ihrer Lautstärke individuell an die eigenen Bedürfnisse anpassen zu können, gute Chancen zu sein, für das deutsche Publikum relevant zu bleiben oder relevanter zu werden.

Abschließend bemerkt wird es spannend werden, zu sehen, wann Deutschland bzw. deutsche Broadcaster dem Beispiel anderer Länder wie zum Beispiel Südkorea und Brasilien folgen und NGA im deutschen Broadcastalltag übernommen wird. Die Bereitschaft der deutschen Bevölkerung, die umfangreicheren und bedeutend tiefergehenden Funktionalitäten und Personalisierungsmöglichkeiten anzunehmen, scheint zweifelslos gegeben.

Nachwort

Den Zeitraum der Erstellung dieser Bachelorarbeit reflektierend sollen in diesem kurzen Abschnitt einige Punkte angesprochen werden, die während der Bearbeitung und Auswertung der Arbeit bzw. der quantitativen Datenerhebung aufgekommen sind.

Den überaus interessanten Themenbereich von Next Generation Audio näher zu beleuchten, weiter kennenzulernen und mich durch zahlreiche Forschungsstudien zu lesen, war mit überaus viel Arbeit und noch mehr investierter Zeit verbunden. Dies überstieg noch einmal bedeutend das, was ich ohnehin bereits erwartet hatte. Auch die benötigte Einarbeitung in die Formalien quantitativer Datenerhebung und die Statistik fiel arbeitsintensiver und vor allem nervenaufreibender aus, als ich dies im Voraus geahnt hatte. Dennoch kann ich diesen drei Monaten viel Positives entnehmen, da ich zwar wiederholt an persönliche Grenzen gekommen bin, diese aber in aller Regelmäßigkeit überwinden konnte. Dabei wurde mir bewusst, dass das Aufbrechen des großen Ganzen in kleine Teile und das Setzen von Meilensteinen, an denen mein Fortschritt im Gesamtkonstrukt ersichtlich wird, eine für mich unabdingbare Herangehensweise an große Aufgaben darstellt. Einige Dinge würde ich mit meinem heutigen Kenntnisstand allerdings anders angehen:

- In meiner Fragebogenentwicklung – vor allem auch aufgrund der Funktion der mehrsprachigen Übertragung – ging ich davon aus, dass ich auch in Deutschland lebende Menschen erreichen könnte, die kein Deutsch sprechen (Gaststudenten, Geflüchtete etc.) und durch den englischen Fragebogen mehr Probanden erreiche. Diese Annahme war fehlerhaft: Der von mir händisch ins Englische übersetzte Fragebogen wurde nur ein einziges Mal aufgerufen. Diesen Mehraufwand würde ich für eine rein bundesdeutsche Befragung daher definitiv nicht mehr betreiben.
- Die strikte Beschränkung auf eine rein quantitative Erhebung würde ich durch einige wenige qualitative Aspekte erweitern. Speziell die Frage, warum die Bewertung der einzelnen Funktionen durch die Probanden ausfiel, wie sie ausfiel, hätte durch einige frei formulierbare Stichwörter tiefgängiger erhoben werden können.

Marvin Schmid

Horb am Neckar, März 2023

Literaturverzeichnis

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM); Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI); Berufsverband Deutscher Markt- und Sozialforscher e. V. (BVM); Deutsche Gesellschaft für Online-Forschung e.V. (DGOF) (2021): Richtlinie für die Befragung von Minderjährigen. Online verfügbar unter <https://www.dgof.de/wp-content/uploads/2021/01/RL-Minderjaehrigen-neu-2021.pdf>, zuletzt geprüft am 18.02.2023.

ARD (2023a): Die Geschichte der Audiodeskription. Online verfügbar unter <https://www.daserste.de/specials/service/barrierefreiheit-geschichte-audiodeskription100.html>, zuletzt geprüft am 15.02.2023.

ARD (2023b): Tonspur Klare Sprache. Bessere Sprachverständlichkeit. Online verfügbar unter <https://www.ard-digital.de/inklusion/klare-sprache>, zuletzt aktualisiert am 31.01.2023, zuletzt geprüft am 31.01.2023.

ARTE (2023): ARTE in Deutschland empfangen | FAQ - ARTE. Online verfügbar unter <https://www.arte.tv/sites/corporate/de/faq/arte-in-deutschland-empfangen/>, zuletzt geprüft am 15.02.2023.

BBC (2020): Next Generation Audio. Online verfügbar unter <https://www.bbc.co.uk/rd/projects/next-generation-audio>, zuletzt aktualisiert am 25.11.2020, zuletzt geprüft am 03.02.2023.

Bleidt, Robert L.; Sen, Deep; Niedermeier, Andreas; Czelhan, Bernd; Fug, Simone; Disch, Sascha et al. (2017): Development of the MPEG-H TV Audio System for ATSC 3.0. In: *IEEE Transactions on Broadcasting* 63 (1), S. 202–236. DOI: 10.1109/TBC.2017.2661258.

Bolt, Marco; Smith, Jacob; Rödén, Jonas; Leteneur, Benoît; Habann, Tobias; Hanschke, Jan-Hendrik et al. (2021): Practical implementation of new open standards for Next Generation Audio production and interchange. In: *EBU Technology & Innovation | Technical Review*.

Bühner, Markus (2006): Einführung in die Test- und Fragebogenkonstruktion. 2., aktualisierte und erweiterte Auflage: Pearson Studium.

Bundesarchiv (2023): Übertragung des ersten Fernsehbildes in Deutschland. Online verfügbar unter <https://www.bundesarchiv.de/DE/Content/Dokumente-zur-Zeitgeschichte/19280211-fernsehen.html>, zuletzt aktualisiert am 26.01.2023, zuletzt geprüft am 26.01.2023.

Centre for Vision, Speech and Signal Processing (CVSSP) (2018): Object-based audio. University of Surrey. Online verfügbar unter https://cvssp.org/data/s3a/public/VISRPluginSuite/html/0.11.2/object_based_audio.html, zuletzt geprüft am 06.02.2023.

Cohen, Jacob (1992): A Power Primer. In: *Psychological bulletin* 112 (1), S. 155–159. DOI: 10.1037//0033-2909.112.1.155.

de Jong, Frans; Driesnack, Dagmar; Mason, Andrew; Parmentier, Matthieu; Sunna, Paola; Thompson, Simon (2019): European Athletics Championships. Lessons from a Live, HDR, HFR, UHD and Next-Generation Audio sports event. In: *Research & Development White Paper* (345).

Dickreiter, Michael; Goeres-Petri, Jürgen (2013): 3 Schallwahrnehmung. In: Michael Dickreiter, Volker Dittel, Wolfgang Hoeg und Martin Wöhr (Hg.): *Handbuch der Tonstudioteknik*. 8. Aufl.: DE GRUYTER, S. 115–135.

Dolby (2018): How Dolby and Telegenic are bringing ringside seats into boxing fans' homes. Online verfügbar unter <https://professional.dolby.com/contentassets/464e561e4b1f46c2996c703af4492162/dolby-telegenic-case-study.pdf>, zuletzt geprüft am 25.03.2023.

Dolby (2021): Dolby® AC-4. Audio delivery for next-generation entertainment services. Online verfügbar unter https://dolbyac4.com/de/dist/pdf/DLB_Atms_AC-4_Whtppr_DRAFT_1Col_v00-09.pdf, zuletzt geprüft am 25.03.2023.

Dolby (2023): Soundbars with Dolby Atmos. Online verfügbar unter <https://www.dolby.com/experience/soundbars/buyers-guide>, zuletzt geprüft am 21.03.2023.

EBU (2019): Next Generation Audio (NGA). In: *Technology Fact Sheet*. Online verfügbar unter https://tech.ebu.ch/docs/factsheets/ebu_tech_fs_nga.pdf, zuletzt geprüft am 16.01.2023.

EBU (2021): 10 things you need to know about Next Generation Audio. In: *Forum for Advanced Media in Europe (FAME)*. Online verfügbar unter https://tech.ebu.ch/docs/other/EBU_NGA_10_Things_You_Need_To_Know.pdf, zuletzt geprüft am 18.02.2023.

ETSI (2018): TS 103 190. Digital Audio Compression (AC-4) Standard (v1.3.1).

ETSI (2019): TS 103 491. DTS-UHD Audio Format; Delivery of Channels, Objects and Ambisonic Sound Fields (v1.2.1).

ETSI (2022): TS 101 154. Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcast and Broadband Applications (v2.7.1).

Filmförderungsanstalt (FFA) (2023): Anzahl der Kinos in Deutschland in den Jahren 2002 bis 2022. In: *Statista*. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/3077/umfrage/anzahl-der-kinos-in-deutschland-seit-dem-jahr-2002/>, zuletzt geprüft am 13.03.2023.

Flick, Uwe (2014): Gütekriterien qualitativer Sozialforschung. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS (Handbuch), S. 411–423.

Fraunhofer Institute for Integrated Circuits (IIS) (2020): Next Generation Audio and Video technologies and the future of broadcast and streaming. Online verfügbar unter <https://www.iis.fraunhofer.de/en/pr/2020/20200918-next-generation-audio-and-video-technologies.html>, zuletzt geprüft am 22.02.2023.

Fraunhofer Institute for Integrated Circuits (IIS) (2022): MPEG-H Audio selected as sole mandatory audio system for Brazil's next-generation TV 3.0 broadcast service. Online verfügbar unter https://www.iis.fraunhofer.de/en/pr/2022/20220118_mpeg-h_brazil.html, zuletzt geprüft am 22.02.2023.

Fraunhofer Institute for Integrated Circuits (IIS) (2023): MPEG-H ADM Profile. Online verfügbar unter <https://www.iis.fraunhofer.de/en/ff/amm/dl/whitepapers/adm-profile.html>, zuletzt geprüft am 12.02.2023.

Friesecke, Andreas (2014): *Die Audio-Enzyklopädie*. Ein Nachschlagewerk für Tontechniker. 2. Auflage. Berlin: De Gruyter Saur.

Gerzon, Michael A. (1973): Periphony: With-Height Sound Reproduction. In: *Journal of the Audio Engineering Society* vol. 21, S. 2–10.

Grewe, Y.; Stuchlik, J.; Ehrhard, M.; Schill, M.; Wagner, M. (2018): The Impact of Object-Based Audio on the Production of a Binaural Audio Play. In: *30th Tonmeister Tagung*, S. 92–95.

Grewe, Yannik; Eibl, Philipp; Rieger, Daniela; Torcoli, Matteo; Simon, Christian; Scuda, Ulli (2021): MPEG-H Audio Production Workflows for a Next Generation Audio Experience in Broadcast, Streaming, and Music. Audio Engineering Society (Convention e-Brief, 649).

Häder, Michael; Häder, Sabine (2014): Stichprobenziehung in der quantitativen Sozialforschung. In: Nina Baur und Jörg Blasius (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS (Handbuch), S. 283–297.

Hedges, Tom (2022): Data & Statistics on Foreign Language Media. GWI. Online verfügbar unter <https://blog.gwi.com/chart-of-the-week/foreign-language-media/>, zuletzt geprüft am 16.03.2023.

Hestermann, S.; Seideneck, M.; Sladeczek, C. (2018): Immersive Object-Based Mastering. In: *30th Tonmeister Tagung*, S. 96–101.

Hochschule der Medien (HdM) (2022): HdM Survey. Hinweise zu Datenschutz und Einwilligungserklärung. Online verfügbar unter https://www.hdm-stuttgart.de/studierende/abteilungen/cld/imagefolder/HdMSurvey_Datenschutz_Einwilligungserkl%C3%A4rung_2022-07-26.pdf, zuletzt geprüft am 18.02.2023.

Institut für Demoskopie Allensbach (IfD) (2022a): Beliebteste Fernsehformate in Deutschland im Jahr 2022. In: *Statista*. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/171208/umfrage/beliebteste-fernsehformate/?locale=de>, zuletzt geprüft am 01.03.2023.

Institut für Demoskopie Allensbach (IfD) (2022b): Bevölkerung in Deutschland nach Einschätzung der eigenen Englischkenntnisse nach Altersgruppen im Jahr 2022. In: *Statista*. Online verfügbar unter <https://de.statista.com/statistik/daten/studie/804875/umfrage/einschaetzung-zu-eigenen-englischkenntnissen-nach-alter/>, zuletzt geprüft am 14.03.2023.

ISO/IEC 23008-3:2015 (2015): Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 3: 3D audio.

ITU-R (2007): Recommendation ITU-R BS.1352-3. File format for the exchange of audio programme materials with metadata on information technology media.

ITU-R (2019a): Recommendation ITU-R BS.2127-0. Audio Definition Model renderer for advanced sound systems.

ITU-R (2019b): Recommendation ITU-R BS.2076-2. Audio definition model.

ITU-R (2019c): Recommendation ITU-R BS.2088-1. Long-form file format for the international exchange of audio programme materials with metadata.

ITU-R (2022a): Recommendation ITU-R BS.2051-3. Advanced sound system for programme production.

ITU-R (2022b): Recommendation ITU-R BS.2125-1. A serial representation of the Audio Definition Model.

ITU-R (2022c): Recommendation ITU-R BS.775-4. Multichannel stereophonic sound system with and without accompanying picture.

Keller, Daniela (2013): Wahl der Skala in Fragebögen. Online verfügbar unter <https://statistik-und-beratung.de/2013/02/wahl-der-skala-in-fragebogen/>, zuletzt geprüft am 01.03.2023.

Krebs, Nadja; Menold, Natalja (2014): Gütekriterien quantitativer Sozialforschung. In: Nina Baur und Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer VS (Handbuch), S. 425–438.

Mäsel, Jakob; Simon, Christian; Choi, Jin; Schulz, Andreas; Silzle, Andreas (2019): Field Test for Immersive and Interactive Audio Production with the Gewandhausorchester Leipzig using MPEG-H. In: *5th International Conference on Spatial Audio*.

Musik zu laut, Dialoge genuschelt (2018). In: *Süddeutsche Zeitung*. Online verfügbar unter <https://www.sueddeutsche.de/kolumne/fernsehen-musik-zu-laut-dialoge-genuschelt-1.4043461>, zuletzt geprüft am 16.03.2023.

Olivieri, Ferdinando; Peters, Nils; Sen, Deep: Scene-Based Audio and Higher Order Ambisonics. A technology overview and application to Next-Generation Audio, VR and 360° Video. In: *EBU Technology & Innovation | Technical Review*.

Pardoe, Laurence; Ward, Lauren; Clawson, Hannah; Moulson, Aimee; Pike, Chris (2020): Investigating user interface preferences for controlling background-foreground balance on connected TVs. In: *Research & Development White Paper* (385).

Rasch, Dieter; Guiard, Volker (2004): The robustness of parametric statistical methods. In: *Psychology Science* Volume 46 (2), S. 175–208.

Riegler, Thomas (2019): Die Geschichte des Fernsehens: Wie alles begann (Teil 1). In: *digitalfernsehen.de*, 30.11.2019. Online verfügbar unter <https://www.digitalfernsehen.de/ratgeber/digital-tv/die-geschichte-des-fernsehens-wie-alles-begann-teil-1-549592/>, zuletzt geprüft am 26.01.2023.

Schätzlein, Frank (2005): Ton und Sounddesign beim Fernsehen. In: Harro Segeberg und Frank Schätzlein (Hg.): *Sound. Zur Technologie und Ästhetik des Akustischen in den Medien*. Marburg: Schüren, S. 184–203.

Schmieder, Jürgen (2010): Noch spielt Tom Hanks in Deutschland Lotto. In: *Süddeutsche Zeitung*, 2010. Online verfügbar unter <https://www.sueddeutsche.de/kultur/synchronisation-noch-spielt-tom-hanks-in-deutschland-lotto-1.801913>, zuletzt geprüft am 16.03.2023.

Simon, Christian; Grewe, Yannik; Faecks, Nicolas; Scuda, Ulli (2018): Field Tests for Immersive and Interactive Broadcast Audio Production using MPEG-H 3D Audio. In: *SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING*, S. 40–46.

Slavik, Karl M. (2013): 15 Film- und Fernseh-ton. In: Michael Dickreiter, Volker Dittel, Wolfgang Hoeg und Martin Wöhr (Hg.): *Handbuch der Tonstudioteknik*. 8. Aufl.: DE GRUYTER, S. 924–1006.

Slavik, Karl M.; Weinzierl, Stefan (2008): Wiedergabeverfahren. In: Stefan Weinzierl (Hg.): *Handbuch der Audiotechnik*. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 609–685.

Statistisches Bundesamt (Destatis) (2021): Ausblick auf die Bevölkerungsentwicklung in Deutschland und den Bundesländern nach dem Corona-Jahr 2020. Erste mittelfristige Bevölkerungsvorausberechnung 2021 bis 2035.

Teufel (2023): Übersicht von Dolby-Atmos-Kinos in Deutschland. Online verfügbar unter <https://blog.teufel.de/uebersicht-dolby-atmos-kinos-in-deutschland/>, zuletzt geprüft am 13.03.2023.

Theile, Günther; Dickreiter, Michael; Graul, Wolfram; Camerer, Florian; Spikofski, Gerhard (2013): 5 Tonaufnahme und Tonwiedergabe. In: Michael Dickreiter, Volker Dittel, Wolfgang Hoeg und Martin Wöhr (Hg.): Handbuch der Tonstudiotechnik. 8. Aufl.: DE GRUYTER, S. 217–369.

Turnwald, Andreas; Simon, Christian; Scuda, Ulli (2018): Eurovision Song Contest 2018 – immersive and interactive. In: *30th Tonmeister Tagung*, S. 88–91.

Verbraucherzentrale (2022): Analoges Kabelfernsehen seit 2019 abgeschaltet: Was wichtig ist. Online verfügbar unter <https://www.verbraucherzentrale.de/wissen/digitale-welt/fernsehen/analoges-kabelfernsehen-seit-2019-abgeschaltet-was-wichtig-ist-23763>, zuletzt aktualisiert am 11.10.2022, zuletzt geprüft am 31.01.2023.

VRTONUNG (2022): MPEG-H Audio vs. "Dolby Atmos" - Es gibt einen Gewinner. Online verfügbar unter <https://www.vrtonung.de/mpeg-h-audio-vs-dolby-atmos/>, zuletzt geprüft am 09.02.2023.

WDR (1963): Neuheiten zur Funkausstellung 1963. Online verfügbar unter <https://www.ardmediathek.de/video/wdr-retro-hier-und-heute/neuheiten-zur-funkausstellung-1963/wdr/Y3JpZDovL3dkci5kZS9CZWl0cmFnLWQyYTZlOGY3LTRmMTgtNDIwZS1iZDkwLTl2ZjkxNTM0M2UzMw>, zuletzt geprüft am 14.02.2023.

Weinzierl, Stefan (2008): Aufnahmeverfahren. In: Stefan Weinzierl (Hg.): Handbuch der Audiotechnik. Berlin, Heidelberg: Springer Berlin Heidelberg, S. 551–607.

ZDF (2022): Tonangebot „Klare Sprache“ bietet verbesserte Sprachverständlichkeit. Online verfügbar unter <https://www.zdf.de/zdfunternehmen/technik-klare-sprache-100.html>, zuletzt aktualisiert am 05.07.2022, zuletzt geprüft am 31.01.2023.

Zotter, Franz; Frank, Matthias (2019): Ambisonics. A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality: Springer International Publishing.

Anhang

Anhang 1:	Tests auf Normalverteilung.....	60
Anhang 2:	Ursprüngliche Altersgruppen.....	61
Anhang 3:	Sekundäre Ergebnisse	62
Anhang 3.1:	Häufigkeit des Konsums.....	62
Anhang 3.2:	Wiedergabesysteme stationär	62
Anhang 3.3:	Wiedergabesysteme mobil.....	63
Anhang 3.4:	Einfluss der Funktionen 1	63
Anhang 3.5:	Einfluss der Funktionen 2	64
Anhang 4:	Fragebogen	65
Anhang 4.1:	Links zu den Videos	65
Anhang 4.2:	Fragebogen in deutscher und englischer Fassung.....	65

Anhang 1: Tests auf Normalverteilung

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Mehrsprachige Übertragung	0,115	223	0,000	0,916	223	0,000
Anpassung von Lautstärkelevels	0,121	223	0,000	0,916	223	0,000
Audiodeskription	0,183	223	0,000	0,838	223	0,000
Positionierbare Objekte	0,084	223	0,001	0,965	223	0,000

a. Signifikanzkorrektur nach Lilliefors

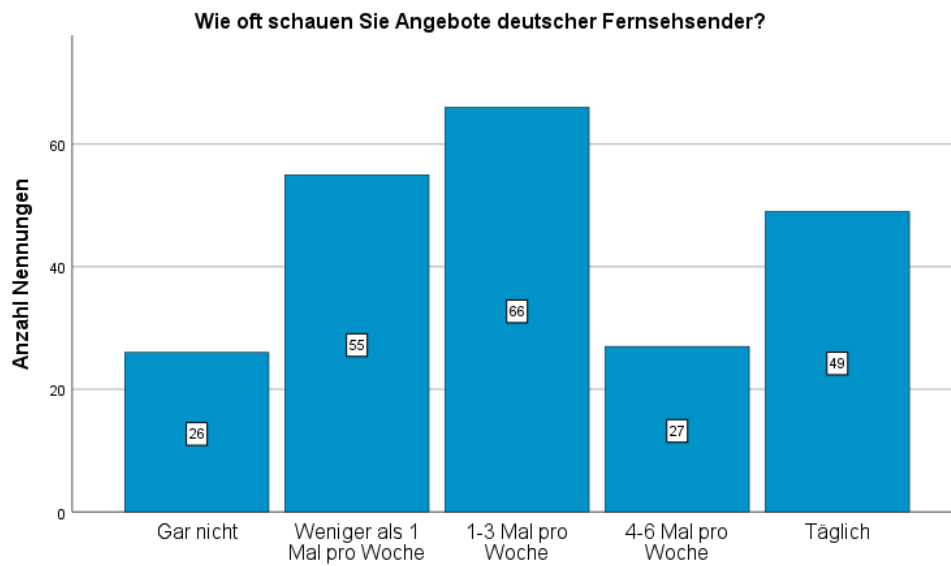
Kommentar: Bei diesen Tests ist die Nullhypothese: „Die Variable ist normalverteilt“. Aufgrund der signifikant ausfallenden Tests muss diese Nullhypothese demnach abgelehnt werden, alle vier Variablen sind nicht normalverteilt.

Anhang 2: Ursprüngliche Altersgruppen

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
16-19 Jahre	1	0,4	0,4	0,4
20-29 Jahre	90	40,2	40,2	40,6
30-39 Jahre	51	22,8	22,8	63,4
40-49 Jahre	29	12,9	12,9	76,3
50-59 Jahre	38	17,0	17,0	93,3
60-69 Jahre	13	5,8	5,8	99,1
70 Jahre oder älter	2	0,9	0,9	100,0
Gesamt	224	100,0	100,0	

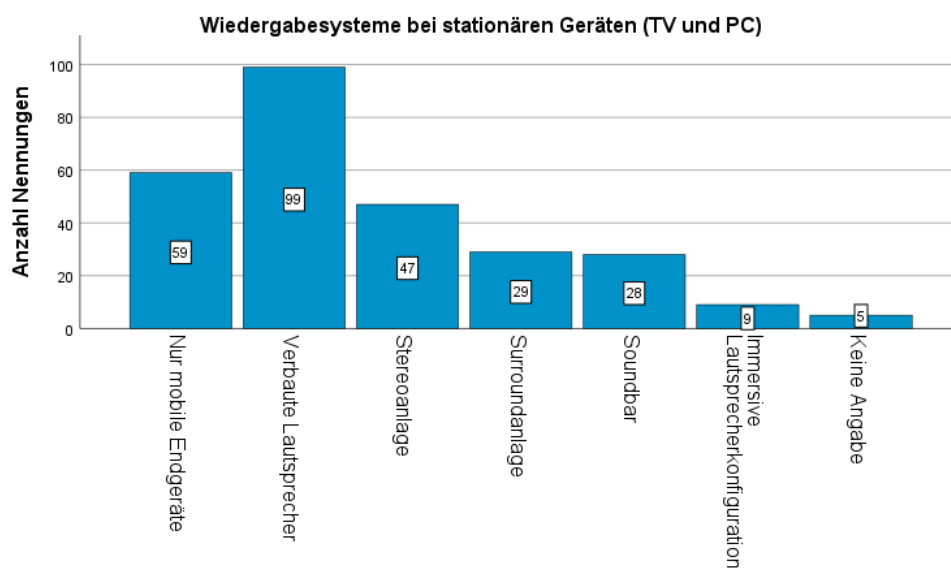
Anhang 3: Sekundäre Ergebnisse

Anhang 3.1: Häufigkeit des Konsums



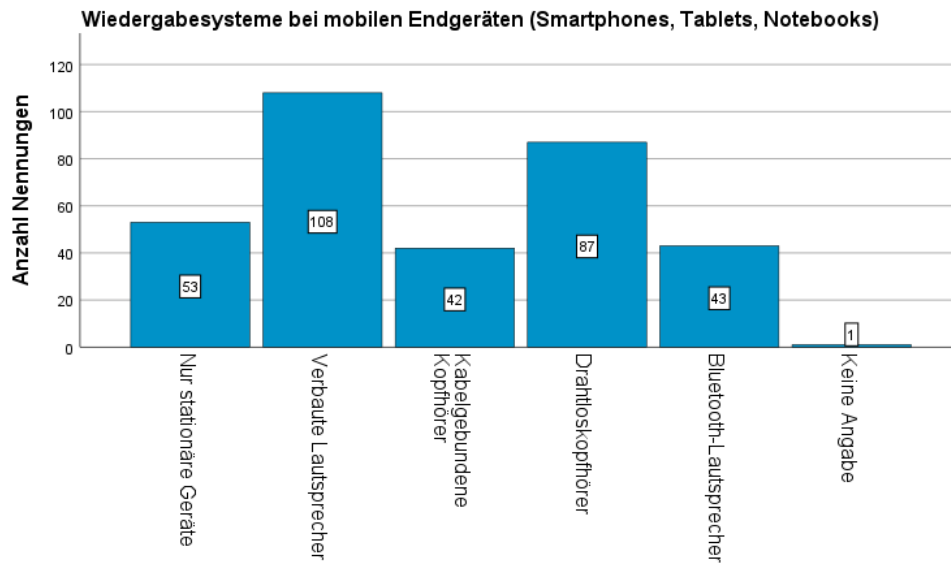
Unter „Andere“ genannt: 13x Kopfhörer, 3x Bluetooth-Lautsprecher

Anhang 3.2: Wiedergabesysteme stationär



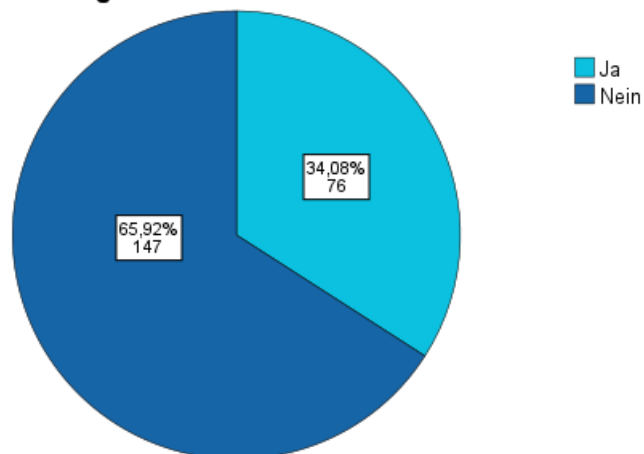
Unter „Andere“ genannt: 2x Stereoanlage

Anhang 3.3: Wiedergabesysteme mobil

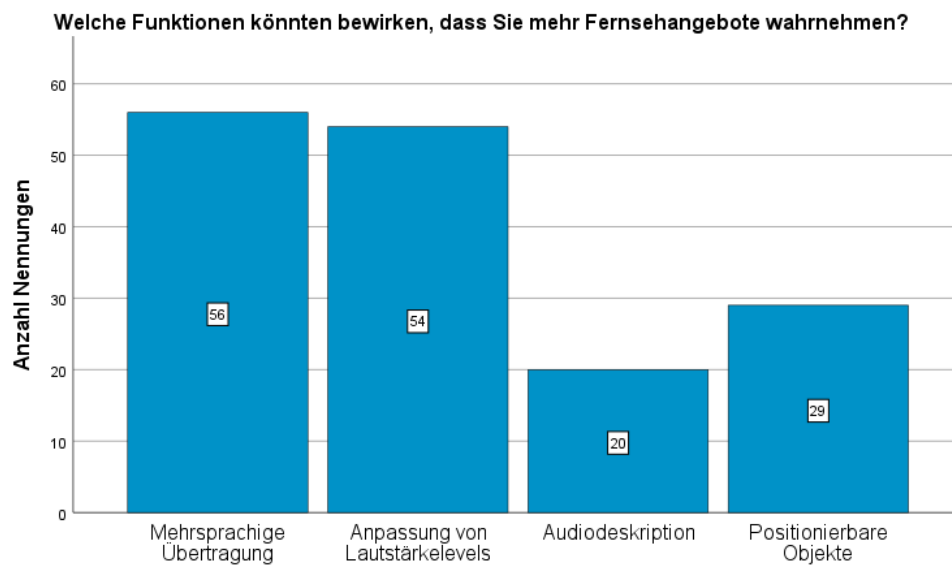


Anhang 3.4: Einfluss der Funktionen 1

Könnte der Einsatz der behandelten Funktionen bewirken, dass Sie mehr Angebote deutscher Fernsehsender schauen?



Anhang 3.5: Einfluss der Funktionen 2



Anhang 4: Fragebogen

Anhang 4.1: Links zu den Videos

Anmerkung: Die folgenden Links sind auf der beiliegenden CD zusätzlich als klickbare Internetverknüpfungen enthalten.

Deutsche Fassung

Mehrsprachige Übertragung – enthalten nur Ausschnitt 0:24 bis 0:44:

Fraunhofer IIS – <https://youtu.be/3Gmn-Gk2Lhg>

Anpassung von Lautstärkelevels – enthalten nur Ausschnitt 0:57 bis 1:22:

Fraunhofer IIS – <https://youtu.be/9fU90uhHBm8>

Audiodeskription – enthalten in voller Länge:

Aktion Mensch – <https://youtu.be/L66Zz0vixRo>

Positionierbare Objekte – enthalten in voller Länge:

Eigenes Video – <https://youtu.be/u30WupWz0iE>

Englische Fassung

Multilingual Transmission – enthalten nur Ausschnitt 0:24 bis 0:44:

Fraunhofer IIS – <https://youtu.be/3Gmn-Gk2Lhg>

Adjustment of Volume Levels – enthalten nur Ausschnitt 0:57 bis 1:22:

Fraunhofer IIS – <https://youtu.be/oErdKT2oS6U>

Audio Description – enthalten in voller Länge:

NSW Public Service Commission – <https://youtu.be/PNIALw-xTqs>

Positioning of Audio Objects – enthalten in voller Länge:

Eigenes Video – <https://youtu.be/jcYDQ5alggw>

Anhang 4.2: Fragebogen in deutscher und englischer Fassung

Relevanz von Funktionen von Next Generation Audio (NGA)

Liebe/r Teilnehmer/in,

danke für Ihr Interesse an dieser Studie!

Dieser Fragebogen ist Teil meiner Erforschung der Relevanz von Funktionen von Next Generation Audio (NGA) für Fernsehzuschauer/innen in Deutschland. Die Befragung findet im Rahmen meiner Bachelorarbeit an der Hochschule der Medien in Stuttgart statt.

Ihre Aufgabe wird es sein, kurze Videos anzuschauen und anschließend Fragen zu beantworten. Wenn möglich, nutzen Sie hierfür bitte Kopfhörer. Die Umfrage dauert etwa 5-8 Minuten und Sie benötigen keinerlei Vorwissen zum genannten Thema.

Bei Rückfragen oder Verständnisschwierigkeiten können Sie sich gerne bei mir melden.

Marvin Schmid
ms548@hdm-stuttgart.de
Hochschule der Medien Stuttgart
Fakultät Electronic Media

Einwilligungserklärung *

Im Folgenden informiere ich Sie über den datenschutzrechtskonformen Umgang mit den erhobenen Daten, Ihre Rechte und bitte um Ihre Zustimmung zur Teilnahme an meiner Studie sowie zur Verwendung Ihrer Daten für die angegebenen Zwecke. Bitte lesen Sie die folgenden Erklärungen sorgfältig durch.

Im Rahmen dieser Befragung werden personenbezogene Daten erhoben (Alter, Geschlecht, Informationen zu sensorischen Beeinträchtigungen). Diese werden nach der Erhebung von der verantwortlichen Person heruntergeladen und nach Art. 6 der DSGVO weiterverarbeitet.

Das bedeutet, dass die Rohdaten

- nur zum angegebenen Zweck weiterverarbeitet werden.
- nicht an Dritte weitergegeben werden.
- auf passwortgeschützten Medien (z.B. nicht auf Google Docs, oder anderen Clouddiensten außerhalb des Geltungsbereichs der DSGVO) gespeichert werden.
- fristgerecht gelöscht oder anonymisiert werden.

Für die personenbezogenen Daten gilt eine Speicherfrist von drei Monaten.

Daten, die weiterverarbeitet und dabei vollständig anonymisiert wurden und deren Nutzung rein wissenschaftlichen Zwecken dient, können an Dritte weitergegeben und veröffentlicht werden. Vollständig anonymisierte Daten werden für sechs Monate auf einer Festplatte gespeichert.

Hiermit erkläre ich,

- dass ich mindestens 16 Jahre alt bin.
- dass ich aktuell in Deutschland lebe.
- dass ich mich freiwillig zur Teilnahme an der Befragung bereiterkläre.
- dass ich über die Inhalte und Ziele der oben genannten Befragung sowie die Verwendung meiner Daten schriftlich aufgeklärt wurde und durch meine Zustimmung mein Einverständnis hierzu erteile.
- dass ich über die Speicherung und Verwendung aller erhobenen Daten schriftlich aufgeklärt wurde und durch meine Zustimmung mein Einverständnis hierzu erteile.
- dass ich die erteilte Einwilligung in die Verwendung meiner Daten jederzeit bei der verantwortlichen Person widerrufen kann, jedoch nur wenn, bzw. so lange, wie eine Zuordnung noch möglich ist, also bis zur Anonymisierung der Daten.
- dass mir bekannt ist, dass ich die Befragung jederzeit, ohne Angabe von Gründen, abbrechen kann und dass mir daraus kein Nachteil entsteht.

☐ Ich stimme zu

☐ Ich stimme **nicht** zu
(Hiermit beenden Sie die Umfrage.)

Wie oft schauen Sie Angebote von deutschen Fernsehsendern?

Gemeint sind ausgestrahlte Programme im TV und im Internet sowie Angebote aus den Internet-Mediatheken

☐ Gar nicht

☐ Weniger als 1 Mal pro Woche

☐ 1-3 Mal pro Woche

☐ 4-6 Mal pro Woche

☐ Täglich

☐ Keine Angabe

Worüber konsumieren Sie zu Hause den Ton dieser Inhalte an ihrem Fernseher oder Desktop-PC?

Mehrfachauswahl möglich

☐ Ich schaue nur auf mobilen Endgeräten

☐ Verbaute Lautsprecher

☐ Stereoanlage

☐ Surroundanlage (z.B. 5.1 oder 7.1)

☐ Soundbar

☐ Lautsprechersystem mit Überkopflautsprechern (z.B. 5.1.2, 5.1.4 oder 7.1.4)

☐ Keine Angabe

☐ Andere

Worüber konsumieren Sie den Ton dieser Inhalte an Ihren mobilen Endgeräten?

Mehrfachauswahl möglich

Gemeint sind Smartphones, Tablets und Notebooks

☐ Ich schaue nur auf meinem Fernseher oder Desktop-PC

☐ Verbaute Lautsprecher

☐ Kabelgebundene Kopfhörer

☐ Drahtloskopfhörer (Bluetooth oder Funk)

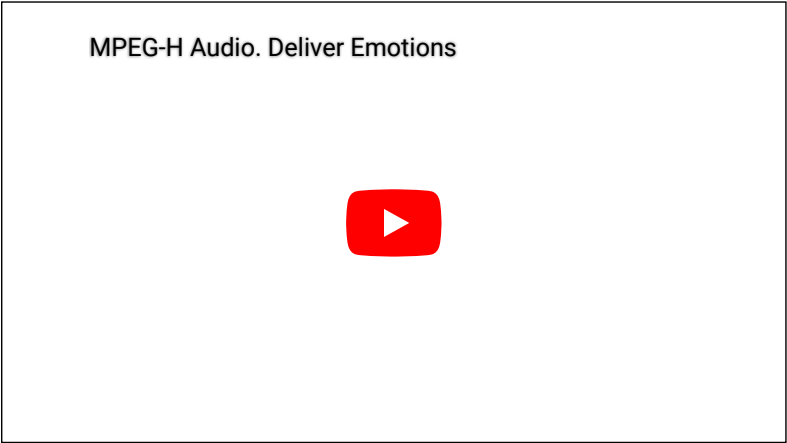
☐ Bluetooth-Lautsprecher

☐ Keine Angabe

☐ Andere

1/4 Mehrsprachige Übertragung

Das folgende Video präsentiert die Funktion verschiedener Wiedergabesprachen bei NGA-Fernsehübertragungen (20 Sekunden).



Quelle: FraunhoferIIS bei YouTube

Sie wollen das Video noch einmal anschauen?
Laden Sie die Seite neu oder scrollen ganz nach unten, klicken "Zurück" und dann wieder auf "Weiter". Das Video wird neu geladen.

Wie bewerten Sie die Option der Wahl verschiedener Sprachen bei Filmen und Serien?

Beispiel: Wahl aus deutscher Synchronisation und englischer Originalsprache

Überhaupt nicht sinnvollAusgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option der Wahl verschiedener Sprachen bei Nachrichtensendungen?

Beispiel: Wahl aus deutscher Sprache und englischer Übersetzung

Überhaupt nicht sinnvollAusgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option der Wahl verschiedener Sprachen bei Sportübertragungen?

Beispiel: Wahl aus deutschem und englischem Kommentar

Überhaupt nicht sinnvollAusgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option der Wahl verschiedener Kommentare nach Vereinszugehörigkeit bei Sportübertragungen?

Beispiel: Wahl aus Heimteam-Kommentar und Auswärtsteam-Kommentar

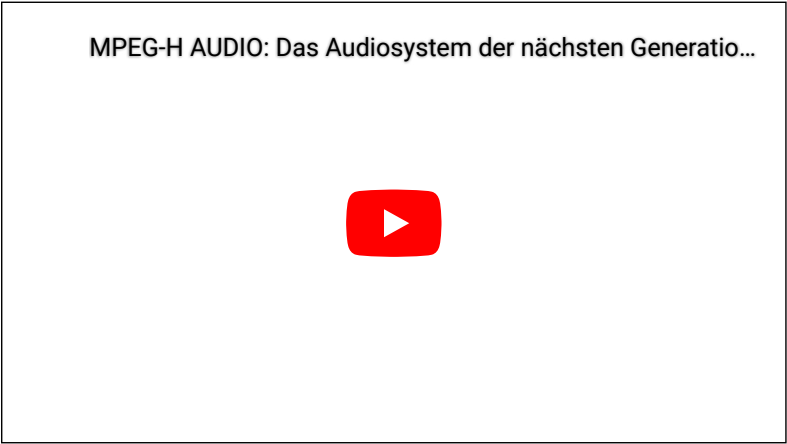
Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

2/4 Anpassung von Lautstärkelevels

Das folgende Video präsentiert die Funktion der Anpassung von Lautstärkelevels einzelner Quellen im gesamten Audio-Mix bei NGA-Fernsehübertragungen (30 Sekunden).



Quelle: FraunhoferIIS bei YouTube

Sie wollen das Video noch einmal anschauen?
Laden Sie die Seite neu oder scrollen ganz nach unten, klicken "Zurück" und dann wieder auf "Weiter". Das Video wird neu geladen.

Wie bewerten Sie die Option der Lautstärkeanpassung von Sprache bei Filmen und Serien?

Beispiel: Sprache/Dialog im Vergleich zum restlichen Audiomix (z.B. Hintergrundmusik) lauter, um die Sprachverständlichkeit zu verbessern

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option der Lautstärkeanpassung von Sprache bei eingespielten Videobeiträgen von Nachrichtensendungen?

Beispiel: Sprache/Dialog im Vergleich zum restlichen Audiomix (z.B. Hintergrundmusik) lauter, um die Sprachverständlichkeit zu verbessern

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option der Lautstärkeanpassung vom Kommentar bei Sportübertragungen?

Beispiel: Kommentar im Vergleich zum restlichen Audiomix (z.B. Fangesänge) lauter, um Sprachverständlichkeit zu verbessern oder leiser/aus, um Stadion- oder Hallenatmosphäre besser zu hören

Überhaupt nicht sinnvoll

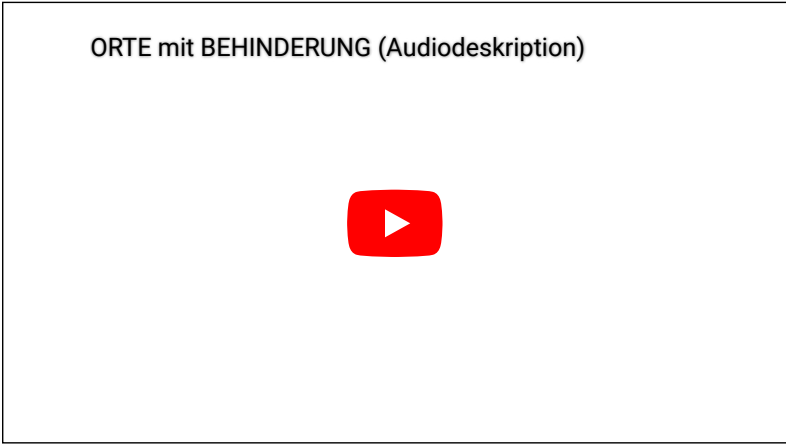
Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

3/4 Audiodeskription

Das folgende Video ist eine Demonstration von Audiodeskription. Eine Audiodeskription ist eine gesprochene Beschreibung von dem, was zu sehen ist (hier: männliche Stimme).

NGA-Fernsehübertragungen können ermöglichen, Audiodeskriptionen zuzuschalten (45 Sekunden).



Quelle: Aktion Mensch bei YouTube

Sie wollen das Video noch einmal anschauen?
Laden Sie die Seite neu oder scrollen ganz nach unten, klicken "Zurück" und dann wieder auf "Weiter". Das Video wird neu geladen.

Wie bewerten Sie die Option zuschaltbarer Audiodeskription bei Filmen und Serien?

Audiodeskription beschreibt in den Zeiträumen ohne Dialog der Figuren die visuelle Handlung

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option zuschaltbarer Audiodeskription bei Nachrichtensendungen?

Audiodeskription beschreibt in Zeiträumen ohne Sprache eingeblendete Bilder und erklärt bei eingespielten Beiträgen die visuelle Handlung

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option zuschaltbarer Audiodeskription bei Sportübertragungen?

Audiodeskription beschreibt den visuellen Ablauf von z.B. einem Fußballspiel detaillierter als der normale Kommentar

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

4/4 Positionierung von Audioobjekten

In diesem letzten Video wird die Funktion der freien Positionierung sogenannter Audioobjekte im dreidimensionalen Raum präsentiert. Eine Umpositionierung eines solchen Audioobjekts beeinflusst dabei den restlichen Audio-Mix nicht.

Die im Video beschriebenen Effekte funktionieren nur mit Kopfhörern (40 Sekunden).



Sie wollen das Video noch einmal anschauen?
Laden Sie die Seite neu oder scrollen ganz nach unten, klicken "Zurück" und dann wieder auf "Weiter". Das Video wird neu geladen.

Wie bewerten Sie die Option individuell positionierbarer Sprache bei Filmen und Serien?

Beispiel: Sie verschieben die Sprache/den Dialog in Richtung des Ihnen am nächsten liegenden Lautsprechers.

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option individuell positionierbarer Sprache bei Nachrichtensendungen?

Beispiel: Sie verschieben die Sprache in Richtung des Ihnen am nächsten liegenden Lautsprechers.

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option individuell positionierbarem Kommentar bei Sportübertragungen?

Beispiel: Sie verschieben den Kommentar in Richtung des Ihnen am nächsten liegenden Lautsprechers.

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Wie bewerten Sie die Option individuell positionierbarer Audiodeskription?

Beispiel: Mehrere Personen schauen einen Film oder eine Sportübertragung. Eine der Personen hat eine Sehbeeinträchtigung. Daher aktivieren Sie die Audiodeskription und verschieben Sie in Richtung des Lautsprechers, der dieser Person am nächsten liegt.

Überhaupt nicht sinnvoll

Ausgesprochen sinnvoll

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Könnte der Einsatz der behandelten Funktionen bewirken, dass Sie mehr Angebote deutscher Fernsehsender schauen?

☐ Ja

☐ Nein

Welche?

Mehrfachauswahl möglich

☐ Mehrsprachige Übertragung

☐ Anpassung von Lautstärkelevels

☐ Audiodeskription

☐ Positionierung von Audioobjekten

Geschlecht

☐ Weiblich

☐ Männlich

☐ Divers

☐ Keine Angabe

Alter

☐ 16-19

☐ 20-29

☐ 30-39

☐ 40-49

☐ 50-59

☐ 60-69

☐ 70 oder älter

☐ Keine Angabe

Haben Sie eine durch die genannten Funktionen beeinflusste sensorische Beeinträchtigung?

Mehrfachauswahl möglich

☐ Hörbeeinträchtigung

☐ Sehbeeinträchtigung

☐ Keine (der genannten)

☐ Keine Angabe

Relevance of Features of Next Generation Audio (NGA)

Dear participant,

thank you for your interest in this study!

This questionnaire is part of my research into the relevance of Next Generation Audio (NGA) features for the television audience in Germany. The survey is carried out as part of my bachelor thesis at the Stuttgart Media University.

Your task will be to watch short videos and answer questions afterwards. If possible, please use some headphones. The survey takes about 5-8 minutes and you are not required to have any prior knowledge on the stated subject.

You can gladly contact me should there be any questions or comprehension difficulties.

Marvin Schmid
ms548@hdm-stuttgart.de
Stuttgart Media University
Electronic Media Faculty

Declaration of consent *

Below, I will inform you about the privacy compliant handling of collected data, your rights and ask for your consent to participation in this study and the usage of your data for the stated purposes. Please read the following statements carefully.

Personal data is collected in this survey (age, gender, information on sensory impairments). The data will be downloaded by the responsible person and processed under article 6 of the DSGVO.

This entails, that raw data

- is only used for the stated purposes.
- is not passed on to third parties.
- is saved on password-protected media (e.g. not on Google Docs or other cloud services outside the scope of the DSGVO).
- is deleted or anonymised in due time.

The retention period for personal data is three months.

Data that is further processed and in that entirely anonymised and of which the use is solely of scientific purpose, can be passed on to third parties and be published. Entirely anonymised data is saved on a hard drive for a period of six months.

Herewith, I declare

- that I am at least 16 years old.
- that I live in Germany.
- that I participate in this survey voluntarily.
- that I was informed about the contents and the objectives of the above mentioned survey and the usage of my data in written form and that I consent by accepting below.
- that I was informed about the storage and usage of all collected data in written form and that I consent by accepting below.
- that I can revoke my given consent to the usage of my data with the responsible person, though only if or as long as there is clear allocation possible, therefore until anonymisation of the data.
- that I am aware that I can cancel the survey at any time without stating reasons and that I will not have any disadvantage as a result.

☐ I consent

☐ I do **not** consent
(The survey will be terminated.)

How often do you watch offerings of German television broadcasters?

Broadcasts on TV and online as well as on-demand content on their online media libraries are meant

☐	Not at all
☐	Less than once a week
☐	1 to 3 times a week
☐	4 to 6 times a week
☐	Daily
☐	Prefer not to say

How do you consume the sound of this content at home on your TV or desktop PC?

Multiple selection possible

☐	I only watch on mobile devices
☐	Built-in speakers
☐	Stereo system
☐	Surround system (e.g. 5.1 or 7.1)
☐	Soundbar
☐	Speaker system with overhead speakers (e.g. 5.1.2, 5.1.4 or 7.1.4)
☐	Prefer not to say
☐	Other

How do you consume the sound of this content on your mobile devices?

Multiple selection possible

Smartphones, tablets and notebooks

☐ I only watch on my TV or desktop PC

☐ Built-in speakers

☐ Wired headphones

☐ Wireless headphones (bluetooth or radio frequency)

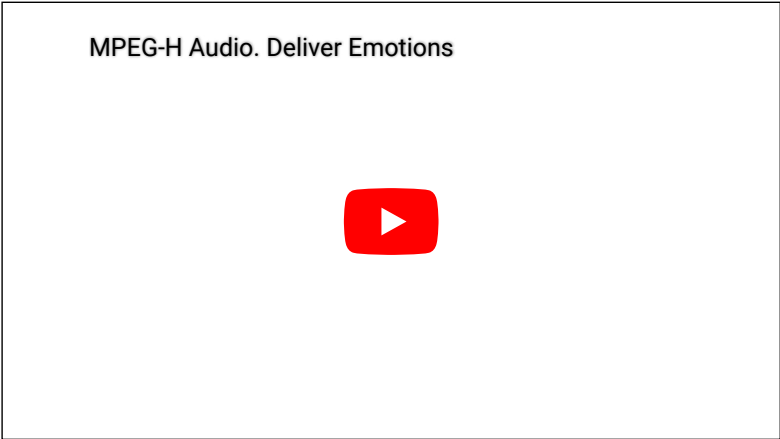
☐ Bluetooth speaker

☐ Prefer not to say

☐ Other

1/4 Multilingual transmission

The following video presents the feature of different playback languages on NGA television broadcasts (20 seconds).



Source: FraunhoferIIS on YouTube

Want to watch the video again?
Reload the page or scroll all the way down, click "Back" and then "Next" again. The video will reload.

How do you evaluate the option of multiple languages in movies or TV series?

Example: Choice between German dubbing and English original language

Not at all useful Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of multiple languages in news broadcasts?

Example: Choice between German original language and English translation

Not at all useful Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of multiple languages in sports broadcasts?

Example: Choice between German and English commentary

Not at all useful Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of multiple commentaries by team affiliation in sports broadcasts?

Example: Choice between home team commentary and away team commentary

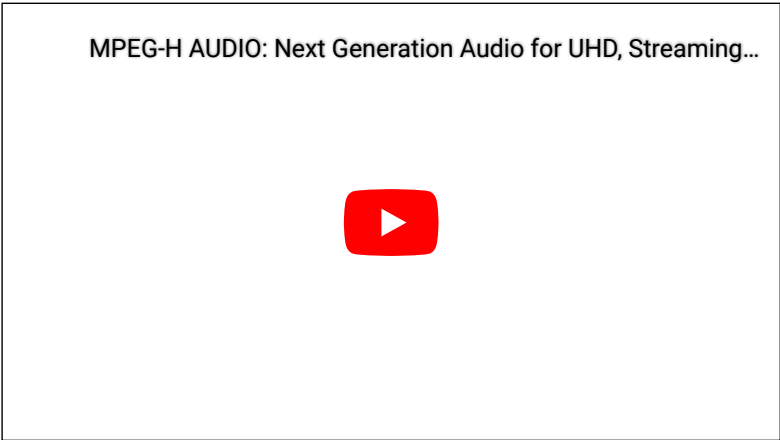
Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

2/4 Adjustment of volume levels

The following video presents the feature of adjustable volume levels of individual sources in the overall audio mix on NGA television broadcasts (30 seconds).



Source: FraunhoferIIS on YouTube

Want to watch the video again?
Reload the page or scroll all the way down, click "Back" and then "Next" again. The video will reload.

How do you evaluate the option of adjusting the volume of speech in movies and TV series?

Example: Speech/dialogue louder as compared to the rest of the audio mix (e.g. background music) to improve intelligibility

Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of adjusting the volume of speech during video segments of news broadcasts?

Example: Speech/dialogue louder as compared to the rest of the audio mix (e.g. background music) to improve intelligibility

Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of adjusting the volume of commentary in sports broadcasts?

Example: Commentary louder as compared to the rest of the audio mix (e.g. fan chants) to improve intelligibility or quieter/off to better hear the atmosphere of the arena

Not at all useful

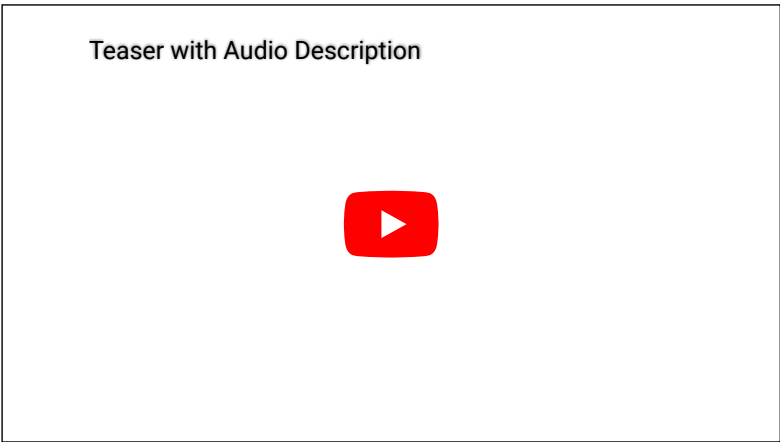
Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

3/4 Audio description

The following video demonstrates audio description. An audio description is a spoken description of what is seen (here: female voice).

NGA television broadcasts can allow the activation of audio descriptions (30 seconds).



Source: NSW Public Service Commission on YouTube

Want to watch the video again?
Reload the page or scroll all the way down, click "Back" and then "Next" again. The video will reload.

How do you evaluate the option of activatable audio description in movies and TV series?

The audio description describes the visual action in periods without dialogue of characters

Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of activatable audio description in news broadcasts?

The audio description describes overlaid pictures on screen and visual action of video segments during periods without speech

Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of activatable audio description in sports broadcasts?

The audio description describes the visual action of e.g. a football game in more detail compared to the normal commentary

Not at all useful

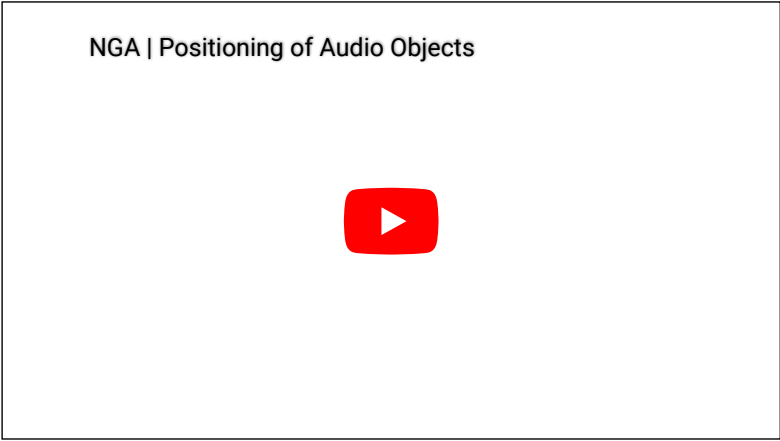
Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

4/4 Positioning of audio objects

In this last video, the feature of free positioning of so-called audio objects in three-dimensional space is presented. The re-positioning of such an audio object does not influence the rest of the audio mix.

The effects described in the video only work with headphones on (40 seconds).



Want to watch the video again?
Reload the page or scroll all the way down, click "Back" and then "Next" again. The video will reload.

How do you evaluate the option of moveable speech in movies and TV series?

Example: You move the speech/dialogue object(s) towards the speaker closest to you.

Not at all usefulVery useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of moveable speech in news broadcasts?

Example: You move the speech object towards the speaker closest to you.

Not at all usefulVery useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of moveable commentary in sports broadcasts?

Example: You move the commentary object towards the speaker closest to you.

Not at all usefulVery useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

How do you evaluate the option of moveable audio description?

Example: Multiple people are watching a movie. One person has a visual impairment. Therefore, you activate the audio description and move it towards the speaker closest to that person.

Not at all useful

Very useful

1	2	3	4	5	6	7
---	---	---	---	---	---	---

Could the use of the discussed features get you to watch more of the offerings of German broadcasters?

☐ Yes

☐ No

Which ones?

Multiple selection possible

☐ Multilingual transmission

☐ Adjustment of volume levels

☐ Audio description

☐ Positioning of audio objects

Gender

☐ Female

☐ Male

☐ Non-binary

☐ Prefer not to say

Age

☐ 16-19

☐ 20-29

☐ 30-39

☐ 40-49

☐ 50-59

☐ 60-69

☐ 70 or older

☐ Prefer not to say

Do you have a sensory impairment influenced by the mentioned features?

Multiple selection possible

☐ Hearing impairment

☐ Visual impairment

☐ None (of the above)

☐ Prefer not to say