Hörversuche zur Entwicklung eines neuartigen Mehrkapsel-Mikrofons

Hendrik Paukert[1], Jonathan Ziegler[1, 2], Andreas Koch[1]
[1]Hochschule der Medien Stuttgart, Germany, Email: paukert@hdm-stuttgart.de
[2]Eberhard Karls Universität Tübingen, Germany Email: zieglerj@hdm-stuttgart.de
[1]Hochschule der Medien Stuttgart, Germany, Email: kocha@hdm-stuttgart.de

Abstract

Für die Entwicklung eines neuartigen digital prozessierten Mikrofonarrays, haben wir im Vorfeld einen Hörversuch zur Untersuchung verschiedener Störgeräusche vorgestellt [1]. Diese dienten dem besseren Verständnis, welchen Faktoren bei der Entwicklung der Algorithmen die größte Beachtung geschenkt werden muss. In diesem Teil werden die Hörversuche, deren Audiodatengenerierung sowie die Ergebnisse zur Einschätzung und Abgrenzung verschiedener Fremd- und Eigenalgorithmen zur Entwicklung des Mehrkapsel-Mikrofonarrays im Haupteinsatzgebiet der Sprache vorgestellt. Hierzu wird das Open Source Tool "WAET" [2] für die menschliche-, sowie die STOI-Methode [3] für die algorithmische Bewertung genutzt.

1. Vorüberlegungen

Der vorausgegangene Hörversuch zur Untersuchung von Störgeräuschen wurde komplett in Max/MSP [2] programmiert und von den Teilnehmern offline an einem Laptop bearbeitet. Um den aktuellen Test nun schneller aufbauen und auch online einer größeren Anzahl an Teilnehmern anbieten zu können, sollte das frei zugängliche „Web Audio Evaluation Tool" - kurz „WAET" [3] genutzt werden. Dieses Tool bietet u.a. die Möglichkeit, einen den ITU-R BS.1534-Empfehlungen [4] entsprechenden Test erstellen zu können. Im Detail kann damit eine zufällige Prüfdatenausgabe, unterschiedliche Skalenbeschriftungen automatische Lautstärkennormalisierung, die Einbindung von Kommentarfeldern, versteckter Anker- und Referenzdatei und das Prüfen auf versehentlich doppelt eingebundene Dateien umgesetzt werden. Auch können weitere Daten wie z.B. Bearbeitungsdauer, Anzahl der Abspielwiederholungen gespeichert und der Hörversuch den Teilnehmern von einem php-fähigen Server browserbasiert angeboten werden.

Auf Grund der zu erwartenden großen Datenmengen, sollte außerdem eine zeitsparende und weitgehend automatisierte Auswerteroutine möglich sein und, um Überbeanspruchung der Probanden zu vermeiden, die Bearbeitungsdauer von 30 min. im Mittel nicht überschritten werden.

1.1 Vor- und Nachteile von Online-Hörversuchen

Durch online zugängliche Hörversuche kann eine große Anzahl an möglichen Teilnehmern erreicht werden. Diese können den Bearbeitungszeitpunkt außerdem frei wählen und den Test bequem an einem beliebigen Ort bearbeiten. Auch für den Versuchsdurchführenden reduziert sich der Aufwand für Auf- und Abbau der Testumgebungen an einem spezifischen Ort, ebenso reduziert sich der Organisationsaufwand.

Nachteile sind jedoch die indirekten Hilfe- bzw. Assistenzmöglichkeiten bei auftretenden Fragen und Problemen, sowie die schlechte Kontrollmöglichkeit des verwendeten Equipments. Ein Online-Versuch sollte also möglichst übersichtlich und selbsterklärend aufgebaut und die Aufgabenstellung dem entsprechend geeignet sein. Durch die in diesem Zusammenhang teilweise geringen Unterschiede unserer Testdaten in Bezug auf Nachhallzeit, Klangveränderung und Bildung von Artefakten, bat sich generell die Nutzung von Kopfhörern an. Auch erwies sich uns die Nutzung einer eingemessenen Abhöre in akustischen optimierter Umgebung als wenig relevant, da in diesem Hörversuch keine frequenz-vollumfänglichen Musik-bearbeitungen, sondern Sprache abgefragt wird. Weiterhin kann davon ausgegangen werden, dass der jeweilige Proband für die Bearbeitung des Hörversuches nur einen Kopfhörer nutzt und somit die Bewertung der Audiodaten zueinander stimmig ist. Die Typenbezeichnung der Kopfhörer fragten wir aus Interesse zusätzlich ab.

2. Zu untersuchende Algorithmen

Haupteinsatzgebiet des neu entwickelten Mikrofonarrays werden Konferenzen sein. Hierzu wurde an der Hochschule der Medien ein Tracking-Algorithmus zur Detektion und Verfolgung jeweiliger Sprecher entwickelt [5], [6]. Eine aus drei Mikrofonkapseln synthetisierte Nieren- oder Superniercharakteristik, kann in Echtzeit ein aktuelles Sprachereignis erfassen, verfolgen und so von unerwünschten Schallereignissen besser freistellen.

Um diesen Freistellungseffekt weiter erhöhen zu können, soll zusätzlich ein Beamformer- bzw. Dereverb-Algorithmus implementiert werden. Hierzu stand prozessiertes Audiomaterial des Beamformers von Bernfried Runow [7], [8], zwei Dereverb-Plug-Ins und ein Spaced-Array-Konferenz-Komplett-System zur Verfügung. Durch die fixe Architektur der verschiedenen Algorithmen, konnte jedoch nur Runows Beamformer alle drei Mikrofonsignale des neuen Mehrkapselmikrofons nutzen und verfügte somit theoretisch über das größte Potential. Ein Plug-In konnte so nur das Summen-Trackingsignal, das andere nur zwei Signale des Mikrofons nutzen. Eine weiterhin geprüfte Spaced-Array-Konferenzanlage, verfügt über ihr eigenes räumliches Arraymikrofon mit 24 Kapseln und dient dem direkten Vergleich zu einem auf dem Markt bereits erhältlichen System.

ISBN
978-3-9812830-9-9

Nach Möglichkeit wurden die verschiedenen Algorithmen auch in verschiedenen Stärkeeinstellungen abgeprüft. Inklusive der Ankerdatei, dem bewusst verschlechterten Signal, und der Referenzdatei, dem optimalen Signal, umfasste der Hörversuch 9, jeweils in zwei Sprachen zu bewertende Audiofiles. Angesichts der gesetzten Bearbeitungsobergrenze von 30 Minuten, erwies sich diese Anzahl bei mehreren Vorversuchen bereits als ausreichend.

Bezeichnung	Beschreibung
Anker	Ankerdatei mit 3,5kHz-Lowpass und einer der Kategorie entsprechenden Degradierung durch hohe Nebengeräusche, Rauschen oder hohen Hallanteil. Dieses Signal muss vom Probanden als „schlecht“ bewertet werden.
Referenz	Nach Möglichkeit optimales Signal (trockene Studiosprachaufnahme), ohne Hall, Nebengeräusche oder Rauschen. Dieses Signal muss als „gut“ bewertet werden.
Kugel	Kugelsignal des Prototypenmikrofons
Runow 50%	B. Runows-Beamforming-Algorithmus in mittlerer Stärke, gespeist mit den drei einzelnen Kapselsignalen und der vom Tracker ermittelten Richtungsinformation.
Runow 100%	B. Runows-Beamforming-Algorithmus in voller Stärke, gespeist mit den drei einzelnen Kapselsignalen und der vom Tracker ermittelten Richtungsinformation.
Plug-In Nr.1 50%	DAW-Plug-In in mittlerer Stärke, gespeist mit zwei geführten Mono-Signalen.
Spaced-Array	Gesamtsystem mit eigenem räumlichen Mikrofonarray und schwacher Einstellung des produkteigenen Algorithmus. Stärkere Einstellungswerte verschlechterten das Signal enorm und mussten daher vom Test ausgeschlossen werden.
Plug-In Nr.2 50%	DAW-Plug-In in mittlerer Stärke, gespeist mit geführtem Mono-Signal.
Plug-In Nr.2 stark	DAW-Plug-In in starker Einstellung, gespeist mit geführtem Mono-Signal. Noch stärkere bzw. maximale Einstellung verschlechterten in unserem Fall die Signalqualität sehr stark.

Tab. 1: Übersicht der genutzten Signale

3. Generierung der Audiodaten

Zur Generierung der Audiodaten nutzten wir den neu entwickelten Aufbau einer reproduzierbaren Konferenzumgebung: über mehrere Iterationen hinweg konnten wir hier, auch durch die detaillierte BA von Robin Hirt [9], eine „virtuelle Konferenz“ gestalten. Mit diesem Mikrofonteststand kann über ein Lautsprechersetup mit definierten Abständen und Zuspielszenarien die Wirkungsweise der verschiedenen Algorithmen bzw. Algorithmen-Revisionen geprüft werden.

Insgesamt werden 10 Lautsprecher in zwei Ringen bzw. Radienabständen vom zu prüfenden Mikrofon aufgebaut. Über die vier Lautsprecher des inneren Rings werden die nahezu reflexionsfreien Sprachdateien mehrerer Personen getrennt und auch parallel ausgegeben. Die Lautsprecher des äußeren Rings erzeugen ein- und mehrkanalige Atmogeräusche wie z.B. Straßenlärm oder das Öffnen und Schließen einer Tür. Weiterhin befindet sich, nahe am Mikrofon, ein einzelner Lautsprecher, welcher z.B. Papier-, Tassen- und Stiftgeräusche ausgibt, sowie in 4 Metern Abstand ein weiterer Lautsprecher zur Generierung von sehr indirekten und räumlichen Signalen. Zur Simulation von Körperschall, z.B. verursacht durch Vibrationen eines Laptops oder Smartphones, kann zur Simulation optional ein kleiner E-Motor am Tisch befestigt werden. Durch eine elastische Aufhängung des Mikrofons ist hier allerdings schon von einer guten Körperschallentkopplung auszugehen.

Zur Nutzung im Hörversuch wurde letztendlich ein 3s kurzer Zeitabschnitt der Aufnahmen in deutscher und englischer Sprache ausgesucht. Längere Abschnitte machten aus Gründen der Vergleichbarkeitsschwäche der menschlichen Klangwahrnehmung und der dadurch folgenden Erhöhung der Versuchsdauer keinen Sinn.

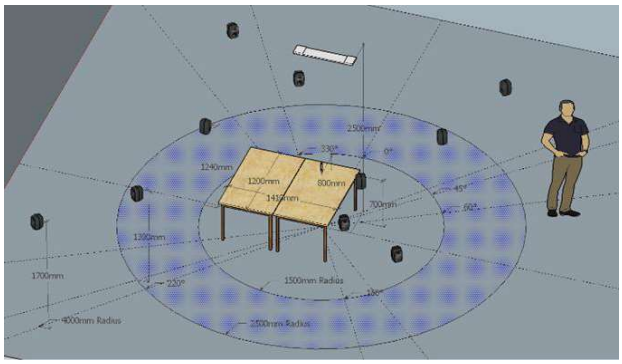


Abb. 1: Aufbauskizze der virtuellen Konferenz [10]: Zu sehen sind die beiden konzentrischen Lautsprecherringe, ergänzende Nah- und Fernlautsprecher, das kleine Prototypenmikrofon in der Mitte sowie das Konferenzsystems in Deckenmontage.

Abb. 2: 360°-Aufnahme des Genelec 1029A [11]-Lausprechersetups der virtuellen Konferenz: Die Raummaßen betragen 8,3x8,2x3,8m, die RT60 2,31s im Bereich von 500-1000Hz 2,31s (leerer und akustisch unbehandelter Raum).

4. Versuchsinterface

Das WAET stellt verschiedene Interface-Arten wie z.B. AB-, ABX- und Checkbox-Verfahren zur Verfügung. Anstatt des klassischen Mehrfachschieberegler-Interface der klassischen

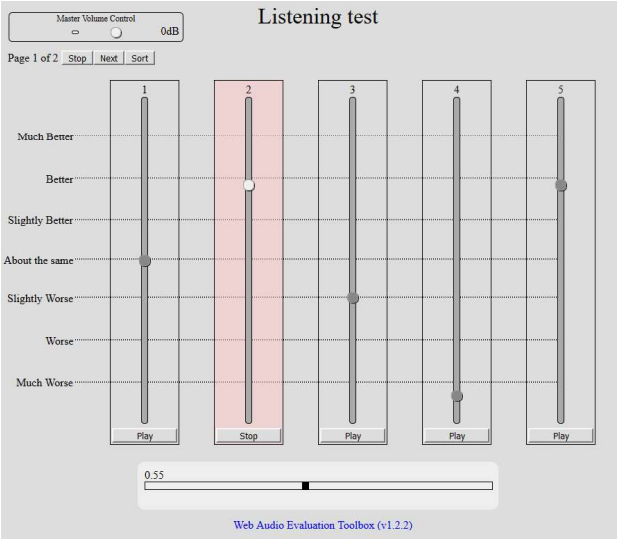


Abb. 3: Mehrfachschieberegler-Interface des Web Audio Evaluation Tool: jedes der fünf angebotenen Signal verfügt über einen eigenen Schieberegler

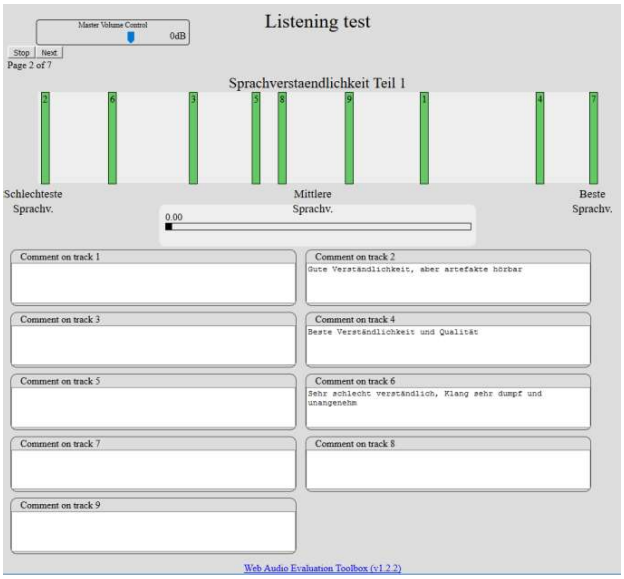


Abb. 4: einachsiges Hörversuchsinterface mit dem Web Audio Evaluation Tool: Jeder der grünen Balken repräsentiert eine auf der Skala zu positionierende Sprachdatei. Weitere Auffälligkeiten können anhand der Kommentarfelder weitergegeben werden.

MUSHRA-Methode (**Abb. 3**), fiel die Entscheidung zu Gunsten der APE-Variante (Audio Perceptual Evaluation) [12]. Wie in **Abb. 4** ersichtlich, wird hierbei nur eine Achse genutzt, auf der alle Soundfiles als Schiebebalken repräsentiert sortiert werden müssen. Das Interface ist dadurch kompakter und lässt pro Prüfabschnitt mehr Platz für Kommentarfelder oder andere Zusatzoptionen. Außerdem schien uns der einachsige Aufbau die Relation der Soundfiles zueinander mehr in den Vordergrund zu stellen, als dies bei getrennten Achsen der Fall wäre [Vgl. 13]. Durch die Textfelder konnten die Probanden auch möglicherweise gar nicht abgefragte, ihnen aber als wichtig erscheinende Informationen weitergeben. Um ein mögliches Biasing bzw. bewussten oder unterbewussten Manipulationen entgegenzuwirken, wurden die Audiodaten pro Abschnitt außerdem zufallsverteilt ausgeben.

Auf Grund des weiten Wertebereichs des einachsigen Aufbaus jedoch, kann bei nicht vollständiger Nutzung der Skala eine Verzerrung der Standardabweichung σ auftreten: ein Proband könnte seine Daten z.B. nur in der unteren Hälfe der Skala anordnen, ein anderer jedoch in der oberen Hälfte. Durch diesen Offset würde die σ aller Daten nun fälschlicherweise ansteigen. Das WAET bietet hier an, gewünschte Soundfiles in einstellbaren (Toleranz-)Bereichen an den Skalenenden anordnen zu müssen. Das nachträgliche Normalisieren der vielen Ergebniswerte kann so vermeiden werden. In unserem Falle musste also die Ankerdatei, welche vom Probanden definitiv als schlechteste Datei erkannt werden muss, am linken Rand, und die Referenzdatei, am rechten Rand angeordnet werden. Diese Maßnahme stellt auch sicher, dass jeder Teilnehmer den Test aktiv und aufmerksam durchführt und ihn bis zum Erkennen und dem korrekten Einordnen der Dateien nicht fortführen kann. Die restlichen Sprachdateien können nun zwischen beiden Werten verteilt werden. Leider konnte die angezeigte Fehlermeldung bei Nichterkennen nicht zu einem verständlichen Hinweis umformuliert werden.

Weiterhin erlaubt die WAET-Testumgebung eine Überprüfung ob jedes Audiodatei komplett abgespielt und bewegt worden ist. Das ist wichtig, da die Dateien, abgesehen von Anker- und Referenzdatei, sonst an ihren zufallsgenerierten Startpositionen stehen bleiben können.

5. Beschreibung der Prüfabschnitte und Ergebnisse

5.1. Einleitende Fragen

Zu Beginn des Hörversuches wurde eine einleitende Beschreibung angezeigt und die verwendeten Kopfhörer, das Alter der Teilnehmer sowie die audiospezifischen Erfahrungsbereiche abgefragt. Wie in **Abb. 5** erkennbar, gaben die meisten Teilnehmer an, ein oder mehrere Instrumente zu spielen und sich mit dem Aufnehmen und Bearbeiten von Musik, Foleys oder Sprache zu befassen. Ein weiterer großer Teil gab an, Musikliebhaber und Konsument von höherwertiger Audiotechnik zu sein. Da auch Mehrfachnennungen möglich waren, wurden oft zwei oder drei Kategorien parallel genannt, siehe **Abb. 6**.

Bei den genutzten Kopfhörern wurden die Marke Beyerdynamic [14], gefolgt von Sennheiser [15] und AKG [16], am meisten genannt (**Abb. 7**). Die vielen Einzelmeldungen unterschiedlicher Marken wurde unter „sonstige Nennungen“ zusammengefasst. Das Durchschnittsalter aller 59 Teilnehmer betrug 39 Jahre, allerdings mit einer hohen Streuung. Jüngster war 19 und ältester Teilnehmer 63 Jahre jung. Auch die durchschnittliche Bearbeitungsdauer schwankte stark, lag im Mittel allerdings bei knapp 22 Minuten.

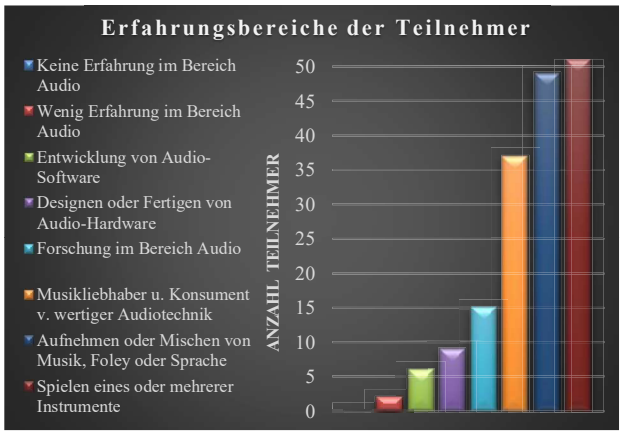


Abb. 5: Erfahrungsbereiche der Teilnehmer: am meisten kennen sich die Versuchsteilnehmer bei Audioproduktionen aus und spielen ein oder mehreren Instrumente.

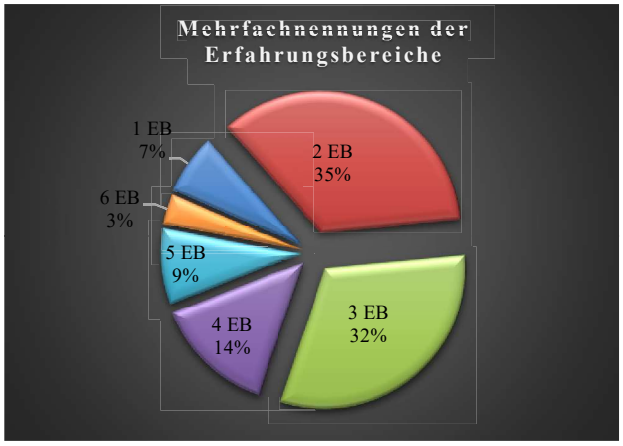


Abb. 6: Mehrfachnennungen der Teilnehmer: 67% nannten max. 3 Erfahrungsbereiche parallel.

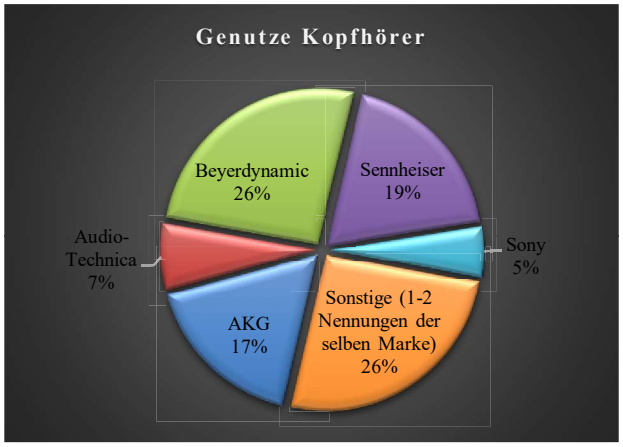


Abb. 7: Genutzte Kopfhörer der Teilnehmer: die größte Blöcke bilden mit insg. 62% Beyerdynamic, Sennheiser und AKG.

5.2 Beschreibung der Balkendiagramme

Die nachfolgenden Diagramme setzen sich aus den Bewertungen der abgeprüften Sprachdateien anhand von blauen Balken, sowie deren Standardabweichung σ anhand von roten Balken zusammen. Der Wert 0,0 steht für eine sehr schlechte und der Wert 1.0 für eine sehr gute Bewertung. Niedrige Werte der roten Balken zeigen eine niedrige σ auf. In diesem Zusammenhang liegen die σ-Werte der Anker- und Referenzdatei generell und trotz des Toleranzbereichs für die Zwangspositionierung an den Skalenenden auf einem sehr niedrigen Niveau. Für den ersten Prüfabschnitt „Training“ steht die 1,0 für eine als sehr hallig empfundenen Bewertung.

5.3 Training

Damit sich die Teilnehmer an die Testumgebung gewöhnen konnten, wurde ein vom Prüfumfang reduzierter Trainingsmodus implementiert. Anstatt 9 wurden nur 6 Soundfiles in einer Sprache, Deutsch oder Englisch, abgefragt. Auch wenn das Training dadurch nicht voll bewertet werden kann, können die Daten auf Grund der Menge der Teilnehmer und der abgefragten Eigenschaft „empfundene Halligkeit“ zumindest für eine erste Abschätzung genutzt werden. Grundsätzlich stellt das Kugelsignal das unbearbeitete und räumlichste Realsignal dar und wird nur von der künstlich verhallten Ankerdatei übertroffen (3s Nachhallzeit). Nachfolgend zur nachhalllosen Referenzdatei, wurde B. Runows Beamformer als sehr trocken bewertet.

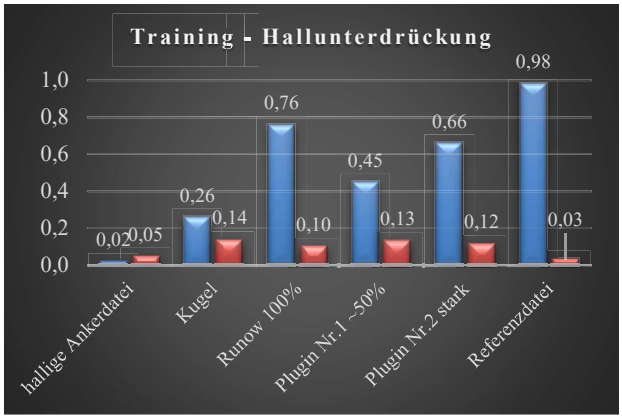


Abb. 8: Ergebnis „empfundene Halligkeit“ im Trainingsmodus: Runows Beamformer wird mit geringster σ und Halligkeit bewertet.

5.4 Sprachverständlichkeit

Eine Haupteigenschaft für (Konferenz-)Mikrofone ist die Verständlichkeit der Sprache. Hierzu wurden nun alle im Versuch vorkommenden Sprachdateien in deutscher und englischer Sprache abgefragt.

Bei beiden Durchläufen wurde B. Runows Beamformer im Maximaleinstellung als bestes bewertet, dicht gefolgt von Plug-In Nr.2. Deutlicher Verlierer ist hier das Spaced-Array-Konferenzsystem, dessen Sprachverständlichkeit deutlich unter einer übermäßigen Signalverschlechterung durch das interne Processing leiden musste.

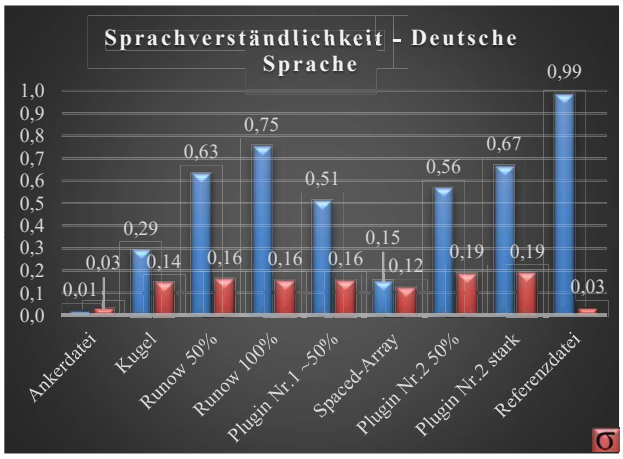


Abb. 9: Ergebnis „Sprachverständlichkeit“ in deutscher Sprache

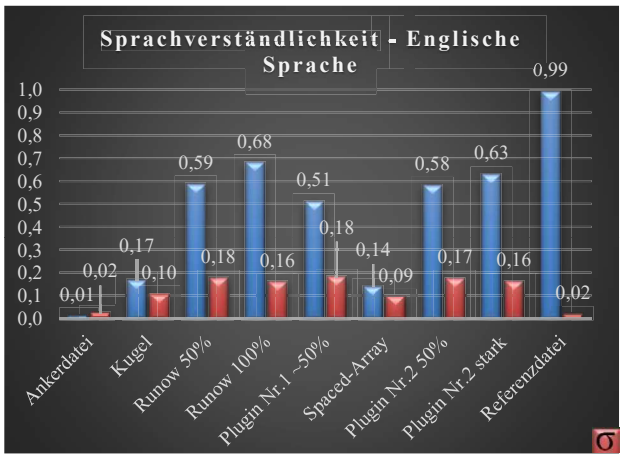


Abb. 10: Ergebnis „Sprachverständlichkeit“ in englischer Sprache

5.5 Unterdrückung der Störgeräusche

In diesem Abschnitt fragten wir die Sprachdateien mit Fokus auf die Störgeräuschunterdrückung ab. Das vorher weit abgeschlagene Konferenzsystem konnte hier deutlich aufholen. Sein starkes Processing versteht es, Nebengeräusche wirksam zu unterdrücken, im Vergleich zur Konkurrenz litt dadurch aber auch erheblich die Qualität, womit die erhöhte σ zusammenhängen könnte. Abgesehen von der störgeräuschsfreien Referenzdatei, führt B. Runows Algorithmus in Maximaleinstellung erneut die Auswertung an.

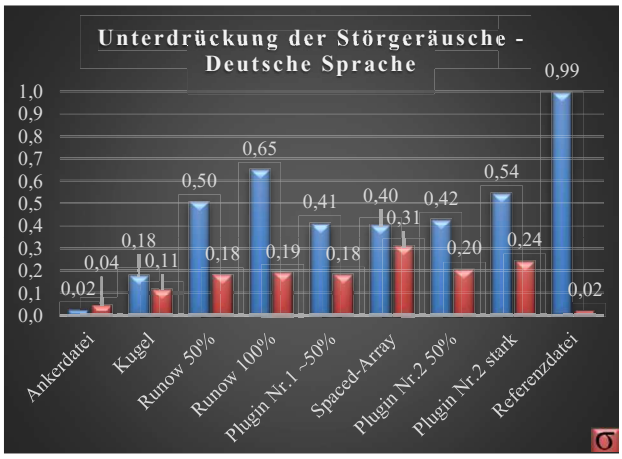


Abb. 11: Ergebnis „Unterdrückung der Störgeräusche“ in deutscher Sprache

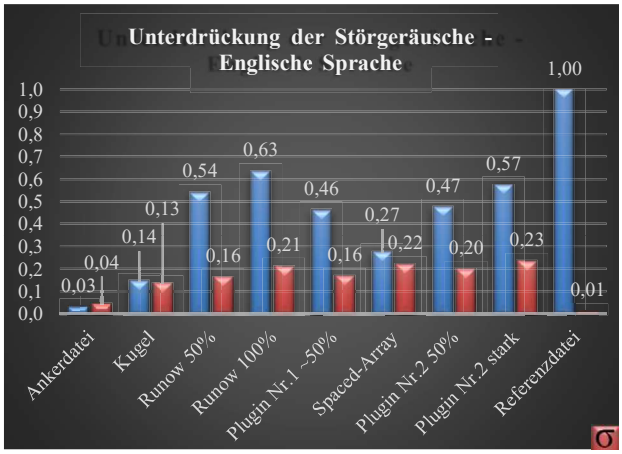


Abb. 12: Ergebnis „Unterdrückung der Störgeräusche“ in englischer Sprache

5.6 Empfundene Qualität

Das Gehör reagiert sehr empfindlich auf Veränderungen des Klangs der Sprache und schon geringe Nuancen werden wahrgenommen. Als weiterer wichtige abzuprüfende Eigenschaft galt daher die empfundene Qualität der bearbeiteten Sprachdateien. Hierbei liegt das Spaced-Array nur knapp über der bewusst mit Rauschen und Bandpassfilterung degradierten Ankerdatei. Trotz hoher Sprachverständlichkeit und Störgeräuschunterdrückung kann B. Runows Beamformer auch hier den ersten Platz behaupten. Da gute Qualität ein individuelles Maß ist, vermieden das Einbinden einer vermeintlich hochwertigen Referenzdatei. Auch in diesem Prüfabschnitt wichen die Ergebnisse in deutscher und englischer Sprache nur gering voneinander ab.

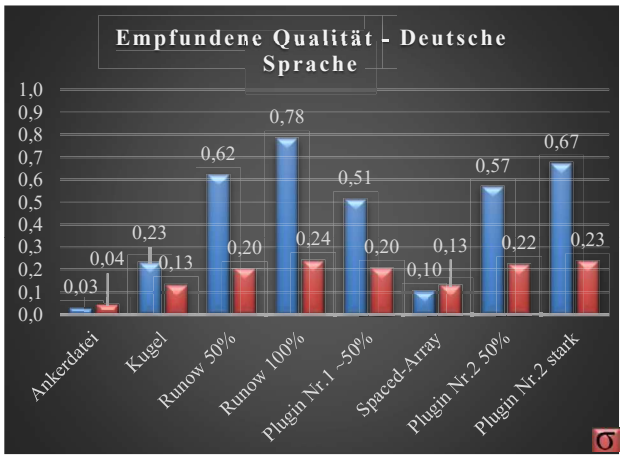
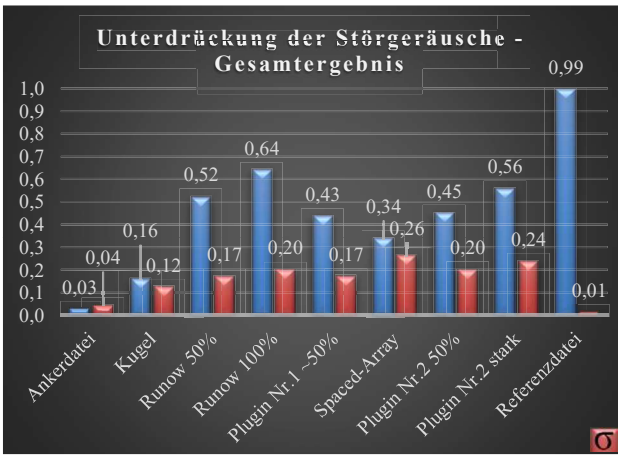


Abb. 13: Ergebnis „empfundene Qualität“ in deutscher Sprache



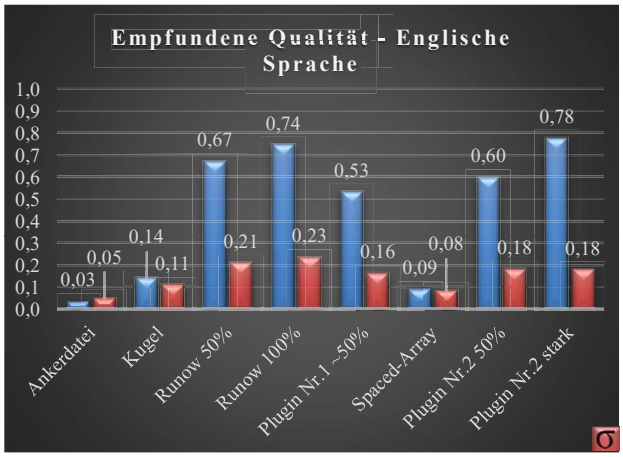


Abb. 14: Ergebnis „empfundene Qualität“ in englischer Sprache

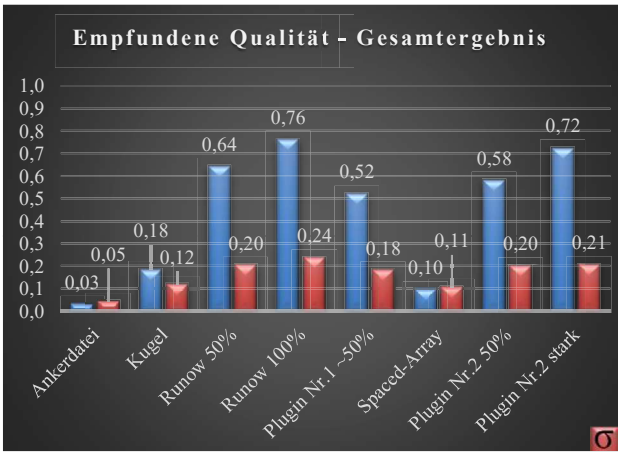


Abb. 17: Gesamtergebnis „empfundene Qualität“

5.7 Zusammenfassende Ergebnisse beider Sprachen

Zu besseren Übersicht werden hier nochmal die zusammenfassenden Ergebnisse beider Sprachen in Summe dargestellt.

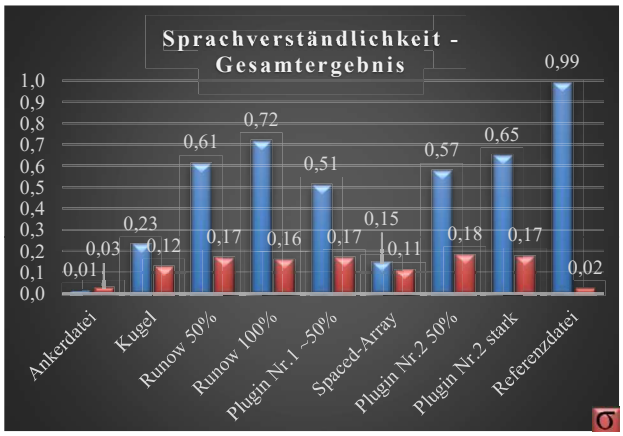


Abb. 15: Gesamtergebnis „Sprachverständlichkeit“

5.8 Sprachverständlichkeitsindex STOI

Zum automatischen Bewerten der Sprachverständlichkeit bietet sich das frei zugängliche Matlabscript „STOI“ [17] an. „STOI“ steht für Short Time Objective Intelligibility Messurement und wurde von C.H. Taal et al an der Delft University of Technology in Holland entwickelt. Das Skript vergleicht in unserem Fall die trockene Studioaufnahme (Referenzdatei) mit denen, über das in der virtuellen Konferenz mit dem Prototypenmikrofon aufgenommen und Dateien. Hohe Ergebniswerte bedeuten eine hohe Sprachverständlichkeit, so dass auch beim Vergleich zweier gleicher Dateien stets eine 1.0 herauskommt. Wir konnten so unsere Sprachdateien erneut abprüfen, gespannt darauf, in wie fern sich die Ergebnisse des Prüfalgorithmus mit unseren bisherigen Ergebnissen decken würden.

Grundsätzlich liegen die Ergebnisse näher beieinander, als bei der durch den Menschen durchgeführten Bewertung. Dadurch sollte den Nachkommastellen mehr Beachtung geschenkt werden. Interessant ist, dass beide Sprachdateien des Spaced-Arrays schlechter als die Ankerdatei bewertet wurde. Wie bei der Ankerdatei sind die Daten des Spaced-Arrays im

Frequenzband stark beschnitten - für den Bewertungs-Algorithmus wahrscheinlich zu stark und in einem als wichtig bewertetem Bereich. Auch das indirekte Kugelsignal wird in beiden Fällen erkannt und in Summe nicht viel besser bewertet als das räumliche Kugelsignal. Interessanterweise schneidet das Tracking Signal bei der STOI-Bewertung auf ähnlich hohem Niveau ab, wie mit kombiniertem Beamformer bzw. Dereverb. Grund dafür könnte die interne Korrelationsmessung bis nur 5000Hz sein, welche a) das Processing der höheren Frequenzen unbewertet lässt und b) die Artefaktbildung korrelationsbedingt höher bewertet als der Mensch, der hier teilweise deutliche Unterschiede hören kann.

Die weiteren Sprachdateien bzw. Algorithmen liegen relativ nahe beieinander, allerdings zeichnet sich auch in diesen geringen Bereich das Führen des Beamformer von B. Runow ab. Deutlich erkennbar ist auch, dass das Plug-In Nr. 1 in dieser Bewertungsrunde stark aufholen kann. Zur Ergänzung wurden zwei weitere Einstellungen des Plug-Ins ausprobiert.

Bezeichnung	Beschreibung
Ankerdatei	Ankerdatei mit 3,5kHz-Lowpass und einer Degradierung durch hohe Nebengeräusche und Rauschen. Dieses Signal muss als „schlecht“ bewertet werden.
Referenz	Sie diente im STOI-Versuch als Vergleichssignal (Wert 1.0). Aus Übersichtsgründen wird dieses Signal daher im Chart nicht dargestellt.
Kugel	Kugelsignal des Prototypenmikrofons
Tracking	Sprachverfolgungs-Algorithmus Hochschule der Medien Stuttgart mit synthetisierter Niere bis Superniere.
Runow 50%	B. Runows-Beamforming-Algorithmus in mittlerer Stärke, gespeist mit den drei einzelnen Kapselsignalen und der vom Tracker ermittelten Richtungsinformation.
Runow 100%	B. Runows-Beamforming-Algorithmus in maximaler Stärke, gespeist mit den drei einzelnen Kapselsignalen und der vom Tracker ermittelten Richtungsinformation.
Plug-In Nr.1 ~25%	DAW-Plug-In in schwacher Einstellung, gespeist mit zwei geführten Mono-Signalen.
Plug-In Nr.1 ~50%	DAW-Plug-In in mittlerer Einstellung, gespeist mit zwei geführten Mono-Signalen.
Plug-In Nr.1 100%	DAW-Plug-In in maximaler Stärke, gespeist mit zwei geführten Mono-Signalen.
Spaced-Array	Gesamtsystem mit eigenem räumlichen Mikrofonarray und schwacher Einstellung des produkteigenen Algorithmus.
Plug-In Nr.1 50%	DAW-Plug-In in mittlerer Stärke, gespeist mit geführtem Mono-Signal.
Plug-In Nr.2 stark	DAW-Plug-In in starker Einstellung, gespeist mit geführtem Mono-Signal.

Tab. 2: Übersicht der ergänzten Signale für die Sprachverständlichkeitsanalyse „STOI“

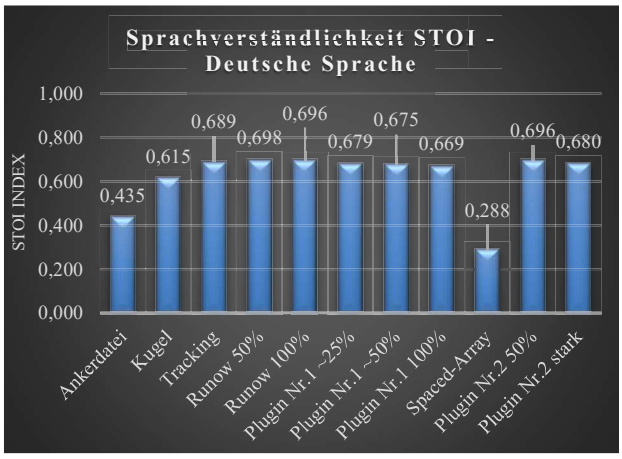


Abb. 18: Ergebnis Sprachverständlichkeitsanalyse „STOI“ in deutscher Sprache

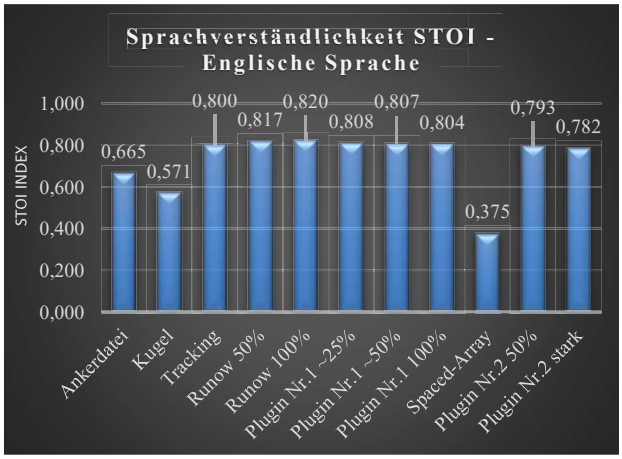


Abb. 19: Ergebnis Sprachverständlichkeitsanalyse „STOI“ in englischer Sprache

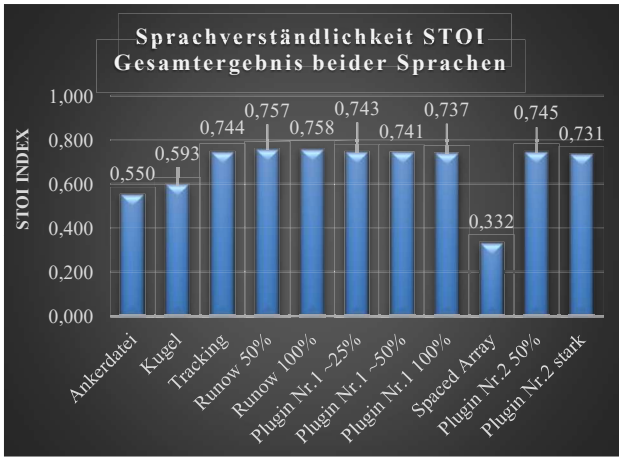


Abb. 20: Gesamtergebnis Sprachverständlichkeitsanalyse „STOI“ beider Sprachen

6. Datenauswertung

Pro Teilnehmer standen über 70, insgesamt über 4000 Werte zur Auswertung bereit. Zur zügigen Verarbeitung der Daten musste beim Erstellen des Hörversuches jedoch beachtet werden, dass WAET die zufällige Reihenfolge der ausgegeben Sprachdateien genauso in der Ergebnisdatei abspeichert. Daher müssen die jeweiligen Sprachdateien bei der Programmierung des Hörversuchs mit Indizes versehen werden um sie nachträglich sortieren zu können. Anschließend konnten die sortierten Datenblöcke in eine Sammelliste übertragen und ausgewertet werden.

7. Zusammenfassung und Ausblick

Durch das Web Audio Evaluation Tool konnte ein individuell programmierbarer und den ITU-Empfehlungen folgender Hörversuch einem großen Teilnehmerkreis online zugänglich gemacht werden.

Insgesamt wurde die Kombination des Trackingalgorithmus mit B. Runows Beamformer in Maximaleinstellung als bestes bewertet. Dieser Beamformer konnte allerdings auch alle drei Kapselsignale nutzen, so dass auch die anderen Algorithmen theoretisch Optimierungspotential besitzen. Plug-In Nr.1 lag mit seinen Ergebnissen oft knapp hinter denen des Plug-Ins Nr.2 und konnte vor allem bei den Ergebnissen der automatischen Sprachverständlichkeitsanalyse STOI deutlich aufholen. Das eigenständige Konferenzsystem wurde von allen prozessierten Dateien in allen Kategorien als schlechtestes bewertet und konnte nur im Bereich der Störgeräuschunterdrückung einigermaßen überzeugen.

Weiterhin können die gesammelten Daten zur fortführenden Mikrofonentwicklung genutzt werden und die Grundlage zur Implementierungen eines Beamformers und dessen Stärkeeistellungen bilden.

8. Quellenangaben

[1] H. Paukert, J. D. Ziegler: “Listening Tests in the Process of Microphone Development”: Tagungsbericht der 29. internationalen Tonmeistertagung des Verbands deutscher Tonmeister e.V., Nov. 2016, Seite 273-280, ISBN 978-3-9812830-7-5, URL: https://www.tonmeister.de/index.php?p=tonmeistertagung/2016/downloads

[2] Cycling ’74, Max 7 perpetual licence, URL: https://www.cycling74.com

[3] N. Jillings, B. De Man, D. Moffat, J. D. Reiss: “Web Audio Evaluation Tool: a browser-based Listening Test Environment”, Centre for Digital Music, Queen Mary University of London, URL: https://github.com/BrechtDeMan/WebAudioEvaluation Tool

[4] “Method for the subjective assessment of intermediate quality level of coding systems”, International Telecommunication Union (ITU) Empfehlung BS.1534-3, 2015, URL: https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en

[5] J. D. Ziegler, A. Koch, A. Schilling: „Speech classification for acoustic source localization and tracking applications using convolutional neural networks”, Audio Engineering Society convention 145, October 2018

[6] J. D. Ziegler, H. Paukert, A. Koch, A. Schilling: “Speaker Tracking with Coincident Microphone Arrays and Convolutional Neural Networks”, IEEE Journal, aktuell ausstehende Bewilligung

[7] B. Runow, O. Curdt: “Mikrofonarrays in der professionellen Audioproduktion”, Tagungsbericht der 28. Internationalen Tonmeistertagung Verband deutscher Tonmeister e.V., Nov. 2014, Seite 263-269, ISBN 978-3-9812830-5-1 URL: https://www.tonmeister.de/index.php?p=tonmeistertagung/2014/downloads

[8] B. Runow, O. Curdt, A. Schilling: „Richtrohrmikrofone versus Mikrofonarrays“, Tagungsbericht der 29. internationalen Tonmeistertagung des Verbands deutscher Tonmeister e.V., Nov. 2016, Seite 221-228, ISBN 978-3-9812830-7-5, URL: https://www.tonmeister.de/index.php?p=tonmeistertagung/2016/downloads

[9] R. Hirt: „Entwicklung einer virtuellen Konferenz unter besonderer Berücksichtigung der Reproduktion von zuvor aufgenommener Sprache“, Bachelorthesis, Hochschule der Medien Stuttgart, 2017

[10] Timbre Inc., URL: https://www.sketchup.com/

[11] Genelec Inc. URL: https://www.genelec.com/support-technology/previous-models/1029a-studio-monitor

[12] B. D. Man, J. D. Reiss, “APE: Audio Perceptual Evaluation toolbox for Matlab”, 136th AES Convention 2014, Berlin, Germany

[13] B. D. Man, J. D. Reiss, “APE: Audio Perceptual Evaluation toolbox for Matlab”, 136th AES Convention 2014, Berlin, Germany, Seite 1

[14] Beyerdynamic GmbH & Co. KG, URL: http://www.beyerdynamic.de

[15] Sennheiser electronic GmbH & Co. KG, URL: https://de-de.sennheiser.com

[16] Harman Deutschland GmbH, URL: https://de.akg.com/

[17] C.H. Taal, R. C. Hendriks, R. Heusdens, J. Jensen: A short-time objective intelligibility measure for time-frequency weighted noisy speech”, Delft University of Technology, Niederlande