

Sprachqualität von Fernseh-Ton

Theresa Liebl¹, Sebastian Goossens¹

¹ *Institut für Rundfunktechnik, 80939 München, E-Mail: liebl@irt.de, goossens@irt.de*

Einleitung

Das Thema Sprachqualität und dabei vor allem die Verständlichkeit von Fernseh-Ton beschäftigt die öffentlich-rechtlichen Rundfunkanstalten schon seit einigen Jahren. Die Beschwerden seitens der Zuschauer in Bezug auf die Verständlichkeit des TV-Programms haben immer weiter zugenommen. Auch in den Medien werden viele TV-Programme immer wieder kritisiert. Schlagzeilen wie „Schlechte Tonqualität im TV – Das große Nuscheln“ [1] und „Wo bleibt der gute Ton im "Tatort"?" [2] nähren die Kritik der Zuschauer. Bei den Rundfunkanstalten besteht hier also dringend Handlungsbedarf und es werden verschiedenste Ansätze verfolgt, um das Problem zu lösen.

Messung der Sprachverständlichkeit

Es existieren viele etablierte Methoden zur Messung von Sprachverständlichkeit, wobei eine Anwendung dieser zur Messung der Verständlichkeit von Fernseh-Programmen nicht möglich ist. Sie liefern hier nur eingeschränkt aussagekräftige Ergebnisse, da im Fernsehbereich in den meisten Fällen der Bereich der messbaren Verständlichkeit verlassen wird. Die Probleme im Fernsehen liegen vermehrt im Bereich der Höranstrengung, was bedeutet, dass eine Verständlichkeit durchaus gegeben sein kann, es aber viel zu anstrengend ist, dem Gesprochenen zu folgen.

Zudem wird die Sprachqualität im Fernsehen von sehr vielen unterschiedlichen Faktoren beeinflusst, die es schwierig machen, an nur einer Stelle in der Kette eine aussagekräftige Messung oder Beurteilung durchzuführen. Diese vielen verschiedenen Einflussfaktoren machen die Verständlichkeit von Fernseh-Ton zu einem sehr komplexen Thema und nur eine Betrachtung aller möglichen Einflüsse kann am Ende eine deutliche Verbesserung erzielen.

Beeinflussung der Sprachqualität im TV

Aufnahmeseitig:

Schon bei der Planung des Drehs sowie der Aufnahme des Tones vor Ort können viele Faktoren die Verständlichkeit der Sprachaufnahme beeinträchtigen.

Bei der Wahl des Drehortes beispielsweise sollten einige Dinge beachtet werden, die Einfluss auf die Sprachverständlichkeit nehmen können. Eine Aufnahme in z.B. einer Kirche mit viel Hall oder an einer Baustelle mit hoher Lärmabstrahlung kann die Aufnahme schon in einer beeinträchtigenden Weise „verunreinigen“. Und natürlich spielt auch die Mikrofonierung und Art der Tonabnahme eine entscheidende Rolle. Der Verzicht auf Ansteckmikrofone, aus weiter Entfernung geangelter Ton, um für Videoaufnahmen auch eine Totale zu ermöglichen, oder eine Tonaufnahme ohne wirkliche Tonprobe auf Grund von Zeitmangel können Probleme verursachen. Aber auch

die Sprecher selbst haben natürlich die Verständlichkeit ihres Gesprochenen in der Hand. Eine genuschelte Unterhaltung, ein Interview in starkem Dialekt gesprochen oder auch nur eine zu schnelle Sprechweise können Schwierigkeiten bei der Verständlichkeit verursachen.

Es gibt also schon aufnahmeseitig einiges zu berücksichtigen und viele Faktoren, die Einfluss auf die am Ende beim Konsumenten ankommende Sprachqualität nehmen.

In der Mischung:

Im Weiteren kann die Verständlichkeit der Sprachaufnahme auch in der Mischung beeinträchtigt werden. Je nach Szene wird die Sprache mit Musik, Geräuschen und Effekten hinterlegt, wobei schon allein der Pegel, bei dem diese Elemente der Sprache zugemischt werden, einen großen Einfluss nimmt. Aber nicht nur der Pegelabstand der einzelnen Elemente, sondern auch die Wahl des Hintergrundes können die Verständlichkeit verschlechtern. Zu dynamische Elemente, die dramatisch wirken sollen, oder Soloinstrumente wie Trompeten oder Schlaginstrumente, sowie Musik mit Gesang wirken sich negativ auf die Verständlichkeit aus. Hinzu kommt, dass es beim Verstehen von Sprache einen Gewöhnungseffekt gibt, was dazu führt, dass eine einmal verstandene Sprachaufnahme für eine Person weiterhin verständlich bleibt, auch wenn Störgeräusche hinzukommen, die die Verständlichkeit soweit beeinträchtigen, dass beim erstmaligen Hören das Gesprochene nicht verstanden wird. Dieser Effekt erschwert die Beurteilung der Verständlichkeit einer Mischung, wenn diese während der Erstellung bereits mehrfach gehört wurde.

Zusätzlichen Einfluss auf die Beurteilung der Verständlichkeit einer Mischung im Studio nehmen auch der dort verwendete Mischpegel und die Ausstattung der Räumlichkeiten. Eine bei hohem Pegel im Studio mit optimalen akustischen Raumbedingungen auf hochwertigen Studio-Monitoren erstellte Mischung kann unter diesen Bedingungen durchaus gut verständlich sein. Bei einer Wiedergabe dieser Mischung mit deutlich geringerem Pegel im Wohnzimmer mit raumakustisch schlechteren Eigenschaften auf einem Flachbildfernseher wird sich die Verständlichkeit der Mischung natürlich verändern. Daher empfiehlt es sich auch im Studio die erstellte Mischung auf einem Konsumergerät bei geringerem Pegel gegen zu hören und sie im Idealfall von „frischen Ohren“ beurteilen zu lassen.

Wiedergabe:

Aber nicht nur senderseitig bei der Erstellung von Beiträgen kann die Sprachverständlichkeit beeinflusst werden, auch bei den Zuschauern zu Hause kann noch einiges passieren.

Schon das Wiedergabegerät, welches vom Konsumenten verwendet wird, spielt hier eine Rolle. Verschiedene

Audiowiedergabe-Geräte liefern deutlich unterschiedliche Qualitäten bei der Sprachwiedergabe. Aber nicht nur die Geräte an sich, sondern auch deren Audioeinstellungen und Platzierung im Wohnzimmer nehmen Einfluss. Beispielsweise kann ein Flachbildfernseher, der seinen Ton nach hinten abstrahlt und eng in einer Schrankwand steht keine gute Verständlichkeit des Inhaltes liefern. Zudem spielen auch der Pegel und die Raumakustik des Wohnzimmers, wie bereits erwähnt, eine entscheidende Rolle. Bei deutlich geringerem Pegel und schlechteren raumakustischen Eigenschaften als im Studio leidet auch die Verständlichkeit.

Ein weiterer Aspekt, der nicht vernachlässigt werden kann, ist der Zuschauer selbst. Altersabhängiger Verschleiß des Gehörs, Hörschädigungen, oder auch eine andere Muttersprache, als die im Fernsehen gehörte, beeinträchtigen die Verständlichkeit von Fernsehinhalten.

Audio-Kette:

All diese Elemente - und noch mehr - können Einfluss auf die Verständlichkeit von Sprache im Fernsehen nehmen. Um eine gute Sprachverständlichkeit beim Endnutzer zu gewährleisten, reicht es also nicht aus nur Teile dieser Kette zu betrachten. Eine optimal verständliche Aufnahme kann beim Zuschauer dennoch unverständlich ankommen, wenn in der Mischung etwas nicht beachtet wird oder die Wiedergabebedingungen beim Zuschauer zu Hause nicht gut sind. Andererseits kann eine schon bei der Aufnahme unverständliche Sprache im Nachhinein nicht wieder gut verständlich gezaubert werden und auch das beste Wiedergabesystem kann hier nicht helfen.

Sprachverständlichkeits-Guideline:

Viele dieser Aspekte und weitere wurden von ARD und ZDF 2014 bereits in einer Empfehlung für Programm und Technik [3] zusammengetragen. Diese Empfehlung ist frei verfügbar und soll vor allem in den Rundfunkanstalten sensibilisieren und motivieren, um das Hörerlebnis für die Zuschauer wieder zu verbessern. Der Ton soll dabei in allen Produktionsstufen als zentraler Programmteil berücksichtigt und in jedem Bereich auf eine gute Verständlichkeit geachtet werden.

Untersuchungsergebnisse

Am IRT wird das Thema Sprachverständlichkeit im Fernsehen schon seit einigen Jahren intensiv untersucht, wobei viele verschiedene Aspekte beleuchtet wurden. Einige dieser Untersuchungen werden im Folgenden kurz dargestellt und für nähere Informationen auf die jeweiligen Veröffentlichungen dazu verwiesen.

„Voice-Over-Voice“ [4]:

Unter Voice-Over-Voice versteht man das gleichzeitige Auftreten zweier Sprachsignale, wie es beispielsweise bei Übersetzungspassagen in Magazinbeiträgen der Fall ist. Bei solchen Passagen gibt es häufig Beschwerden der Zuschauer bezüglich der Sprachverständlichkeit des Off-Sprechers. Es wurden daher Untersuchungen durchgeführt, die sich speziell mit dem Mischungsverhältnis von Off-Sprecher und O-Ton beschäftigen.

Im Rahmen eines Hörversuchs wurden mehrere mögliche Einflussfaktoren auf die Sprachverständlichkeit von Übersetzungs-Passagen beleuchtet. Die Ergebnisse des Hörversuchs zeigten, dass es nicht möglich ist eine allgemein gültige Aussage zu treffen mit welchem Pegelabstand zwei Sprachspuren idealerweise gemischt werden sollten (siehe Abb. 1). Vielerlei unterschiedliche Einflussfaktoren bestimmen die Verständlichkeit der Mischungen.

So sind der Geschmack und die Toleranz eines jeden Zuschauers anders, wodurch sich große Unterschiede in den gewünschten Pegelabständen ergeben. Deutlich zu erkennen war aber auch der Einfluss unterschiedlich stark störender O-Töne. Nicht jede Originalsprache ist gleich störend für die Verständlichkeit der Übersetzung und auch die Abhörlautstärke beeinflusst das gewünschte Ergebnis.

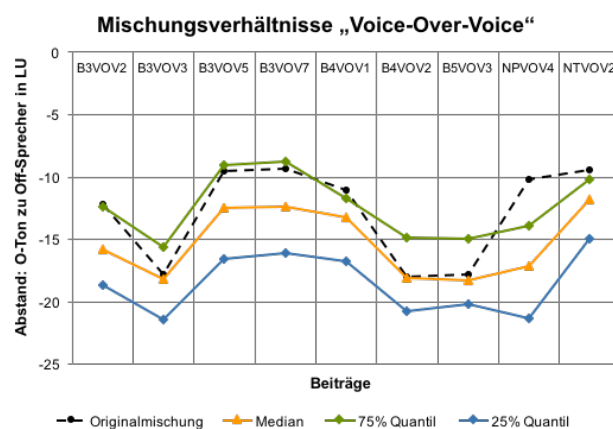


Abbildung 1: Gegenüberstellung der O-Ton-Absenkung aus der Originalmischung zum Median, sowie den 25%- und 75%-Quantilen der ermittelten Werte aus dem Hörversuch, gemittelt über alle Versuchsteilnehmer.

Clean Audio Service für hörgeschädigte Nutzer [5, 6]:

Besonders für hörgeschädigte Personen ist die Verständlichkeit des gesprochenen Dialogs im Fernsehen ein wichtiger Punkt. In einem EU-Projekt (HBB4ALL [7]) wurde speziell der barrierefreie Zugang zu Medien für alle Bürger über HbbTV behandelt. Ziel des Projektes war es unter anderem die Sprachverständlichkeit von TV-Produktionen für Hörgeschädigte zu verbessern.

Dabei wurde zuerst eine Umfrage mit der Zielgruppe durchgeführt, um zu ermitteln, ob es spezielle Faktoren gibt, welche die Sprachverständlichkeit im aktuellen TV-Programm besonders beeinflussen und welche Maßnahmen zur Verbesserung ergriffen werden sollten. Außerdem wurde ein Verfahren entwickelt, welches die Verständlichkeit von typischen und bereits fertig produzierten Stereo- und 5.1-Fernsehsignalen verbessern soll.

Bei Hörversuchen zur Evaluierung der Effektivität dieses Verfahrens zeigte sich, dass es keine allgemeingültige Bearbeitung gibt, welche alle Anforderungen von Hörbeeinträchtigten abdeckt und eine Verbesserung der Verständlichkeit für alle bewirkt. Dementsprechend wurden verschiedene Bearbeitungen erstellt und mehrere Hörversuche durchgeführt. Allerdings konnten auch mit den

verschiedenen Bearbeitungsvarianten nur leichte Verbesserungen der Verständlichkeit erzielt werden (Abb. 2). Die Effektivität der Bearbeitung hängt auch hier sehr stark vom jeweiligen Ausgangssignal ab und kann nicht automatisch auf jedes Material angewandt werden.

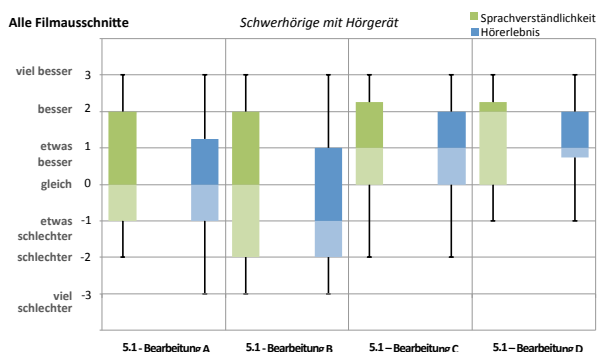


Abbildung 2: Bewertung der Sprachverständlichkeit sowie des Hörerlebnisses verschiedener Bearbeitungsvarianten im Vergleich zur Originalmischung am Beispiel von 5.1-Signalen und Schwerhörigen mit Hörgerät als Probanden.

3.0 Tonkanalformat [8]:

Bei den Bemühungen die Sprachverständlichkeit im Fernsehen zu verbessern wurde auch die Einführung eines Tonkanalformats 3.0 diskutiert. Bei einer 3.0-Mischung sollte der Center-Kanal möglichst nur Sprache enthalten. Die Überlegung war, ob allein dadurch, dass die Sprachanteile separat im Center-Kanal liegen, eine verbesserte Verständlichkeit auf dem richtigen Wiedergabesystem erzielt werden kann. Es wurden verschiedene Aspekte von 3.0 untersucht und betrachtet, ob beim Zuschauer zu Hause eine Verbesserung der Verständlichkeit erreicht werden kann.

Die Ergebnisse zeigten, dass eine 3.0-Mischung im Vergleich zu Stereo nicht eindeutig besser bewertet wurde, was die Verständlichkeit betrifft. Es konnte jedoch gezeigt werden, dass der separate Sprachkanal einer 3.0-Mischung einen großen Vorteil bietet, da dieser auf dem richtigen Wiedergabegerät vom Zuschauer individuell in der Lautstärke verändert und somit vom Rest der Mischung hervorgehoben werden kann. Diese Möglichkeit wurde von allen Probanden genutzt, um die Verständlichkeit zu erhöhen. Zu beachten ist allerdings, dass nur wenige untersuchte Geräte diese Möglichkeit der Ansteuerung des Center-Kanals überhaupt anbieten.

Wiedergabelautstärke [9]:

Im Rahmen der Untersuchungen zum Tonkanalformat 3.0 wurde auch der Einfluss verschiedener Pegel, bei denen gemischt bzw. abgehört wird, genauer untersucht. Dabei wurde betrachtet, ob der Abhörpegel den gewählten Center-Kanal-Pegel einer 3.0-Mischung unter Betrachtung einer gut verständlichen Mischung beeinflusst. Als feste Pegel wurden 50, 60 und 70 dB(A) gewählt, da Untersuchungen [4] gezeigt haben, dass 50 dB(A) die durchschnittliche Abhörlautstärke für TV-Inhalte in einer Wohnzimmerumgebung sind und 70 dB(A) die mittlere Abhörlautstärke beim Mischen im Produktionsstudio.

Bei dieser Untersuchung konnte gezeigt werden, dass die verschiedenen Abhörpegel einen deutlichen Einfluss auf das gewünschte Mischungsverhältnis zwischen Sprache und Hintergrund haben, wenn eine gute Verständlichkeit von Fernsehinhalt gewährleistet werden soll. Je höher der Abhörpegel, desto leiser wurde der Center-Kanal mit Sprache im Verhältnis zur Atmo eingestellt. Mischungsverhältnisse sind also nicht für alle Wiedergabelautstärken gleichwertig. Eine bei höherem Abhörpegel erstellte Tonmischung kann im Studio gut verständlich sein, bei einer leiseren Wiedergabe im Wohnzimmer ist jedoch ein lauterer Center-Kanal-Pegel nötig, um eine gleichwertige Höranstrengung zu gewährleisten.

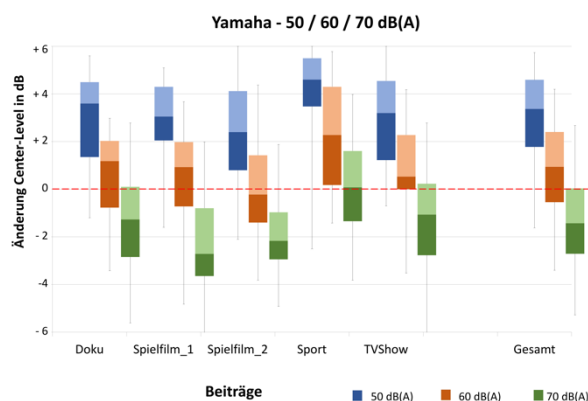


Abbildung 3: Eingestellter Center-Kanal-Pegel für die drei untersuchten Abhörpegel 50, 60 und 70 dB(A).

Fazit der Untersuchungen:

Unter Betrachtung der zuvor aufgezeigten Audio-Kette und der möglichen Faktoren, welche die Verständlichkeit entlang dieser Kette beeinflussen können, sowie der Ergebnisse der hier dargestellten Untersuchungen ergibt sich die Schlussfolgerung, dass nur individuelle Lösungen eine effektive Verbesserung der Sprachverständlichkeit im Fernsehen liefern können. Eine Möglichkeit dies umzusetzen, wäre die Übertragung eines separaten Sprachkanals, welcher ausschließlich Sprache enthält und von den Zuschauern zuhause individuell im Pegel verändert werden kann. Dies umzusetzen wäre aktuell allerdings schwierig, da ein separater Sprachkanal von Produktionen meist nicht vorhanden ist und auch nicht einfach gesendet werden kann. Zudem ist die manuelle Steuerung des Center-Kanals einer 3.0- oder auch 5.1-Mischung bisher nur auf wenigen Endgeräten möglich. Eine Lösung für diese Problemstellungen wäre eine Umsetzung mit objektbasierter Audio-Produktion.

Objektbasierte Audio-Produktion

Aktuell finden Fernsehproduktionen rein kanalbasiert statt. Dabei werden mehrere Audiospuren in einer DAW gemischt und in einer festen Kanalbelegung (z.B. Stereo oder 5.1) gespeichert und zum Zuschauer übertragen.

Bei einer objektbasierten Audioproduktion (Abb. 4) wird eine in der DAW erstellte Mischung nicht mehr in einer festen Kanalbelegung gespeichert, sondern in Form von Einzelspuren, sog. Objekten, mit verknüpften Metadaten als

objektbasierte Audioszene. Die Metadaten enthalten die nötigen Informationen für die erstellte Mischung, wie Pegel und Position der einzelnen Objekte. Ein Renderer übernimmt am Ende dann die Zusammensetzung der gespeicherten Szene in die gewünschte Mischung, angepasst auf das jeweils angeschlossene Wiedergabesystem. Diese Produktionsform bietet die Möglichkeit, auch nach der Erstellung der Mischung noch individuelle Anpassungen vorzunehmen.

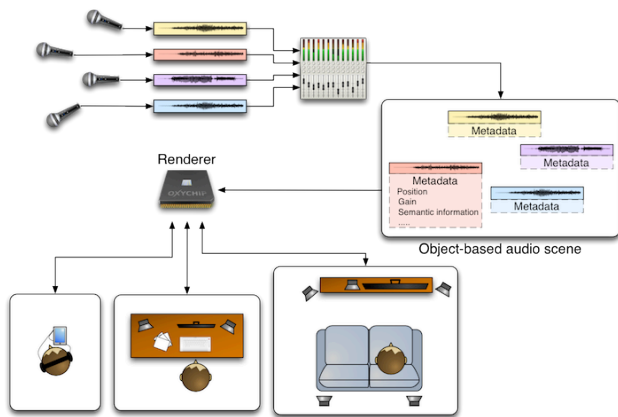


Abbildung 4: Schematische Darstellung einer objektbasierten Audioproduktion.

Diese Art der Produktion und Übertragung wäre in Hinblick auf die Sprachverständlichkeit sehr hilfreich, da das Mischungsverhältnis von separat übertragenen Sprach- und Atmo-Objekten, versehen mit Metadaten, beim Konsumenten zu Hause individuell verändert werden könnte. Der Ersteller der Mischung könnte in den Metadaten auch festlegen, in welchem Bereich eine solche Veränderung der Mischung durchgeführt werden darf. Was eine schnelle Umsetzung dieser Technik jedoch erschwert, ist, dass eine Anpassung in der Aufnahme und Produktion stattfinden müsste. Außerdem ist die Möglichkeit den Pegel von Einzelobjekten zu verändern von den Endgeräten abhängig.

Aktuelles

Am IRT wird weiter daran gearbeitet, eine objektbasierte Audioproduktion und Übertragung im Rundfunk einzuführen. Daneben werden auch weiterhin verschiedenste Untersuchungen zum Thema Sprachverständlichkeit durchgeführt. Darunter wird beispielsweise angestrebt, ein Verfahren zu entwickeln, welches Sprache aus komplexen Audiosignalen extrahiert. Diese generierte Sprachspur soll anschließend zur Erzeugung von in der Sprachverständlichkeit verbesserten Mischungen verwendet werden. Hierzu wird aktuell eine breite Recherche zu bestehenden Tools zu Quelltrennung und Sprachseparation durchgeführt und vielversprechende Tools getestet.

Wiedergabeseitig wird daran gearbeitet HbbTV 2 zu nutzen, um speziell aufbereitete Tonvarianten zeitsynchron zum eigentlichen Broadcast-Ton wiederzugeben. Dadurch könnte den Zuschauern schon jetzt die Möglichkeit geboten werden, alternative Tonvarianten zu nutzen.

Neben eigenen Entwicklungen wird auch untersucht, welche Möglichkeiten es bereits gibt, die Sprachverständlichkeit

wiedergabeseitig zu verbessern. Einige Endgerätehersteller bieten bereits individuelle Anpassungsmöglichkeiten. Diese Endgeräte werden untersucht, um festzustellen inwieweit sie einen Mehrwert für die Zuschauer bieten können.

Fazit

Die Betrachtung der gesamten Audio-Kette von der Aufnahme über die Mischung bis hin zum Endnutzer zeigt, wie komplex das Thema Sprachverständlichkeit im Fernsehen ist, und dass es viele Möglichkeiten gibt diese zu beeinflussen. Bei dem Bestreben die Verständlichkeit im Fernsehen zu verbessern, sollte daher immer die komplette Kette bedacht werden.

Es wird deutlich, dass eine effektive Verbesserung nur durch individuelle Lösungen erreicht werden kann, was im Idealfall durch die Übertragung einer sauberen separaten Sprachspur möglich wäre. Eine Umsetzung wäre am besten durch die Etablierung der objektbasierten Audioproduktion möglich, da diese alle Möglichkeiten mit sich bringt, um den Zuschauern zu Hause individuelle Lösungen zu bieten.

Literatur

- [1] Bremmer, A., Schlechte Tonqualität im TV: Das große Nuscheln
URL: <http://www.tagesspiegel.de/medien/schlechte-tonqualitaet-im-tv-das-grosse-nuscheln/13457678.html> (März 2019)
- [2] Gasteiger C., Wo der gute Ton im "Tatort" bleibt
URL: <https://www.sueddeutsche.de/medien/schlecht-verstaendliche-krimis-wo-bleibt-der-gute-ton-im-tatort-1.3013234> (März 2019)
- [3] ARD/ZDF - Sprachverständlichkeit im Fernsehen Empfehlungen für Programm und Technik zur Sprachverständlichkeit im Fernsehen (Juli 2014)
URL: <https://www.irt.de/publikationen/technische-richtlinien/technische-richtlinien-download/>
- [4] Liebl, T., Goossens, S., Krump, G., „Verbesserung der Sprachverständlichkeit, speziell bei Voice-Over-Voice-Passagen“. Tonmeisterstagung, Köln, 2014
- [5] Weitnauer M., Goossens S., Liebl T.: „Anpassbarer Fernsehton für Hörgeschädigte über HbbTV 2.0“. DAGA, Nürnberg, 2015
- [6] Liebl, T., Weitnauer M., Meier, M.: „Evaluierung eines Ansatzes zur Verbesserung der Sprachverständlichkeit von Stereosignalen im Rundfunk“. DAGA, Aachen, 2016
- [7] HBB4ALL: URL: <http://www.hbb4all.eu/>
- [8] Liebl, T., Heinz, S., Hartmann, C., „Kann durch das Tonkanalformat 3.0 die Sprachverständlichkeit im Fernsehen verbessert werden?“. DAGA, Kiel, 2017
- [9] Liebl, T., Hartmann, C., „Verständlichkeit von Fernsehton bei unterschiedlichen Abhörpegeln und veränderbarem Sprache-Hintergrund-Verhältnis“. DAGA, München, 2018