Evaluierung eines Ansatzes zur Verbesserung der Sprachverständlichkeit von Stereosignalen im Rundfunk

Theresa Liebl, Michael Weitnauer, Michael Meier

Institut für Rundfunktechnik, 80939 München, E-Mail: liebl@irt.de; weitnauer@irt.de; meier@irt.de

Einführung

Die Verständlichkeit gesprochener Dialoge im Fernsehen ist, sowohl für Normalhörende als auch insbesondere für hörgeschädigte Personen, von grundsätzlicher Bedeutung. Das europäische Kooperationsprojekt HBB4ALL[1] (Hybrid Broadcast Broadband for All) [1] beschäftigt sich mit dem barrierefreien Zugang zu Medien für alle Bürger – unter anderem über den HbbTV-Standard (Hybrid Broadcast Broadband TV). Eine der Herausforderungen dabei ist die Bereitstellung von Inhalten über verschiedene Kanäle hinweg, mit einer Spezifizierung für unterschiedliche Zielgruppen. Die Personalisierung von interaktiven Fernsehinhalten und Dienstleistungen ist eine der zentralen Aufgabenstellungen von HBB4ALL.

Aus vielerlei Gründen wird die Sprachverständlichkeit aktueller TV-Produktionen vor allem von älteren Personen und Hörgeschädigten oft als unzureichend beurteilt. Darüber hinaus sind in den letzten Jahren die Beschwerden in Bezug auf eine schlechte Sprachverständlichkeit im Fernsehen deutlich gestiegen. Mögliche Gründe dafür sind neben den mangelhaften Frequenzgängen von typischen Flachbildschirmen auch Sprachmischungen des Fernsehtons, welche nicht an die Bedürfnisse der älteren Bevölkerungsteile angepasst sind. Gerade darauf sollte allerdings speziell geachtet werden, da der Anteil der Personen mit altersbedingter Hörminderung durch die Verschiebung der Altersverteilung in der deutschen Bevölkerung immer höher wird. Die Deutsche Gesellschaft der Hörgeschädigten fordert unter anderem daher auch eine schrittweise Erhöhung des Programmanteils von „Clear-Audio“ im Rundfunk bis auf 80%. Clear-Audio setzt dabei auf einen SNR (Verhältnis von Sprache zu Nebengeräuschen) von mehr als 15 dB.

Auf Grund dieser und weiterer Gegebenheiten ist es ein Ziel des HBB4ALL-Projektes, die Sprachverständlichkeit von TV-Produktionen - insbesondere für Hörgeschädigte - zu verbessern. Die Hauptaufgabe des IRT ist es dabei einen personalisierbaren Clean Audio Service für hörgeschädigte Benutzer zu entwickeln. Clean Audio ist in diesem Zusammenhang nicht mit Clean Speech gleichzustellen, was eine reine Darbietung von Sprache bedeuten würde, sondern mit dem bereits erwähnten Clear-Audio. Clean Audio bedeutet hier also, dass eine allgemeine Verbesserung der Sprachverständlichkeit vorgenommen wird und das Sprache-Restsignal-Verhältnis mindestens 15 dB betragen sollte.

Das am IRT entwickelte Clean Audio Tool ist ein Verfahren, mit dem bereits fertig gemischte Rundfunk-Signale bezüglich ihrer Sprachverständlichkeit verbessert werden sollen. Dabei werden die Stereo- und 5.1-Signale speziell für Hörgeschädigte angepasst [2, 3]. Unterschieden werden dabei auch die beiden Zielgruppen von Hörgeschädigten mit und ohne Hörhilfe, wobei für beide Gruppen unterschiedliche Signalbearbeitungen durchgeführt werden.

[1] Das Projekt HBB4ALLwird gefördert im Rahmen des 'Competitiveness and innovation framework programme 2007-2013' der Europäischen Union

Erste Hörversuche

Es zeigte sich bereits sehr früh im Projekt, dass es keine allgemeingültige Bearbeitung gibt, welche alle Anforderungen von Hörbeeinträchtigten abdeckt. Dementsprechend wurden verschiedene Bearbeitungen erstellt und mehrere Hörversuche durchgeführt. In den Versuchen wurden immer mehrere Bearbeitungsvarianten für die beiden Gruppen, Hörbeeinträchtigte mit und ohne Hörhilfe, getestet.

Eine erste Reihe an Hörversuchen wurde nach ITU-R BS. 1284-1 [4] erstellt und durchgeführt. Dabei wurde ein ABAB-Vergleich durchgeführt, was bedeutet, dass den Probanden ein TV-Ausschnitt im Original und in einer bearbeiteten Variante im direkten Vergleich, zweimal hintereinander, dargeboten wurde. Die jeweils aktuell gezeigte Variante wurde dabei im Bild immer klar als Original oder Bearbeitung gekennzeichnet. Anhand der in Abbildung 1 gezeigten Skala sollte dann die Höranstrengung und auch das gesamte Hörerlebnis der gezeigten Bearbeitung im Vergleich zum Original eingestuft werden.

	viel schlechter	schlechter	etwas schlechter	gleich	etwas besser	besser	viel besser
Höranstrengung	- 3	-2	-1	0	1	2	3
Hörerlebnis	- 3	-2	-1	0	1	2	3

Abbildung 1: Bewertungsskala der Hörversuche nach dem ABAB-Vergleichsprinzip

Die Hörversuche wurde in einem Raum mit guter Wohnzimmerakustik, über externe Abhörmonitore und einem Flachbildschirm für die Bildwiedergabe durchgeführt. Innerhalb der Versuche wurden mit dem selben Ausgansmaterial verschiedene Bearbeitungsvarianten getestet. Die genaue Versuchsbeschreibung, Durchführung und Auswertung dieser ersten Hörversuche können in [2, 3] nachgelesen werden.

Bei diesen ersten Versuchen konnten speziell für Stereo-Material keine aussagekräftigen Ergebnisse erzielt werden. Ein Grund dafür ist unter anderem, dass einige der Versuchsteilnehmer große Probleme mit der angewandten Methodik hatten. Aufgrund der Rückmeldungen wurde erkannt, dass speziell untrainierte Teilnehmer Schwierigkeiten hatten, die gezeigten Varianten in einem direkten Vergleich mit der Referenz zu beurteilen. Auch die häufige Wiederholung der Inhalte war für die Teilnehmer problematisch. Es fiel ihnen schwer die Ausschnitte zu beurteilen nachdem sie den Inhalt

bereits mehrfach gehört und auch verstanden hatten was gesprochen wurde. Einigen viel es dabei auch schwer dem Inhalt überhaupt noch aufmerksam zu folgen und nicht gleich bei einer bereits zuvor gebildeten Meinung zu bleiben.

Aus diesen Gründen wurde die Notwendigkeit eines anderen Hörversuchskonzepts deutlich.

Neues Hörversuchskonzept

Für die neue Hörversuchsmethodik wurden bestimmte Rahmenbedingungen festgelegt. Es sollte kein direkter Vergleich zwischen dem Original und einer bearbeiteten Variante mehr durchgeführt werden, sondern eine rein qualitative Bewertung einzelner Clips in einem Blindtest. Durch den Blindtest sollte sichergestellt werden, dass die Probanden die Clips ohne das Wissen über die jeweils gezeigte Variante (Original oder Bearbeitung) bewerten. Ein möglicher Einfluss auf die Beurteilung sollte dadurch vermieden werden. Auch der Erinnerungseffekt sollte in dem neuen Versuch verhindert werden, weshalb jeder zu bewertende Inhalt nur einmal abgespielt werden sollte. Wie bereits in den ersten Versuchen sollte auch hier wieder mit realen TV-Inhalten gearbeitet und keine künstlich erzeugten Signale verwendet werden. Das Ziel war es, die Originalversion und drei verschiedene Bearbeitungsvarianten bewerten zu lassen und diese dann in der Auswertung miteinander zu vergleichen.

Durchgeführt wurde der Versuch letztendlich nach dem in Abbildung 2 schematisch dargestellten Prinzip.

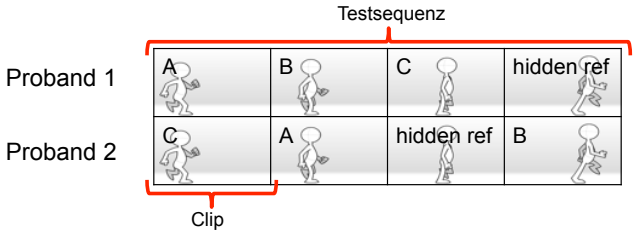

Abbildung 2: Schematischer Aufbau des durchgeführten Hörversuchskonzeptes; A, B und C stehen für drei verschiedenen Bearbeitungsvarianten

Es wurden Testsequenzen von mindestens 25 Sekunden Länge ausgewählt. Diese Mindestlänge war wichtig, damit die einzelnen Clips lang genug warum, um den Probanden die Möglichkeit zu geben sich in den dargebotenen Inhalt einzuhören und eine sinnvolle Beurteilung treffen zu können. Die Testsequenzen wurden auf vier gleich lange Teile aufgeteilt, wobei jeder dieser entstehenden Clips im Versuch unterschiedlich bearbeitet wurde. Für diese Vorgehensweise war es wichtig, dass die Testsequenzen über die komplette Länge eine gleichbleibende Mischung aufwiesen, d.h. möglichst den gleichen Sprecher, sowie die gleiche Hintergrundmusik und -atmo. Damit wurde gewährleistet, dass die vier Clips einer Sequenz ähnliche Inhalte aufwiesen und die Bewertungen dieser Clips innerhalb einer Testsequenz am Ende miteinander verglichen werden konnten. Für jeden Probanden wurde jede Testsequenz nur einmal abgespielt, wobei die Reihenfolge der Bearbeitungen und der hidden reference innerhalb der Sequenz für jeden Probanden zufällig angeordnet wurde. Nach jedem Clip pausierte die Sequenz und die Probanden sollten den gehörten Inhalt anhand der in Abbildung 3 gezeigten vier Bewertungsmöglichkeiten beurteilen. Nach jeder gegebenen Bewertung lief die Sequenz automatisch weiter.

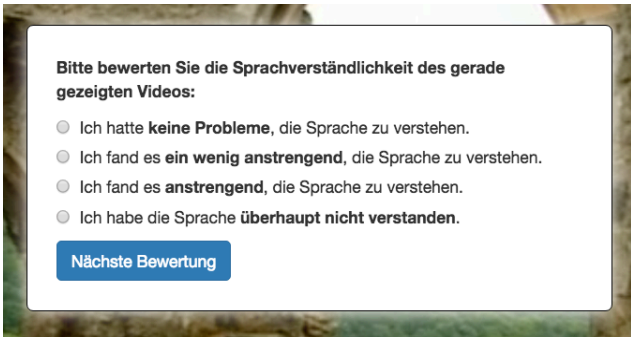

Abbildung 3: Bewertungsmöglichkeiten im Hörversuch für jeden einzelnen dargebotenen Clip

Das im Hörversuch verwendete Material wurde aus TV-Sendungen ausgewählt, die von den Rundfunkanstalten auf Grund von Zuschauerbeschwerden bezüglich der schlechten Sprachverständlichkeit für die Untersuchungen zur Verfügung gestellt wurden. Aus diesem Material wurden einige Sequenzen mit mindestens 25 Sekunden Länge zusammengeschnitten. Von einer kleinen Gruppe Hörgeschädigter wurden aus dieser Vorauswahl an Inhalten 25 Sequenzen als nicht optimal verständlich eingestuft, welche dann für den Hörversuch verwendet wurden. Bei allen Testsequenzen wurde darauf geachtet, dass über die komplette Länge der gleiche Sprecher und die gleiche Hintergrundmusik vorliegen, dass keine kurzzeitigen Signalveränderungen und keine längeren Sprechpausen auftreten, um für alle vier entstehenden Clips eine ähnliche Verteilung der Sprachanteile zu gewährleisten. Bei den Inhalten handelte es sich hauptsächlich um Dokumentationen, da dabei die Voraussetzungen am besten erfüllt werden konnten. Zudem wurden noch Live-Sequenzen und Ausschnitte aus einem Fernsehfilm ausgewählt. Die 25 ausgewählten Testsequenzen wurden den Teilnehmer im Hörversuch in zufälliger Reihenfolge dargeboten.

Vor dem eigentlichen Versuch wurde mit jedem Probanden auch ein kurzes Training mit vier Testsequenzen durchgeführt. Dadurch sollten die Teilnehmer ein Gefühl für die Methodik bekommen und den Inhalt kennenlernen. Die hier verwendeten Sequenzen wurden im eigentlichen Versuch nicht noch einmal benutzt, hatten aber einen ähnlichen Inhalt.

Auswertung der Hörversuchsergebnisse

An dem Hörversuch haben 22 Personen mit Hörbeeinträchtigung ohne Hörgerät teilgenommen. Das Alter der Teilnehmer lag zwischen 32 und 77 Jahren. Für die Auswertung der im Hörversuch erhaltenen Ergebnisse wurden die gegebenen Bewertungen („keine Probleme“, „ein wenig anstrengend“, „anstrengend“ und „überhaupt nicht verstanden“) aller Teilnehmer für alle Clips in den vier Varianten (hidden reference und Bearbeitungen A / B / C) aufsummiert. Dadurch ergibt sich die in Abbildung 4 dargestellte Verteilung.

Bewertung aller Clips

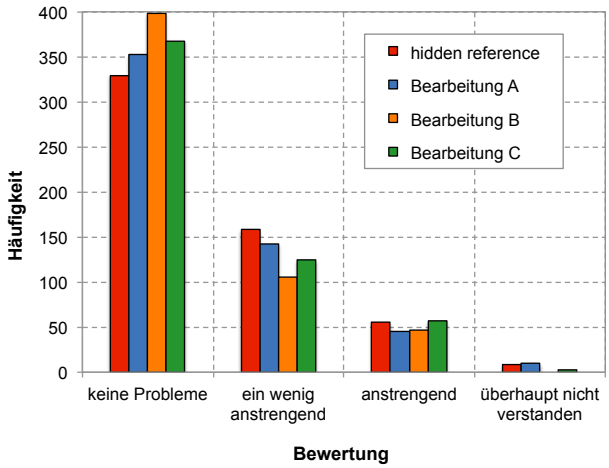

Abbildung 4: Absolute Verteilung der Antworten aller Versuchspersonen für alle Clips, aufgeteilt auf die vier Bewertungsmöglichkeiten

In dieser Darstellung ist zu erkennen, dass für alle vier Varianten, inklusive der hidden reference, größtenteils die Bewertung „keine Probleme, die Sprache zu verstehen“ gegeben wurde. Das zeigt, dass ein Großteil des verwendeten Materials auch ohne Bearbeitung bereits sehr gut verständlich war, aber auch alle drei Bearbeitungsvarianten vorwiegend sehr gut bewertet wurden. Es ist auch zu sehen, dass die Bearbeitungsvariante B bei „keine Probleme“ die meisten Wertungen und bei den restlichen drei Kategorien in der Summe die geringsten erhalten hat. Dies führt zu dem Rückschluss, dass diese Bearbeitung insgesamt für die beste Sprachverständlichkeit sorgt.

Mit Hilfe statistischer Auswertungen, wie dem Chi-Quadrat-Test, konnten unterschiedliche Verteilungen der Bewertungen für die verschiedenen Varianten ermittelt werden. Zwischen der hidden reference und den Bearbeitungen B und C ergab sich im Vergleich ein deutlicher Unterschied in der Verteilung der Bewertungen (p-Wert < 0,01). Die beiden Bearbeitungsvarianten wurden dabei eindeutig besser als die hidden reference bewertet. Zwischen der hidden reference und der Bearbeitung A ist im Gegensatz dazu kein signifikanter Unterschied erkennbar (p-Wert = 0,43). Auch die beiden Bearbeitungen B und C zueinander weisen keinen eindeutigen Unterschied in ihrer Verteilung auf (p-Wert = 0,12). Die Reihenfolge, in der die vier Varianten innerhalb einer Testsequenz angeordnet wurden hatte keinen Einfluss auf die gegebene Bewertung. Dies wurde mit Hilfe eines Unabhängigkeitstests nachgewiesen.

Da in dieser Darstellungsvariante nicht hervorgeht, ob durch eine Bearbeitung eine Verbesserung, oder gar eine Verschlechterung der Verständlichkeit erzielt wurde, wurde eine weitere Auswertung durchgeführt. Dafür wurde den vier Varianten je eine Zahl von 0 für „keine Probleme“ bis 3 für „überhaupt nicht verstanden“ zugewiesen. Anschließend wurden die Differenzen zwischen den Bewertungen für die hidden reference und den drei Bearbeitungen für jede Testsequenz gebildet. Somit konnte für die drei Bearbeitungsvarianten der Unterschied in der Bewertung im Vergleich zur unbearbeiteten Variante innerhalb einer Testsequenz gezeigt werden. Die Ergebnisse dieser Berechnung sind in Abbildung 5 dargestellt.

In dieser Darstellung bedeutet die 0 auf der x-Achse eine gleiche Bewertung der Bearbeitung und der hidden reference innerhalb einer Testsequenz. Die positiven Werte bedeuten eine Verbesserung um je eine, zwei oder drei Bewertungsstufen und die negativen eine Verschlechterung im Vergleich zur hidden reference.

Änderung zur „hidden reference“ aller Sequenzen

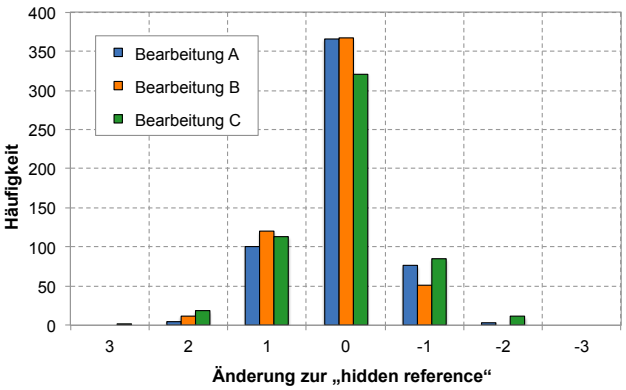

Abbildung 5: Absolute Verteilung aller errechneten Differenzen zwischen der hidden reference und den drei Bearbeitungen innerhalb einer Testsequenz; 0 bedeutet eine gleiche Bewertung der Bearbeitung und der hidden reference, positive Werte bedeuten eine Verbesserung, negative eine Verschlechterung der Verständlichkeit durch die Bearbeitung im Vergleich zur hidden reference

Bei dieser Auswertung ist zu erkennen, dass bei einem Großteil der Sequenzen alle drei Bearbeitungen die gleiche Bewertung wie die hidden reference, also die Differenz 0, erhalten haben. Insgesamt konnten zudem mehr Verbesserungen als Verschlechterungen erzielt werden, wobei vor allem die Bearbeitung B mehr positive als negative Werte erhalten hat. Die Häufung der Differenzergebnisse bei 0 ist wieder auf die in Abbildung 4 zu sehende häufige Bewertung bei „keine Probleme“ zurückzuführen. Da der Großteil der Clips in allen vier Varianten bereits die beste Bewertung erhalten hat, ist eine Verbesserung im Vergleich zur hidden reference gar nicht mehr möglich. Somit ist die Häufung bei der Differenz 0 durchaus als positives Ergebnis einzustufen, da für den Großteil der bereits sehr gut verständlichen Originale durch die Bearbeitung zumindest keine Verschlechterung bewirkt wurde.

Um nun nur jene Sequenzen zu betrachten, bei welchen eine Verbesserung der Verständlichkeit im Vergleich zum Original überhaupt möglich war, sind in Abbildung 6 die Differenzen für die Sequenzen dargestellt, bei denen die hidden reference mit „ein wenig anstrengend“ oder schlechter bewertet wurde.

In dieser Darstellung ist nun gut zu erkennen, dass für einen Großteil der Sequenzen eine Verbesserung erzielt wurde. Nur bei einem geringen Anteil wurde die Verständlichkeit durch die durchgeführte Bearbeitung verschlechtert. Besonders durch die Bearbeitungsvarianten B und C wurden deutlich mehr Verbesserungen als gleichbleibende Bewertungen

oder gar Verschlechterungen im Vergleich zum Original erzielt.

Änderung zur „hidden reference“ für Sequenzen, in denen die „hidden reference“ nicht mit „keine Probleme“ bewertet wurde

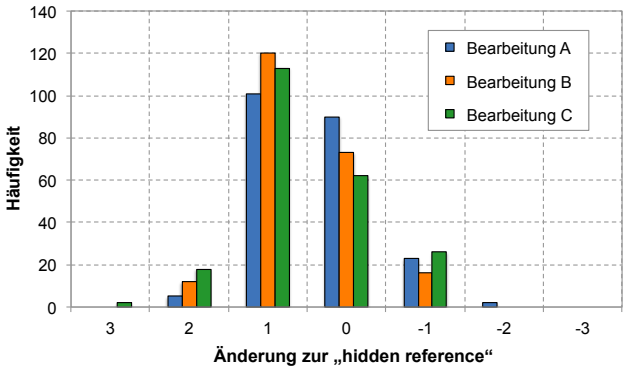

Abbildung 6: Absolute Verteilung der errechneten Differenzen zwischen der hidden reference und den drei Bearbeitungen innerhalb einer Testsequenz für die Sequenzen, bei denen die hidden reference nicht mit „keine Probleme“ bewertet wurde; 0 bedeutet eine gleiche Bewertung der Bearbeitung und der hidden reference, positive Werte bedeuten eine Verbesserung, negative eine Verschlechterung der Verständlichkeit durch die Bearbeitung im Vergleich zur hidden reference

Eine erneute statistische Betrachtung dieser Ergebnisse zeigte, dass auch bei dieser Auswertung die Varianten B und C einen deutlichen Unterschied in ihrer Verteilung aufweisen als die Variante A (p-Wert < 0,01) und im Schnitt auch besser beurteilt wurden. Die beiden Varianten B und C zueinander zeigen auch hier keinen signifikanten Unterschied in ihrer Verteilung (p-Wert = 0,15).

Schlussfolgerung

Zusammenfassend lässt sich festhalten, dass durch die nachträgliche Bearbeitung fertig gemischter TV-Inhalte mit dem Clean Audio Tool eine Verbesserung der Sprachverständlichkeit für Stereosignale erzielt werden kann. Für die hier ausgewerteten Ergebnisse der Hörgeschädigten ohne Hörgerät wurden die Bearbeitungsvarianten B und C signifikant besser bewertet als das Original und die Bearbeitung A. Zwischen den beiden Varianten B und C zueinander konnte kein eindeutiger Unterschied ermittelt werden.

Bei einer kritischen Betrachtung der durchgeführten Versuchsmethodik muss man feststellen, dass ein schlecht verständliches Ausgangsmaterial Grundvoraussetzung für einen erfolgreichen Test ist. Optimal wäre eine gleichmäßige Verteilung der Bewertungen der Original-Clips über alle vier Bewertungskategorien. Dadurch könnte eine mögliche Verbesserung, Verschlechterung oder gleichbleibende Sprachverständlichkeit für unterschiedlich gut verständliches Ausgangsmaterial untersucht werden.

Positiv an der neuen Versuchsmethodik ist, dass die Probanden deutlich weniger Probleme mit der gestellten Aufgabe hatten, als bei der ABAB-Vergleichsmethode. Es fiel ihnen leichter eine einfache qualitative Beurteilung der einzeln angespielten Clips abzugeben, als einen expliziten Vergleich zwischen zwei hintereinander abgespielten Varianten anzustellen. Da jeder Inhalt nur einmal abgespielt wurde, war auch die Aufmerksamkeit der Probanden bei jedem Clip höher und mehr auf den Inhalt gerichtet, als bei einer häufigen Wiederholung der gleichen Inhalte. Die neue Methodik wurde als weniger ermüdend und einfacher in der Bewertung beschrieben.

Neben der einfacheren Handhabung für die Probanden gestaltete sich bei der neuen Methodik im Vergleich zur vorher durchgeführten allerdings die Suche nach geeignetem Versuchsmaterial als deutlich schwieriger. Die hohen Anforderungen an das Material, die notwendig sind, um den Versuch sinnvoll durchführen zu können, werden von vielen TV-Inhalten nicht erfüllt.

Alles in allem bietet die neue Versuchsmethodik eine gute Möglichkeit verschiedenen Bearbeitungsvarianten bezüglich ihrer Sprachverständlichkeit bewerten zu lassen und in der Auswertung miteinander zu vergleichen. Es zeigt sich, dass die Methodik auch für unerfahrene Probanden geeignet ist und zu aussagekräftigen Ergebnissen führt.

Literatur

[1] Das Projekt HBB4ALL erhält im Rahmen des EU-ICT-PSP-Programms unter ID 621014 Zuwendungen der Europäischen Gemeinschaft - http://www.hbb4all.eu/

[2] Weitnauer M., Bohne M.: Automatic and customizable improvement of the speech intelligibility from TV signals for hearing impaired people. TMT, Köln, 2014

[3] Weitnauer M., Goossens S., Liebl T.: Anpassbarer Fernsehton für Hörgeschädigte über HbbTV 2.0. DAGA, Nürnberg, 2015

[4] ITU-R BS. 1284-1 General methods for the subjective assessment of sound quality - https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1284-1-200312-I!!PDF-E.pdf