Höranstrengung von TV-Mischungen in Abhängigkeit von charakteristischen Hintergrundsignalen

J. Rennies[1] , H. Baumgartner[1], A. Volgenandt[1], Nathan Wiedemann[1], M. Kahsnitz[2]

[1]*Fraunhofer Institute for Digital Media Technology IDMT, 26129 Oldenburg, E-Mail:Hannah.Baumgartner@idmt.fraunhofer.de*

[2]*RTW GmbH, 50829 Köln, E-Mail: mkahsnitz@rtw.de*

Einleitung

In Schleswig-Holstein initiierten der Landesseniorenrat und die Ärztekammer eine „Kampagne für deutliche Sprache“. Leserbriefe und Beschwerden wegen schlechter Sprachverständlichkeit im Rundfunk häufen sich bei den Programmverantwortlichen. Besonders aufwendig und budgetintensiv produzierte Spielfilme wie z.B. „Tatort“ oder „Polizeiruf“ stehen häufig in der „Sprachverständlichkeits-Kritik“. Die Problematik ist komplex, die Ursachen nicht eindeutig. Es gibt keine etablierte objektive Messung für Sprachverständlichkeit, keine Anzeige, die die Qualität einer TV-Mischung bezüglich der Verständlichkeit ihrer Dialoge überprüft. Computergestützte Hörmodelle könnten eine objektive Vorhersage von Sprachverständlichkeit in Zukunft ermöglichen. Es ist bekannt, dass die Sprachverständlichkeit im Allgemeinen mit Verbesserung des SNR steigt und sich die Höranstrengung verringert [1]. Darüber hinaus zeigen diese Effekte aber starke Abhängigkeiten von der Art des Hintergrundgeräusches. Stationäre oder tonale Störgeräusche zeigen bspw. andere Störwirkungen auf die Sprachverständlichkeit als instationäre oder rauschhafte Hintergrundgeräusche, z.B. [2, 3].

Im Anwendungsfeld Film und Fernsehen sind Störgeräusche bzw. Hintergrundsignale (Atmos) annähernd so vielseitig wie die echte „akustische Welt“. Das Ziel dieser Studie war die Erfassung der subjektiv empfundenen Höranstrengung für Fernsehclips mit diversen Hintergrundarten und systematisch veränderten Mischungsverhältnissen mit normal- und schwerhörenden Probanden.

Experiment

In dieser Studie wurde die Höranstrengung von TV-Ausschnitten mit diversen Hintergründen (Atmos) in unterschiedlichen Mischverhältnissen erfasst. Die Höranstrengungsdaten wurden in Bezug auf die Auswirkungen signalspezifischer Störeigenschaften in unterschiedlichen Signal-zu-Hintergrund-Abständen analysiert. Grundsätzliche Forschungsfragen waren dabei:

- Wie bewerten die Probanden die Höranstrengung (Listening Effort) bezüglich der unterschiedlichen Hörproben und Konditionen?
- Wird die Höranstrengung für reine Audiosignale anders bewertet als für audio-visuelle Signale?
- Wie unterschiedlich bzw. ähnlich bewerten unterschiedliche Probanden (mit der gemeinsamen Eigenschaft „normalhörend“ / „schwerhörend“) Höranstrengung?
- Welchen Einfluss bzgl. der empfundenen Höranstrengung zeigen unterschiedliche Mischungsverhältnisse (SNR) und unterschiedliche Mischsignale (Sprachsignal/ Hintergrund)?
- Zusätzlich: Was sind die individuell bevorzugten Abhörpegel (leise, angenehm, laut) und wie sehr variieren diese Pegel interindividuell bzw. intraindividuell?

Probanden: Insgesamt 20 normalhörende Versuchspersonen (Alter: 20-30 Jahre, 10w/10m; Median = 20Jahre) und 10 schwerhörende Probanden (Alter: 61 - 83 Jahre, 2w/8m; Median = 74 Jahre) nahmen an den Hörversuchen teil. Mit allen Versuchspersonen wurde ein aktuelles Audiogramm gemessen. Der durchschnittliche Pure Tone Average (PTA, gemittelter Hörverlust bei den Frequenzen 500, 1000, 2000 und 4000 Hz) der normalhörenden Kohorte lag bei etwa 1dB HL, der durchschnittliche PTA der schwerhörenden Gruppe bei 42dB HL.

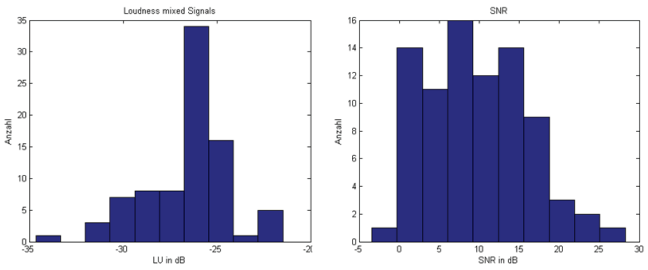

Abbildung 1: Histogramme zu Loudness-Verteilung (links) und Sprache-zu-Hintergrund-Verhältnissen (rechts) in dB LU der verwendeten TV-Clips.

Testmaterial: Als Testmaterial dienten 83 Clips von etwa 12s Dauer. Die Clips wurden aus Original-Fernsehproduktionen erstellt (Dokumentarfilm, Sportstudio, Polizeiruf). Die ausgesuchten Clips bestanden immer aus Sprachsequenzen gekoppelt mit Atmo. Manche Clips enthielten klar und neutral gesprochene Sprache im Stil von Berichterstattung (Interviews, Voice Over), andere eher emotional gefärbte Sprache (Flüstern, Schreien, Nuscheln, etc.). Die Clips wurden so zusammengestellt, dass weibliche und männliche Stimmen gleichermaßen vorkamen. Auch die Hintergründe waren divers: Musik, Geräusche, Stimmen, etc.

Versuchs-Konditionen: Die Clips wurden in jeweils fünf unterschiedlichen Konditionen getestet – jede Versuchsperson musste also insgesamt 415 Clips bzgl. der Höranstrengung bewerten.

- Kondition1: Clips mit SNR aus ursprünglicher TV-Mischung ohne Bild („Orig-Audio“)
- Kondition2: Clips mit um 6dB verschlechterten SNR als Original-Mischung ohne Bild („-6dB-Audio“)

- Kondition3: Clips mit einem festen SNR von -3dB ohne Bild („minus 3dB-Audio“)
- Kondition4: Clips mit SNR aus ursprünglicher TV-Mischung mit TV-Bild („Orig-AV“)
- Kondition5: Clips mit um 6dB verschlechterten SNR als Original-Mischung mit TV-Bild („-6dB-AV“)

Die Lautheit der Quellen entsprach den Empfehlungen der EBU zur Lautheitsnormalisierung [4]. Die SNR wurden aus der Differenz der Lautheit der Sprache und der Lautheit der Atmo berechnet und eingestellt (Sprache-zu-Hintergrund-Verhältnisse in dB LU). Die Verteilung der SNR der einzelnen Clips zeigt das Histogramm in Abbildung 1. Die Reihenfolge der Darbietungen innerhalb einer Session war für jede VP randomisiert.

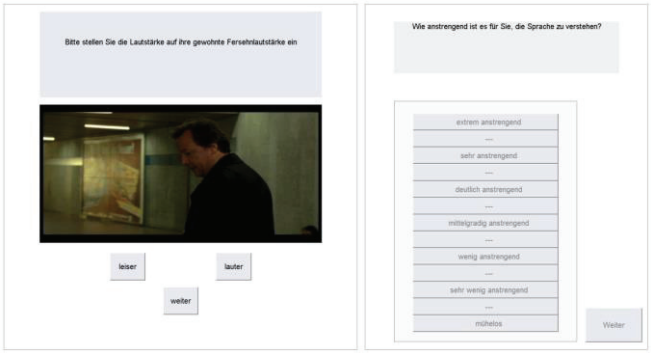

Abbildung 2: Nutzerschnittstelle des Höranstrengungstests.

Durchführung: Der Hörversuch bestand aus zwei Terminen von etwa zwei Stunden Dauer, jeder Termin ließ sich wiederum in zwei Sessions unterteilen. Zu Beginn des ersten Termins wurde das aktuelle Audiogramm aufgenommen. Die Probanden wurden instruiert, nach jeder Signaldarbietung einzuschätzen, wie anstrengend das Verstehen der Sprache für sie war. Hierfür stand eine 13-teilige kategoriale Höranstrengungsskala zur Verfügung, deren Kategorien von „mühelos“ bis „extrem anstrengend“ reichen (Abbildung 2). Jeder dieser Kategorien war dabei ein für die Probanden nicht sichtbarer numerischer Wert in der Einheit „Effort Scaling Categorical Unit (ESCU)“ [1] zugeordnet, wobei 1ESCU „mühelos“ und 13ESCU „extrem anstrengend“ entspricht. Das Experiment fand unter Laborbedingungen in einer Hörkabine und mit Kopfhörer statt. Jede VP wurde aufgefordert im Vorfeld des eigentlichen Versuchs sich mit Hilfe dreier Testsignale einen „leisen“, einen „lauten“ und einen „angenehmen“ Wiedergabepegel einzustellen. Für die Bewertung der Höranstrengung wurde die Wiedergabelautstäke individuell auf den „angenehmen“ Pegel justiert.

Ergebnisse

Individuell bevorzugte Abhörpegel: Mittlere Einstellpegel (Interquartilbereiche) der normalhörenden Probanden liegen für

- „leise“ zwischen 50-59dB SPL,
- für „angenehm“ zwischen 65-69dB SPL und
- für „laut“ zwischen 69-79dB SPL.

Zwischen „laut“ und „leise“ liegen im Mittel etwa 18dB. Die Korrelationskoeffizienten für die Einstellvorgänge der zwei Sessions liegen zwischen 0.7 und 0.8, die über alle Pegeleinstellungen bei 0.9 – was einer guten bis sehr guten Korrelation entspricht. Die individuellen Abweichungen von Einstellung zu Einstellung sind für die „leise“-Pegel mit durchschnittlich ~7dB am höchsten.

Bei den schwerhörenden Probanden liegen die durchschnittlichen Einstellpegel (Interquartilbereiche) für leise deutlich höher:

- „leise“ zwischen 61-67dB,
- für „angenehm“ zwischen 67-69dB und
- für „laut“ zwischen 69-73dB.

Zwischen „laut“ und „leise“ liegen im Mittel nur etwa 7dB. Die verbleibende Dynamik der SH-Gruppe zwischen laut und leise ist deutlich reduziert. Die Korrelations-koeffizienten für die Einstellvorgänge der zwei Sessions liegen zwischen 0.3 für „laut“ und 0.9 für „leise“, die über die Gesamtheit aller Pegeleinstellungen bei 0.8 – was einer mittleren Korrelation entspricht. Die individuellen Abweichungen von Einstellung zu Einstellung sind für die „angenehm“-Pegel mit im Schnitt ca. 5dB am höchsten.

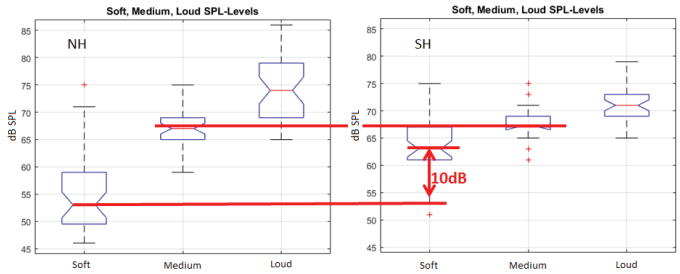

Abbildung 3: Zwischen „laut“ und „leise“ liegt für die normalhörende Gruppe (links) im Mittel 19dB. Für die schwerhörende Gruppe (rechts) ist die verbleibende Dynamik zwischen Laut und Leise mit im Mittel nur etwa 8dB deutlich reduziert. Die tatsächliche Abhörlautstärke (medium) unterscheidet sich für beide Gruppen kaum.

Bewertung der Höranstrengung:

Interindividuelle Unterschiede: Die Bewertungen der einzelnen Probanden streuen erheblich: über alle Ratings ergeben sich mittlere LES-Werte für einzelne Probanden zwischen 1 und 7 ESCU. In der Gruppe der schwerhörenden Probanden lassen sich für einzelne Probanden LES-Medianwerte zwischen 1 und 9 ESCU beobachten. Die Bewertung der Höranstrengung zeigt sich in beiden Gruppen sehr individuell.

Only-Audio vs. Audio-visuell: Vergleicht man die Bewertungen der normalhörenden Probanden für die Konditionen1&2 „nur Audio“ mit den entsprechenden Datensätzen für die Konditionen4&5 „Audio-visuell“ lässt sich kein nennenswerter Unterschied (mittlerer Unterschied <0,1 ESC-Unit) zwischen mit und ohne Bild ausmachen. Auch für die schwerhörenden Probanden ist der Unterschied zwischen Kondition1 und Kondition4, also der Originalmischung ohne bzw. mit Bild marginal (mittlerer Unterschied ~0,1 ESC-Unit), beim um 6dB verringerten SNR in den Konditionen2&5 werden die Nur-Audio-Signale bzgl. ihrer Höranstrengung im Mittel um etwa ~0,6 ESC-Units höher bewertet als die AV-Signale. Insgesamt zeigen

sich also keine bis geringe Unterschiede der Höranstrengungsbewertungen mit und ohne Fernsehbild.

Bewertung der Höranstrengung durch normalhörende und schwerhörende Probanden: Abbildung 4 zeigt die Häufigkeitsverteilungen der der Mediane der unterschiedlichen Höranstrengungskategorien für die beiden Probandengruppen und die fünf Konditionen. Diese legen nahe, dass die Versuchskonditionen sinnhaft gewählt wurden. Kondition3_ („minus 3dB-Audio“) wird als deutlich anstrengender bewertet als die Original-Konditionen 1&4, in den um 6dB verschlechterten Konditionen 2&5 finden annähernd alle LE-Bewertungsmöglichkeiten zwischen 1 und 13 Verwendung.

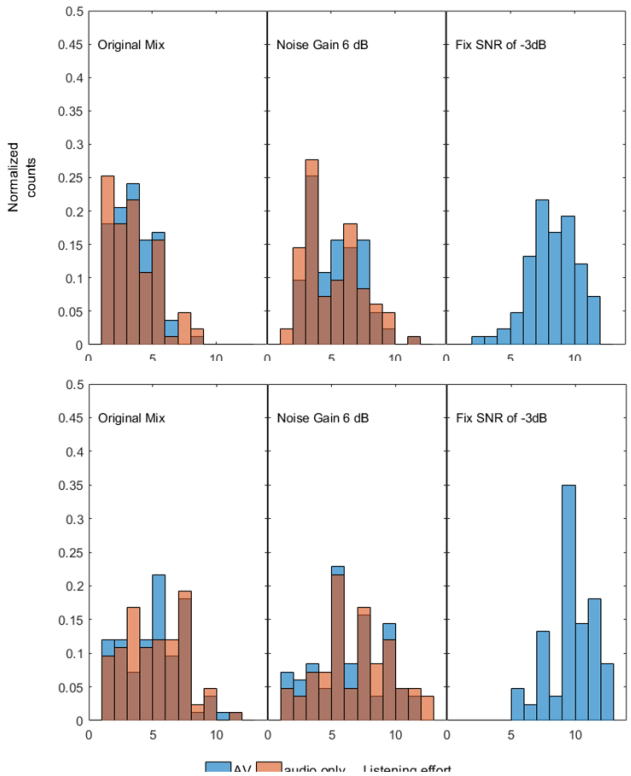

Abbildung 4: Histogramm der Median-Höranstrengungswerte über alle Clips und alle Probanden für die verschiedenen Konditionen – in blau mit Bild (AV), in rot nur Audio für die normalhörende Gruppe (oben) und die schwerhörende Gruppe (unten).

Bewertung der Höranstrengung durch normalhörende Probanden: Die „-6dB“-Mischungen aus Kondition2&5 führen verglichen mit dem Original aus Kondition1&4 in etwa 80% der Fälle zu einer leichten Zunahme der Höranstrengung: für die Gruppe der Normalhörenden beträgt der Anstieg im Mittel etwa einen Skalenwert. Kondition3 („minus3dB“) führt in fast 100% der Fälle zu einer Zunahme der Höranstrengung, im Mittel um etwa 4-5 Skalenwerte im Vergleich zur Originalmischung. Trotz festem SNR in der „fixSNR-3dB“-Kondition schwanken die Medianbewertungen aber erheblich. Eine pauschale SNR-Empfehlung als Maßstab guter Verständlichkeit oder eines geringen Höraufwandes scheint kein gangbarer Weg zu sein.

Bewertung der Höranstrengung durch schwerhörende Probanden: Die „-6dB-Audio“-Kondition2 führt verglichen mit der „Orig-Audio“-Kondition1 in etwa 65% der Fälle zu einer Zunahme der Höranstrengung: nach Bereinigung von Floor-Effekten im Mittel um etwa 2 ESC-Units. Die „-6dB-AV“-Kondition5 führt verglichen mit der „Orig-AV“-Kondition4 in nur 43% der Fälle zu einer Verschlechterung der Höranstrengung, im Mittel um weniger als 1 Skaleneinheit. Kondition3 („minus3dB“) führt in 96% der Fälle zu einer Zunahme der Höranstrengung, im Mittel steigt die Höranstrengung um etwa 5-6 Skalenwerte. Auch für die schwerhörenden Probanden variieren die LE-Bewertungen bei einem festen SNR erheblich.

Abbildung 5 veranschaulicht mittlere Bewertungsunterschiede normalhörender und schwerhörender Probanden: Für alle Konditionen kann ein Bewertungsverschiebung von etwa 1,5 bis 2 Skaleneinheiten ausgemacht werden. Für die schwerhörenden Probanden ist die Original-Kondition bereits ähnlich anstrengend wie für normalhörende Probanden Kondition 2&4 (-6dB).

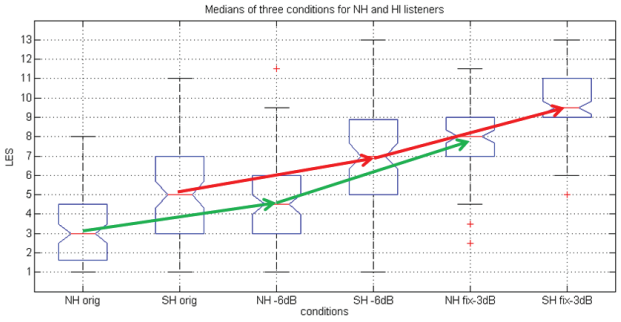

Abbildung 5: Verteilungen der Höranstrengung für normalhörende und schwerhörende Probanden.

Einfluss unterschiedlicher Mischungsverhältnisse (SNR) und unterschiedlicher Mischsignale (Sprachsignal/ Hintergrund) auf die Höranstrengung: Bezüglich weiblicher oder männlicher Stimmen zeigen sich die Ergebnisse annähernd symmetrisch. Das Geschlecht der Sprecher scheint hier keinen Einfluss auf die Bewertung des Höraufwandes zu haben.

Untersuchung unterschiedlicher Mischungsverhältnisse: Je größer die Verschlechterung des SNR im Vergleich zu den Originalmischverhältnissen, desto größer sollten auch die Unterschiede der Höranstrengungsbewertung sein. Dies trifft in dieser Studie jedoch nicht auf alle Bewertungen und Bewertungsunterschiede zu, die sich über alle Clips weitestgehend unabhängig von SNR-Unterschieden der Konditionen zeigen: Die Korrelationskoeffizienten von SNR und Medianen der Höranstrengungswerte sind für alle Konditionen kleiner als 0,01. Korrelationskoeffizienten zwischen SNR-Unterschieden und Medianen der Bewertungsunterschiede für die Konditionen2&3 sind zwischen 0,05 und 0,1.

Untersuchung unterschiedlicher Mischungssignale: Die SNR-Untersuchungen legen nahe, dass es eher wahrscheinlich ist, dass spezifische Kombinationen aus Sprach- und Hintergrundsignalen sich auf die Bewertung der Höranstrengung auswirken. Als mögliche Hintergrundkategorien wurden Babble (Sprachgewirr), Speech (deutliche

Sprache) und Music, Active (percussiv, dynamisch, abwechslungsreich) und Static (stationär) bestimmt.

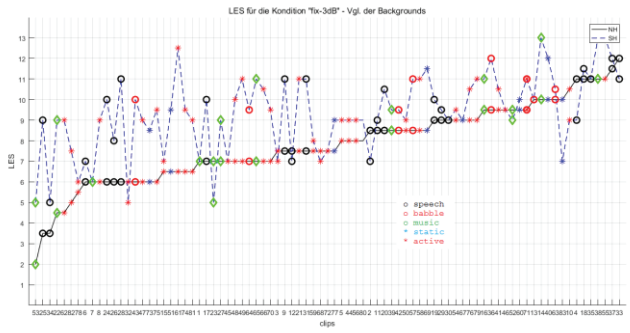

Abbildung 6: Mediane der LE-Bewertungen von normalhörenden (schwarze Linie) und schwerhörenden (blau gestrichelt) Probanden für Clips in Kondition3 („minus 3dB-Audio“) - sortiert nach Medianen der Normalhörenden. Die Symbole kodieren Hintergrundtypen.

Abbildung 6 zeigt die Mediane der Höranstrengungs-bewertungen von normalhörenden (schwarze Linie) und schwerhörenden (blau gestrichelt) Probanden für Clips in Kondition3 („minus 3dB-Audio“) - sortiert nach mittleren Clip-Bewertungen für Normalhörende. Die Symbole kodieren die verschiedenen Hintergrundkategorien. Durch diese grobe Kategorisierung der Hintergrundtypen lässt sich kein Zusammenhang mit der gemessenen Höranstrengung für Clips mit konstantem SNR zeigen. Einflussfaktoren der Zielsprache (Klarheit, Dialekt) wurden hier nicht systematisch betrachtet.

Zusammenfassung

Zwischen „laut“ und „leise“ liegt für die normalhörende Gruppe im Mittel 19dB. Für die schwerhörende Gruppe ist die verbleibende Dynamik um etwa 10dB geringer. Bezüglich der tatsächlichen Abhörlautstärke (angenehmer Pegel) unterscheiden sich Normal- und Schwerhörende in dieser Studie kaum.

Die Bewertung der Höranstrengung zeigt sich in beiden Gruppen sehr individuell. Die Gruppe der schwerhörenden Probanden bewertete bereits die Originalmischung knapp schlechter als die Normalhörenden-Gruppe die Hörproben mit um 6dB verringerten SNR. Das Geschlecht der Sprecher schien in beiden Gruppen keinen Einfluss auf die Bewertung des Höraufwandes zu haben. Diese Ergebnisse stimmen qualitativ mit [5] überein, in denen ebenfalls Höranstrengung mit anderen TV-Audiosamples und unterschiedlichen Probandengruppen gemessen wurde. Darüber hinaus zeigten sich in dieser Studie deutlich für beide Gruppen keine bis geringe Unterschiede bei der Bewertung von nur Audio-Material und AV-Material.

Eine pauschale SNR-Empfehlung als Maßstab guter Verständlichkeit oder eines geringen Höraufwandes scheint kein gangbarer Weg zu sein. Die Bewertungen und Bewertungsunterschiede zur Originalmischung korrelieren nicht mit dem SNR bzw. dem Ausmaß der Verschlechterung des SNR, so dass für eine objektive Messung der Sprachverständlichkeit elaboriertere Modellansätze erforderlich sind.

Von den Clips, die von den Probanden als eher höraufwendig bewertet wurden, weisen manche Eigenschaften auf, welche erwarten lassen, dass diese die Höranstrengung steigern: Mehrere Sprecher zur selben Zeit, schlechte Artikulation, Hintergrundmusik mit Gesang, laute oder verhallte Umgebung, störende Hintergrundgeräusche wie Sirenen oder Straßenverkehr und manchmal auch Kombinationen der genannten Aspekte. Diese „auditorischen Szenen“ weisen ein hohes Ablenkungspotential auf und sie zu verstehen bedarf es Aufmerksamkeit und Fokussierung – so könnte es schwerer fallen, das Objekt „Sprache“ zu isolieren und der Höraufwand ist entsprechend höher. Aber auch viele Clips mit geringer geschätztem Höraufwand zeigen diese Eigenschaften – so lässt sich auch aus den genannten Charakteristiken kein einfaches Gesetz zur Bestimmung der Höranstrengung ableiten.

Danksagung

Die Studien wurden im Rahmen des Kooperationsprojekt „Objektive Analyse, Visualisierung und Korrektur von Sprachverständlichkeit in Broadcastanwendungen für Normal- und Schwerhörende“ durchgeführt und gefördert durch das Bundesministerium für Wirtschaft und Energie BMWi aufgrund eines Beschlusses des Deutschen Bundestages (FK: ZF4072002SS5).

Literatur

[1] Rennies, J. et al., (2014). “Listening effort and speech intelligibility in listening situations affected by noise and reverberation.” in The Journal of the Acoustical Society of America 136, 2642-2653.

[2] ANSI (1997). “ANSI S3.5-1997 Methods for calculation of the speech intelligiblility index”, Standards Secretariat, Acoustical Society of America, New York, USA.

[3] Rhebergen, K. S. and Versfeld, N. J. (2005).“A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners”, The Journal of the Acoustical Society of America 117, 2181–2192.

[4] EBU Tech Doc 3341 ‘Loudness Metering: ‘EBU Mode’ metering to supplement loudness normalisation in accordance with EBU R 128

[5] Wächtler M., Rennies J. & Kollmeier B. (2015). „Automatische Erkennung von Abschnitten mit kritischer Sprachverständlichkeit für Normal- und Schwerhörende in Film und Fernsehen.“ In DAGA 2015 – Fortschritte der Akustik, S. 301-304, Nürnberg, Deutschland, März 2015.