Abschlussbericht

zum *KMU-innovativ* Verbundprojekt

SITA -
Universell einsetzbares Softwaresystem zur automatisierten, individualisierbaren Verbesserung von Sprachverständlichkeit für Broadcast und Empfänger

Projektlaufzeit 09/2017 - 06/2020

Förderkennzeichen: 01IS17017 A-C

Beteiligte Institutionen

RTW GmbH & Co. KG, Köln
Fraunhofer IDMT Institutsteil Hör-, Sprach- und Audiotechnologie, Oldenburg
ebee Engineering GmbH, Dresden (vormals TechniSat Dresden GmbH)

Köln, 25.03.2021

Projektleitung: RTW GmbH & Co. KG, Herr Dipl.-Ing. Andreas Tweitmann

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IS17017 A-C gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Kurzdarstellung des Projektes

1.1 Aufgabenstellung

Für Menschen mit reduziertem Hörvermögen ist das Sprachverstehen beim Fernsehen bisweilen anstrengend bis unmöglich. Der Einsatz von Musik, Effekten und Nebengeräuschen ist manchmal ein Problem - nicht nur für ältere oder schwerhörende Menschen. Regelmäßig erhalten die Rundfunkanstalten entsprechende Beschwerden.

Das vom *Bundesministerium für Bildung und Forschung* (*BMBF*) geförderten Verbundprojektes *SITA - Speech Intelligibility Transformation & Autocorrection* setzt sich aus den Partnern RTW, Fraunhofer IDMT und ebee Engineering zusammen.

Das Projektziel war die Entwicklung einer Technologie zur objektiven Messung der Sprachverständlichkeit und zu deren automatischen und zielgruppenspezifischen Verbesserung. Dabei sollte das System nicht nur im professionellen Umfeld zum Einsatz kommen, wie beispielsweise bei Rundfunk- und Fernsehanstalten, sondern auch dem Endkunden am heimischen Fernsehgerät ein besseres Hörerlebnis ermöglichen.

Das Projekt gliederte sich dabei in die zwei Teilgebiete A) SITA für den Endverbraucher und B) SITA für den professionellen Audiobereich. In Abstimmung mit dem Projektverbund setzten die Projektpartner ihre einzelnen Aufgabengebiete um.

Dem Unternehmen RTW oblag die Aufgabe der Implementierung der Softwarelösungen sowie der Entwicklung eines Hardwareprototyps zur Messung und Visualisierung der Sprachverständlichkeit für den professionellen Audiobereich.

Das Fraunhofer IDTM verantwortete die Untersuchung der Sprachmodelle und die Entwicklung der Algorithmen zur Bewertung der Sprachverständlichkeit. Dies beinhaltete insbesondere Verfahren des maschinellen Lernens zur blinden Quellentrennung komplexer Audiosignale.

ebee implementierte die Technologien zur Verbesserung der Sprachverständlichkeit beim Endverbraucher.

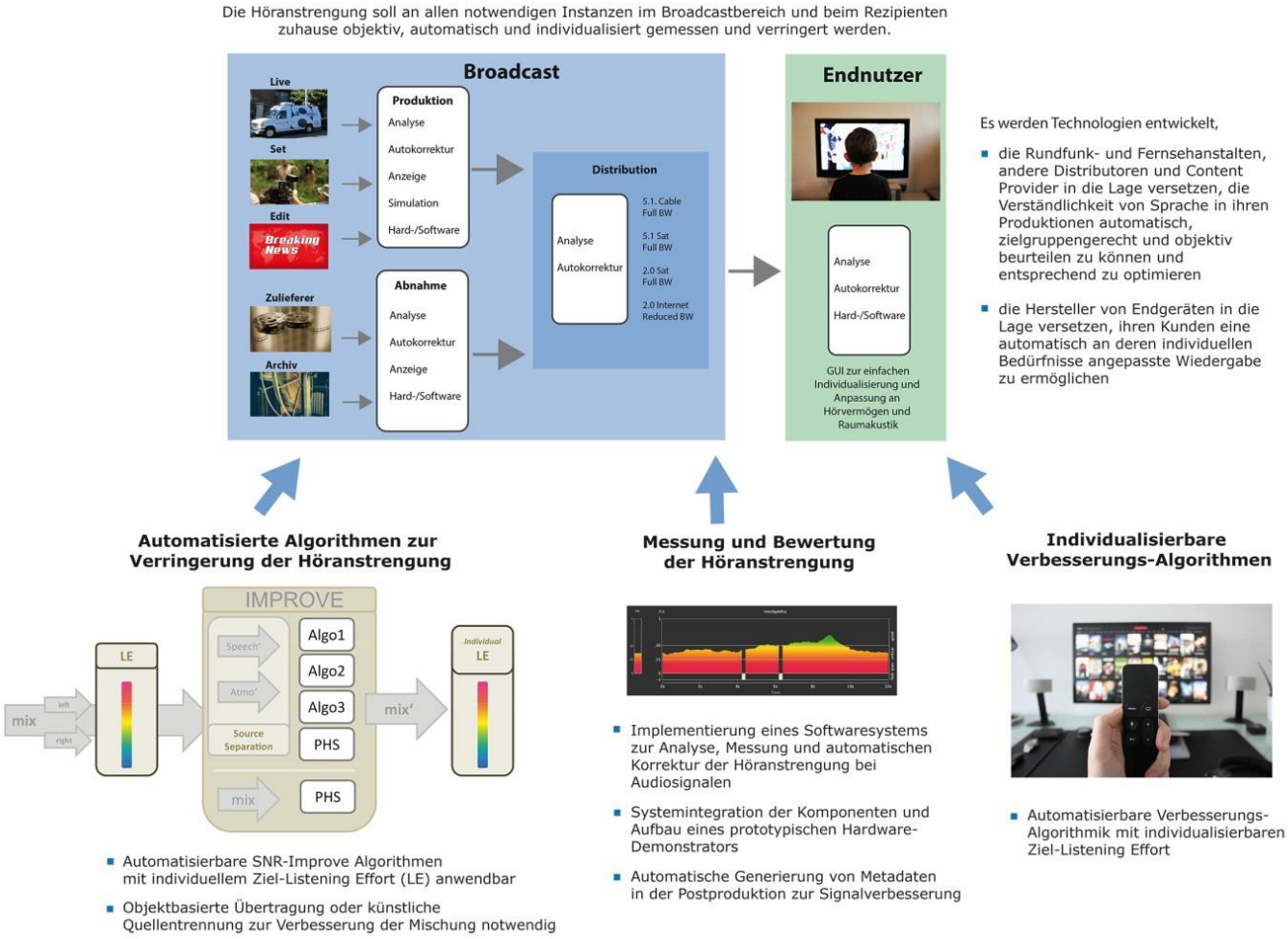

Abbildung 1: Thematischer Inhalt des SITA- Forschungsprojektes

1.2 Voraussetzungen des durchgeführten Vorhabens

Am 17. Juli 2017 erfolgte die Genehmigung des SITA-Verbundprojektes. Die inhaltliche Durchführung des Vorhabens wurde in fünf Arbeitspakete gegliedert, wobei die Koordination des Projektes durch das Unternehmen RTW erfolgte.

RTW

Das Unternehmen RTW mit Sitz in Köln blickt auf eine langjährige Unternehmensgeschichte zurück. Gegründet wurde das mittelständische Unternehmen als spezialisierter Service-Betrieb für Audiogeräte und entwickelte sich in den Folgejahren zu einem der wichtigsten Lieferanten für professionelle Broadcast-Technologie und Audio-Messinstrumente. Seit nunmehr 50 Jahren begleitet RTW den kontinuierlichen Technologiewandel in der professionellen Audiobranche mit innovativen Instrumenten und Technologien zur visuellen Audio-Signalkontrolle in Broadcast, Produktion, Nachbearbeitung und Qualitätsüberwachung.

RTW hat in den letzten Jahren verschiedene Forschungs- und Entwicklungsprojekte erfolgreich umgesetzt und vermarktet. Damit verfügt das Unternehmen sowohl über erfahrenes, hoch-qualifiziertes Personal als auch über die erforderliche technische Ausstattung zur Umsetzung des Projektes.

Die wichtigste Vorarbeit von RTW für das hier beantragte Projekt stellt das gemeinsam mit dem Fraunhofer IDMT durchgeführte Projekt *SI4B - Speech Intelligibility for Broadcast* dar. RTW entwickelt hier ein Steuerungs- und Visualisierungsmodul zur Bewertung und Korrektur von Sprachverständlichkeit im Broadcast-Bereich.

Fraunhofer IDMT

Der 2008 als Projektgruppe gegründete Institutsteil Hör-, Sprach- und Audiotechnologie (HSA) des Fraunhofer-Instituts für Digitale Medientechnologie IDMT ist Teil der Oldenburger Hörforschung, welche seit über 20 Jahren Grundlagenforschung zur menschlichen Hörwahrnehmung betreibt. Ihre Stärken liegen insbesondere in den Bereichen der subjektiven und modellbasierten Charakterisierung von individuellen Hörprofilen und Hörstörungen, der digitalen und statistischen Audiosignalverarbeitung sowie der Hörrehabilitation (wissenschaftliche Exzellenz u.a. belegt durch Vielzahl von Fachpublikationen, Exzellenzcluster Hearing4All, Deutscher Zukunftspreis 2012).
Zu den Technologien und Verfahren, die in diesen Bereichen entwickelt und integriert werden, gehören u.a. die für dieses Vorhaben relevanten Themen der Hörgerätealgorithmen (z.B. [1]) und der assistiven Technologien zur Unterstützung der alternden Gesellschaft mit Schwerpunkt auf Nutzerschnittstellen und Hörstörungen [2,3,4,5].
HSA hat sich zu einem interdisziplinären Forschungs- und Entwicklungs-Team bestehend aus Physikern, Akustikern, Ingenieuren, Psychologen und Produktdesignern entwickelt, um alle Aspekte eines nutzerzentrierten Entwicklungsprozesses abzudecken. Für diesen Prozess greift HSA auf umfangreiche Erfahrungen mit Nutzerstudien und nutzerzentrierter Entwicklung, ethischen Fragestellungen, auf eine umfassende Datenbank von normalhörenden und schwerhörigen Probanden (N>2400) sowie umfangreiche Infrastruktur (u.a. Mess-Hörkabinen, reflexionsarmer Raum, Hallraum, Kommunikationsakustik-Simulator) zurück. Der Institutsteil HSA verfügt über umfangreiche Projekterfahrungen (Industrieprojekte u.a. in den Märkten Medizintechnik, Automotive und Consumer Elektronik; diverse öffentlich geförderte Projekte), wobei insbesondere die Vorarbeiten aus dem BMBF-Projekt *SI4B - Speech Intelligibility for Broadcast* [6] eingebracht wurden.

ebee Engineering

Die Ingenieure der ebee Dresden GmbH entwickelten seit 1990 Consumer-Elektronik-Produkte für die TechniSat-Firmengruppe mit Hauptsitz in Daun (Eifel). Der Standort Dresden verfügt über ca. 70 Entwicklungsingenieure, die über umfangreiches Know-how auf den Gebieten Embedded-Systems, SW-Security-Architekturen, Verschlüsselungstechnologien, SW-Update-Mechanismen, drahtloser und drahtgebundener Kommunikation, Bild- und Audiooptimierung, HF-Entwicklung sowie Produktionsüberleitung verfügen. Vorrangig werden Produkte für die eMobility, Krankenhaustechnik sowie für die elektrische Sicherheit entwickelt. Zusätzlich liegt auch weiterhin ein Focus auf Radios, Receiver, LCD-TV-Geräte, HF-Verteiltechnik und Komponenten für Hausautomatisierung.

Im SITA-Projekt beschäftigte sich ebee vornehmlich mit dem Bereich der Sprachverständlichkeitsverbesserung beim Endverbraucher.

1.3 Planung und Ablauf des Vorhabens

Die Planung und Durchführung des Projektes gliederte sich entsprechend des Projektantrages in die unter Abschnitt 1.3.1 beschriebenen Arbeitspakete.

Aufgrund der nachfolgenden Erläuterungen wurde das Projekt kostenneutral um vier Monate bis zum 30.06.2020 verlängert:

Im Jahr 2018 gab es vorübergehend Vakanzen im Projektteam. Diese konnten zwar im Verlauf des Jahres wieder geschlossen werden, dennoch mussten einzelne Arbeiten in das Jahr 2019 verlegt werden (s.a. Abschnitt 2.1, AP 4.4).

Des Weiteren ergaben sich Abweichungen vom Projektplan, die aufgrund der Beschränkungen der COVID-19-Pandemie eingetreten sind. In diese Zeit fiel die kostenneutrale Verlängerung des SITA Projektes, in der noch abschließende Arbeiten durchgeführt werden sollten (s.a. Abschnitt 2.1, AP 5).

1.3.1 Ablauf und Organisation

Der Ablauf des Vorhabens war in fünf Phasen gegliedert. Abbildung 2 zeigt den groben Ablauf des SITA-Projektes. Die gelb markierten Felder stellen dabei den Zeitpunkt der jeweiligen Meilensteine dar:

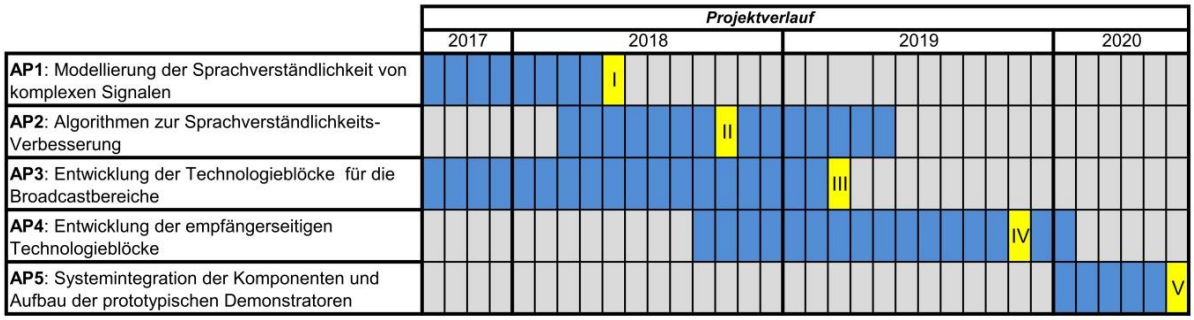

Abbildung 2: Übersicht zum Ablauf des SITA Projektes

In jeder der fünf Phasen war ein Meilenstein umzusetzen:

I	Monat 9	Quellentrennung	Softwaremodul „Modellierung der Sprachverständlichkeit von komplexen Signalen in komplexen Situationen“	Ende AP1
II	Monat 14	Korrekturprozess	1. Meilensteintreffen mit PT: Finalisierung des Automatisierungskonzeptes	Ende AP2
III	Monat 19	Hardware Broadcast	Integration aller Technologieblöcke im Broadcastbereich	Ende AP3
IV	Monat 25	Hardware Empfänger	2. Meilensteintreffen mit PT: Integration aller Technologieblöcke (Broadcast- und Empfängerbereich)	Ende AP4
V	Monat 34	Demonstratoren	Prototypische Demonstratoren für die Funktionalität	Ende AP5

Abbildung 3: Die fünf SITA-Meilensteine

Das Teilprojekt wurde im September 2017 mit AP1 gestartet. Ein Kick-off Meeting mit allen Verbundpartnern wurde am 20. September 2017 bei RTW in Köln durchgeführt. Ebenfalls in Köln fand im Februar 2018 ein Workshop „SITA-Beratertreffen“ statt, zu welchem potentielle und assoziierte Partner aus dem professionellen Umfeld Broadcast/ Postproduktion angeschrieben und eingeladen wurden. Im November 2018 folgte das erste Meilensteintreffen im Rahmen der Tonmeistertagung 2018 in Köln. Das Projekt präsentierte sich auf der Tonmeistertagung im Rahmen der Session „Sprachverständlichkeit in Film und Fernsehen“, initiiert von HSA-Mitarbeiterin Hannah Baumgartner.

Grundlage des regelmäßigen Austausches waren Telefonkonferenzen mit allen Partnern oder bei Bedarf in einem engeren Kreis. Des Weiteren fanden projektbegleitend mehrere Projekttreffen per anno entlang der Meilensteine sowie ein Abschlusstreffen statt.

Veranstaltung	**Datum**	**Ort**
Projekttreffen: „Kick-off“	September 2017	RTW, Köln
Projekttreffen	November 2017	Fraunhofer IDMT, Oldenburg
Treffen mit assoziierten Partnern	Februar 2018	RTW, Köln
Projekttreffen	Mai 2018	Fraunhofer IDMT, Oldenburg
Meilensteintreffen I	November 2018	RTW, Köln
Projekttreffen	Januar 2019	RTW, Köln
Meilensteintreffen II	November 2019	Fraunhofer IDMT, Oldenburg

Abbildung 4: Liste der Schlüsselveranstaltungen im Forschungsprojekt SITA

1.3.2 Inhalt der Arbeitspakete

In **AP1** widmete sich der Partner Fraunhofer IDMT der Modellierung von Sprachverständlichkeit von komplexen Signalen in komplexen Situationen. Einer der Schwerpunkte war eine erfolgreiche Signalklassifikation und Quellentrennung, um für die Sprachverständlichkeitsverbesserung Sprache und Hintergrund zu separieren.

In **AP2** wurden verschiedene Algorithmen zur Verbesserung der Sprachverständlichkeit implementiert und untersucht. Diese Algorithmen finden zum einen in der Postproduktion ihre Verwendung durch professionelle Anwender (AP2.1), zum anderen kommen sie automatisiert und individualisiert bei der Signaladaption im heimischen Umfeld zum Einsatz (AP2.2). Die in AP1 erforschten Modelle dienen dabei als Steuergrößen. Im AP2 lag ein Großteil der Entwicklungsarbeit bei Fraunhofer IDMT. Mögliche Metadatenkonzepte sollten von RTW überprüft werden, um in den Algorithmen eingesetzt zu werden.

In den APs 3 und 4 wurden entsprechende spezifische Technologieblöcke für den Broadcast- und für den Heimanwenderbereich aufgebaut.

In **AP3** wurden alle Abläufe, die während der Postproduktion zu einer zielgruppenspezifischen Mischung führen, automatisiert. Die Abhörmöglichkeiten in der Postproduktion wurden um eine Simulation des beim Rezipienten ankommenden Mix-Signals erweitert - in Abhängigkeit von Raumgröße, Raumbeschaffenheit und Hörfähigkeiten (AP3.1). Die gesamte Technologie musste sich in die herrschenden Workflows der Produktionsbetriebe einfügen und sich ohne Mehraufwände

anwenden lassen (AP3.2). RTW sollte dabei einen komplexen Hintergrundprozess zur Erfassung von Regeldaten konzeptionieren und aufbauen (AP3.3), mit welchem die Abläufe vereinfacht oder automatisiert (AP3.4) werden können. Schwerpunkt des AP3 war der Aufbau eines Hardware-Demonstrators zur Analyse und Visualisierung der Sprachverständlichkeit (AP3.6). Zur softwareseitigen Steuerung des Demonstrators und zur Visualisierung der Messergebnisse war die Entwicklung eines intuitiven Benutzer-Interfaces notwendig (AP3.5). Evaluation der einzelnen Komponenten wurde in AP3.7 umgesetzt.

Die Schwerpunkte des **AP4** waren die Ermittlung der Abhörbedingungen beim Empfänger und die technische Integration der Sprachverständlichkeitsverbesserung beim Endverbraucher zu Hause. Dabei waren Faktoren wie die vorhandene Raumakustik (AP4.1) und die an die individuellen Höreigenschaften des Rezipienten angepassten Korrekturparameter zu ermitteln (AP4.2), welche die Algorithmen der Sprachverständlichkeitsanalyse und -korrektur voreinstellen (AP4.3). Der Projekt-partner ebee sollte dazu seine Expertise in Bezug auf Nutzerschnittstellen und Technologie-integration in das Arbeitspaket mit einbringen und so die von Fraunhofer IDMT entwickelten Algorithmen für den Einsatz beim Endverbraucher integrieren.

In **AP5** sollte nach weiteren Systemtests und Usability-Studien eine abschließende Systemintegration der prototypischen Demonstratoren durchgeführt werden.

Entscheidend für den Ablauf des SITA-Projektes war die enge Zusammenarbeit zwischen den Projektpartnern sowie eine starke Orientierung in Richtung Fernseh- und Rundfunkanstalten, um diese während der gesamten Laufzeit des Projektes mit einzubeziehen. Diese Teilhabe wurde u.a. durch Projekttreffen, Workshops und Tagungen abgesichert. Eine Übersicht der wichtigsten Veranstaltungen im Projekt ist in Abbildung 4 sowie unter Punkt 2.6 zu finden.

1.4 Stand der Wissenschaft und Technik zu Beginn des Projektes

Für die Umsetzung des Projektes mussten verschiedene Technologien und algorithmische Ansätze entwickelt und in geeigneter Weise zusammengeführt werden, wie es zum Projektstart in keinem System möglich war. Die unterschiedlichen Technologienblöcke waren:

1. Messung und Bewertung von Sprachverständlichkeit
2. Abhörbedingungen im Broadcast-Bereich
3. Algorithmen zur Verbesserung der Sprachverständlichkeit für die Postproduktion
4. Individualisierte automatisierte Sprachverständlichkeitsverbesserung für den Empfänger zu Hause
5. Automatische Quellentrennung und Signalklassifikation basierend auf Machine-learning-Verfahren.

Neu ist der Ansatz einer automatisierten Verbesserung komplexer Audioaufnahmen bezüglich der Sprachverständlichkeit - ein Prozess, der bisher rein manuell in der Postproduktion stattfinden konnte. Im Folgenden wird ein Überblick zum Stand der Technik zu Beginn des Projektes gegeben. Eine detaillierte Beschreibung der einzelnen Punkte ist in der Vorhabenbeschreibung zum SITA-Projekt zu finden.

1.4.1 Stand der Technik: Messung und Bewertung von Sprachverständlichkeit

Die Wahrnehmung von Sprache ist ein hochgradig komplexer Prozess, der von diversen Faktoren abhängt. Diese lassen sich folgendermaßen unterteilen:

Eigenschaften des Sprachsignals	• Intensität bzw. Lautstärke • spektraler Gehalt • zeitliche Struktur • Qualität der Aussprache etc.
Eigenschaften des „Störgeräuschs“ - Beim Film: Musik, Atmo, Voice Under, Effekte, etc.	• Intensität relativ zum Sprachsignal • Spektralgehalt relativ zum Sprachsignal • Räumliche Anordnung relativ zum Sprachsignal • Zeitliche Struktur • Informationsgehalt etc.
Eigenschaften der Signalübertragung	• Bandbegrenzung oder Artefakte • Encoding/ Decoding • Endgeräte etc.
Individuelle Gegebenheiten beim Rezipienten	• Abhörraum • Hörverlust • Sprachkenntnisse • Kontextwissen wie Bildinformationen • Etwas zum zweiten Mal hören etc.

Abbildung 5: Einflussgrößen für Sprachverstehen und ihre Verortung in der Übertragungskette

Während die Rolle einzelner Faktoren vielfach wissenschaftlich untersucht wurde und für einzelne Abhörsituationen quantitativ durch Vorhersagemodelle berechnet werden kann, sind die zugrundeliegenden perzeptiven Mechanismen der Sprachwahrnehmung sowie die Interaktion einiger dieser Faktoren noch nicht hinreichend verstanden.

- Es existieren Modellansätze, die die Verarbeitungsstufen des menschlichen Gehörs effektiv nachbilden und so ohne spezifische Anpassungen Sprachverständlichkeit in verschiedenen Abhörbedingungen vorhersagen können.
- Es existieren der Sprachverständlichkeitsindex (engl. speech intelligibility index, SII; [15]) und der Sprachübertragungsindex (engl. speech transmission index, STI; [16]) als standardisierte Methoden zur instrumentellen Berechnung von Sprachverständlichkeit.
- Es existiert noch kein instrumentelles Verfahren, welches alle Faktoren berücksichtigt und eine objektive Bewertung von Sprachverständlichkeit zulässt.
- Alle genannten Ansätze erfordern ein getrenntes Vorliegen von Sprachsignal und Nebengeräuschen. Bestehende „blinde“ Ansätze, die Verständlichkeit aus gemischten Signalen abschätzen (z.B. [17]), sind in ihrer Anwendbarkeit und Zuverlässigkeit noch weiter eingeschränkt.

Die Projektpartner Fraunhofer IDMT und RTW entwickelten im gemeinsamen Projekt *SI4B* ein Analyse- und Anzeigeverfahren zur objektiven Beurteilung der Verständlichkeit in der Postproduktion bei Rundfunkanstalten, um möglichst vor der Ausstrahlung geeignete Maßnahmen zur Verbesserung vorzuschlagen. Das Projekt zeigte:

- Eine modellbasierte Messung der Verständlichkeit in der Postproduktion ist grundsätzlich möglich.

- Die Sprachverständlichkeit kann mit bestehenden Verfahren nur beurteilt werden, solange der Mix in getrennten Spuren vorliegt (somit nur anwendbar in der Postproduktion und dort auch nur, wenn die Zielsprache mit guter Qualität separat produziert / aufgenommen wurde).
- Eine Verbesserung der Sprachverständlichkeit findet ausschließlich manuell durch den Tonmeister statt, es erfolgt keinerlei Automatisierung.

1.4.2 Stand der Technik: Abhörbedingungen im Broadcast-Bereich

Die Sprachverständlichkeit von Film- und Fernsehprodukten wird weltweit rein subjektiv erfasst. Die Modellierung diverser Nutzergruppen oder realer Abhörbedingungen sind dabei nicht üblich.

Zur Simulation von Hörschädigungen existieren VST-Plug-ins, die eine Schwerhörenden-Simulation ermöglichen; deren Einsatz ist aber sehr umständlich und die Markakzeptanz ist gering.

Es existieren Methoden zu Raumsimulation, allerdings dienen diese als gestaltendes Mittel in einer Mischung und sind nicht für die Integration in den Produktionsablauf vorgesehen.

1.4.3 Stand der Technik: Sprachverständlichkeitsverbesserung für die Postproduktion

Es gibt Software-Plug-ins oder Effekte, die sich zur Aufbereitung von Sprachsignalen anwenden lassen - Equalizer, Kompressoren, etc. Allerdings setzt der Einsatz der unterschiedlichen Algorithmen Zeit und Geschick der Tonverantwortlichen voraus und erfolgt nicht automatisiert oder modellgesteuert.

- Keiner der Algorithmen ist auf die Verbesserung der Sprachverständlichkeit spezialisiert.
- Der Einsatz ist auf den Bereich Musikmischung fokussiert.
- Ein weiterer Einsatzbereich ist innerhalb der Forensik

1.4.4 Stand der Technik: Individualisierte Sprachverständlichkeitsverbesserung für den Empfänger zu Hause

Einstellmöglichkeiten der TV-Geräte

Es ist üblich, dass TV-Geräte über optimierte Voreinstellungen für die Tonwiedergabe verfügen. Standardfunktionen sind zumeist individualisierbare Equalizer-Einstellungen (EQ), ohne eine dynamische Anpassung an das aktuell vorliegende Broadcast-Signal.

Die Algorithmen beschränken sich auf einfache Signalmodifikationen (insb. EQ) und nutzen die Möglichkeiten aktueller Sprachverbesserungsalgorithmen, die auch die Signaldynamik berücksichtigen, nicht aus. Hinzu kommt, dass der Nutzer nur begrenzte Möglichkeiten hat, die akustischen Eigenschaften des Raumes zu berücksichtigen.

Personalisierte Verarbeitung des Sendesignals

Personalisierte Klanganpassung in TV-Kopfhörern: Die Sennheiser electronic GmbH & Co. KG und das Fraunhofer IDMT haben einen Kopfhörer (Modell RS 195) entwickelt, der durch Hörgerätealgorithmen personalisierte Klanganpassung altersbedingte Hörminderungen kompensiert.

Hörunterstützung über Smartphones: Die Sennheiser-Lösung MobileConnect ermöglicht Live-Audio-Streaming auf mobile Endgeräte via WLAN, um das Sprachverstehen in verschiedenen Anwendungsgebieten zu optimieren.

1.4.5 Stand der Technik: Automatische Quellentrennung und Signalklassifikation

Objektbasierte Audiocodecs: Der MPEG-Standard SAOC (Spatial Audio Object Coding) ermöglicht eine effiziente Übertragung einzelner Audio-Objekte innerhalb eines kanalbasierten Mixes. Durch Metadatierung einzelner Audioobjekte (wie Sprache, Hintergrund, Musik, etc.) können die Verhältnismäßigkeiten des übertragenen Mixes beim Empfänger nochmal verändert bzw. personalisiert werden: einzelne Audio-Objekte können verstärkt oder stumm geschaltet werden. Dabei ist SAOC kompatibel mit nahezu allen Wiedergabemöglichkeiten (Stereo, Surround, etc.).

Verfahren zur blinden Quellentrennung: Es existieren in der Audiosignalverarbeitung verschiedene Verfahren zur blinden Quellentrennung, die intensiv in unterschiedlichen Forschungsgebieten untersucht werden. (z.B. Non-negative matrix factorization, NMF).

1.5 Zusammenarbeit mit anderen Stellen

Im Rahmen des Projektes fanden Treffen und Workshops mit assoziierten Partnern aus dem Umfeld Film und Fernsehen statt. Diese hatten im Verlauf des Projektes die Gelegenheit, den SITA-Ansatz anhand eines produktnah realisierten Software-Plug-ins zu testen. Für weitergehende Informationen wird an dieser Stelle auf die Erläuterungen in Kapitel 2, Abschnitt AP 3.2 (Beratertreffen mit den assoziierten Partnern) sowie auf den Abschnitt des AP 3.7 (Software-Evaluation) verwiesen.

Im Verlauf des Projektes erfolgte außerdem eine Kooperation mit der Carl-von-Ossietzky - Universität Oldenburg im Bereich Algorithmik für blinde Quellentrennung. Eine gemeinsame Veröffentlichung zu Quellentrennungs-Algorithmen im Anwendungsbereich Broadcast ist in Arbeit.

2 Eingehende Darstellung des Projektverbundes

Das SITA Forschungsprojekt gliederte sich in die beiden Teilbereiche A) SITA für den professionellen (SITA-Pro) Bereich und B) SITA für den Heimanwender (SITA-Home). Dabei gab es thematisch keine strikte Abgrenzung zwischen den Teilbereichen. So konnten beispielsweise die in SITA-Pro entwickelten Lösungen auch in SITA-Home eingesetzt werden und umgekehrt.

Die jeweiligen Bereiche wurden in unterschiedliche Teilarbeitspakete untergliedert, die von den Projektpartnern entsprechend bearbeitet wurden. In Abbildung 6: Schwerpunktthemen des SITA-Projektessind die inhaltlichen Schwerpunkte des Projektes zusammengefasst.

Abbildung 6: Schwerpunktthemen des SITA-Projektes

Im Wesentlichen konnten alle Aufgaben der jeweiligen Arbeitspakete umgesetzt werden. Allerdings mussten gewisse Abstriche im Projekt gemacht werden, die fehlendem Personal und der COVID-19-Pandemie geschuldet waren.

Im Folgenden werden die erzielten Ergebnisse der jeweiligen Arbeitspakete erläutert und der geplanten Aufgabenstellung des Projektes gegenübergestellt.

2.1 Ergebnisse

AP 0 - Projektmanagement

In diesem Arbeitspaket wurde die Leitung des Projektes durch den Partner RTW GmbH vorgenommen, in der insbesondere die Organisation des Austausches der beteiligten Partner einen Schwerpunkt darstellte. Dazu gehört darüber hinaus die Abstimmung mit dem Projektträger und den assoziierten Partnern.

AP 1 - Modellierung der Sprachverständlichkeit von komplexen Signalen

In AP1 modellierte der Projektpartner Fraunhofer IDMT die Sprachverständlichkeit von komplexen Audiosignalen. Es wurden drei Module entwickelt: eine Sprachaktivitätserkennung, eine modellbasierte Abschätzung der Höranstrengung und eine blinde Quellentrennung zum Erhalt von Sprache und Hintergrund. Alle drei Module basieren auf neuronalen Netzen.

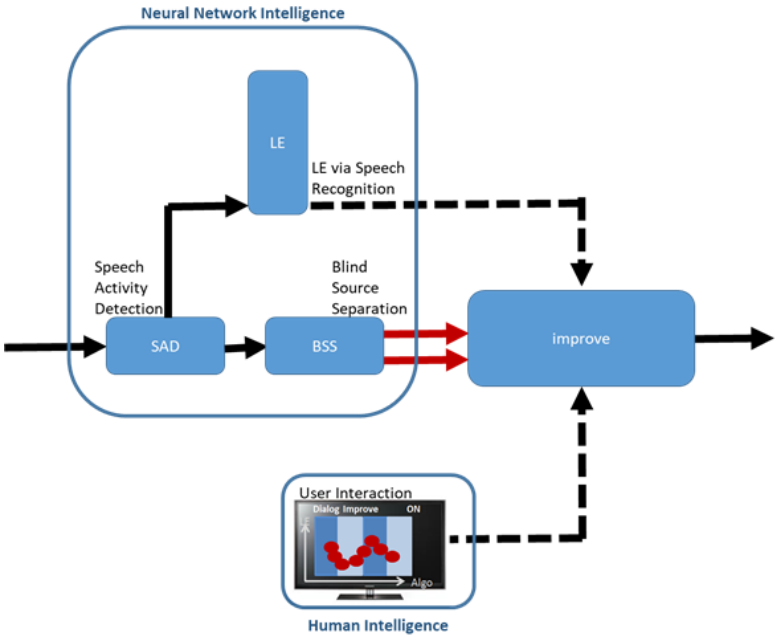

Abbildung 7: Verwendung neuronaler Netze zur Erkennung von Sprachanteilen im Signal.
Die Sprachaktivitätserkennung (SAD) analysiert das eingehende Signal auf die Anwesenheit von Sprache, der akustische Teil eines Spracherkenners (Speech Recognition) schätzt die Wahrscheinlichkeit einzelner Phoneme und erschließt so die Höranstrengung des Kontexts. Die Signaltrennung in Sprache und Hintergrund (Blinde Quellentrennung/ BSS) ermöglicht eine neue verständlichkeitsverbessernde Mischung.

Ergebnisse des Teilarbeitspaketes AP 1.1: Modellbasierte Schätzung der Sprachverständlichkeit

Mit Hilfe einer Sprachaktivitätserkennung (SAD) werden Signalanteile mit Sprache detektiert. Nur diese werden bezüglich ihrer Höranstrengung bzw. Verständlichkeit analysiert. Für die Modellierung der eigentlichen Sprachverständlichkeit wurde ein referenzfreies Modell zur Vorhersage der empfundenen Höranstrengung implementiert. Dieses Verfahren verwendet zur Schätzung der Sprachverständlichkeit von gemischten Signalen sogenannte „Posteriorgramme“ [7,8], welche üblicherweise am Ausgang von (tiefen) neuronalen Netzen automatischer Spracherkennungssysteme anliegen. Diese Posteriorgramme beinhalten die Wahrscheinlichkeiten für die Aktivität verschiedener Sprach-Phoneme, aufgetragen über die Zeit (Abbildung 8). Störungen von Sprache wie z.B. Hintergrundgeräusche „verschmieren“ diese Phonemwahrscheinlichkeiten.

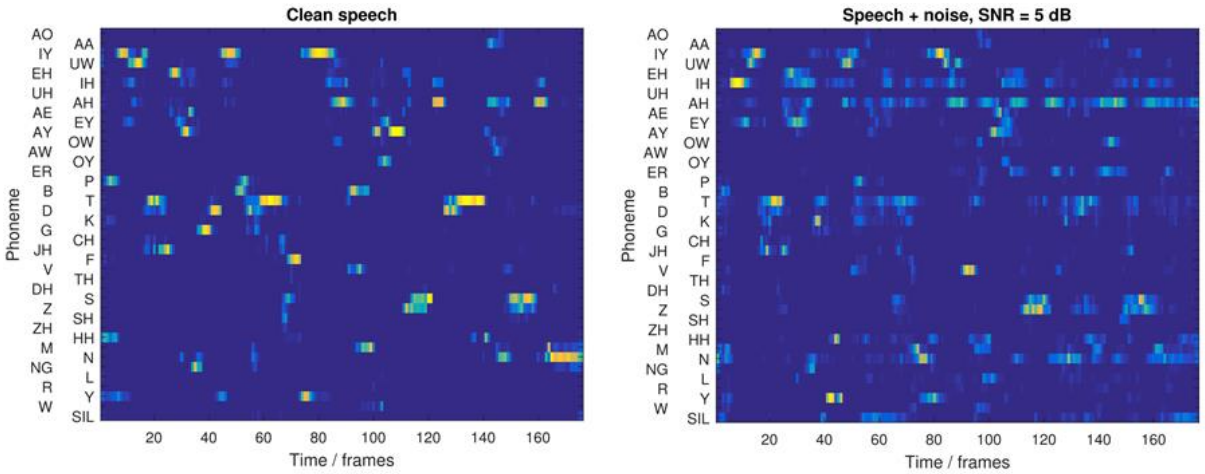

Abbildung 8: Verwendung von Phonem-Wahrscheinlichkeiten ("Posteriorgramme") eines DNN zur automatischen Spracherkennung.
Links: Das Posteriorgramm für ungestörte Sprache. Rechts: Posteriorgramm für Sprache mit Störgeräusch im Hintergrund. Der Grad der Posteriorgramm-"Verschmierung" – also wenig eindeutige Phonemwahrscheinlichkeiten, stattdessen verteilte geringere Wahrscheinlichkeiten - der Phoneme „entspricht" dem schlecht verstehen.

Der Grad der Verschmierung wird mit Hilfe eines geeigneten Maßes („M-Measure" von Hermansky et al., 2013 [9]) quantifiziert. Dieses Maß korreliert mit der subjektiven Beurteilung von Höranstrengung bewertet anhand einer Listening Effort Skala (Abbildung 9). Die subjektive Beurteilung von Höranstrengung wurde mit einer Vielzahl von Probanden in zahlreichen Hörversuchen evaluiert.

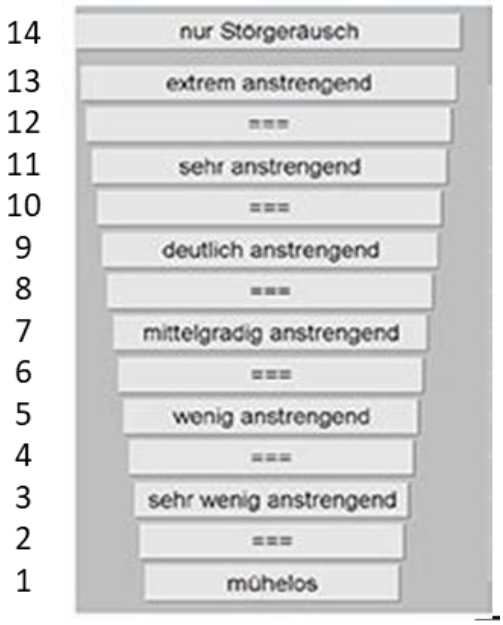

Abbildung 9: subjektive Bewertung der Höranstrengung anhand einer Skala.

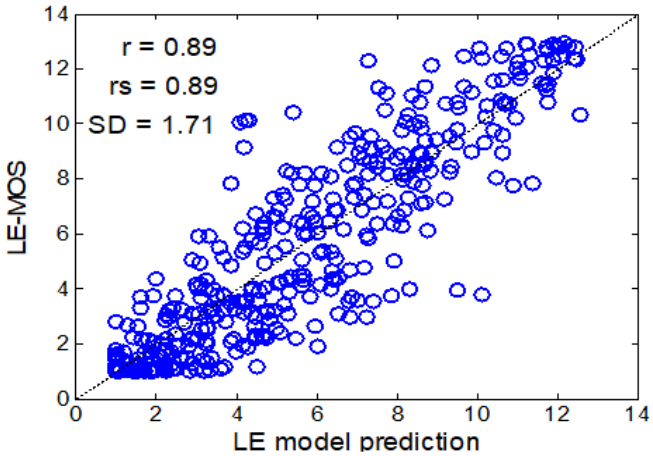

Abbildung 10: Scatterplot für die Entsprechung der Modellwerte zum Mean Opinion Score der Listening Effort Bewertung über alle Probanden.
Die Punkte stehen für unterschiedliche Hörbeispiele.

Das so implementierte Höranstrengungsmodell konnte Korrelationen (Korrelationskoeffizient nach Pearson) von 0.68 (Studie 1) bzw. 0.89 (Studie 2) zwischen subjektiv bewerteten Höranstrengungen und entsprechenden Modellvorhersagen erzielen und ist echtzeitfähig (Abbildung 10).

AP1.1 konnte entsprechend der Planung abgeschlossen werden. Im weiteren Projektverlauf wurde das LE-Modell zunehmend erweitert und evaluiert (Englisch, Spanisch, Voice-over-Voice).

Ergebnisse des Teilarbeitspaketes AP 1.2: Quellentrennung komplexer Signale

Wichtige Voraussetzung für die in SITA entwickelten Algorithmen zur Sprachverständlichkeits-verbesserung (AP2) ist die Trennung von Sprache und Hintergrund, auf deren Basis ein neuer Mix erstellt wird, der die Höranstrengung für den Nutzer auf ein (subjektiv) angenehmes Maß verringert (vgl. Abbildung 7).

Zur Quellentrennung komplexer Signale wurden mehrere Methoden auf Performanz und Anwendbarkeit untersucht. Mehrere zunächst vielversprechende Ansätze stellten sich im Laufe der Evaluation als für SITA nicht geeignet heraus. Der „Durchbruch“ bzgl. einer ausreichenden Qualität der quellengetrennten Signale gelang schließlich - mit leichter Verspätung relativ zur Projektplanung - mit Ansätzen recurrenter neuronaler Netze [10,11,12]. Um Streamingfähigkeit und kausale Verarbeitung zu ermöglichen wurde ein BLSTM-Ansatz (bidirectional long short-term memory network approach) zur Quellentrennung im Projekt auf ein LSTM Framework reduziert.

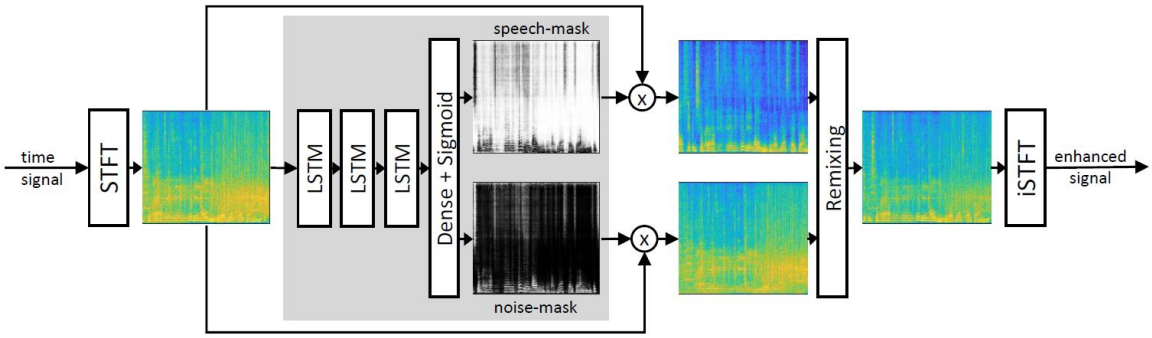

Abbildung 11: Schematik des entwickelten BSS-Systems.
Das Zeitsignal wird in den Zeit-Frequenzbereich transformiert, das LSTM-Netz berechnet Zeit-Frequenz-Masken für Sprach- und Hintergrundsignal. Das Remix-Signal mit deutlicherem Sprachanteil wird zurück in den Zeitbereich transformiert.

Das LSTM-System wurde für die Echtzeitverarbeitung beim Empfänger (SITA-Home) ausgelegt. Für Anwendungen im professionellen Bereich (SITA-Pro) ist der Einsatz des nicht-echtzeitfähigen BLSTM-Ansatzes als Offline-System mit höherer Qualität weiterhin denkbar.

AP1.2 wurde mit einiger Verspätung zur Planung abgeschlossen. Es wurden einige Ansätze und Versionen erprobt, bis mit den LSTM-Netzen eine zufriedenstellende Lösung gefunden werden konnte.

AP 2 - Algorithmen zur Sprachverständlichkeitsverbesserung

Dieses Arbeitspaket beinhaltete die Entwicklung und Untersuchung verschiedener Methoden und Algorithmen, die zur Verbesserung der Sprachverständlichkeit eingesetzt werden können. Diese Algorithmen finden zum einen durch professionelle Anwender in der Postproduktion ihre Verwendung (AP3), zum anderen sollten sie automatisiert bzw. individualisiert bei der Signaladaption im heimischen Umfeld eingesetzt werden (AP4). Das in AP1.1 erforschte Höranstrengungsmodell liefert dabei die Steuergröße. Für den (automatisierten) Remix ist es wichtig, dass sich das Klangbild der Mischung durch die Algorithmen nicht verschlechtert. Der Großteil der Entwicklung wurde in diesem Arbeitspaket durch das Fraunhofer IDMT umgesetzt.

Ergebnisse der Teilarbeitspakete AP 2.1 und AP 2.2: Automatisierte Algorithmen für die Broadcastbereiche und individualisierbare Verbesserungsalgorithmen

Ein im Vorgängerprojekt SI4B entstandenes Höranstrengungsmodell [13] nutzt bei seiner Schätzung der Höranstrengung das frequenzabhängige SNR-Verhältnis von Sprache und Hintergrund. Dieses eigentlich überholte Modell wurde in SITA für die Anwendung zur Verbesserung der Höranstrengung angepasst und optimiert. Die automatisierbaren Algorithmen greifen auf bestimmte Funktionen des SI4B-Modells zurück, um Mischungsverhältnisse der quellengetrennten Signale automatisch so zu gestalten, dass eine vorgegebene Höranstrengung nicht überschritten wird. In Sita wurden zwei automatisierte Strategien zur Verbesserung der Höranstrengung verfolgt - ein adaptiver Algorithmus, welcher nur eingreift, wenn die Höranstrengung des Originals eine gewisse Schwelle überschreitet und ein einfacher, statischer Remix-Ansatz, indem das Mischungsverhältnis pauschal zu Gunsten der Sprache verbessert wird.

Die getrennten Spuren können in der Postproduktion auch manuell justiert werden, mit Hilfe des LE-Meters können kritische Stellen im Signal einfach gefunden werden.

Der Erfolg der vorgeschlagenen Strategien bei optimal getrennt vorliegenden Sprach- und Hintergrundsignalen ist sehr gut und beschreibt die bestmögliche Signal-Verbesserung, sozusagen den „Gold Standard“ der Verbesserungswirkung. Die mit den künstlich getrennten Signalteilen tatsächlich erzielte Höranstrengungsverringerung der vorgeschlagenen Strategien im Verhältnis zu diesem Gold Standard, konnte deshalb auch immer als Qualitätsmaß für die Güte der künstlich getrennten Signale bzw. zur Evaluation der Quellentrennungsalgorithmen genutzt werden.

**Abbildung 12: Einfache Realisierung der „Listening-Effort-controlled“ Remix-Funktion durch einfache Eingabe eines Zielwertes.
Eine aufwendigere Realisierung erfolgt über eine zweidimensionale Touch-Oberfläche auf der zwischen verschiedenen Algorithmen und Ziel-Listening-Efforts navigiert werden kann.**

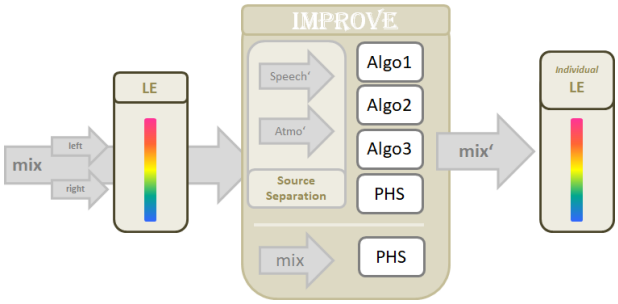

**Abbildung 13: Die Verbesserungswirkung der Algorithmen ist stark abhängig von der Güte der Trennsignale bzw. dem verwendeten Blinden Quellentrennungsverfahren.
Die maximale Verbesserung lässt sich erreichen bei ideal getrennten Signalen oder objektbasierter Übertragung.**

In AP2.2 werden ähnliche, allerdings individualisierbare Algorithmen angewandt. Individualisierbar bedeutet in diesem Kontext: von z.B. Hörumgebung, Audiogramm oder persönlichen Hörvorlieben abhängige Verbesserungsalgorithmen.

AP2 wurde entsprechend der Planung bearbeitet.

Ergebnisse des Teilarbeitspaketes AP 2.3: Metadatenkonzepte

Im Teilarbeitspaket 2.3 sollte untersucht werden, ob erfasste Metadaten dazu genutzt werden können, eine Verbesserung der Prozesse, beispielsweise bei der Anpassung der Parameter oder als Unterstützung bei der blinden Quellentrennung, zu erzielen.

Auf den verschiedenen Veranstaltungen und bei Besuchen von assoziierten Partnern wurde die Nutzung von Metadaten erörtert. Grundsätzlich sollte der Einsatz von Metadaten die notwendige Komplexität des Regelprozesses vereinfachen bzw. die Bearbeitung beschleunigen. Dazu wäre es notwendig, bereits bei der Erstellung der Produktion Messdaten zu erfassen und bis zum Endverbraucher als Metadatenstrom durchzuschleusen. Sowohl die Erzeugung als auch die Durchschleusung dieser Metadaten stellen jedoch erhebliche Anforderungen an eine Integration sowohl im professionellen als auch im Consumer-Umfeld.

Auch Recherchen zur Verwendung anderer Systeme, wie beispielsweise Cloud-basierte Ansätze, führten bisher zu keinem gewünschten Ergebnis, da die Integration solcher Systeme sich als generell nicht praxistauglich erwiesen hat.

In Abstimmung mit dem Projektpartner Fraunhofer IDMT wurde die Entscheidung getroffen, die Bearbeitung des AP 2.3 zu einem späteren Zeitpunkt fortzuführen. Diese Entscheidung hat keine Auswirkung auf das Erreichen des Projektzieles, da die Metadaten eine unterstützende Funktion haben sollen und die Algorithmen bereits ohne vorhandene Metadaten ihre Funktion erfüllen.

AP 3 - Entwicklung der Technologieblöcke für die Broadcastbereiche

In AP3 wurde auf Basis der Technologien zur Verbesserung der Sprachverständlichkeit eine Software entwickelt (SITA-Pro-Software).
Zur Erweiterung der Abhörmöglichkeiten in der Postproduktion wurde ein Software-Modul entwickelt, mit dem die akustische Situation im heimischen Umfeld beim Empfänger simuliert werden kann.

Abschließend fand die Entwicklung eines Hardware-Prototyps statt, welcher die Messung und Visualisierung der Sprachverständlichkeit im Broadcastworkflow ermöglicht. Herzstück des Hardwaregerätes ist die von den Projektpartnern Fraunhofer IDMT und RTW entwickelte SITA-Pro-Software, welche die Analyse und Visualisierung der Sprachverständlichkeit ermöglicht.

Ergebnisse des Teilarbeitspaketes AP 3.1: Simulation der Abhörbedingungen beim Empfänger

Auf der Seite von RTW wurde im Rahmen einer Masterarbeit ein Software-Plug-in entwickelt, welches Toningenieuren mittels binauralem Rendering über Kopfhörer die Möglichkeit bietet, die Sprachverständlichkeit der Tonmischung bereits vorab durch eine simulierte Raumakustik beim Konsumenten abzuschätzen.
Bestandteil des Arbeitspaketes war die Implementierung eines Algorithmus zur Simulation der Raumakustik. Außerdem musste ein binaural-Renderer und der *Scattering Delay Network* Basis-Algorithmus in die Software integriert werden. Dabei mussten zwei wichtige Aspekte berücksichtigt werden: Zum einen wurden Filter implementiert, die die Wandmaterialien (bspw. Tapete, Beton, usw.) des Raumes und somit die akustische Absorption der Wand berücksichtigen. Zum anderen wurde die Simulation von Lautsprechern integriert, indem bekannte Impulsantworten von Lautsprechern auf die virtuellen Schallquellen übertragen worden sind.

In der Software können Faktoren wie Wandmaterialien und Raumgröße sowie die Position der Audioquellen und die des Rezipienten variabel und in Echtzeit angepasst werden. Somit kann durch die Simulation der Abhörbedingungen bereits bei der Mischung das Ergebnis in Bezug auf die Sprachverständlichkeit im heimischen Wohnzimmer beurteilt werden. Basis der Software bildet das bei RTW eingesetzte C++ Framework, wodurch die Weiterentwicklung und der Einsatz der Software in zukünftigen RTW-Produkten gewährleistet sind.

Im Rahmen der Untersuchungen wurde festgestellt, dass eine Raumsimulation in der geplanten Art über die in der Regel vorhandenen Lautsprechersysteme in Kontrollräumen nicht möglich ist. Der Fokus wurde daher verändert, sodass eine Simulation des Abhörraumes beim Rezipienten mit Hilfe von binauralem Rendering über Kopfhörer stattfindet.
Zusätzlich hat das bisher geplante En- und Decoding der Eingangssignale mit gängigen Kompressionsverfahren wie AAC und MP3 nach den abgeschlossenen Untersuchungen nur einen geringen Einfluss auf die Sprachverständlichkeit gezeigt und stellte aus diesem Grund keinen Hauptbestandteil des Arbeitspaketes mehr dar.
Zum Abschluss wurde ein grafisches Benutzerinterface entwickelt, welches die Steuerung der Simulation zulässt bzw. es ermöglicht, Parameter des simulierten Raumes wie bspw. Raumgröße, Wandmaterialien und die Positionen der Audio-Quellen und die des Hörers im Raum zu setzen (s. Abbildung 7). Die Fertigstellung der Software hatte sich durch die Komplexität des Algorithmus zeitlich verzögert, konnte aber im Dezember 2018 erfolgreich abgeschlossen werden.

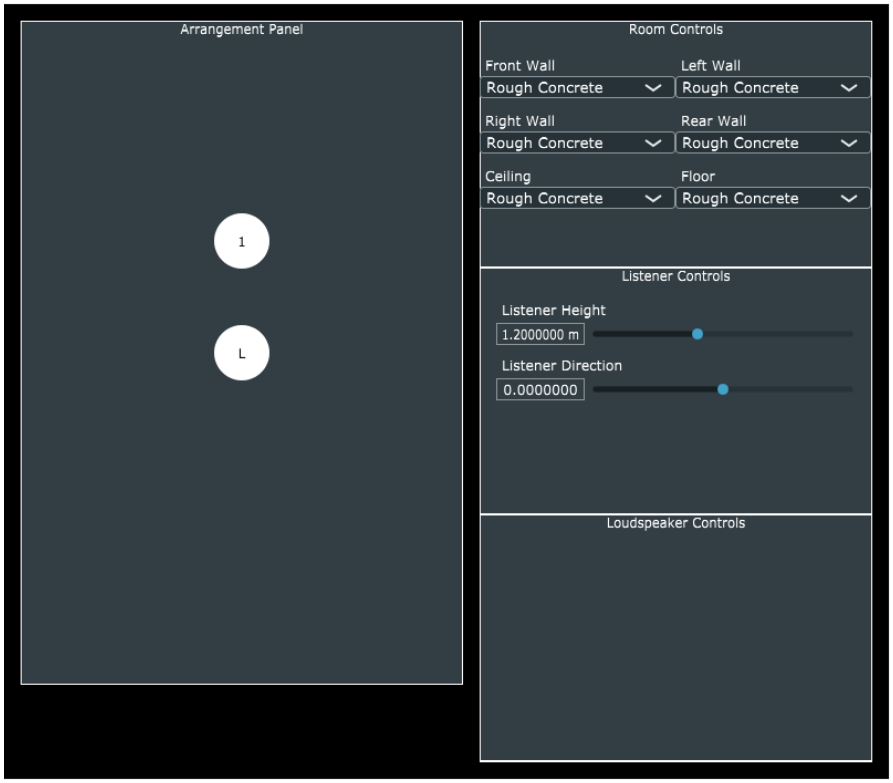

Abbildung 14: GUI der Software zu Simulation der Abhörbedingungen beim Empfänger

ebee als Endgeräte-Spezialist war in der Lage, durch seine Erfahrungen und das vorhandene Equipment neben theoretischen Aspekten, wie das Erarbeiten eines Skriptes zur Berechnung von verschiedenen Nachhallzeiten beim Empfänger, auch praktische Messungen an Endgeräten vorzunehmen.

Dabei hat sich im Laufe des Arbeitspakets immer mehr das Ergebnis abgezeichnet, dass die Probleme für die schlechte Sprachverständlichkeit vermehrt im schlechten Ton von Endgeräten zu suchen sind und die Umgebungsbedingungen beim Empfänger eine eher untergeordnete Rolle spielen. Ein großer Erkenntnisgewinn war die übergeordnete Rolle der schlechten Schallausbreitung von so genannten „Downfire“ – Lautsprechern, die in den meisten aktuellen TV-Geräten verbaut sind. Dies sollte die Sichtweise auf die Erarbeitung vieler verschiedener zukünftiger Aufgaben einschlägig beeinflussen.

Ergebnisse des Teilarbeitspaketes AP 3.2: Integration der Technologieblöcke für den Broadcastworkflow

Im Rahmen von AP3.2 fand am 28.02.2018 ein Treffen mit assoziierten Partnern und Interessierten bei RTW in Köln statt. Neben den Projektpartnern nahmen an diesem Treffen insgesamt 19 Personen aus dem professionellen Audiobereich teil. Der Personenkreis umfasste Damen und Herren von privaten wie öffentlich-rechtlichen Sendeanstalten (u.a. aus der Produktion, der Programmleitung, Integrationsbeauftragte, Vertreter von Verbänden). Das Thema stieß auf reges Interesse. Im ersten Schritt wurden alle Teilnehmer auf einen gleichen Wissensstand zum Projekt, seinen Zielen und den

bereits erfolgten Arbeitsschritten gebracht. Im weiteren Verlauf wurden die Problemstellungen des Projektes verdeutlicht, um anschließend mögliche Lösungsansätze zu erarbeiten.

Die Gruppe wurde im Tagesverlauf in zwei Arbeitsgruppen aufgeteilt. Themenbereich 1 beschäftigte sich mit der Sprachverständlichkeit in der Postproduktion, Themenbereich 2 mit den Anforderungen beim Konsumenten.

Über diese beiden Arbeitsgruppen sollten die Anforderungen der jeweiligen Nutzer und deren Arbeitsweise bzw. beim Endnutzer vor allem dessen technisches Verständnis und die damit verbundenen Möglichkeiten, Parametrierungen vorzunehmen, evaluiert werden, um daraus klare Definitionen für das Anforderungsprofil der Lösungsansätze für die beiden Nutzergruppen zu erhalten.

Themenkreis 1 (Broadcast)

Diverse Anforderungen für den Produktions- und Sendebetrieb wurden definiert, es wurden aber auch die Erwartungen der Integrationsbeauftragten zum Thema Sprachverständlichkeit besprochen. Ein wichtiges Ergebnis der Diskussionen war, dass es bei den Broadcastern hauptsächlich darum gehen muss, ein für die jeweilige Zielaltersgruppe entsprechend aufbereitetes Signal zur Verfügung zu stellen. Im Klartext, bei ARD/ZDF mit seiner Zielaltersgruppe 50+ ist die mittlere, altersgerechte Hördegeneration das Maß der Dinge, während bei RTL diese Bewertung für die Degenerationsklasse 25-35 Jahre erfolgen muss. Abgesehen von Spartensignalen wird also auf diese Art und Weise ein für die jeweilige Klientel der Sender aufbereitetes, allgemein verständliches Produkt erstellt. Es bedarf hier auch der Entwicklung eines Normsignals zum Test und der Verifikation.

Themenkreis 2 (Konsument)

Als ein Ergebnis der Verbesserung der Sprachverständlichkeit beim Konsumenten wurde die Klassifizierung der Audiowiedergabequalität von Geräten vorgeschlagen. Somit kann bereits beim Kauf der Endgeräte eine Weiche in Richtung guter Klang im heimischen Umfeld gestellt werden. Dabei sollte vergleichbar mit den Energieeffizienzklassen eine für jeden Verbraucher bei Gerätekauf eindeutige Markierung auf den Geräten sein, die die Audioqualität des Systems beschreibt.

Weitere Überlegungen betrafen den Einsatz der SITA-Technologie in den Endgeräten des Rezipienten. Diese sollten in der Lage sein, die akustische Situation beim Konsumenten zu erfassen und entsprechend die Algorithmen im Endgerät auf die Raumsituation anzupassen, um eine bessere Sprachverständlichkeit zu erreichen.

Ein weiterer Aspekt war das grafische Benutzerinterface, um die Parametrierung der persönlichen Umgebung und Hörfähigkeiten vorzunehmen. Es sollte möglichst einfach gehalten werden und smarte Bedienelemente enthalten, um auf unkomplizierte Weise verschiedenen Nutzgruppen eine Klanganpassung zu ermöglichen.

Die Erfassung des individuellen Hörvermögens und die Möglichkeit, verschiedene Nutzprofile auf dem Endgerät zu speichern, war ein weiterer Aspekt.

Die bei dem Workshop gemachten Überlegungen und die daraus resultierenden Lösungsansätze hatten einen bedeutenden Einfluss auf die in den Arbeitspaketen 3 und 4 entwickelten Technologien.

Im ersten Schritt sollte durch dieses Treffen den Broadcastern die Einsatzgebiete der SITA-Technologie nähergebracht und das Vertrauen der Mitarbeiter in eine Messung geschaffen werden, um im nächsten Schritt eine manuelle oder automatische Korrektur der Signale durchzuführen.

Die Aussagen des Workshops bestätigten die zum SITA-Projekt gemachten Überlegungen: Eine Sprachverständlichkeitsmessung in der Postproduktion mit einem übersichtlichen Benutzer-interface wurde als prinzipiell hilfreiches Werkzeug anerkannt. Insbesondere Simulationsmöglichkeiten für das Hören von Schwerhörenden und Senioren wurde als beim Mischen hilfreich angesehen.

Die Sprachverständlichkeitsverbesserung wurde kritischer gesehen: Sie müsse eine hohe Qualität aufweisen, eine Automatisierung schien allerdings nur denkbar bei speziellen Mischungen für Schwerhörige.

Ergebnisse des Teilarbeitspaketes AP 3.3: Aufbau eines Hintergrundprozesses zu Erfassung von Regeldaten

Im Teilarbeitspaket 3.3 sollte ein Hintergrundprozess geschaffen werden, der eine automatische Erstellung von Metadaten und deren Speicherung ermöglicht.

Wie bereits in AP2.3 beschrieben, lässt sich die Erfassung, Verarbeitung und Übertragung der gewonnenen Metadaten nur bedingt bzw. gar nicht in den bestehenden Workflow der Rundfunkanstalten integrieren. Es müsste ein eigener Standard zum Erfassen und Übertragen der Metadaten geschaffen werden, der sowohl auf Seiten der Rundfunkanstalten als auch auf Seiten der Unterhaltungselektronik-Hersteller Anwendung findet. Im Rahmen des SITA-Projektes ist diese Aufgabe nicht zu bewerkstelligen.

Des Weiteren ist eine direkte Erfassung von Regeldaten und Prozessabläufen, die innerhalb einer Digital Audio Workstation (DAW) ablaufen, nicht generell gegeben. Zwar ist es möglich, auf einzelne Informationen der DAW zurückzugreifen, insgesamt haben diese Informationen für das SITA-Framework jedoch keine Relevanz und können somit nicht zur Verbesserung der Prozessabläufe bei der Sprachverständlichkeitsmessung beitragen.

Aus diesen Gründen wurde der zeitliche Aufwand für das AP 3.3 reduziert und auf die Arbeitspakete 3.4 und 3.5 verlagert. Diese APs hatten im Hinblick auf den weiteren Verlauf des Projektes Priorität.

Ergebnisse des Teilarbeitspaketes AP 3.4: Automatisierung der Sprachverständlichkeits-verbesserung für die Broadcastbereiche

Für eine Automatisierung der Prozesse ist eine Sprachaktivitätserkennung [14] zur automatischen Erkennung von Sprache/Dialog im Signal die Grundlage. Höranstrengung wird nur gemessen und angezeigt, wo Sprache im Signal vorhanden ist. Die in AP2 erweiterten Verbesserungsalgorithmen für die Postproduktion wurden als Dynamic Link Library (DLL) implementiert.

Das Plug-in wurde an die SITA-assoziierten Partner zur Evaluation im praktischen Einsatz ausgeliefert. Die Sprachaktivitätserkennung wurde bzgl. Performanz und Erkennungsgüte für unterschiedliche Ausgangssprachen und Material-Cluster/TV-Genres evaluiert.

Weitere Ansätze für eine automatische Voice-over-Voice-Erkennung und eine automatische Sprachenerkennung wurden erarbeitet und evaluiert, sind aber noch nicht ausreichend entwickelt für die Einbindung in die Library.

Ergebnisse des Teilarbeitspaketes AP 3.5: Entwicklung eines Benutzerinterfaces für die Broadcastbereiche

Im AP3.5 wurde das grafische Benutzerinterface für die Broadcastbereiche (graphical user interface (GUI)) erstellt. Dieses dient als Steuerelement der in AP3.6 entwickelten SITA-Software und ermöglicht alle grundlegenden Einstellmöglichkeiten zur Messung und Visualisierung der Sprachverständlichkeit (SV).

Die Entwicklung des grafischen Benutzerinterfaces wurde vom Projektpartner RTW durchgeführt. Auf Basis des Workshops mit den assoziierten Partnern (vgl. Abschnitt des AP3.2) wurden die notwenigen GUI-Komponenten implementiert.

In Abbildung 15 ist das Ergebnis der Umsetzung der grafischen Benutzeroberfläche zu sehen. Die Benutzeroberfläche beinhaltet im Wesentlichen folgende Elemente:

Abbildung 15: Darstellung der graphischen Benutzeroberfläche der SITA-Pro-Software

1. Anzeige zur Darstellung des aktuellen Sprachverständlichkeitswertes für Normalhörende.
2. Anzeige der numerischen Werte der Sprachverständlichkeit für normalhörende und hörbehinderte Personen.
3. Statistische Auswertung der aufgetretenen Sprachverständlichkeitswerte.
4. Chart: Darstellung der Sprachverständlichkeitswerte in einem Graphen über einen kurzen Zeitbereich.
5. Logging: Darstellung der Sprachverständlichkeitswerte über einen langen Zeitbereich.
6. Peak Program Meter (PPM) des Audiosignals.
7. Speicherung der zu berechnenden Sprachverständlichkeitswerte in einer externen Datei. Zurücksetzen der statistischen Berechnung.
8. Auswahl der Sprache prozessierbaren Sprachmodelle
9. Auswahl *Voice-over-Voice*-Szenario *on/off.*

Ergebnisse des Teilarbeitspaketes AP 3.6: Design und Implementierung einer Hardware für die Plattform

Im Arbeitspaket 3.6 wurde der Hardware-Demonstrator für den Broadcastbereich entwickelt. Dies beinhaltete sowohl die Entwicklung der Hardwarekomponenten als auch der Firmware des Demonstrators. Des Weiteren wurde die SITA-Pro-Software als Applikationssoftware für den Demonstrator implementiert.

Die SITA-Pro-Software beinhaltet die einzelnen Teilmodule des AP2 und AP3. Um den Entwicklungsaufwand des Projektes und die zukünftige Anwendung möglichst effizient zu gestalten, wurde ein generischer, plattformunabhängiger Ansatz gewählt. D.h. es wurde ein Framework entwickelt, welches es erlaubt, aus einer Basis heraus verschiedene Plattformen - auch im Hinblick für den zukünftigen Einsatz in unterschiedlichen Systemen - bedienen zu können.

Zu Beginn wurde eine einfache generische Software implementiert, welche die Schnittstelle zu den von Fraunhofer IDMT entwickelten Softwarekomponenten bildete und erste Tests der Sprachverständlichkeitsmessung ermöglichte. Parallel dazu wurde an der Entwicklung der Komponenten für das grafische Benutzerinterface gearbeitet (AP3.5), sodass eine reibungslose Integration der GUI mit der Softwarebasis garantiert werden konnte.

Im weiteren Verlauf wurde der Aufbau des Hardware-Prototypen, welcher im Broadcastbereich zur Analyse und Visualisierung der Sprachverständlichkeit eingesetzt werden konnte, verwirklicht. Die Basis des SITA-Hardware-Demonstrators bildet ein eingebettetes System, welches aus verschiedenen Hardwarekomponenten und einer speziell darauf zugeschnittenen Firmware besteht. Die Basiskomponenten des eingebetteten Systems wurden im späteren Verlauf auch im AP 4.4 eingesetzt.

Es wurden weitreichende Evaluationen zur Auswahl der Basishardware durchgeführt. Da die SITA-Pro-Software komplexe Algorithmen, wie bspw. neuronale Netze zur Detektion von Sprachsignalen einsetzt, musste eine entsprechend performante Plattform zum Einsatz kommen. Eine grundlegende Entscheidung im Hardwaredesign war, dass der Demonstrator basierend auf einem System-on-Modul (SoM)-Baustein umgesetzt werden sollte. Der Vorteil solcher Konzepte liegt in deren modularem Aufbau, wodurch in verhältnismäßig kurzer Zeit ein flexibles und skalierbares Hardwaresystem

aufgebaut werden kann. Zudem ermöglichen diese Systeme, eine an die Bedürfnisse angepasste Linux-basierte Distribution für das System entwickeln zu können.

Es wurden verschiedene Hardwarekomponenten und SoM-Systeme auf ihre Eignung für den Einsatz im Hardware-Demonstrator untersucht. Des Weiteren wurde ein Konzept erstellt, wie die SITA-Software-Frameworks zum Aufbau des Hardware-Demonstrators integriert werden können. In Abbildung 16 ist der schematische Aufbau dargestellt.

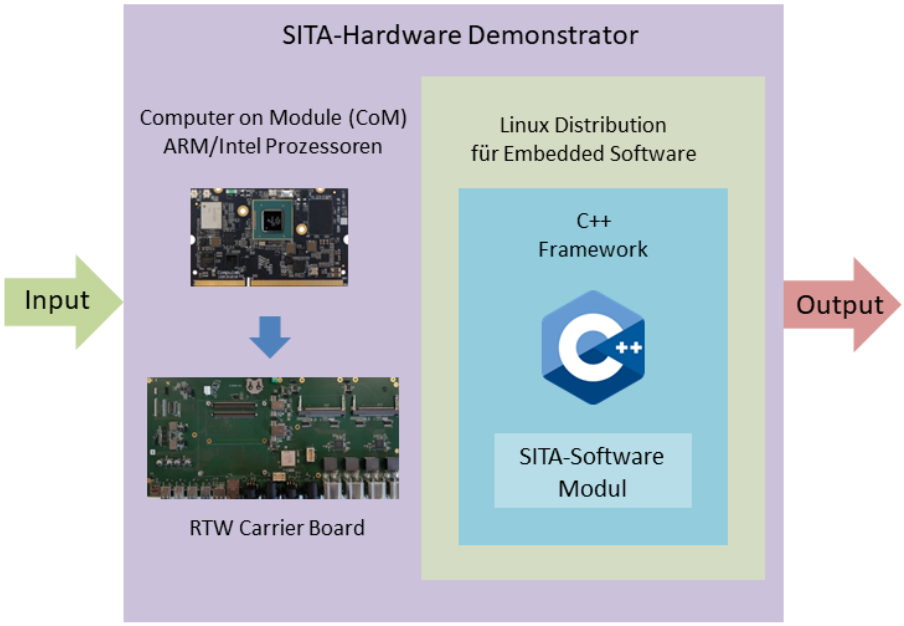

Abbildung 16: Schematische Darstellung des Aufbaus des Demonstrators

Während der Evaluation gab es Rückschläge, u.a. was die Performance der ausgewählten Prozessormodule betraf. Zusätzlich zeigte sich, dass keines der bisher am Markt verfügbaren SoM-Systeme alle für das Projekt notwendigen Aufgaben abdecken konnte. RTW entschied sich aus diesem Grund, selbst ein Basisboard zu entwickeln, welches alle notwendigen Funktionalitäten für den Einsatz im Demonstrator integriert.

Zu den weiteren Ergebnissen des AP3.6 zählt die Erstellung der Firmware des eingebetteten Systems. Hierzu wurde von RTW eine auf dem Linux-Kernel basierte Software erstellt, welche speziell für die zuvor genannten Hardware-Komponenten entwickelt wurden und die softwareseitige Grundlage bildet, um die SITA-Pro-Applikationssoftware betreiben zu können.

Alle grundlegenden Komponenten des Demonstrators wurden erstellt, sodass diese in dem AP5 in das prototypische Gesamtsystem integriert werden konnten.

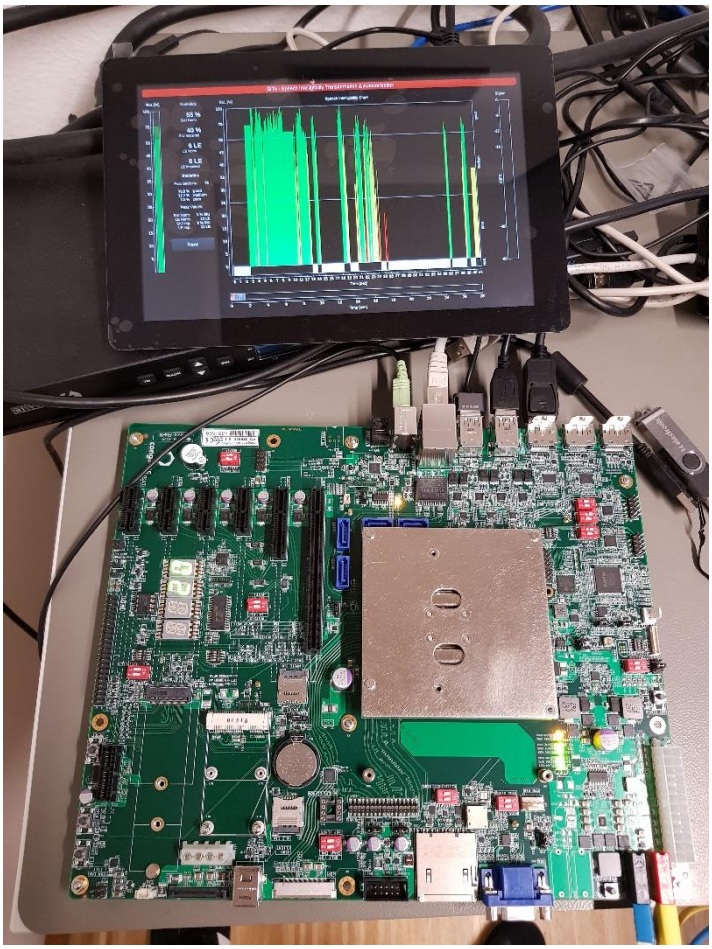

Abbildung 17: Erster prototypischer Aufbau der SITA-Hardware inkl. der SITA-Pro Applikationssoftware

Ergebnisse des Teilarbeitspaketes AP 3.7: Evaluation

Im September 2018 wurde einem ausgewählten Personenkreis der assoziierten Partner ein erster Software-Demonstrator zur Verfügung gestellt. Der Demonstrator wurde als Software-Plug-in umgesetzt, sodass er bei den Rundfunkanstalten in den Workflow mit DAWs eingebunden werden konnte. Zu diesem Zeitpunkt war eine Analyse von komplexen Audiosignalen, also von bereits gemischten Audiosignalen, noch nicht möglich, die Spuren „Sprache“ und „Hintergrund“ mussten noch getrennt vorliegen, um eine Aussage zur Verständlichkeit der anvisierten Mischung geben zu können. Die Nutzer hatten die Möglichkeit, über einen Zeitraum von mehreren Wochen zu testen.

In dieser Versuchsphase ging es RTW in erster Linie darum, zu sehen, wie die Software in den Workflow der Rundfunkanstalten eingebunden werden kann und ob die vorgeschlagenen Interfaces bezüglich Usability und Workflow die potentiellen Nutzer überzeugen. Der Projektpartner Fraunhofer IDMT evaluierte, wie gut die subjektiv wahrgenommene Höranstrengung mit den Modellergebnissen übereinstimmt.

Im Anschluss an die Evaluation konnten die Testnutzer ihre Kritik und Verbesserungsvorschläge an die Projektpartner frei oder in Form von Antworten auf Fragebögen übermitteln. Die Rückmeldungen dienten als Grundlage für die weitere Entwicklung an den grafischen Benutzerschnittstellen.

Der zweite bedeutende Test der Software fand im Juni 2020 statt. Die annähernd finale Software war nun in der Lage, die Sprachverständlichkeit komplexer Audiosignale zu analysieren. Wie bei der Evaluation im Jahr 2018 wurde auch dieses Mal ein Plug-in an assoziierte Partner zur testweisen Nutzung gegeben. Bei dieser Evaluation lag der Schwerpunkt der Testung auf der finalen Optimierung der Software.

AP 4 - Entwicklung der empfängerseitigen Technologieblöcke

Die Schwerpunkte des AP4 lagen darin, die Abhörbedingungen beim Empfänger zu untersuchen und die Sprachverständlichkeitsverbesserung im heimischen Umfeld zu integrieren. Es wurden alle Rahmenbedingungen zur Umsetzung der Soft- und Hardware erarbeitet.

Es wurde ein Hardwaresystem entwickelt und implementiert, welches die Integration auf Empfängerseite im TV-Gerät ermöglichen sollte: Von Fraunhofer IDMT wurde eine Nutzerschnittstelle entworfen, welche insbesondere Laien eine einfache Einstellung der unterschiedlichen Höranpassungsstrategien ermöglicht (Ziel-Höranstrengungswert, Klang). Das Empfängersystem wurde ausführlich getestet und evaluiert.

Zu Beginn wurden Untersuchungen zu der im heimischen Umfeld vorhandenen akustischen Situation gemacht, durch deren Erkenntnisse dann im weiteren Verlauf die Verfahren zur automatisierten Verbesserung der Sprachverständlichkeit beim Empfänger entwickelt werden konnten. Abschließend konnte die Verfahren erfolgreich in einem TV-Endgerät eingesetzt und vom Benutzer individuell gesteuert werden. Die folgenden Beschreibungen erläutern die einzelnen Arbeitsschritte genauer.

Ergebnisse des Teilarbeitspaketes AP 4.1: Modellierung individueller Abhörbedingungen

Im AP 4.1 wurde ein Modell entwickelt, das die individuellen Abhörbedingungen auf der Empfängerseite, also beim Endverbraucher zu Hause, abbildet. Das Hörvermögen, die vorhandene Raumakustik (z.B. Nachhallzeit) und die Art des Wiedergabegeräts (Soundbar, TV, Lautsprecher usw.) sind Faktoren, die in dieses Modell einflossen.

Die beteiligten Parteien ebee und RTW trafen im ersten Schritt Überlegungen bezüglich der zu verwendenden Hardware und Schnittstellen, welche die Erfassung der Daten ermöglichen sollten:

- Erfassung der Daten bezüglich der vorhandenen Raumakustik
- Erfassung der Daten zum vorhanden Wiedergabegerätes des Empfängers (TV, Lautsprecher, Soundbar etc.)
- Eingabe der präferierten Ziel-Höranstrengung und der Klangauswahl (GUI-Einstellungen)

Dabei wurde vor allem darauf geachtet, möglichst viele Varianten in den Modellen zu berücksichtigen, sodass die Breite der Modelle möglichst viele verschiedene reale Szenarien abdecken konnten.

Die erfassten Daten konnten durch die detailreiche Betrachtung für das RTW-Simulationsmodul und für die Implementierung der GUI in Arbeitspunkt 4.2. aufbereitet werden.

Ergebnisse des Teilarbeitspaketes AP 4.2: Entwicklung einer Benutzerinterfaces zur Erfassung individueller Korrekturparameter

Das Ziel in Arbeitspaket 4.2. war die Erarbeitung eines Benutzer-Interfaces zur interaktiven Erfassung individueller Korrekturparameter. Dabei sollten verschiedene Muster zur Aufnahme von Daten erstellt werden. Wichtigster Bestandteil war ein Dialog zur Erfassung von Nutzungsparametern über ein generisches Software-Interface. Des Weiteren sollten durch eine Raumakustik-Einmessung die aktuellen Standortbewertungsmuster aufgenommen werden. Außerdem waren verschiedene Formen der Verarbeitung dieser Parameter gefordert, um diese an den signalverarbeitenden Teil der Applikation übergeben zu können und diese dort zu Nutzen.

Zur besseren Veranschaulichung werden die Ergebnisse im Folgenden in kleinere Teilbereiche unterteilt.

Dialog zur Erfassung notwendiger Daten und Entwicklung eines generischen Software-Interfaces

Zur Erfassung der Parameter des Rezipienten ist die Entwicklung eines generischen Software-Interface durchgeführt worden. Durch das positive Feedback der ebee-Connect-App und der Erfahrungen mit dieser, ist die Wahl auf ein Touchscreen-Interface auf Basis einer Android-App für Tablets gefallen. Wie in der Abbildung 11 gezeigt, können dort Profile angelegt und grundlegende Einstellungen vorgenommen werden. Durch einfaches Navigieren durch das Menu können in den einzelnen Menus verschiedene Parameter angepasst werden und so der Algorithmus gesteuert werden. Diese App ist die Grundlage für die spätere Umsetzung eines Demonstrators, mit dem die Funktionsweise überprüft und Evaluierungen durchgeführt werden können.

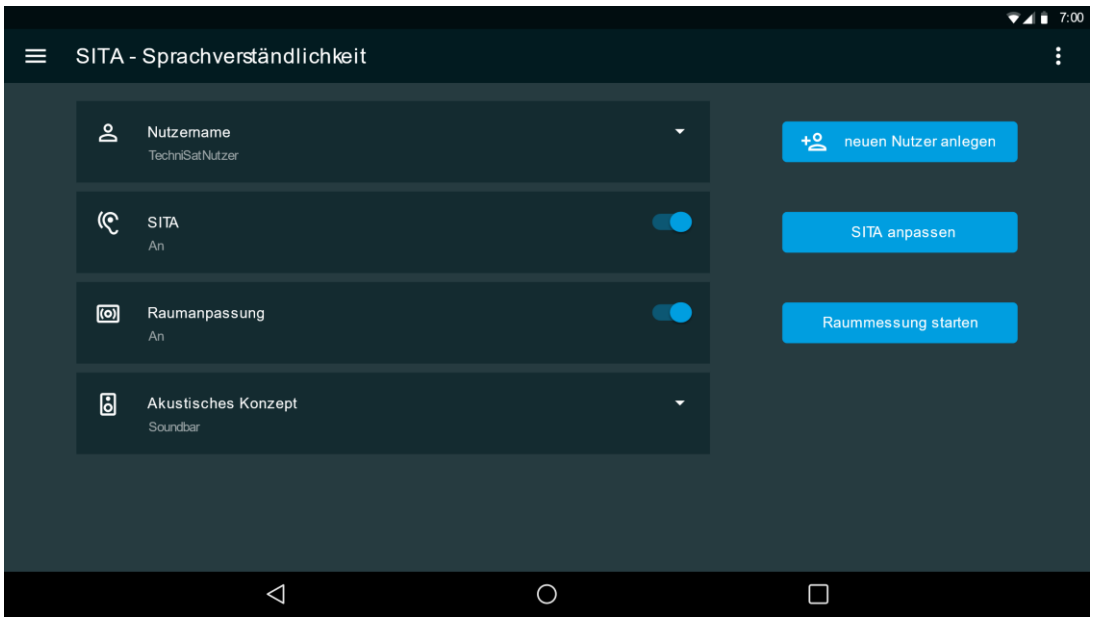

Abbildung 18: Menü-Konzept der SITA-Android-Applikation

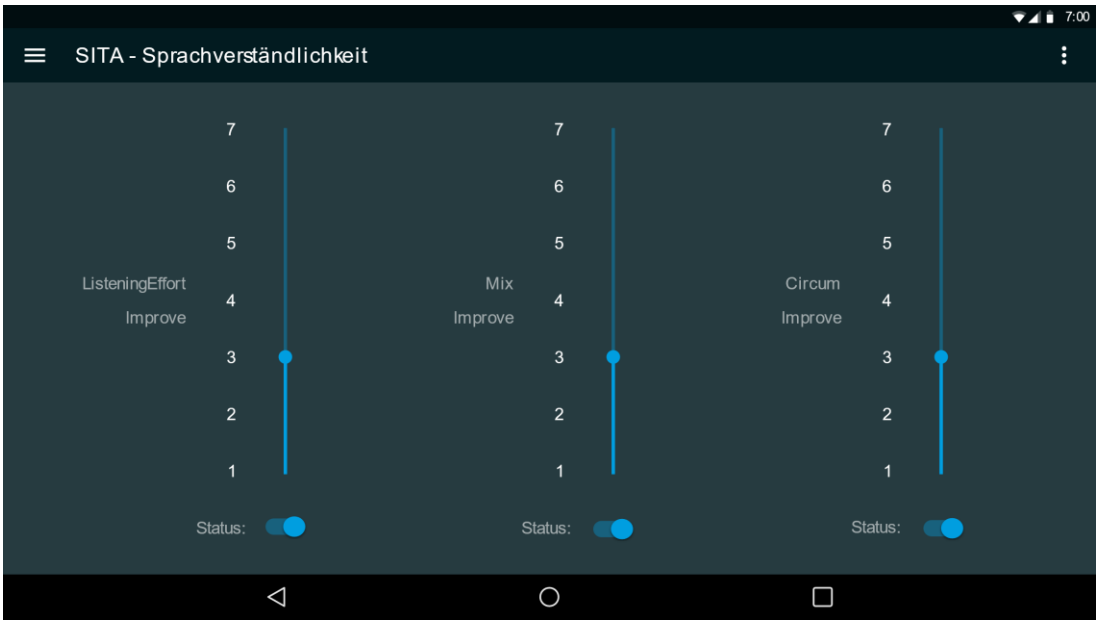

Abbildung 19: Primäres Reglerkonzept der SITA–Android-Applikation

Erfassung der Raumakustik und Entwicklung geeigneter Standardhörbewertungsmuster

Die Erfassung der Raumakustik als Bestandteil von Arbeitspaket 4.2. wurde durch ebee in Form eines MATLAB-Skriptes und entsprechender Mess- und Debug-Hardware für die zu verwendeten TV-Geräte erarbeitet. Der Ansatz durch eine sich an die Raumumgebung anpassende Filterbank erzielte Ergebnisse, die den Klang der Geräte beeinflussen konnte. Dieses Verfahren war angelehnt an die von Lautsprecher-Herstellern verwendete Methode. Jedoch konnte durch Evaluierung der Sprach-verständlichkeit vor und nach der Filteranpassung festgestellt werden, dass die Verbesserung, wenn überhaupt, sehr klein ist. Deshalb wurde beschlossenen, keine weiteren Ressourcen in die Implementierung dieses Algorithmus in die SITA-Bibliothek vorzunehmen, da sich dadurch kein Mehrwert für das Projekt versprochen wurde.

Ergebnisse des Teilarbeitspaketes AP 4.3: Automatisierung der Sprachverständlichkeits-verbesserung beim Empfänger

Das Arbeitspaket 4.2 und das Arbeitspaket 4.3 überschnitten sich in vielen Punkten. Die Aufnahme von Korrekturparametern ist eine Verbesserung der in Arbeitspaket 4.2. geschaffenen Aufnahmemethode der GUI, wodurch die Nutzererfahrung für die Steuerung der Algorithmen enorm gesteigert werden konnte.

Fraunhofer IDMT schlug zur Bedienung eine intuitive Bedienoberfläche vor, die keinerlei Vorkenntnisse von NutzerInnen verlangt. Kriterium zur Algorithmenauswahl ist allein die Hör-Präferenz der NutzerInnen. Die für die Signaladaptionen nötigen Nutzereingaben erfordern keine Unterbrechung des Fernsehkonsums und sind ähnlich einer Lautstärkeregelung zu bedienen. Die nötigen Nutzereingaben beschränken sich auf:

- On / off
- Eingabe der maximalen Ziel-Höranstrengung nötig (einmalig gesetzt – kann nachjustiert werden)
- Auswahl des präferierten Algorithmus/Klangbildes (einmalig gesetzt – kann nachjustiert werden)

Die Bedienoberfläche ermöglicht als „2D-Touch“-Realisierung die Einstellung durch Verschiebung eines Punktes/Kreuzes auf einer zweidimensionalen Fläche. Der Punkt auf dem Feld steuert verschiedene Parameter der Algorithmen. Im betrachteten Beispiel des „Dialog Improve“ - Algorithmus (Abbildung 20) lässt sich so eine Ziel-Höranstrengung einstellen, sowie die Funktionsweise des Algorithmus anpassen. Durch diese „Drag and Drop“ - Steuerung kann sich der Rezipient je nach persönlichen Klangvorlieben den besten Punkt auf dem Touchscreen auswählen. Ein tieferes Wissen über die Algorithmen im Hintergrund ist nicht nötig, Kriterium für die Einstellung ist die individuelle Hör-Präferenz.

Hat der Zuschauer bereits Eingaben zu seinen gewählten Präferenzen gemacht, reicht es aus, die Verständlichkeitsverbesserung an- oder abzuschalten. Dieses „SITA-improve“-Modul wurde den Projektpartnern als dynamische Programmbibliothek (DLL) zur Verfügung gestellt und enthält folgende SITA-Module:

- Höranstrengungs-Schätzung
- Quellentrennung
- Improve-Algorithmen

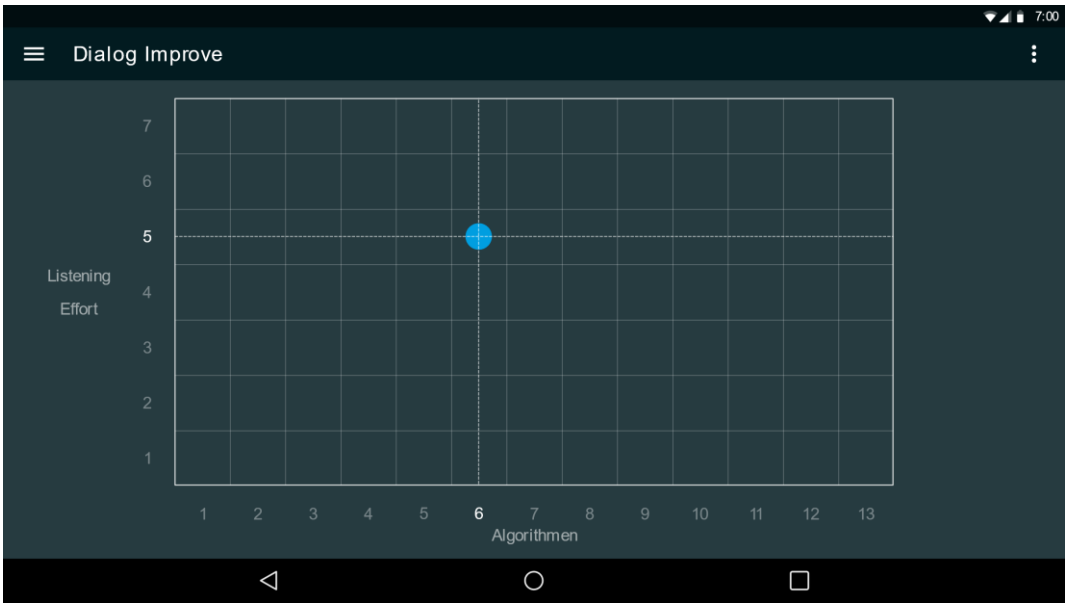

Abbildung 20: Beispielhafter Aufbau eines 2D-Touchfeldes

Ergebnisse des Teilarbeitspaketes AP 4.4: Design und Implementierung eines Hardwaremoduls zur Integration in Endgeräte

In Arbeitspaket 4.4 wurde der prototypische Aufbau des Demonstrators für den Heimbereich entwickelt. RTW erstellte auf Grundlage der in AP3.6 umgesetzten Arbeiten in das RTW-Carrier-Board, welches zur Evaluierung der Algorithmen diente sowie die Basishardware des späteren Demonstrators darstellen sollte (Abbildung 22).

Zu den Tätigkeiten zählten u.a. die Erstellung des Layouts des Basisboards und die notwendige Portierung der SITA-Software für das Linux-basierte System des Hardware-Demonstrators. Als Grundlage des Ganzen diente ein von RTW erstelltes Embedded Linux.

Das Design und die Entwicklung des Hardware-Boards mussten aufgrund von Personalmangel zu einem späteren Zeitpunkt als geplant durchgeführt werden. Aufgrund der daraus resultierenden Verzögerungen in der Entwicklungszeit wurde alternativ ein PC-basierter Ansatz zur Implementierung des initialen Prototypen gewählt. Dieser sollte sowohl die Signalverarbeitung als auch die Handhabung und Darstellung der Nutzeranwendung beinhalten. Diese wurde als Software-Demonstrator bezeichnet.

Im Vorfeld eines Entwicklertreffens wurde von ebee die Applikation entwickelt, die sowohl die Verbindung der im Vorfeld von ebee entwickelten App mit dem Server als auch die Parameter für die Algorithmen auswertet und an diese übermittelt. In Abbildung 21 ist der grundlegende Aufbau des SITA-Software-Demonstrators dargestellt.

Der Aufbau des Demonstrators setzt sich wie folgt zusammen:

Set-Top-Box:	Empfang von beliebigem A/V - Content
Touchscreen/Tablet:	Steuerung des Demonstrators/Nutzerinterfaces
Windows PC:	SITA-Server, Initialisierung des Funktionsumfangs
ebee-Device:	Anzeige des Bildes und Abspielen des Tons

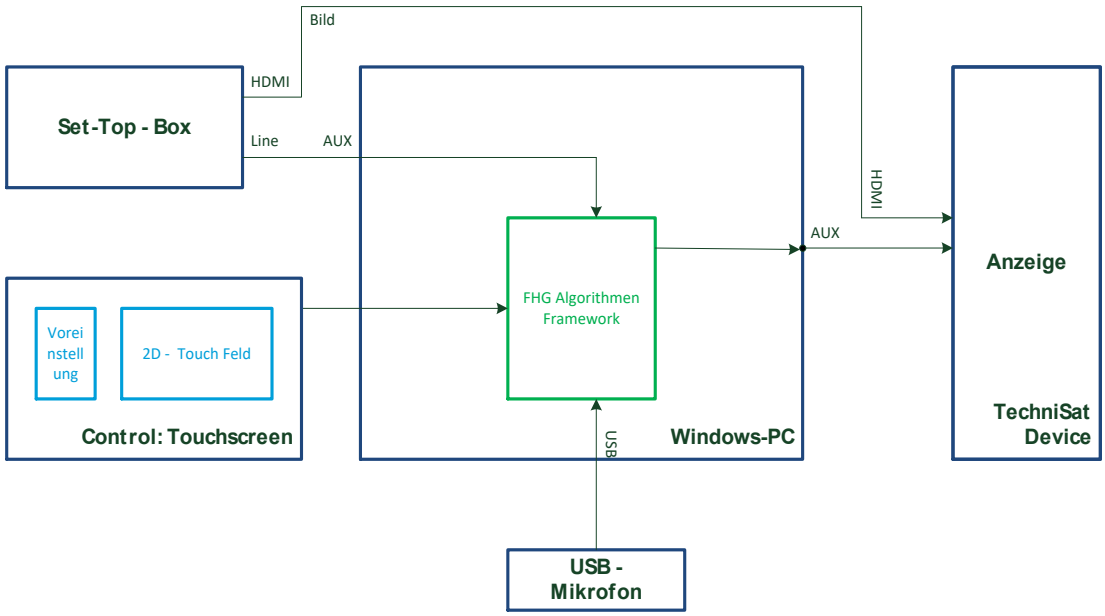

Abbildung 21: Grundlegender Aufbau des SITA-Software-Demonstrators

Der Demonstrator wurde im Zuge des Meilensteintreffens im November 2019 vorgeführt. Die Mehrzahl der Algorithmen war bereits funktionsfähig implementiert, sodass man eine erste, sehr positiv ausfallende Evaluation der Effekte durch die Verarbeitung durchführen konnte. Einzig die Anbindung an Umgebungsgeräusche fehlte noch. Diese sollte im Zuge der Weiterentwicklung der Demonstratoren noch ergänzt werden.

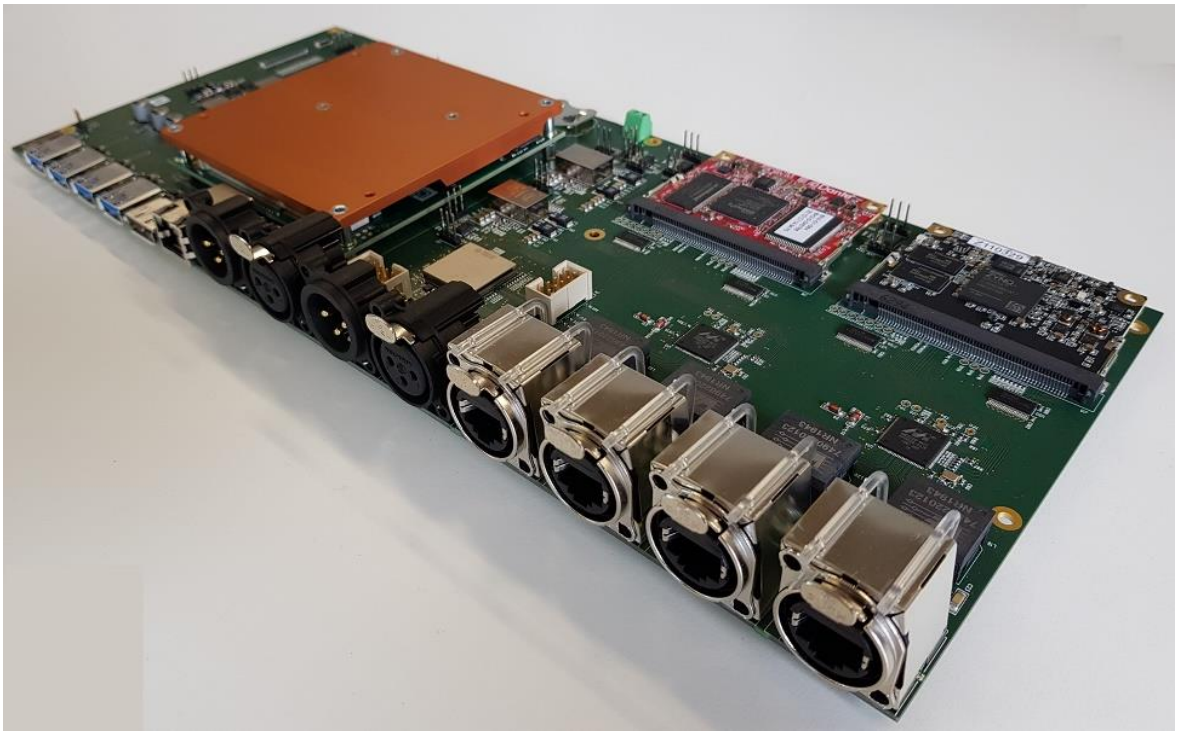

Abbildung 22: Finale Version des RTW Hardwareboards für die Hardware-Prototypen

Ergebnisse des Teilarbeitspaketes AP 4.5: Evaluation

Für die Evaluation der Demonstratoren aus AP4 sollten in Kombination mit den in AP5 vorgesehenen Evaluationen Usability-Studien durchgeführt werden. Allerdings konnte die Evaluation aufgrund der im Frühjahr 2020 aufgetretenen Beschränkungen anlässlich der Covid-19-Pandemie nicht mehr planmäßig durchgeführt werden. Aufgrund der Situation mussten Änderungen im Projektplan vorgenommen werden und stattdessen als Alternative firmeninterne Tests durchgeführt werden (vgl. Erläuterungen zum AP 5).

AP 5 - Systemintegration der Komponenten und Aufbau der prototypischen Demonstratoren

In AP5 fanden Systemtests und Usability-Studien statt, die zur abschließenden Systemintegration der prototypischen Demonstratoren führte. Aufgrund der in dieser Zeit fallenden gesellschaftlichen Einschränkungen durch die COVID-19-Pandemie konnten nicht alle Untersuchungen vollständig abgeschlossen werden.

Ergebnisse des Teilarbeitspaketes AP 5.1: Systemintegration und Evaluation der Demonstratoren (Im Broadcastbereich und beim Empfänger

Zu Beginn des SITA-Projekts wurde geplant, die im Arbeitspaket 4 erarbeiteten Demonstratoren so in ebee-Produkte einzubauen, dass diese als Testgerät wie im späteren Endgerät vorliegen könnten. Durch die beschriebene erhöhte Rechenleistung der Signalverarbeitung in der SITA-Bibliothek war jedoch recht schnell klar, dass man von diesem Ziel abgehen musste. Es wurde jedoch auch kein

handlicher Hardware-Aufbau erarbeitet, weil dessen Fertigstellung sich zu sehr verzögert hatte und in Absprache mit den Projektpartnern die Funktionalität des Software-Demonstrators als ausreichend angesehen wurde.

Ergebnisse des Teilarbeitspaketes AP 5.2: Usability der entwickelten Technologieblöcke

Aufgrund der COVID-19-Beschränkungen konnten keine größer angelegten Nutzertests, wie ursprünglich in AP5 geplant, durchgeführt werden.

Auch der von ebee in eigenen Entwicklungsprozessen geplante Heimtest konnte nicht durchgeführt werden. Als Ersatz wurden die Demonstratoren innerhalb des Unternehmens intensiv getestet und die Software-Demonstratoren verbessert. Es wurden:

- durch elektrische Messungen die Signalverarbeitung von allen Media- und Audioquellen nachgewiesen
- die Signalverbesserung der Medien durch die Algorithmen durch eine Anzahl von Kollegen bestätigt
- Software-Tests durchgeführt

So wurde das komplette Spektrum der ebee-Endgeräte genutzt. Sowohl Kontent über Massenspeichermedien als auch über Übertragungswege, wie Internet, DVB-C usw. wurden getestet.

Ergebnisse des Teilarbeitspaketes AP 5.3: Überarbeitung der Engineering-Muster und finale Systemintegration

Mithilfe finaler Systemtests wurden sämtliche SITA-Funktionalitäten überprüft und finalisiert. Durch Auflagen und Einschränkungen seit März 2020 kam es zu weiteren Verzögerungen bei der Hardware-Integration (RTW). Die SITA-Partner einigten sich deshalb auf eine abschließende Experten-Usability-Evaluation per SITA-Plug-in und Serverlösung: So konnten die Funktionalitäten der SITA-Lösungen abschließend und produktnah – wenn auch als Software-Tool und nicht als Hardware – von Profis und Laien bewertet werden und im Nachgang optimiert werden.

2.2 Wichtigste Positionen des zahlenmäßigen Nachweises

Während der Projektdurchführung sind Kosten in Form von Materialkosten, Personalkosten und Reisekosten entstanden, über die im Detail im Verwendungsnachweis berichtet wird.

Die wichtigsten Positionen des zahlenmäßigen Nachweises bei RTW waren die Personalkosten sowie die Ausgaben für das notwendige Material zum Aufbau der Prototypen, insbesondere für die Evaluationsboards, die Prozessormodule, für Platinen und Bauteile.

Auf Seiten von FHG waren die wichtigsten Positionen des zahlenmäßigen Nachweises die Personalkosten zur Planung und Durchführung der Forschungs- und Entwicklungs-Arbeiten und Nutzerstudien.

Die wichtigsten Positionen des zahlenmäßigen Nachweises bei ebee waren Personalkosten, Materialkosten sowie Reisekosten.

2.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Das beantragte Projekt war für die beteiligten Unternehmen mit erheblichen technischen Risiken verbunden, insbesondere sind hier die Latenz zur Berechnung und Visualisierung der Messergebnisse hervorzuheben. Die konkreten Lösungen der technischen Probleme waren zu Beginn des Projektes noch unklar, wodurch auch das Risiko eines Scheiterns des Projektes gegeben war.

Die für die Förderung bereitgestellten Mittel ermöglichten es bei allen drei Projektpartnern, bereits im Unternehmen beschäftigte Mitarbeiter im Projekt zu integrieren bzw. neue Mitarbeiter mit entsprechendem Fachwissen einzustellen. Das Projekt unterteilte sich in verschiedene Teilgebiete, in denen jeweils spezifisches Fachwissen notwendig war. Somit war die Einbindung von Experten aus unterschiedlichen Fachrichtungen notwendig.

Ein Großteil des Projektes befasste sich mit der Entwicklung von Hardwaregeräten, wodurch erfahrungsgemäß hohe Kosten entstehen. Um ein geeignetes Hardwaresystem für das Projekt zu finden, mussten verschiedene Hardware-Komponenten und Prozessoren beschafft und evaluiert werden. Des Weiteren war im Rahmen der Entwicklung von elektronischen Baugruppen die Zusammenarbeit mit Dritten notwendig, um die Leiterplatinen herzustellen.

Die wissenschaftlich-technischen Arbeitsziele des Gesamtvorhabens waren nur in der speziellen Konstellation eines Verbundprojektes mit dem jeweiligen Know-how der drei Projektpartner erfolgreich umzusetzen. Darüber hinaus war für die geplante Verwertung eine Zusammenarbeit mit den Projektpartnern sinnvoll und notwendig.

Die geleisteten Arbeiten wurden vollständig für die im Teilvorhabenantrag formulierte Forschungsfragen aufgebracht und waren damit angemessen.

Der in SITA verfolgte Ansatz bietet eine umfassende Möglichkeit, Sprachverständlichkeit sowohl pauschal als auch individualisiert zu verbessern. Im Sinne des Anspruches einer barrierefreien Informationsversorgung zeigt das Ergebnis nicht nur wirtschaftliche, sondern auch sozioökonomische Relevanz.

2.4 Nutzen und Verwertbarkeit der Ergebnisse

Der Einsatz von Technologien zur Verbesserung der Sprachverständlichkeit birgt ein großes Potenzial im gesamten medialen Umfeld, sei es bei den Fernseh- und Rundfunkanstalten, bei Streaming-Diensten oder bei Unterhaltungselektronik-Herstellern.

Dies resultiert zu großen Teilen aus der bereits beschriebenen gesellschaftlichen Bedeutung hinsichtlich Barrierefreiheit und Sprachverständlichkeit. Auf Grund der veränderten Produktions-bedingungen und des demografischen Wandels stellt die Sprachverständlichkeit eine wichtige Herausforderung an die Informationsgesellschaft dar, sowohl auf nationaler als auch internationaler Ebene.

Die Forschungs- und Entwicklungs-Tätigkeiten in SITA wurden konsequent mit Blick auf eine Umsetzung in der Praxis ausgelegt und legten damit bereits während des Projektverlaufs den Grundstein für eine wirtschaftliche Verwertung: Aktuell wird im Rahmen eines Verwertungs-Projektes gemeinsam mit einer öffentlich-rechtlichen TV-Anstalt und RTW als Hardwarehersteller der Einsatz der Technologie zur vollautomatisierten Erstellung einer zweiten, leichter verständlichen Dialogspur, welche parallel zur Originalmischung gesendet und empfangen werden kann, geprüft.

2.4.1 Verwertungsplan des Projektpartners RTW

Für das Unternehmen RTW liegt das Marktpotential in zwei Kernbereichen. Zum einen wird RTW das Geschäft mit Komplettprodukten im Bereich Software und Lizenzen für Geräte erweitern. In Analogie zu heutigen Produkten sollen eigenständige Softwareprodukte, Messgeräte bzw. Software-Lizenzen für Messgeräte und Prozessoren im Broadcast-Bereich angeboten werden. Das Kernprodukt in diesem Bereich soll ein Prozessor zur Sprachverständlichkeitsverbesserung sein, der auf die Bedürfnisse von Rundfunkanstalten zugeschnitten ist und im Sendebetrieb eingesetzt werden kann. Darüber hinaus soll die bestehende Produktpalette von Softwareprogrammen zur Audiosignalbewertung um eine Applikation zum Messen der Sprachverständlichkeit für komplexe Signale erweitert werden.

Zielkunden in diesem Bereich sind u.a. die Rundfunk- und Fernsehanstalten und deren Zulieferer sowie die Content-Provider und Produzenten für Audio-Streaming.

Zum anderen ist geplant, dass RTW das bestehende Angebot an OEM-Produkten ausbaut, d.h. als Add-on-Komponente für die Verwendung durch Dritte unter deren Markennamen. Zielkunden sind hier einerseits Gerätehersteller, die wie RTW für die Rundfunk- und Fernsehindustrie Geräte entwickeln und vermarkten und ihrerseits den Bedarf einer Sprachverständlichkeitsanalyse in ihren Produkten haben (z.B. Hersteller von Mischpulten, Systemintegratoren). Andererseits sind dies Endgerätehersteller wie der Projektpartner ebee, die die Technologie beim Empfänger zur individualisierten Verbesserung der Sprachverständlichkeit einsetzen wollen (z.B. Premium-Geräte im Bereich Fernsehen und Hi-Fi).

2.4.2 Verwertungsplan des Projektpartners Fraunhofer IDMT

Die Fraunhofer-Gesellschaft verwertet ihre Forschungs- und Entwicklungsergebnisse in Form von Lizenzeinnahmen und im Rahmen der industriellen Auftragsforschung und -entwicklung. Darüber hinaus verbreitet sie ihre Forschungsergebnisse im wissenschaftlichen und öffentlichen Bereich.

Eine Lizenzierung des Höranstrengungs-Meters für eine DAW eines internationalen Anbieters ist bereits erfolgt.

Für weitere mögliche Anschlussprojekte, welche die Einsatzbereiche der SITA-Technologien erweitern und stärken, werden derzeit in diversen Technologiefeldern, auch außerhalb von Film und Broadcast, entsprechende Partner akquiriert.

Wissenschaftliche Publikationen wurden bereits zur Projektlaufzeit forciert (Fachkonferenzen, Journale). Gleichsam sind Nutzerschnittstellen und Verfahren zur Personalisierung der Übertragung

breit anwendbar, so dass die Projektergebnisse die Technologiebasis von Fraunhofer IDMT stärken und auch die Möglichkeiten der weiteren Verwertung in anderen Bereichen verbessert (z.B. Consumer Electronics, Automotive, Kommunikationssysteme).

2.4.3 Verwertungsplan des Projektpartners ebee

Durch die Ergebnisse des SITA Projektes erhofft sich ebee langfristig eine gute Ausgangsposition zur Anwendung in seinen eigenen Produkten und zur besseren Vermarktung der Produkte. Des Weiteren erhofft sich ebee aus der Vermarktung von Lizenzen an Drittanbieter für deren eigene Audio-Produkte einen Zugewinn.

Die Ergebnisse und Lösungen zur Sprachverbesserung aus dem SITA-Projekt sollen deshalb in alle zukünftigen Premium-Produkte von ebee, wie TV-Geräte, Set-Top-Boxen, Soundbars und Radios einfließen. So erhofft sich ebee durch den demografischen Wandel in der Bevölkerung und dem deshalb stetig an Bedeutung gewinnenden Thema der Spracherkennung und Sprachverbesserung in den Endgeräten einen Wettbewerbsvorteil.

2.5 Fortschritt bei anderen Stellen

Sprachverständlichkeit ist seit geraumer Zeit ein Thema im Umfeld von Film und Fernsehen. Insbesondere die öffentlich-rechtlichen TV-Anstalten stehen in der Kritik. Mit einer Zielgruppe, die im Durchschnitt zwischen 60 und 65 Jahren alt ist, stehen die Sender in der Verantwortung bezüglich Barrierefreiheit und Sprachverstehen zu handeln.

Der Markt „schläft“ nicht. Auf der Tonmeistertagung 2018 hatte die Firma Izotope ein Verständlichkeitsmetering vorgestellt und das Thema Sprachverstehen klar adressiert.

Ein Plug-in der Firma Audionamix widmet sich der Aufwertung von Dialogverständlichkeit. Das Fraunhofer IIS testet im Moment ebenfalls gemeinsam mit dem WDR eine Dialog+-Spur, um die Interessen der Senioren entsprechend zu bedienen. Das Fraunhofer IDMT und das Fraunhofer IIS stehen diesbezüglich in unmittelbarer Konkurrenz.

Ein anderes Konzept stellt das Unternehmen *Mimi* mit seinem Produkt *Mimi Defined vor*. In Kooperation mit dem Unterhaltungselektronik-Hersteller *Loewe* wurde ein System entwickelt, welches in der Lage ist, den Ton des TV-Gerätes auf das individuelle Hörvermögen des Zuschauers anzupassen. Im Unterschied zu SITA konzentriert sich der Ansatz auf die Verbesserung der Sprachverständlichkeit beim Empfänger und nicht auf eine Verbesserung der Sprachverständlichkeit, die bereits bei den Rundfunkanstalten stattfindet könnte, wie es bei SITA der Fall ist. Somit wird nicht die gesamte Kette von der Erzeugung bis hin zum Empfang der Audiodaten am TV-Gerät untersucht.

Die öffentlich-rechtlichen Fernsehanstalten haben einen Arbeitskreis gegründet, der sich speziell mit der Thematik und entsprechenden Lösungsansätzen auseinandersetzt. SITA hat sich bereits während des Projektes immer wieder auch selbst bekannt gemacht und „in Szene" gesetzt - so wird der SITA-Ansatz derzeit auch in der AG Audio als Lösungsansatz diskutiert und getestet. Klarer Marktvorteil

von SITA ist die „ganzheitliche“ Herangehensweise: Die Sprachverständlichkeit wird gemessen und kontrolliert und kann nach Bedarf oder pauschal vollautomatisiert und in Echtzeit verbessert werden und als alternative Mischung genutzt werden.

2.6 Vorträge und Veröffentlichungen

Der Schlussbericht wird nach Freigabe durch die Projektpartner auf der Webseite von RTW unter https://www.rtw.com/de/unternehmen/forschungsprojekte/sita.html veröffentlicht.

Während der Laufzeit des Projektes fanden verschiedene Veranstaltungen statt, auf denen die Inhalte von SITA und Vorträge zum Thema Sprachverständlichkeit der Öffentlichkeit präsentiert werden konnten:

- TMT 2018: Hannah Baumgartner & Wolfgang Hoeg, „Session: Sprachverständlichkeit in Rundfunk und Film“, mit: Felix Andriessens, Harald Fuchs, Mike Kahsnitz, Sebastian Goossens, Rainer Huber
- DAGA 2019: Rainer Huber et al., “Erfassung der Höranstrengung fertiger TV-Mischungen”
- 2019 Pro Light and Sound/ Broadcast & Production Forum: Christian Rollwage, Hannah Baumgartner, „Sprachverständlichkeit ist messbar!”, 05.April 2019
- 2019 MDR-Workshop Inklusion und Medien. Lisa Bodenseh, Hannah Baumgartner, „Wie Sprache besser verstanden werden kann“, 06. Mai 2019
- DAGA 2020: Rainer Huber et al., „Single-ended Prediction of Listening Effort for English Speech“
- VDT-live 2020 - Streaming Event: „Tatort Sprachverständlichkeit: Neuronale Netze in der Audioverarbeitung und Bewertung“, https://tonmeister.org/de/vdt-live/beitraege/
- Forum Acusticum 2020: Rainer Huber et al., “ASR-Based, Single-Ended Modeling of Listening Effort – A Tool for TV Sound Engineers”
- FKTG Fachtagung 2020, Hannah Baumgartner „Modellierung der Sprachverständlichkeit von Audiomischungen“

Referenzen

[1] Hohmann, V. & Kollmeier, B. (2006). A nonlinear auditory filterbank controlled by sub-band instantanenous frequency estimates. International Symposium on Hearing - ISH 2006, Cloppenburg, Springer.

[2] Xiong, F., Schneider, D., Goetze, S., Ewert, S., Rohdenburg, T. & Appell, J.-E. (2011). Hearing-Loss Compensation in a Telephone System. In Proc. der 37. Jahrestagung der Deutschen Gesellschaft für Akustik, DAGA 2011, 377-378.

[3] Oetting, D. & Appell, J.-E. (2013). Technische Möglichkeiten der Hör- und Audiounterstützung für altersgerechte Lebenswelten. 16. Jahrestagung der Deutschen Gesellschaft für Audiologie, Rostock.

[4] Rennies, J., Goetze, S. & Appell, J.-E. (2011). Personalized Acoustic Interfaces for Human-Computer Interaction. In: M. Ziefle und C. Röcker (Eds.), Human-Centered Design of E-Health Technologies: Concepts, Methods and Applications, IGI Global, 180-207.

[5] Goetze, S., Xiong, F., Rennies, J., Rohdenburg, T. & Appell, J.-E. (2010). Hands-free telecommunication for elderly persons suffering from hearing deficiencies. In 12th IEEE International Conference on E-Health Networking, Application and Services (Healthcom'10), Lyon, Frankreich.

[6] Rennies, J. et al., Höranstrengung von TV-Mischungen in Abhängigkeit von charakteristischen Hintergrundsignalen, DAGA 2017.

Oder: „Objektive Analyse, Visualisierung und Korrektur von Sprachverständlichkeit in Broadcastanwendungen für Normal- und Schwerhörende/ Speech Intelligibility for Broadcast" - kurz: SI4B durchgeführt, und gefördert durch das Bundesministerium für Wirtschaft und Energie [Förderkennzeichen ZF4072002SS5].

[7] R. Huber, A. Pusch, N. Moritz, J. Rennies, H. Schepker, and B. T. Meyer, B.T.: "Objective Assessment of a Speech Enhancement Scheme with an Automatic Speech Recognition-Based System." Proceedings ITG Conference on Speech Communication (2018), 86-90

[8] R. Huber, J. Ooster, and B. T. Meyer, "Single-ended Speech Quality Prediction Based on Automatic Speech Recognition", J. Aud. Eng. Soc., vol. 66, no. 10. pp. 759-769, 2018.

[9] H. Hermansky, E. Variani, and V. Peddinti, "Mean temporal distance: predicting ASR error from temporal properties of speech signal," in Proc. IEEE Conf. Acoust. Speech, Signal Process. (ICASSP), Vancouver, Canada, May. 2013, pp. 7423-7426.

[10] Y. Isik, J. Le Roux, Z. Chen, S. Watanabe, and J. Hershey, "Single channel multi-speaker separation using deep clustering," in Interspeech 2016, San Francisco, 09 2016, pp. 545–549.

[11] D. Yu, M. Kolbæk, Z.-H. Tan, and J. Jensen, "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017, pp. 241–245.

[12] M. Kolbæk, D. Yu, Z.-H. Tan, and J. Jensen, "Multitalker speech separation with utterance-level permutation invariant training of deep recurrent neural networks," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 25, no. 10, pp. 1901–1913, 2017.

[13] H. Baumgartner, A. Volgenandt, J. Rennies-Hochmuth: „Verringerung der Höranstrengung von TV-Mischungen durch Vorverarbeitung einzelner Spuren während der Mischung", Fortschritte der Akustik. DAGA 2016, S.960-963; DEGA, Berlin, 2018

[14] N. Moritz, J. Drefs, H. Baumgartner, J. Rennies: „Sprachaktivitätserkennung basierend auf Deep Neural Networks für Anwendung in Film und Fernsehen," Fortschritte der Akustik. DAGA 2016, S.960-963; DEGA, Berlin, 2016

[15] ANSI (1997). ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index", American National Standards Institute, New York.

[16] International Electrotechnical Commission IEC 60268-16 Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index, Fourth edition, 2011.

[17] Rennies, J. et al. (2010). „Automatic Live Monitoring of Communication Quality for Normal-Hearing and Hearing-Impaired Listeners." in K. Miesenberger et al. (Eds.): ICCHP 2010, Part II, LNCS 6180, pp. 568–575, 2010. Springer-Verlag Berlin Heidelberg 2010