

## Barrierefreie Audiokommunikation – Von der Aufnahme bis ins Wohnzimmer (Barrier-free Audio Communication from On-site Recording to the Living Room)

H. Baumgartner<sup>1</sup>, W. Hoeg<sup>2</sup>, A. M. Koolwaay<sup>3</sup>, J. RENNIES<sup>1</sup>, M. Wächtler<sup>1</sup>, W. Schukrafft<sup>4</sup>, E. Bodenseh<sup>5</sup>

<sup>1</sup> Fraunhofer Institut für Digitale Medientechnologie, Oldenburg, Email: [hannah.baumgartner@idmt.fraunhofer.de](mailto:hannah.baumgartner@idmt.fraunhofer.de)

<sup>2</sup> Audio Consultant, Berlin, Email: [wolfgang.hoeg@tonmeister.de](mailto:wolfgang.hoeg@tonmeister.de)

<sup>3</sup> Deutscher Schwerhörigenbund Landesverband NRW, Email: [infodsblvnrw@aol.com](mailto:infodsblvnrw@aol.com)

<sup>4</sup> Fa. Satisfactory e.K. Nemitz, Email: [ws@satisfactory.de](mailto:ws@satisfactory.de)

<sup>5</sup> Freie Tonmeisterin, Email: [elisabeth.bodenseh@gmail.com](mailto:elisabeth.bodenseh@gmail.com)

### Abstract

Modernes Audiodesign bedeutet meist hohe Dynamik und effektvolle Klangwelten. Doch des einen Freud, des anderen Leid – vor allem - aber nicht nur - hörgeschädigte Zuhörer finden wenig Gefallen an großen Lautheitsunterschieden und effektvollen Tonmischungen. Insbesondere sensorineuraler Hörverlust führt zu einer veränderten Lautheitswahrnehmung und einer Verringerung des wahrnehmbaren Dynamikbereichs: Leise Passagen werden nicht gehört, laute Stellen hingegen als unangenehm laut empfunden. Doch nicht nur Schwerhörigkeit beeinflusst die tatsächliche Wiedergabedynamik. Auch die akustischen Voraussetzungen einer realen Abhörumgebung können die Sprachverständlichkeit von TV- und Hörfunk-Programmen schmälern.

Wo hört gestalterische Freiheit auf und wo beginnt die Pflicht der Barrierefreiheit? Wie kann man allen Ansprüchen gerecht werden? Um den Verlust an Wiedergabedynamik zu kompensieren, wird u. a. an Technologien gearbeitet, welche eine Personalisierung dieser Größe entsprechend den individuellen Bedürfnissen der Empfänger ermöglichen. Das Anliegen dieses Roundtables ist es, für das Thema der barrierefreie Audiokommunikation zu sensibilisieren und einen guten Überblick zu verschaffen.

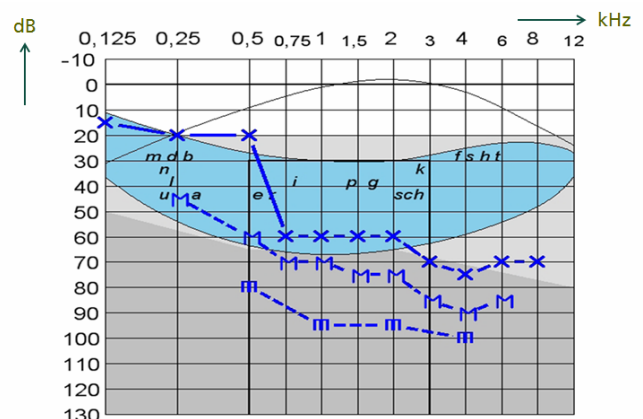
## 1. Einführung

Der Roundtable diskutierte den Stand der Entwicklung und demonstrierte vielfältige Möglichkeiten: Zunächst wurde die Problematik aus Sicht einer Betroffenen kurz geschildert und anhand von Beispielen illustriert. Ein State-of-the-Art Kapitel gab einen Überblick über existierende Technologien und Ansätze. Eine Software wurde vorgestellt, welche die Sprachverständlichkeit einer Mischung automatisch und objektiv überprüft. Ein Tonmeister referierte über Stolpersteine und Hürden am Set und zeigte Lösungen auf. Der letzte Beitrag war der Entwicklung von Produktionsrichtlinien zur Verbesserung der Sprachverständlichkeit im Fernsehen gewidmet.

Hannah Baumgartner und Wolfgang Hoeg organisierten und moderierten den Roundtable. Die Reihenfolge der Autoren im Paper entspricht der Folge der Vorträge während der Veranstaltung. Die Verantwortung zum Inhalt der Einzelbeiträge liegt bei den jeweiligen Autoren.

## 2. Betroffen

Anna Maria Koolwaay war als Bauingenieurin vorwiegend im Bereich Brückenbau tätig. Bereits seit ihrem 28. Lebensjahr nahm ihr Hörvermögen krankheitsbedingt stetig ab, sie hat im Laufe ihres Lebens alle Grade von Hörverlusten durchlebt, mittlerweile trägt sie Cochlear Implantate. Seit 2006 ist sie im Deutschen Schwerhörigenbund, Landesverband NRW, Referatsleiterin für „Barrierefreies Planen und Bauen“, seit 2013 widmet sie sich der Thematik „Barrierefreier Rundfunk/ Fernsehen“.



**Abb. 1:** Sprachbanane nach Fant (1959): Schematische Darstellung des gemittelten Sprachspektrums innerhalb eines Audiogramms als Hauptbereiche für Formanten. Hauch-, Zisch- und Explosivlaute, die im hochfrequenten Bereich liegen, werden bereits bei moderaten Hörverlusten nicht mehr gehört.

Hören und Verstehen sind zwei ganz verschiedene Dinge. Hören kann man vieles, beispielsweise Musik, Straßenlärm, Flüstern, Vogelstimmen - aber Sprache zu verstehen ist oft eine echte Herausforderung. Aus den vielen Hintergrundgeräuschen - Musik gehört in bestimmten Situationen auch dazu - die Sprache herauszufiltern, ist eine hohe Kunst. Genau das müssen Hörgeräte und Sprachprozessoren (Cochlear Implantate, CIs) täglich leisten. Daher ist durchaus nachvollziehbar, dass Hörgeräte oder CIs nur Hilfsmittel sind, ein gesundes Ohr ersetzen sie jedoch nicht.

Die meisten schwerhörigen Menschen haben einen Hörverlust im Hochtonbereich (vgl. **Abb. 1**). Sie hören nicht nur alles leiser, sondern Hauch-, Zisch- und Explosivlaute wie h, f, s, ß, z, sch, b, k, p, t, die im hochfrequenten Bereich liegen, werden nicht mehr gehört und müssen durch Kombinieren (und teilweise durch Lippenabsehen) ausgeglichen werden [1]. Hörgeräte können einen Ausgleich schaffen, aber nicht vollständig und nicht in jeder Lebenssituation. Radio und Fernsehen sind auch mit einem Hörgerät nicht ausreichend zu verstehen.

## 2.1. Ursachen schlechter Verständlichkeit

Undeutliche (nuschelnde) oder zu schnelle Aussprache der Schauspieler und Moderatoren beeinträchtigen das Verstehen der Dialoge für alle Menschen. Eine klare, vernehmbare Aussprache sollte während der Sprachausbildung intensiv trainiert werden.

Bei O-Tönen oder Interviews erschweren Hintergrundgeräusche, z. B. Verkehrs- oder Baustellenlärm, nicht nur schwerhörigen Zuschauern unnötig das Verstehen der Sprache. Des Weiteren wird durch stark frequenzabhängige Hörstörungen die Wahrnehmung einer Klangmischung verändert: so ergibt sich für Schwerhörige nicht nur eine Informationsbarriere, sondern oftmals auch Hör-Stress. Durch Recruitment (veränderte Lautheitswahrnehmung bei sensorineurealem Hörverlust) wird der Lautheitsbereich zwischen „ganz leise“ und „ganz laut“ auf teilweise nur 30 dB komprimiert. Hörgeschädigte reagieren deshalb stärker als Guthörende auf Pegelschwankungen.

Niemand möchte die Hintergrundmusik (Atmosphäre) verbieten, sie ist wichtig für den Film, Aber sie sollte nur dezent eingesetzt werden, damit die Sprache immer noch ausreichend verstanden werden kann. Selbst bei Dokumentationen und in reinen Informationssendungen werden heute zum Text dramatisierende Musik oder laute Hintergrundgeräusche beigemischt, so dass die eigentliche Sprachinformation völlig überdeckt wird und selbst „Guthörende“<sup>1</sup> Schwierigkeiten haben, zu verstehen.

## 2.2. Untertitelungen

Erfreulicherweise hat das Angebot an untertitelten Fernsehsendungen im öffentlich rechtlichen Fernsehen in den letzten Jahren deutlich zugenommen und soll auch laut NDR in Zukunft weiter verbessert werden.

Ich begrüße grundsätzlich das verbesserte Angebot der öffentlich rechtlichen Fernsehsender, sehen aber noch einen erheblichen Nachholbedarf an Untertitelungen in den Bildungskanälen Phoenix und ARD-Alpha.

ARD-Alpha bietet Sendungen aus den Themenbereichen Wissenschaft, Geschichte, Kunst, Kultur, Musik, Religion

und Zeitgeschehen nur vereinzelt mit Untertitel an. Von den angebotenen Sprachkursen sind Kinder und Jugendliche, die auf Untertitel angewiesen sind, ausgeschlossen.

Phoenix, der Sender der Bürgerinnen und Bürgern Informationen anbietet, damit sie sich eine politische Meinung bilden können, bietet keine Untertitel an und ist somit für hörbehinderte Menschen nicht zugänglich.

Genauso verfahren viele private Fernsehsender und verhindern so die Inklusion hörgeschädigter Menschen.

Besondere Schwierigkeiten bereiten dem Zuschauer aber auch die unterschiedlichen Wege zum Zuschalten von Untertiteln. Abhängig vom Fernsehhersteller und Fernsehsender können Untertitelungen über Videotext oder digital zugeschaltet werden. Viele Zuschauer sind damit hoffnungslos überfordert und müssen sich erst mit dem Fernsehgerätehersteller auseinandersetzen. Eine klare einheitliche Regelung, am besten mit einer UT-Taste auf der Fernbedienung, ist für alle Fernsehgeräte anzustreben.

Filme aus der Mediathek mit Untertitel zu laden, ist eine weitere Herausforderung: Je nach Medium sind sie zuschaltbar oder nicht. Tablets oder Smartphones erfordern nur ein einfaches Anklicken auf das UT-Symbol. Beim Fernseher ist es wieder vom Hersteller abhängig, ob UT zu sehen sind oder nicht. Warum das so kompliziert ist, kann man als Nutzer einfach nicht nachvollziehen. Daher sollte auch hier schnellstens Abhilfe geschaffen werden!

## 3. State of the Art

*Jan Rennies untersuchte in seiner Promotion Lautheit und Sprachverständlichkeit mit dem Fokus auf räumliches Hören und komplexe Hörsituationen. Hannah Baumgartner schloss ihr Studium mit einer Arbeit zur Personalisierung von Audioübertragung, insbesondere für Schwerhörige ab und arbeitet seitdem im etwas weiter gefassten Themenbereich assistive Technologien. Beide sind Mitarbeiter der Oldenburger Fraunhofer IDMT Gruppe „Hör-, Sprach- und Audiotechnologie“ und engagieren sich u.a. innerhalb des Themenkomplexes „Barrierefreie Audiokommunikation“.*

Im Allgemeinen klagen Schwerhörige über eine schlechte Hörsamkeit (*audibility*) des Hörerlebnisses und eine schlechte Sprachverständlichkeit (*intelligibility*). Eine hohe Programmdynamik, Lautheitssprünge zwischen Sendern und Formaten oder extreme Mischungsverhältnisse können vom geschädigten Gehör nur unzureichend verarbeitet werden und führen zu Stresserfahrung und mangelhafter Rezeptionsqualität beim Hörer [2].

### 3.1. Produktionsseitige Maßnahmen

Bezüglich der meist störenden Lautheitssprünge überschneiden sich die Interessen der Hörgeschädigten mit den Themen der „Lautheitsdebatte“, welche die Idee der Lautheitsnormalisierung im Rundfunk etabliert hat.

Bezüglich der meist störenden Lautheitssprünge überschneiden sich die Interessen der Hörgeschädigten mit

<sup>1</sup> Der DSB bemüht sich seit einiger Zeit, den Begriff „Normalhörende“ durch den Begriff „Guthörende“ zu ersetzen. Guthörende haben eine Hörschwelle in der Nähe von 0 dB. Normalhörigkeit ändert sich hingegen in Abhängigkeit vom Alter.

den Themen der „Lautheitsdebatte“, welche die Idee der Lautheitsnormalisierung im Rundfunk vorantreibt. Eine Verbesserung der Sprachverständlichkeit lässt sich z.B. effektiv durch die Verbesserung des Signal-Rausch-Abstandes (signal-to-noise ratio, SNR) bewirken. Innerhalb des Programmes wird der SNR u.a. bestimmt durch das Verhältnis von Sprache zu Effekten, Atmo und Musik. Für Guthörende liegt ein angemessener SNR bei etwa 6 dB, bei Schwerhörigen jedoch oft über 10 dB [3]. Im Interesse der Schwerhörenden liegt demnach ein adaptierbares Mischungsverhältnis von Sprachkanal und Effektkanal, realisierbar zum Beispiel mit Hilfe eines neben-geräuschfreien Nur-Sprach-Kanals (*clean speech track*).

Objektbasierte Audiocodierungen ermöglichen neue Wege: Die Fraunhofer-Technologie *Dialogue Enhancement* [4] ermöglicht zum Beispiel, die relative Lautstärke der Sprachanteile für eine individuelle Tonmischung selbst festzulegen. Mit Hilfe parametrischer Daten wird beschrieben, welche Anteile im Summensignal zu welchen Ursprungssignalen gehören. Diese Metadaten werden mit übertragen und ermöglichen so am Endgerät, die Mischungsverhältnisse an persönliche Bedarfe anzupassen. *Dialogue Enhancement* wird momentan in DVB (Digital Video Broadcasting) standardisiert.

Gehörlose, hochgradig Schwerhörige, aber auch Menschen mit einer Fremdsprachenbarriere, können durch die zusätzliche Generierung und Übertragung von Untertiteln oder visuellen Informationen unterstützt werden. Bei der automatischen Untertitelung eines Programmes wird Sprache zu Text transkribiert. Deutlich gesprochene Sprache ohne Störgeräusche erzielt dabei die beste Qualität – daraus ergibt sich ein weiterer Einsatzbereich für einen clean speech track. Fremdsprachenbarrieren können mit Multi-Lingual-Kanälen abgebaut werden. Für Blinde bietet die Übertragung einer zusätzlichen Tonspur, welche die sprachliche Beschreibung der Szenerie enthält, mehr Anschaulichkeit (Hörfilm, *audio description*).

### 3.2. Maßnahmen auf der Multiplex-und-Netzwerk-Ebene

Maßnahmen auf der Multiplex-und-Netzwerk-Ebene (Service Provider) sind die Integration und Intersektion der zusätzlichen Audiospuren (*clean speech track*, *audio description*, Multi-Lingual-Kanälen) und Daten (Metadaten, Untertitelung, Visualisierung), dementsprechend auch das Bereithalten nötiger Sendekapazitäten und die Entwicklung geeigneter Kodierungsverfahren. Netzbetreiber sollten die Langzeitlautheit (24 Stunden) und die Metadaten auf Einhaltung und Wahrheitsgehalt loggen und überprüfen und bei Bedarf regulierend eingreifen.

### 3.3. Empfängerseitige Maßnahmen

Häufig werden nützliche Zusatzinformationen und Metadaten in modernen digitalen Rundfunksystemen zwar übertragen, scheitern aber an Barrieren bei der Nutzung. So sind neben der Bereitstellung der technischen Möglichkeiten auch barrierefreie Nutzungsschnittstellen und die

Aufklärung der Konsumenten und Hersteller zu vorhandenen Möglichkeiten wichtige Schritte auf dem Weg zur Barrierefreiheit. Existierende Metadaten, die zur Optimierung von Audiomaterial auf die Bedürfnisse von Schwerhörigen verwendet werden könnten, sind z.B. Dynamic Range Control (DRC, Dynamikbereich-Steuerung), Music-Speech-Flag (Kennzeichnung von Sprache und Musik), Pitch-Control (Tonhöhen-Steuerung) oder Speed-Control (Steuerung der Wiedergabegeschwindigkeit).

Für eine Personalisierung der Audioübertragung muss die individuelle Hörbeeinträchtigung ausgemessen werden und persönliche Schwellen erhoben werden. Ein entsprechendes Fittingsystem für Empfangsgeräte, welches die verschiedenen Schritte bei der Hörgeräteanpassung nachstellt, muss einfach und ohne Vorwissen bedienbar sein und vom Nutzer am heimischen TV- oder Audiogerät selbständig durchgeführt werden können. Die Frage, wie Hörgerätealgorithmen in Audio-Geräte integriert werden können, ist Gegenstand aktueller Forschung. Variationen sind möglich beim Grad der Individualisierung (Gruppierung von Hörverlust-Typen, Auswahl aus spezifischen Voreinstellungen), bei der Erfassung des Hörverlustes (Auswahl der Testsignale, Art und Anzahl der im Fitting erhobenen Schwellen, Anzahl der adaptierbaren Parameter, Einbeziehung von Metadaten) und bei der Berechnung der Kompressionsparameter (Anzahl und Grenzen der Kompressionsbänder, Zeitkonstanten, ein- oder ausgangskontrollierte Kompression).

Hochgradig Schwerhörende beziehen die Möglichkeit, an Lippen abzulesen, häufig in ihre Wahrnehmung mit ein. So wurden z.B. die Möglichkeiten untersucht, die ein synthetisches Gesicht (vgl. **Abb. 2**) eröffnet, indem es zusätzlich zum Hören das Lippenlesen ermöglicht [5]. Auch hier bietet der clean speech track die beste Grundlage zur Realisierung. Ein ähnlicher Ansatz bei der Produktion von Gebärdendolmetscher-Inserts ist vorstellbar.



**Abb. 2:** Synthetisches Gesicht als Avatar/Insert einer schwedischen Forschergruppe [5], welches zusätzlich zum Hören das Lippenlesen ermöglichen soll.

Neben dem originären SNR des Programmes trägt auch der SNR, der sich durch die Umgebungsgeräusche am Hörort ergibt, zum Sprachverstehen bei. Die Signalverarbeitung kann im Hintergrund das Nutzerverhalten unter Beobachtung der vorgenommenen Einstellungen und der Messung von

Parametern aus der Umgebungsakustik modellieren. Eine von der Akustik der Umgebung abhängige automatische Dynamikanpassung sowie andere Möglichkeiten, das Störgeräusch in der Wahrnehmung zu unterdrücken (Kopfhörer, Richtstrahler, Raumakustik), sind vorstellbar.

#### 4. Zur Historie der Barrierefreiheit

*Wolfgang Hoeg, vorm. Deutsche Telekom (Forschung), befasst sich seit vielen Jahren u. a. mit Themen wie lautheitsgerechter Aussteuerung, Dynamic Range Control und barrierefreier Audiokommunikation. Zudem ist er Autor/Herausgeber von Fachbüchern wie „Digital Audio Broadcasting“ und „Handbuch der Tonstudioteknik“ [2].*

Das Problem der Lautstärke von Musik, Sprache und Hintergrundgeräusch vs. Textverständlichkeit im Rundfunk (Hörfunk, Fernsehen) ist so alt wie das Medium selbst. Mitte der 20er Jahre des letzten Jahrhunderts fanden erste öffentliche Rundfunkübertragungen statt, TV-Übertragungen folgten ca. 10 Jahre später. Im gleichen Zeitraum etablierte sich der Tonfilm. Die elektro-akustische Beschallung von Veranstaltungen (PA-Systeme) entwickelte sich parallel mit der entsprechenden Technik. Erste wissenschaftliche Fragestellungen zu Lautstärkedifferenzierung und Sprachverständlichkeit finden sich bereits Anfang der 40er Jahre, daraus resultierten zahlreiche Untersuchungen zu Entstehung und Wirkung verschiedener Formen der Schwerhörigkeit vs. Textverständlichkeit.

Elektrische Hörgeräte zur Versorgung von Hörgeschädigten entstanden bereits um die Jahrhundertwende des 19. zum 20. Jahrh., diese Technik entwickelt sich bis heute stetig weiter.

##### 4.1. Frühe Aufgabenstellungen und Lösungen

Erste Richtlinien zur (manuellen) lautheitsgerechten Aussteuerung von Tonprogrammen wurden bereits in den 70er Jahren erlassen (IRT, OIRT), jedoch nicht immer konsequent befolgt. Internationale Studienprogramme und Lösungsvorschläge zum Problem „Hearing Impaired“ beim Rundfunkempfang existieren seit den frühen 90er Jahren (ITU-R SG10, BBC, Deutsche Telekom), wurden jedoch lange Zeit seitens der Programmanbieter ignoriert.

Zwischenzeitlich entstanden verschiedene internationale Standards als Voraussetzung einheitlicher Lösungen, u. a. Definition von Metadaten zur Übertragung von Dynamikwerten in Audiodatenströmen (MPEG-2, MPEG-4 AAC, Dolby etc.), insbesondere auch ein Verfahren zu Dynamic Range Control für DAB (Eureka147). Diese wurden allerdings von der Industrie bisher kaum in entsprechenden Gerätelösungen umgesetzt.

##### 4.2. Wichtige nationale/internationale Zielstellungen und Projekte

Ein früher Projektvorschlag „Audible“ (BBC, DTAG, IRT u. a. m., 1997) scheiterte mangels Interesse des Managements auf allen Seiten. Erst eine Initiative der European Commission (EC) zu „eAccessibility“ von 2005 [6] änderte

die Einstellung vieler Administrationen, darunter auch die der ARD, zur Notwendigkeit der Unterstützung und Einführung neuer Technologien zur barrierefreien Informationsversorgung.

Die intensiven Arbeiten der EBU Projekt Gruppe P/LOUD zur lautheitsbezogenen Aussteuerung zur Vermeidung von lästigen Lautheitssprüngen innerhalb und zwischen Programmen, die in der Recommendation EBU R128 [7] sowie weiteren Dokumenten und Aktivitäten mündete, entsprach den Interessen von Schwerhörenden ebenso wie denen von Guthörenden.

Innerhalb des europäischen Projektes DTV4all (2008-2011) wurde die Thematik Barrierefreiheit endlich bzgl. existierender Barrieren und entsprechender Lösungsansätze systematisch aufbereitet [8], um entsprechende Empfehlungen an die zuständigen Stakeholder machen zu können.

#### 5. Hörmodellbasierte Überprüfung der Sprachverständlichkeit

*Moritz Wächter ist Student der Hörtechnik und Audiologie und schreibt seine Abschlussarbeit zum Thema „Automatische Erkennung von Abschnitten mit kritischer Sprachverständlichkeit für Normal- und Schwerhörende in Film und Fernsehen“ am Fraunhofer IDMT in Oldenburg.*

Die Tonverantwortlichen arbeiten im Studio meist unter optimalen Abhörbedingungen und werden darüber hinaus während ihrer Arbeit zunehmend vertraut mit dem verwendeten Audiomaterial. Diese Faktoren können eine zuverlässige und objektive Beurteilung der Sprachverständlichkeit erschweren, insbesondere wenn auch alters- oder hörverlustbedingte Einschränkungen der Zielgruppe Beachtung finden sollen. Eine automatisierte Vorhersage der Sprachverständlichkeit bietet hier eine Hilfestellung, da sie eine zeiteffiziente, objektivierbare und zielgruppenspezifische Beurteilung der Sprachverständlichkeit ermöglicht. Dieser Beitrag stellt ein Werkzeug vor, welches während der Tonmischung automatisch Abschnitte im Audiomaterial detektiert, in denen die Sprachverständlichkeit als kritisch beurteilt wird. Die objektive Beurteilung erfolgt hierbei durch den Einsatz eines angepassten Hörmodells.

##### 5.1. Speech Intelligibility Index

Für die Analyse werden die Audiosignale zunächst in kurze Abschnitte unterteilt. In einem ersten Schritt trifft ein Sprachaktivitätsdetektor basierend auf der Leistung des Signals die Entscheidung darüber, ob der betreffende Abschnitt Sprache enthält. Nur wenn dies der Fall ist, erfolgt die weitere Analyse mit dem Hörmodell. Als Modell wird hier der Speech Intelligibility Index [9] herangezogen. Beim SII handelt es sich im Wesentlichen um einen gewichteten Mittelwert der auf das Intervall von 0 bis 1 abgebildeten Signal-Rausch-Abstände über mehrere Frequenzbänder. Die Gewichtung erfolgt hierbei entsprechend der Wichtigkeit,

den der jeweilige Frequenzbereich für das Sprachverstehen hat. Der SI-Index stellt ein Maß dafür dar, welcher Anteil der für das Verstehen relevanten Signalanteile (Cues) in der Sprache dem Hörer zur Verfügung steht. Somit beschreibt der SII ausschließlich energetische Signaleigenschaften, inwiefern der einzelne Hörer von diesen grundlegenden Signaleigenschaften Gebrauch machen kann, hängt dann von vielen weiteren Faktoren, wie z.B. von Kontextwissen oder Sprachkenntnissen, ab. Die Einbeziehung von Hörverlusten in die Modellierung ist ebenfalls möglich.

Unterschreitet der Wert des SI-Index einen kritischen Wert, so wird die Sprache als potentiell nicht voll verständlich betrachtet und der betreffende Abschnitt wird visuell in der vom Programm angezeigten Wellenform hervorgehoben. Dieser kritische Wert wurde auf Basis von Hörversuchen mit guthörenden sowie hörbeeinträchtigten Personen festgelegt. Eine zu niedrige Wahl des kritischen Wertes kann dazu führen, dass Abschnitte nicht hervorgehoben werden, obwohl sie für viele Probanden ein unzureichendes Sprachverstehen bieten. Bei einem zu hoch gewählten Wert kommt es hingegen zu vielen Fehlalarmen. Ein guter Kompromiss ist hier also von großer Wichtigkeit.

## 5.2. Modellierung von Zielgruppen

Für die Mischung kann das Analysewerkzeug auf verschiedene Zuschauergruppen angepasst werden. So können etwa anhand des Alters und Geschlechts durchschnittliche Audiogramme gemäß der ISO (1999) [10] berechnet werden. Auf Wunsch kann auch die manuelle Eingabe eines Audiogramms erfolgen.

Eine Limitierung der vorgestellten Methode besteht in der Verwendung des SII, der ausschließlich sogenannte energetische Merkmale der Sprache berücksichtigt. Der Einfluss phonetischer Faktoren, wie der Deutlichkeit der Aussprache oder der Sprechgeschwindigkeit, findet (noch) keine Berücksichtigung.

## 6. Sprachverständlichkeit beginnt am Set

*Wolfgang Schukrafft ist freiberuflicher Filmtonemeister. Im Laufe von 35 Jahren entstanden viele Kino- und Fernsehfilme, u.a. mit Detlev Buck, Wolfgang Becker, Tom Tykwer und Leander Haußmann. Seit 2010 unterrichtet er Filmtone-Gestaltung am Institut für Musik und Medien der Robert Schumann Hochschule Düsseldorf im Modul Musik und AV Produktion.*

### 6.1. Akustische Umweltverschmutzung

Unser Alltag wird mit zunehmender Tendenz „zugemüllt“ mit Tönen, Klängen, Geräuschen, Lärm. Der „Verlust der Stille“ ist nicht nur ärgerlich, der Körper reagiert mit psychosomatischen Störungen.

In diesem Ambiente einen sauberen Filmtone aufzunehmen, grenzt oft an Zauberei, so man als Tonemeister überhaupt die Zeit und die Möglichkeit bekommt, seine Kompetenz

einzusetzen, Dialoge sauber aufzunehmen. Hat der Set-Tonemeister die Möglichkeit, Einfluss auf die Motivwahl, die Drehorte zu nehmen und deren akustische Beschaffenheit vor Drehbeginn zu optimieren, ist das schon mal die halbe Miete. Bekommt er darüber hinaus seitens des Herstellungsleiters einen zweiten Angler oder versierten Tonpraktikanten zugestanden, ist das Glück fast perfekt. Wenn dann noch der Regisseur ‚audiophil‘ ist, dem Team tagtäglich vermittelt, dass ihm ein guter Tone so wichtig ist wie das Bild - dann bekommt er als Gegenleistung mindestens 95% brillanten Original-Tone!

### 6.2. Der ökonomische Druck erntet faule Früchte

Noch in den 80er-Jahren waren zwei Minuten Kinofilm pro Drehtag üblich (also für 120 Minuten Spielfilm 60 Drehtage), heute werden 5-10 Minuten pro Tag kalkuliert, komplette Fernsehfilme im extremsten Fall an einem Tag in die Kiste gebracht. Gedreht wird mit zwei oder mehr Kameras mit unterschiedlichen Einstellungsgrößen. Eine Praxis, die dem Tone-Team jegliche Chance nimmt, den zur Einstellungsgröße korrespondierenden Dialog adäquat zu angeln.

Bodymikes an der Brust eines jeden Schauspielers bringen die vermeintliche Präsenz, alles platt, perspektivisch falsch weil ganz nah, ohne Abbildung der Räumlichkeit. Da Bodymikes fast ausnahmslos Kugel-Charakteristik haben, nehmen sie den akustischen Umgebungsmüll mit auf; hinzu kommen Phasenprobleme, Störungen und Kammfiltereffekte in der Funkübertragung.

Die Schauspieler bekommen zunehmend nur noch eine einzige Chance, die Szene wird in Mehrkameratechnik durchgedreht, das war's... Umbau zur nächsten Szene.

„We fix it in the post!“ ist dann das tröstliche Schulterklopfen für das frustrierte Tone-Team. Wer in der Postproduktion arbeitet weiß zu gut, dass aus einem mangelhaften O-Tone keine brillante Zauberflöte geschnitzt werden kann. Der redliche Versuch der „Reinigung“ im O-Tone-Schnitt bringt allenfalls ‚tendenzielle‘ Verbesserung der Sprachverständlichkeit, erzeugt aber auch Artefakte im Frequenzspektrum der Sprache und Sprünge in der Atmo, die dann mit Musik oder Geräuschen ‚kaschiert‘ werden müssen. Ein Teufelskreis, will man nicht die Szene, gar den ganzen Film synchronisieren.

### 6.3. Tonemeisterlicher Dreisprung am Drehort

Was zeichnet eine/n gute/n Set-TonemeisterIn aus, was muss er/ sie mitbringen um die Hauptaufgabe „saubere und verständliche Dialoge“ am Drehort aufzunehmen, zu meistern? – es sind drei Dinge:

1/3 **angemessene richtige Technik** – inklusive der Kompetenz diese sinnvoll einzusetzen: in der Regel ist es nicht das Neueste, Größte, Universellste, Multi-Channel... eher die Beschränkung auf das Wesentliche: „Ein einziges Mikrofon zur richtigen Zeit am richtigen Ort...“ (Zitat meines Lehrmeisters Milan Bor, †1998)

1/3 **Virtuosität & Mut** auf diesem Instrumentarium wie ein Pianist auch konzeptionell Neues zu wagen: Das bedeutet bereits vor Drehbeginn Entscheidungskompetenz zu entwickeln und diese im Interesse eines optimalen Originaltons zusammen mit Regie und Produktion konzeptionell umzusetzen, wie zum Beispiel bei

- „Wir können auch anders“ (zu einer Zeit, in der nahezu alle Filme in Deutschland (noch) monophon produziert wurden, konnte ich den Produzenten überzeugen, meinen ersten Dolby-Stereo-Film zu wagen...)
- „Gloomy Sunday“ (welcher Flügel - Bösendorfer oder Schimmel, stummgeschaltet oder tönend, Handdouble, Liveaufnahme oder vorproduziertes Playback, ...?)
- „Das Leben ist eine Baustelle“ (wie wird aus einer tonnenförmigen Baseball-Halle mit Schwingboden ein Volltonstudio...?)
- „Herr Lehmann“ (Aufbau einer 4 x 10 Meter großen Glaswand auf dem Bürgersteig, um die Bedingung des Regisseurs zu erfüllen, die Showdown-Szene im türkischen Imbiss in Originalton zu drehen, statt sie zu synchronisieren...)

... auch wenn solche Entscheidungen manchmal mit höherem finanziellen Aufwand verbunden sind.

1/3 **Psychologie am Drehort:** inclusive der Psychologie des ersten Drehtags: „Was ich am ersten Tag durchgehen lasse ist zementiert, damit muss ich bis zum Drehende leben.“

- Wie bewältige ich ohne qualitative Abstriche die Pflicht (meint saubere Aufnahmen des Dialogs, Dialog und nochmals Dialog!!!)?
- Wie meistere ich trotz Stress und 16 Stundentag am Set die Kür (Atmosphären, Geräusche, Effekte, Komparserie)?
- Wie lerne ich das unwiederbringlich Wesentliche vom (reproduzierbaren) Unwesentlichen zu unterscheiden?
- Wie gewinne ich das Vertrauen der Schauspieler, so dass sie meine auditive Kompetenz als Katalysator für ihre Performance erfahren und nicht als anmaßende Einmischung in ihr Spiel (Stimmung, Emotion, Tempore, Coloration, ...)?

## 6.4. Apocalypse Now

Das deutsche Kino war bis Mitte der 80er-Jahre MONO!

Der Film „Apocalypse Now“ hat uns aus dem Dornröschenschlaf geschossen, Action- statt Erzählkino. Der Subwoofer hielt Einzug in den Kinos und später auch in den Wohnzimmern, Surround-Sound musste ringsum wummern, in Deutschland wie in USA.



**Abb. 3:** Monophonie: - „Ein einziges Mikrofon zur richtigen Zeit am richtigen Ort...“

Die englische Sprache im Film ist da als Vokalsprache klar im Vorteil, um ca. 6 dB verständlicher als die deutsche Konsonantensprache, das bedeutet: Ambience und Musik können in englischsprachigen Filmen etwa doppelt so laut sein, bei gleicher Sprachverständlichkeit. Unsere Hörgewohnheiten, der „Zuschauerwunsch“ wurde so ‚amerikanisch‘ geprägt, wir versuchen dem gerecht zu werden, letztlich um marktfähig zu bleiben; Zuschauerzahlen im Quotenzwang - Daumen unten oder oben - Mainstream, wissentlich mit Verlusten in der Sprachverständlichkeit.

## 6.5. Workflow und Delivery-Anforderungen

Nach nunmehr 15 Jahren digitalem Filmtone ist der Workflow zwischen Aufnahme am Drehort, Postproduction und Mischatelier definiert und konsolidiert. Somit gehören Qualitätsverluste in der Verarbeitungskette der Vergangenheit an.

Dringender Klärungsbedarf besteht jedoch an der Schnittstelle zwischen Tonstudio und Sendeanstalten. Ein Konvolut der diffusen Delivery-Anforderungen spiegeln oft die mangelnde Kompetenz auf Seiten der Entscheidungsträger, was angeliefert werden soll: DD-Downmix, 2.0 TV Mix, 5.1-Mix, AC3, PCM-Stereofassung, LT RT TV Mix, DME, Dolby E, ...???

Zusätzlich zu den Formatfragen ist die gewünschte Dynamik meist ungeklärt: original dynamic, R 128, simple digital normalized, airline version maximized, Sonderdynamik, ...???

All diese Faktoren haben gravierenden Einfluss auf die Sprachverständlichkeit. Was bei der Abnahme im Mischatelier noch brilliant und klar erklingt, bleibt – worst case - im Nadelöhr der Übergabe hängen. Was dann aus den Lautsprechern im Wohnzimmer dringt, erzeugt bei den ‚Erzeugern‘ nacktes Grauen; hier besteht dringender Klärungs-, Normierungs- und Schulungsbedarf.

## 7. Entwicklung einer Production Guideline

Elisabeth Bodenseh studierte in Wien an der Universität für Musik und darstellende Kunst das Fach Tonmeister mit dem Schwerpunkt Filmtone. Ihre Diplomarbeit zum Thema „Sprachverständlichkeit im Fernsehen“ schrieb sie in Zusammenarbeit mit dem SWR Stuttgart. Derzeit arbeitet sie als freie Tonmeisterin und ist u.a. in einem Filmtonestudio als Mischtonmeisterin tätig.

Sprachverständlichkeit im Fernsehen ist in den letzten Jahren zu einem brisanten Thema gewachsen. Obwohl es einen Hinweis bzgl. Sprachverständlichkeit und Mischungsverhältnis in den „Technischen Richtlinien – HDTV zur Herstellung von Fernsehproduktionen für ARD, ZDF und ORF“ gibt, vermehren sich die Beschwerden der Zuschauer. Zusammen mit dem SWR wurde eine Diplomarbeit erstellt, die das Thema für Tonmeister und Toningenieure aufarbeitet und die Basis zu einer neu entstehenden „Production Guideline“ [11] zum Thema Verbesserung der Sprachverständlichkeit bietet. Diese Guideline spricht technische Empfehlungen aus, welche zu einer besseren Sprachverständlichkeit in Sendungen führen soll. Im Zuge der Diplomarbeit wurden zunächst Beschwerden der Sender ARD, WDR, ZDF, NDR, BR und SWR analysiert und die genannten Probleme ausgewertet. Das Ergebnis zeigt, dass das meist genannte Problem in Fernsehsendungen die zu laut gemischte Musik unter Sprache ist.

### 7.1. Hörversuch und Ergebnisse

Darauf basierend wurde ein Hörversuch mit 21 Probanden durchgeführt. Das Alter der Probanden reichte von 14 – 64 Jahren. Die Testpersonen mussten mit Hilfe eines Fragebogens vier Filmausschnitte aus den Genres Spielfilm, Dokumentation, Unterhaltungsshow und Sport bewerten (Abb. 4). Gefragt wurde nach der Höranstrengung und der Beurteilung der Lautstärkeverhältnisse zwischen Sprache und Musik. Jedem Ausschnitt wurden 4 verschiedene Mischvarianten unterlegt. Diese unterschieden sich im Lautstärkeverhältnis zwischen Sprache und Musik bzw. Sprache und Geräuschen. Die Sprache blieb dabei immer gleich laut bei 0 LU, die Musik/Geräusche wurden in vier unterschiedlichen Lautstärkestufen dazu gemischt. Daraus entstanden zu jedem Filmausschnitt folgende Mischungsvarianten:

- Sprache 0 LU – Musik/Geräusche -2LU
- Sprache 0 LU – Musik/Geräusche -2LU + EQ
- Sprache 0 LU – Musik/Geräusche -7LU
- Sprache 0 LU – Musik/Geräusche -10LU

Die insgesamt 16 Ausschnitte wurden den Probanden randomisiert vorgespielt.



Abb. 4: Hörversuch im Hörspielstudio des SWR in Stuttgart

Die Ergebnisse des Fragebogens zeigten, dass die verschiedenen Varianten von den Probanden kaum differenziert werden konnten, obwohl die verschiedenen Mischungsverhältnisse aus technischer Sicht sehr unterschiedlich waren. Auffallend war allerdings, dass die älteren Probanden (über 50 Jahre) bei den Varianten, in denen die Musik nur 2 LU leiser als die Sprache war, mehr Schwierigkeiten hatten als die Probanden unter 50 Jahre. Man kann also daraus den Schluss ziehen, dass die älteren Zuschauer aufgrund einer anzunehmenden Altersschwerhörigkeit bei zu laut gemischter Musik mehr Probleme haben als jüngere Zuschauer.

### 7.2. Empfehlungen für die Guideline

Da das Durchschnittsalter der meisten Fernsehsender bei geschätzten 60 Jahren liegt, stellt sich also die Frage, in wie weit man die Mischungen einer Sendung an das „Zielpublikum“ anpassen sollte. Aus dem Ergebnis des Hörvergleichs kann man schließen, dass es in Bezug auf das Mischungsverhältnis einen sehr großen Spielraum gibt, in welchem der „guthörende Zuschauer“ Unterschiede kaum wahrnimmt. So könnte also die Musik unter der Sprache um wenige LU reduziert werden, so dass auch ältere Zuschauer mit einem gewissen Grad an Altersschwerhörigkeit die Sprache besser verstehen können.

Diese Empfehlung wurde unter anderen in die neue Production Guideline aufgenommen und soll zu einer Verbesserung der Situation führen.

## 8. Zusammenfassung

Das Problembewusstsein für Sprachverständlichkeit, insbesondere bei Schwerhörigkeit entwickelt sich bereits seit mehr als 50 Jahren - Lösungsvorschläge innerhalb der realen technischen Möglichkeiten wurden allerdings nur wenig genutzt, die praktische Umsetzung geeigneter Lösungen und Hilfestellungen ging nur punktuell voran.

Mit der eAccessibility 2005 [6] wurde der globale politische Wille zur Durchsetzung der Prinzipien dokumentiert und wird seither auch von vielen internationalen und nationalen Institutionen unterstützt.

Die Umsetzung der Lautheitsnormalisierung war ein erster Schritt in Richtung Zuschauer-Zufriedenheit, nun gilt es, diese „...auch im Hinblick auf eine durchgängig gute Sprachverständlichkeit zu gewinnen.“ Innovative Werkzeuge zur objektiven Überprüfung von Sprachverständlichkeit im Produktionsprozess können dabei ihren Beitrag leisten.

Ein wichtiger Schritt ist auch die Entwicklung bzw. Standardisierung des künftigen MPEG-H-3D-Audio-Standards. Die kanal-/objektbasierte Technologie des Fraunhofer IIS ermöglicht den Hörern ein interaktives „Sound Mixing“ und somit die Möglichkeit der Personalisierung von Audioübertragung. Das kommt nicht nur Schwerhörenden zu Gute. Der endgültige Standard soll im Frühjahr 2015 fertiggestellt werden [12].

So bleibt nur mehr zu hoffen, dass die entsprechenden Empfangsgeräte auch bedienfreundlich genug konzipiert werden, um den Nutzern die theoretischen Möglichkeiten des Standards auch praktisch zu erlauben.

## 7. Literatur

- [1] Ruhe, C., 50 dB Hörverlust - Was heißt das, was und wie hört man dann?, Informationsschrift Deutscher Schwerhörigenbund (2011)
- [2] Baumgartner, H. u. Hoeg, W., Barrierefreie Audio-kommunikation, in: Dickreiter/Dittel/Hoeg/Wöhr (Hrsg.), Handbuch der Tonstudioteknik, 8. Aufl., Kap. 17.7, DeGruyter (2014)
- [3] Mathers, C. D., A Study of Sound Balance for the Hard of Hearing. BBC Research Department, Kingswood Warren (1991)
- [4] Fuchs, H. et al., Dialogue Enhancement – Technology and Experiments. EBU Technical review (2012)
- [5] Salvi, G. et al., SynFace - Speech-Driven Facial Animation for Virtual Speech-Reading Support. In EURASIP Journal on Audio, Speech, and Music Processing (2009)
- [6] European Commission (EC): eAccessibility (2005)
- [7] EBU R128-2014. Audio loudness normalisation and permitted maximum level (2014)
- [8] <http://www.psp-dtv4all.org/>
- [9] ANSI (1997). *Methods for calculation of the speech intelligibility index*. American National Standard S3.5-1997 (Standards Secretariat, Acoustical Society of America, New York, USA).
- [10] ISO (1990). *Acoustics - Determination of occupational noise exposure and estimation of noise-induced hearing impairment*. International Standard ISO 1999 (International Organization for Standardization, Genf, Schweiz).
- [11] Eberhard M. et al., Sprachverständlichkeit im Fernsehen - Empfehlungen für Programm und Technik, ARD/ZDF FSBL-K (2014)
- [12] [http://www.iis.fraunhofer.de/de/pr/2013/20130903\\_Audio\\_3D\\_MPEG.html](http://www.iis.fraunhofer.de/de/pr/2013/20130903_Audio_3D_MPEG.html)