Forschung für eine bessere akustische Kommunikation

Perzeptive Prinzipien, Algorithmen und Anwendungen

Volker Hohmann, Jörn Anemüller, Thomas Biberger, Matthias Blau, Thomas Brand, Simon Doclo, Stephan Ewert, Giso Grimm, Julia Schütze, Bernhard Seeber, Kirsten C. Wagener, Anna Warzybok-Oetjen

Menschliche Sprachkommunikation ist die Grundlage unserer Kultur und der Schlüssel zur aktiven Teilhabe. Sie wird von einer Reihe von Faktoren wie schwierigen Hörsituationen, komplexer Raumakustik, Lärm und Nachhall herausgefordert. Diese Faktoren erschweren die Kommunikation insbesondere bei Menschen mit Hörbeeinträchtigung, und trotz der Fortschritte der Signalverarbeitung bieten aktuelle elektroakustische Hörhilfen nur einen begrenzten Vorteil. Ein wesentlicher Grund dafür ist, dass die Wechselbeziehung zwischen diesen Faktoren, der Gerätefunktion und den individuellen Hördefiziten nicht ausreichend verstanden ist. Insbesondere wird bei der Entwicklung aktueller Geräte passives Hören angenommen, wohingegen die Kommunikation im realen Leben ein aktives Zuhören erfordert. Die akustische Kommunikation erfolgt demnach in einer Schleife, die das Schallfeld, das Gerät sowie die Wahrnehmung und die Aktivität des Nutzenden umfasst. Moderne Methoden der Hörakustik beziehen diese Kommunikationsschleife systematisch mit ein und gehen damit über die bisherigen Methoden hinaus. Aktuelle Forschungsergebnisse dazu zeigen Möglichkeiten zur besseren Unterstützung der akustischen Kommunikation in realen Umgebungen durch elektroakustische Geräte, bessere Prinzipien der Mensch-Maschine-Interaktion in der Unterhaltungselektronik, sowie eine umfassende Basis für eine verbesserte Entwicklung und Evaluation von Hörgeräten auf und sind hoch relevant für unsere alternde Kommunikationsgesellschaft.

Basic research for better acoustic communication

Human speech communication is the basis of our culture and the key to active participation. It is challenged by a number of factors such as difficult listening situations, complex room acoustics, noise and reverberation. These factors make acoustic communication difficult, especially for people with hearing impairments, and despite advances in signal processing, current electroacoustic hearing aids offer only a limited benefit. A key reason for this is that the interrelationship between these factors, device function and individual hearing deficits is not well understood. In particular, the development of current devices assumes passive hearing, whereas communication in real life requires active listening. Acoustic communication therefore takes place in a loop that includes the sound field, the device and the user's perception and activity. Modern research methods in hearing acoustics consider this communication loop systematically and thereby go beyond established methods. This article presents recent work in that area. The research results, i. e. better support of acoustic communication in real environments by electroacoustic devices, better principles of human-machine interaction in consumer electronics, as well as a comprehensive basis for improved development and evaluation of hearing aids, are highly relevant for our ageing communication society.

Einleitung

Akustische Kommunikation ist allgegenwärtig und für die soziale Interaktion und die Informationsbeschaffung in vielen verschiedenen akustischen Umgebungen unerlässlich, zu Hause, bei gesellschaftlichen Zusammenkünften, am Arbeitsplatz und in öffentlichen Räumen wie z. B. Hörsälen, Konzertsälen, Bahnhöfen oder Supermärkten. Elektroakustische Systeme, die die akustische Kommunikation in diesen Umgebungen unterstützen, haben sich in den letzten Jahrzehnten rasant entwickelt, insbesondere durch den umfassenden Einsatz digitaler Signalverarbeitung in Kombination mit der ständig wachsenden Rechenleistung von universellen und anwendungsspezifischen Signalprozessoren.

Verschiedene Arten solcher kommerziell relevanten Systeme sind im Einsatz oder werden derzeit entwickelt, z. B. in Beschallungsanlagen, Mobiltelefonen, Home-Entertainment-Systemen mit Lautsprechern und Kopfhörern, Hörgeräten und tragbaren Hörhilfen mit Audio-Ohrstücken („Hearables"). Sie alle müssen in diesen sehr unterschiedlichen akustischen Umgebungen und für Nutzer und Nutzerinnen mit unterschiedlichen Anforderungen funktionieren,

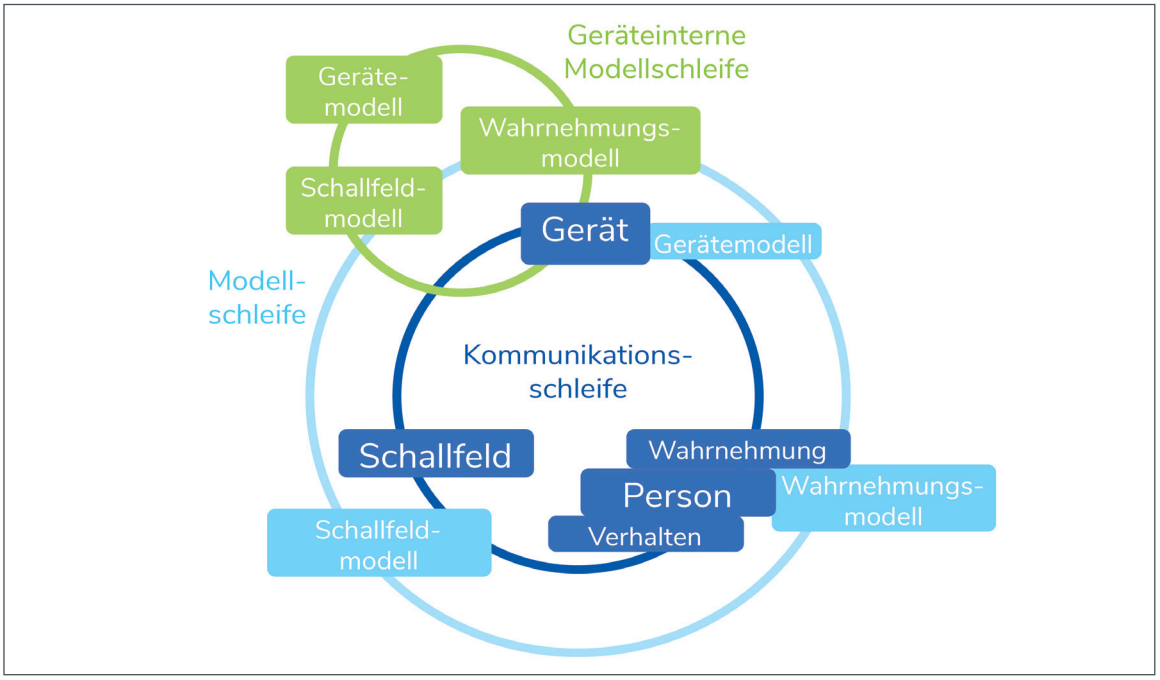

Abb. 1: Akustische Kommunikationsschleife, die die dynamische Interaktion zwischen dem Schallfeld, dem Hörgerät und der Person mit ihrer Wahrnehmung und Aktivität bzw. ihrem Verhalten beschreibt. Die Modellschleife zielt darauf ab, diesen Prozess quantitativ zu simulieren und bildet damit die Grundlage für die Verbesserung der Gerätefunktion durch die Einbeziehung einer geräteinternen Modellschleife.

z. B. für junge und ältere Hörerinnen und Hörer sowie für Personen mit normalem oder beeinträchtigtem Hörvermögen.

Frühere vielversprechende Ergebnisse zeigen, dass die Probleme dieser Geräte, eine mühelose Kommunikation unter schwierigen Hörbedingungen zu ermöglichen, im Prinzip lösbar sein sollten. Aktuelle Forschungsergebnisse zeigen, dass es dazu jedoch nicht ausreicht, ausschließlich passives Hören als Modell für die Signalverarbeitung und die experimentellen Methoden zu verwenden. Vielmehr muss als Basis für die Forschungsansätze eine akustische Kommunikationsschleife verwendet werden, die sich aus dem Schallfeld mit seinen eingebetteten Schallquellen, dem Hörgerät und dem Nutzer bzw. der Nutzerin mit seiner bzw. ihrer Wahrnehmung und Aktivität als Reaktion auf die Umgebung und die Kommunikationsziele der Person ergibt. Diese Schleife, die im inneren Kreis von Abbildung 1 skizziert ist, bildet einen hochdynamischen Prozess, der aufgrund seiner Komplexität und fehlender Technologien bis vor einigen Jahren noch nicht umfassend untersucht werden konnte. Die hier vorgestellten Forschungsarbeiten nutzen aktuelle Technologien wie die audiovisuelle virtuelle Realität und Methoden des maschinellen Lernens zur quantitativen und modellgestützten Charakterisierung der akustischen Kommunikationsschleife, so dass die gewonnenen Erkenntnisse kontinuierlich in die akustischen Kommunikationsgeräte der Zukunft einfließen können. Diese sollen in der Lage sein, ihre Verarbeitung und Leistung zu optimieren, indem sie sich auf ein im Gerät eingebautes Modell der Kommunikationsschleife stützen („geräteinterne Modellschleife“). Zur Erreichung dieses anspruchsvollen Ziels ist eine Kombination aus Grundlagen- und angewandter Forschung notwendig. Konkret benötigt wird Grundlagenforschung im Bereich der Wahrnehmungsmodelle und Wahrnehmungsprinzipien der menschlichen Sprachverarbeitung (Closed-Loop-Modelle) sowie im Bereich der Schallerfassung, -verarbeitung und -verbesserung (Closed-Loop-Algorithmen). Darauf aufbauende angewandte Forschung widmet sich der Schallwiedergabe und den Bewertungsmethoden für Hörgeräte und elektroakustische Geräte (Closed-Loop-Geräte). Die folgenden Abschnitte stellen aktuelle Ansätze sowie beispielhafte Ergebnisse in diesen Forschungsbereichen vor.

Wahrnehmungsmodelle und Wahrnehmungsprinzipien der menschlichen Sprachverarbeitung (Closed-Loop-Modelle)

Die Grundprinzipien der Wahrnehmung akustischer Signale, beispielsweise Sprache oder Musik, lassen sich mittels Wahrnehmungsmodellen und entsprechenden „kritischen“ Experimenten tiefergehend erforschen. Um ein optimales modellbasiertes Verfahren für Closed-Loop-Modelle zu finden, testen wir verschiedene Modellansätze im Zusammenhang mit dem akustischen Kommunikationsgerät der

Zukunft. Hierbei sind vielfältige Wahrnehmungsaspekte und ihre Modelle wie binaurales Hören, Verständlichkeit von Sprache, Anstrengung beim Hören, Qualität der Audiodaten und Identifikation von Musikinstrumenten zu berücksichtigen.
Ein solcher Ansatz verfolgt die Entwicklung eines allgemeingültigen Modells, das unterschiedliche wahrnehmungsbezogene Experimente (Psychoakustik, Verständlichkeit von Sprache, Audio-Qualität) prognostiziert. Um ein internes Modell der Kommunikationsschleife in einem Gerät zu realisieren, wird die Entwicklung von echtzeitfähigen, nicht-intrusiven Wahrnehmungsmodellen benötigt, welche die Signalverarbeitung in einem Hörgerät unterstützen. Dabei wird ein Ansatz erforscht, der durch geschätzte binaurale Parameter Sprachverständlichkeit und Anstrengung beim Hören simuliert. Des Weiteren werden die Wahrnehmungsmodelle für modellbasierte Anpassungen von Hörgeräten unter Berücksichtigung des Sprachverstehens in akustisch komplexen Umgebungen genutzt.
Um eine hohe Akzeptanz von Hörgeräten wie Kopfhörern, Hearables oder Hörgeräte-Prototypen zu erreichen, ist es grundlegend, die natürliche Wahrnehmung akustischer Reize wie Sprache oder Musik mithilfe dieser Systeme zu verstehen. Hierfür führen wir sowohl experimentelle als auch numerische Simulationen durch, welche die Grundlage für akustische Transparenz mit Hörgeräten bilden.
Neben der Sprache nehmen wir auch die Wahrnehmung von Musik in Betracht, wofür wir perzeptive Experimente zur menschlichen Identifikationsleistung in einem Szenario eines Symphonieorchesters unter realistischen akustischen Bedingungen (wie simulierte Konzertsaalakustik) durchführen.
Insgesamt eröffnen die beschriebenen Methoden die Machbarkeit der Integration von perzeptiven Modellen in die Signalverarbeitung von Hörgeräten, was deren Nutzung als objektive Maße menschlicher Wahrnehmung in akustisch komplexen Szenen mit und ohne Hörversorgung ermöglicht. Im Folgenden werden drei Beispiele für diese Ansätze erläutert.

<u>Model in the Loop – Vorhersage des Sprachverstehens und der Höranstrengung in binauralen Situationen mit einem blinden Modell in Echtzeit</u>

Hintergrundgeräusche und Nachhall können die Sprachwahrnehmung stark beeinträchtigen. Das auditorische System des Menschen kann in solchen Situationen oft einen beträchtlichen Nutzen aus dem binauralen (beidohrigen) Hören ziehen, insbesondere wenn die Zielsprache und die störenden Signale aus unterschiedlichen Richtungen kommen. Liegt jedoch eine Schwerhörigkeit vor, ist das Sprachverstehen im Störgeräusch häufig beeinträchtigt bzw. anstrengend. Hier setzen Signalverarbeitungsalgorithmen zur Verbesserung des Sprachverstehens an. Für solche Algorithmen ist es sehr hilfreich, wenn sie über eine Schätzung der Sprachverständlichkeit bzw. der empfundenen Höranstrengung in der aktuellen Hörsituation verfügen, so dass sie ihre Signalverarbeitung entsprechend anpassen können.
Hierfür wurde ein Modellrahmen entwickelt, der sowohl die Sprachverständlichkeit als auch die subjektiv empfundene Höranstrengung basierend auf dem gestörten Sprachsignal und dem bekannten Hörverlust der Versuchsperson vorhersagt. Das Framework basiert auf einer Kombination aus einer blinden binauralen Verarbeitungsstufe [1], die einen Equalization-and-Cancellation-Mechanismus verwendet, und einem blinden Backend, das auf einem automatischen Spracherkenner basiert [2]. Weder das Frontend noch das Backend erfordern zusätzliche Informationen wie die Richtungen der Quelle, das Signal-Rausch-Verhältnis oder die Anzahl der Quellen, weshalb dieser Ansatz blind genannt wird und eine Vielzahl von Anwendungen ermöglicht.
Von dem blinden Modell wurde eine Echtzeitversion entwickelt und auf der Plattform des Master-Hearing Aids (MHA, [3]) implementiert, was die direkte Interaktion mit den anderen auf dieser Plattform laufenden Hörgerätealgorithmen ermöglicht, um die für die jeweilige Hörsituation optimale Verarbeitungsstrategie auszuwählen. In diesem Sinn befindet sich nun das Vorhersagemodell „in-the-loop", da es direkt in die Schleife aus akustischer Situation, Hörgerätealgorithmus und schwerhörender Person eingebunden ist.
Das blinde Modell wurde anhand verschiedener Datensätze validiert, bei denen Sprachverständlichkeit und wahrgenommene Höranstrengung für eine Reihe akustischer Bedingungen gemessen wurden, die sich in Nachhall und binauralen Hinweisen unterscheiden [4]. Die Vorhersagen des blinden Modells entsprechen weitgehend denen eines nicht-blinden Referenzmodells, d. h. eines Modells, das perfektes Wissen über die Zielsprache und das Störgeräusch hat und daher präzisere Vorhersagen macht, dafür allerdings nicht für unbekannte Situationen anwendbar ist. Das blinde Modell erklärt 94 % der Varianz für die Sprachverständlichkeit und das nicht-blinde Modell 98 %. Bei der Vorhersage der Höranstrengung zeigen beide Modelle eine geringere Vorhersagegenauigkeit, erklären jedoch immer noch signifikante Anteile der beobachteten Varianz (88 % und 71 % für das nicht-blinde bzw. blinde Modell).
Allerdings hat das gegenwärtige blinde Modell auch Limitationen, da die Annahme des Modells, dass Zielsprache und Störgeräusch sich hinsichtlich ihrer Amplitudenmodulationen unterscheiden, nicht zutrifft, wenn das störende Signal ebenfalls Sprache ist.

Aktuelle Arbeiten erweitern das Modell für störende Sprachsignale.

Objektive Maße zur Vorhersage von monauralen und binauralen Audioqualitätsaspekten in Hörsystemen

Neben Sprachverständlichkeit spielt die Sprach- und Audioqualität bei modernen Hörsystemen eine immer wichtigere Rolle. Ein Beispiel für solche Systeme sind Hearables, moderne Kopfhörer, die im Ohr getragen werden und eine Vielzahl unterschiedlicher Algorithmen verwenden, die auch in klassischen Hörgeräten zum Einsatz kommen. Hierzu zählen Rückkopplungsunterdrückung und Störgeräuschunterdrückung, um unerwünschte Schallquellen zu unterdrücken, aber auch Anpassungen des Klanges um akustische Transparenz zu erreichen, bei der Nutzer im Idealfall keinen Unterschied in der Wahrnehmung der akustischen Umgebung mit und ohne Hearables feststellen können. Solche Algorithmen können jedoch das ursprüngliche Signal so verändern, dass Nutzer monaurale oder binaurale (räumliche) Signalverzerrungen wahrnehmen und die Audioqualität des Signals vermindert wird.

Um den Einfluss solcher Verzerrungen auf die wahrgenommene Audioqualität objektiv und reproduzierbar zu untersuchen, werden Audioqualitätsmodelle entwickelt. Derartige Modelle können helfen, die durch die Signalverzerrung beeinträchtigten auditorischen Cues zu identifizieren und zukünftige Hörsysteme zu verbessern. Um zu untersuchen, welche monauralen und binauralen Audioqualitätsmodelle sich für die Vorhersage von typischen Signalverzerrungen in Hearables eignen, wurden verschiedene monaurale, binaurale und kombinierte monaural-binaural Modelle für drei Datenbanken mit unterschiedlichen Signalverarbeitungsalgorithmen getestet [5]. Die hier getesteten Algorithmen beeinträchtigten hauptsächlich die Abbildung monauraler, spektraler Cues, weshalb rein monaurale Audioqualitätsmodelle wie das Generalized Power Spectrum Model for Quality (GPSMq, [6]), das Natürlichkeitsmaß [7] oder das Hearing-Aid Speech Quality Index version 2 (HASQIv2; [8]) im Mittel die beste Vorhersageleistung erzielten. Der Einfluss binauraler Verzerrungen auf die Gesamtqualität war in den untersuchten Algorithmen deutlich geringer. Jedoch können andere Algorithmen, wie etwa eine binaurale Rauschreduktion, zu deutlich wahrnehmbaren binauralen Signalverzerrungen (z. B. einer Änderung der wahrgenommenen Position oder Quellenbreite) führen und somit einen starken Einfluss auf die wahrgenommene Audioqualität haben. Es ist daher sinnvoll, in Audioqualitätsmodellen sowohl monaurale als auch binaurale Qualitätsaspekte abzubilden. Ein Vertreter solcher kombinierten Audioqualitätsmodelle ist das Model for Combined Assessment of

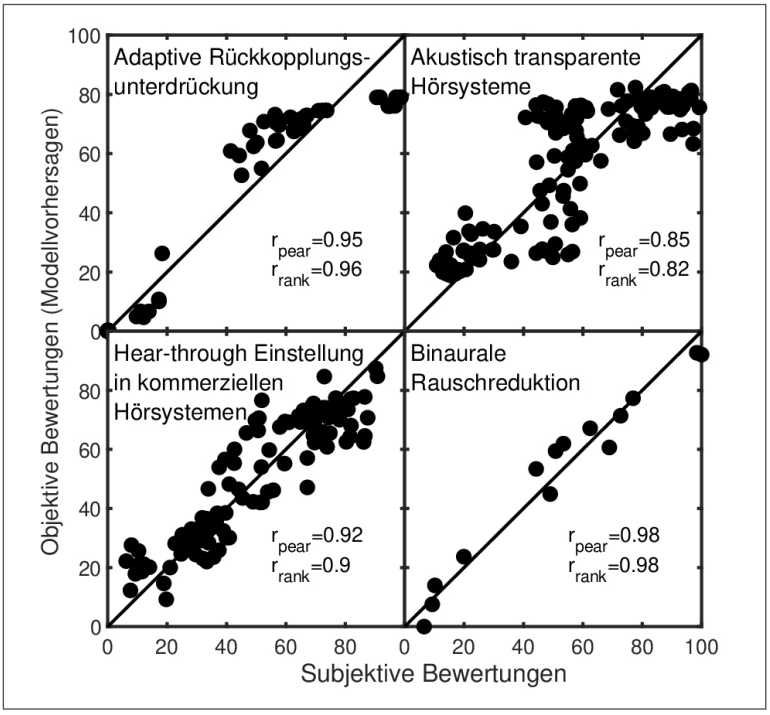

Abb. 2: Die vier Panele zeigen subjektive Audioqualitätsbewertungen und Modellvorhersagen des überarbeiteten MoBi-Qs für die in [5] verwendeten drei Datenbanken „Adaptive Rückkopplungsunterdrückung" [12], „Akustisch transparente Hörsysteme" [13], und „Hear-through in kommerziellen Hörsystemen" [14] sowie einer weiteren Datenbank mit Algorithmen zur binauralen Rauschreduktion [11].

Monaural and Binaural Audio Quality (MoBi-Q, [9]), das das monaurale GPSMq mit dem binauralen Qualitätsmodell Binaural Auditory Model for Audio Quality (BAM-Q, [10]) kombiniert, indem die Audioqualitätsvorhersagen beider Teilmodelle verglichen werden und die geringste Audioqualität die Gesamtqualität bestimmt. Verglichen mit den oben genannten monauralen Modellen, führt diese Teilmodellkombination im MoBi-Q in dieser Studie zu einer etwas geringeren Vorhersagegenauigkeit. Eine Überarbeitung der Teilmodellkombination, in der monaurale und binaurale Verzerrungen unterschiedlich gewichtet und additiv kombiniert werden, führt, wie Abbildung 2 zeigt, zu einer deutlich höheren Vorhersagegenauigkeit für die in dieser Studie verwendeten drei Datenbanken und eine weitere Datenbank mit Algorithmen zur binauralen Rauschreduktion [11]. Dies unterstreicht, dass zur Vorhersage von Gesamtaudioqualität einerseits relevante monaurale und binaurale Qualitätsaspekte abgebildet werden müssen und diese andererseits auch sinnvoll miteinander kombiniert werden müssen.

Tool für interaktive, d. h. vom Benutzer einstellbare Darstellung der Spracherkennungsleistung in alltäglichen Umgebungen

Hierbei wird die Spracherkennung in einer komplexen Szene dargestellt, und der Einfluss unterschied-

licher Faktoren auf die Spracherkennungsleistung kann analysiert werden [15]. Berücksichtigt werden Faktoren wie die Kopfrichtung des Zuhörers, Position und Anzahl von Störgeräuschquellen, Präsenz von Nachhall, individuelles Hörvermögen und Kompensation eines Hörverlustes mit einem Hörgerät.
Dabei wird eine Kette von drei verschiedenen Modellen verwendet. Hierbei handelt es sich um eine Kombination aus einem akustischen Raummodell, einem Hörgerätemodell und einem Hörermodell, die zur Generierung eines Datensatzes zu Demonstrationszwecken verwendet wird. Die akustische Umgebung ist durch eine Wohnzimmerszene mit drei Störgeräuschquellen repräsentiert: 1) Ein Fernseher, der eine Wettervorhersage wiedergibt, 2) ein informelles Gespräch, das aus einem über eine Tür verbundenen Nachbarraum stammt, und 3) ein Geschirrspüler, der im Nachbarraum in Betrieb ist. Die Szene wurde mit der Toolbox for Acoustic Scene Creation and Rendering (TASCAR) implementiert und gerendert [16].
Zur Vorhersage der Spracherkennungsleistung wurde das Simulation Framework for Auditory Discrimination Experiments (FADE) benutzt [17, 18]. FADE basiert auf einem Ansatz der automatischen Spracherkennung und sagt vorher, welcher Sprachpegel für eine bestimmte räumliche Anordnung des Sprechers und Zuhörers in der Szene benötigt wird, damit 50 % der Sprache korrekt erkannt wird. Diese Größe wird auch in empirischen Studien standardmäßig erfasst, so dass Vergleiche zwischen Modellvorhersagen und Messungen ermöglicht werden. Da FADE in der Lage ist, eine Schwerhörigkeit zu berücksichtigen [18], bietet das Tool auch die Möglichkeit, die Spracherkennungsleistung eines schwerhörenden Zuhörers in der Szene vorherzusagen. Außerdem sind durch die Anbindung des open Master Hearing Aid (openMHA, [19]) auch versorgte Spracherkennungsvorhersagen implementiert. Damit lässt sich der Erfolg eines Hörgerätes in akustisch komplexen Szenen vorhersagen.
Insgesamt bildet dieses Tool eine Basis, um Vorhersagen von unterschiedlichen Sprachverständlichkeitsmodellen in ökologisch relevanten Konditionen zu vergleichen und zu validieren.

Schallerfassung, -verarbeitung und -verbesserung (Closed-Loop-Algorithmen)

Ein Ziel der Entwicklung und Implementierung des Kommunikationsschleifenmodells ist es, die algorithmische Grundlage für die Audioverarbeitung und -verbesserung in Hörgeräten zu schaffen. Ein breites Spektrum von Methoden, einschließlich grundlegender und angewandter Methoden des maschinellen Lernens und wissensbasierter Signalverarbeitungsmethoden, wird derzeit eingesetzt, um insbesondere die Frage zu klären, wie die Wahrnehmung der Schallquelle(n), auf die sich der Benutzer / die Benutzerin gerade konzentriert („Hörwunsch"), verbessert werden kann. Ein wesentlicher Ansatz dazu ist die Kombination von Methoden der Computational Auditory Scene Analysis (CASA) mit Online-Biosignalanalyse (z. B. Kopf- und Augenbewegungen) und akustischer Analyse. Diese Kombination erlaubt grundsätzlich eine Schätzung des Hörwunsches und eine nachfolgende Verbesserung der im Aufmerksamkeitsfokus stehenden Schallquelle [20]. Zur Realisation der für diesen als immersives Hörgerät bezeichneten Ansatz benötigten Closed-Loop-Algorithmen bieten sich wissensbasierte CASA-Algorithmen an, die Wissen aus der auditorischen Verarbeitung und der Mehrkanalsignalverarbeitung berücksichtigen und mit Methoden des maschinellen Lernens (ML) kombinieren. Da CASA ein weites Feld mit mehreren konkurrierenden und bislang unausgereiften Ansätzen der Signalverarbeitung und des maschinellen Lernens ist, bemühen sich aktuelle Forschungsarbeiten um die vergleichende Entwicklung und Bewertung dieser Ansätze. Ein Beispiel dafür sind hierarchische Strukturen unter Verwendung von U-Netzen und Variations-Auto-Encodern zur Darstellung akustischer Szenen entlang einer Hierarchie von Repräsentationen verschiedenen Abstraktionsniveaus. Die Anwendung von U-Netzen auf akustische Daten ermöglicht es, die probabilistische Natur der auditiven Komponenten besser zu repräsentieren. Insgesamt zeigen die beschriebenen Ansätze die Machbarkeit von Closed-Loop-Hörgeräten, die eine auditive Szenenanalyse beinhalten und Sensordaten nutzen, um den Aufmerksamkeitsfokus des Nutzers bzw. der Nutzerin zu schätzen und die Signalverarbeitung entsprechend anzupassen. Für deren Evaluierung bieten sich Messparadigmen mit interaktiver Telepräsenz in einer audiovisuellen virtuellen Umgebung an [21].

Immersives Hörgerät

Das menschliche Gehör ermöglicht es, uns in Kommunikationssituationen unter schwierigen Bedingungen mit mehreren gewünschten Schallquellen und Hintergrundgeräuschen aktiv auf eine bestimmte Schallquelle zu konzentrieren, die wir hören wollen, während alle anderen Schallquellen in den Hintergrund rücken. Dieser Aufmerksamkeitsfokus kann willentlich auf eine andere Schallquelle verlagert werden, z. B. bei einem Gespräch an einem Tisch mit mehreren gleichzeitig sprechenden Personen. Zur Unterstützung von Hörsystemträgerinnen und -trägern bei der Fokussierung ihrer Aufmerksamkeit auf eine bestimmte Schallquelle muss das Hörsystem wissen, auf welche der gerade aktiven Schallquellen sich die Aufmerksamkeit richtet, um diese selektiv verstärken zu können. Das Problem besteht darin, dass der Auf-

merksamkeitsfokus nicht einfach aus dem akustischen Signal allein abgeleitet werden kann.

Das immersive Hörgerät ist ein neues Hörgerätekonzept, das derzeit in mehreren Labors weltweit untersucht wird und in Zukunft das Problem der selektiven Aufmerksamkeit lösen könnte. Die Grundidee besteht darin, das Verhalten und die Aktivität des Nutzers bzw. der Nutzerin mit Hilfe eines multimodalen Signalverarbeitungsansatzes zu messen, wie im Blockdiagramm in Abbildung 3 dargestellt. Zu diesem Zweck werden verschiedene Sensoren am Hörgerät angebracht (linke Spalte der Verarbeitungsblöcke in Abbildung 3), insbesondere ein Bewegungssensor (IMU) zur Messung der Kopfbewegung, eine Reihe von Elektroden in der Nähe des Ohrs oder im Gehörgang zur Messung der Blickrichtung und möglicherweise mehrere weitere Elektroden um das Ohr herum (cEEGrid) zur Messung der Gehirnaktivität (EEG). Die akustische Szene wird anhand der Mikrofonsignale analysiert, um das Vorhandensein, die Aktivität und die räumliche Position von Schallquellen zu bestimmen (rechte Spalte der Verarbeitungsblöcke). Die Szenenanalyse kann durch visuelle Daten einer kleinen Kamera, die am Hörgerät oder an einer Brille angebracht ist, verbessert werden, z. B. durch die Erkennung von Mund-/Lippenbewegungen [22]. Eine Entscheidungseinheit kombiniert dann die akustischen Informationen über die Hörsituation mit den Sensordaten, um für den Nutzer bzw. die Nutzerin wichtige von unwichtigen Quellen zu trennen. Wichtige Quellen können dann selektiv verstärkt und unwichtige abgeschwächt werden. Die Entscheidungseinheit verwendet Techniken des maschinellen Lernens, um die Beziehung zwischen der akustischen Szene und den Sensordaten aus großen Mengen von Testdaten zu erlernen und so eine hohe Genauigkeit bei der Schätzung der Aufmerksamkeitsquelle zu erreichen. Schließlich verbessert ein Signalverbesserungsblock das erkannte Signal und präsentiert es dem Benutzer. Derzeit werden Signalverbesserungsverfahren entwickelt, die auf maschinellem Lernen basieren (z. B. [23], [24]) und eine bessere Signalverbesserung und Klangqualität als herkömmliche Verfahren ermöglichen. Auch Richtmikrofone mit mehreren simultanen Keulen bieten Vorteile gegenüber konventionellen steuerbaren Richtmikrofonen, da sie eine artefaktärmere Signalverarbeitung ermöglichen und sich gut in das Konzept des immersiven Hörgeräts integrieren lassen. Alle Verarbeitungsblöcke des in Abbildung 3 dargestellten Konzepts werden noch erforscht, aber erste vollständige Systeme wurden bereits demonstriert, z. B. blickbasierte Aufmerksamkeitssteuerung ([20]) und neurogesteuerte Hörgeräte ([25], [26]).

Bei der Entwicklung verhaltensgesteuerter Hörsysteme kommt der Evaluationsumgebung eine besondere

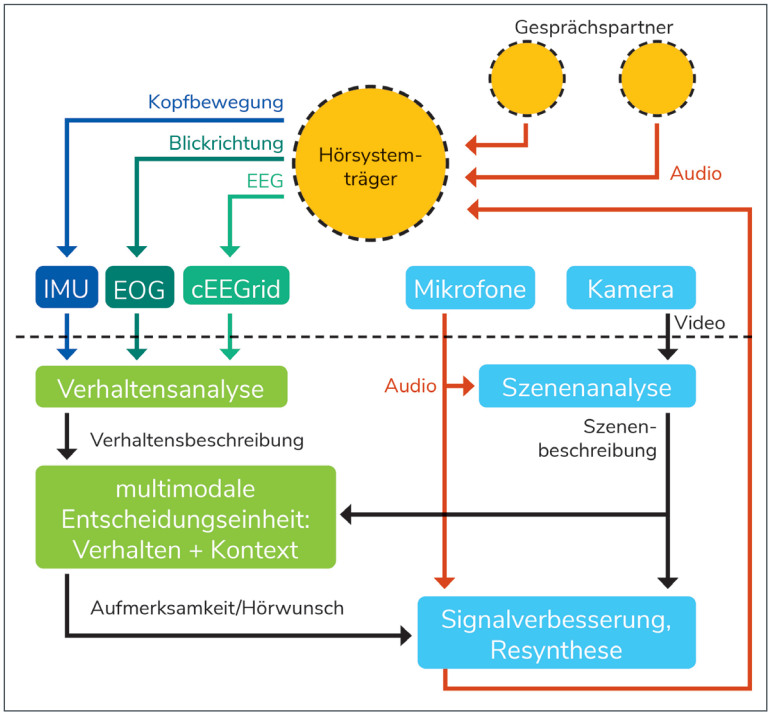

Abb. 3: Blockdiagramm eines verhaltensgesteuerten Hörsystems. IMU: inertiale Messeinheit (inertial measurement unit), EOG: Elektrookulographie, EEG: Elektroenzephalografie, cEEGrid: c-förmiges Elektrodenarray zur Messung des EEG

Bedeutung zu. Sie muss so gestaltet sein, dass die Testpersonen darin ein Bewegungs- und Kommunikationsverhalten zeigen, welches sie auch in entsprechenden natürlichen Kommunikationssituationen zeigen würden. Eine Voraussetzung dafür ist, dass sowohl die akustischen als auch die visuellen Eigenschaften den natürlichen Umgebungen entsprechen. Dies wird durch die Nachbildung typischer Kommunikationssituationen in der virtuellen Realität unterstützt ([27], [28]). Darüber hinaus müssen die in den Evaluationsumgebungen verwendeten Messparadigmen so gestaltet sein, dass typische Verhaltensmuster auftreten. Dies kann z. B. durch die Verwendung freier Kommunikation anstelle von Sprachtests erreicht werden, siehe Abbildung 4 ([29]). Durch die Verwendung virtueller Realität kann die Reproduzierbarkeit und die Kontrolle über die akustische Szene erhöht und gleichzeitig ökologisch valide Testparadigmen realisiert werden.

Einkanalige Signalverbesserung mit U-Netzwerken

Einen Schwerpunkt im methodisch-theoretischen Bereich der Closed-Loop-Algorithmen bildet die Forschung an tiefen neuronalen Netzen für Anwendungen der ein- und mehrkanaligen Audiosignalverarbeitung. Die Fortschritte der Audiosignalverarbeitung in den letzten Jahren waren beträchtlich und sind zu großen Teilen eine Folge der Verschmelzung der ehemals getrennten Disziplinen der Signalverarbeitung

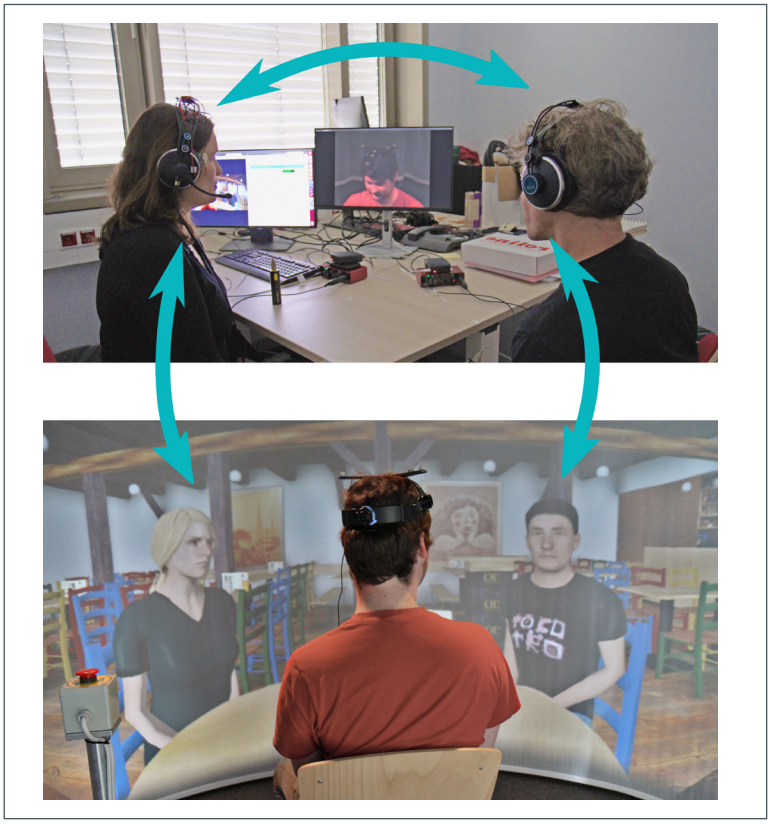

Abb. 4: Interaktive Kommunikation in virtueller Realität mittels Telepräsenz

und des maschinellen Lernens. Etwa bei der Separation mehrerer akustischer Objekte in einer Signalmischung – wichtig für Spracherkennung, Hörgeräte und Videokonferenzen – können nun auch bei nur einem einzigen Mikrofonkanal reale Hintergrundgeräusche zuverlässig von Sprache getrennt werden.

Dabei widersetzte sich dieses Problem lange einer wirksamen Lösung: In der Spektrogrammdarstellung einer akustischen Szene mit zeitlicher und spektraler Auflösung führt die akustische Überlagerung mehrerer Schallquellen zu einer additiven Superposition aller komplexwertigen Fourieranteile der verschiedenen Quellen. Im Gegensatz zur visuellen Modalität, in der Objekte zwar teilweise verdeckt aber immerhin räumlich getrennt sind, lassen sich in der Akustik einzelne Zeit-Frequenzpunkte nicht eindeutig einer Schallquelle zuordnen.

Klassische Ansätze, etwa basierend auf Wiener-Filtern, berechnen Gewichtungsfaktoren, die Bereiche der Zeit-Frequenzebene teils einer zu extrahierenden Signalquelle, teils unerwünschtem Hintergrundsignal zuordnen.

Besonders für dynamisch veränderliche Störquellen erweisen sich aktuell Signalverarbeitungsarchitekturen, die auf Methoden des maschinellen Lernens basieren, als besser geeignet. Reale akustische Szenen enthalten oft dynamische Störquellen, etwa beschleunigende Fahrzeuge, Vogelzwitschern oder auch Hintergrundsprache von nicht attendierten Sprechern. Tiefe Neuronale Netze lernen während einer Trainingsphase die akustischen Strukturen von Sprachsignalen und unerwünschten Störquellen. Dazu werden dem Netz mehrfach hunderte oder tausende Stunden akustischen Materials präsentiert, deren statistische Eigenschaften in gelernte Verbindungen zwischen Millionen von neuronalen Verarbeitungseinheiten eingehen und diese iterativ in Richtung optimaler Werte anpassen.

Somit erlernt das Netzwerk die statistische Struktur akustischer Szenen und der verschiedenen in ihnen vorkommenden akustischen Objekte. Ein dem trainierten Netzwerk neu präsentiertes Audiosignal kann danach, ganz analog der klassischen Theorie der akustischen Szenenanalyse, automatisch analysiert und entsprechend dem jeweils gewünschten Ziel modifiziert werden.

In der Architektur der U-Netzwerke [30] werden in einer modularen Struktur von tiefen Faltungsnetzen die Schritte der Signalanalyse und der Signalrekonstruktion verbunden, so dass sie sich zur Untersuchung der automatischen akustischen Szenenanalyse und Sprachsignalverbesserung besonders eignen. U-Netzwerke kodieren ein Eingangssignal in einem Enkodierungszweig über mehrere hierarchische Verarbeitungsstufen in eine niedrigdimensionale Darstellung, ein Prozess, der als eine Informationsreduktion motiviert werden kann, bei der aber relevante Information über die Struktur der vorkommenden akustischen Objekte erhalten bleibt.

Um hieraus ein verbessertes Audiosignal hoher Qualität rekonstruieren zu können, wird es in einem zweiten Signalzweig, dem Dekodierzweig, ebenfalls über hierarchische Verarbeitungsstufen wieder auf die Zeit-Frequenz-Auflösung des Eingangssignals expandiert. Dabei steht dem Netzwerk auf jeder Hierarchiestufe durch laterale Verbindungen die gelernte Enkodierungsrepräsentation auf der jeweiligen Stufe zur Verfügung (siehe Abbildung 5).

Die in [31] entwickelten U-Netzwerke zeichnen sich dadurch aus, dass sie zur Signalrekonstruktion die statistische Verteilung der gelernten Repräsentation abtasten („variational inference“ [32]), anstatt eine rein deterministische Berechnung zu nutzen.

Im Ergebnis lassen sich so hohe Signalverbesserungen gerade auch für Sprache in realen, stark fluktuierenden Hintergrundgeräuschen erzielen. Dort beträgt, je nach verwendetem Datensatz, die Signalverbesserung typischerweise zwischen 10 dB und 20 dB SI-SDR (Skalen-invarianter Signal-Verzerrungs-Abstand) [31].

Von besonderem Interesse ist ein Verständnis der in einem solchen Netzwerk gelernten akustischen Repräsentationen. Dies kann dadurch erzielt werden, dass die Struktur des trainierten U-Netzwerkes mit Methoden der erklärbaren KI („explainable artificial

intelligence“) untersucht wird, wie eine erste Analyse für Audiosignale in [33] zeigt.
Die Kombination klassischer Signalverarbeitung mit Methoden des maschinellen Lernens hat zum Feld der Audiosignalverarbeitung und -analyse neue Algorithmen beigetragen, die aktuell und für die Zukunft den Stand des technisch Machbaren kontinuierlich verbessern und neue Anwendungen eröffnen werden.

Schallwiedergabe und Bewertungsmethoden für Hörgeräte und elektroakustische Geräte (Closed-Loop-Geräte)

Interaktive audiovisuelle Szenen in virtueller Realität bieten neue Möglichkeiten für eine ökologisch valide Hörforschung sowie zur Bewertung des realen Nutzens von Algorithmen und Hörsystemen. Ziel ist es, ein tieferes Verständnis der perzeptiven Unterschiede zwischen virtuellen akustischen Umgebungen und ihren realen Pendants zu gewinnen. Darauf aufbauend werden neue, interaktive subject-in-the-loop-Bewertungsmethoden entwickelt, bei denen die Person („subject“) aktiv in den Bewertungsprozess eingebunden wird, um den individuellen Nutzen von Hörsystemen oder elektroakustischen Geräten zu beurteilen. Ein besonderer Schwerpunkt liegt dabei auf der ökologischen Validität von Untersuchungsmethoden im Labor, die darauf abzielen, den Alltag von Hörgeschädigten so realitätsnah wie möglich abzubilden.
Darüber hinaus wird derzeit an Verfahren zur Schallwiedergabe in akustisch nicht optimalen Umgebungen gearbeitet, wie beispielsweise in halligen Räumen. Ziel ist es, die negativen Einflüsse der Raumakustik zu minimieren und das Schallfeld so abzubilden, dass es perzeptiv möglichst authentisch wirkt. Hierbei handelt es sich um eine Art Raumtransformation, bei der der Wiedergaberaum akustisch in den ursprünglichen Aufnahmeraum transformiert wird, um eine originalgetreue Klangwiedergabe zu erreichen.
Des Weiteren werden neuartige ohrnahe Hörsysteme erforscht, die eine akustisch transparente Sprachkommunikation durch den Einsatz eines aktiven Ohrpassstücks mit integrierten Mikrofonen und Receivern ermöglichen sollen. Dies erfordert individualisierte Lösungen zur Schalldruckentzerrung, zur Rückkopplungsunterdrückung sowie zur aktiven Reduktion von Störgeräusch und Okklusionseffekten.

<u>Virtuelle akustische Umgebungen für ökologisch valide Hörforschung</u>

Die Verwendung von synthetischen Stimuli in abstrakten Hörsituationen hat eine lange und erfolgreiche Geschichte in der Hörforschung. Es besteht ein zunehmendes Interesse daran, die verbleibende Lücke zur Hörumgebung im wirklichen Leben zu schließen, indem Situationen mit hoher ökologischer Validität

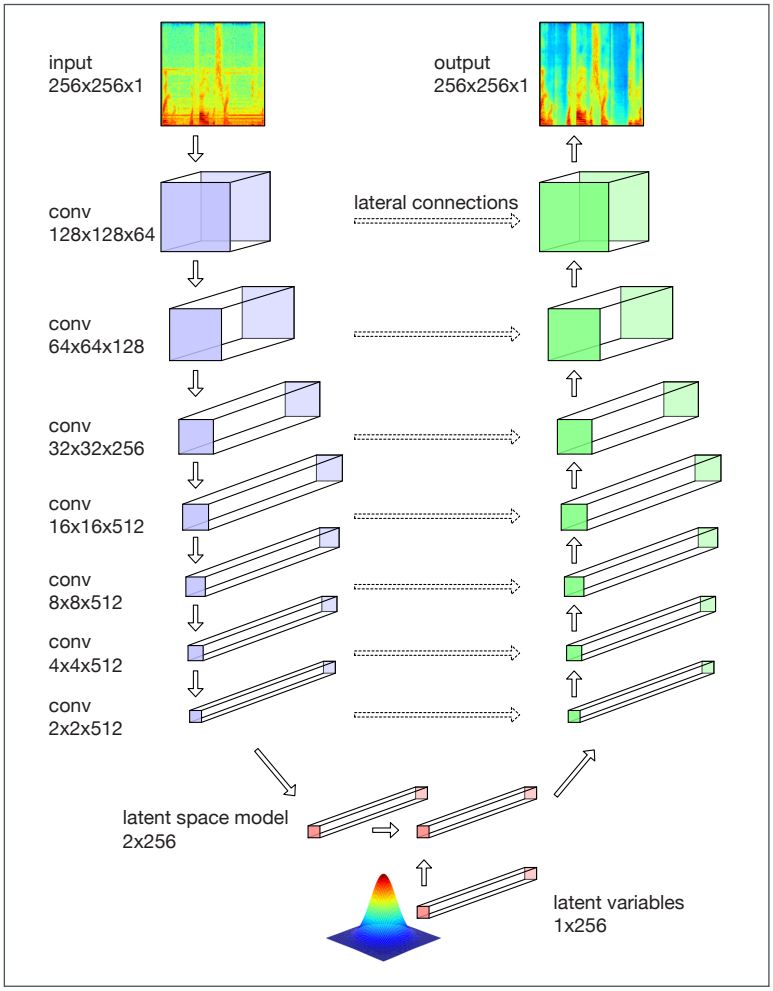

Abb. 5: Schematische Darstellung eines tiefen U-Netzwerkes zur Verbesserung von Sprachschallquellen. Die Signalanalyse wird durch den Enkodierzweig (links) mit einer Hierarchie von Faltungstransformationen durchgeführt. Der Dekodierzweig (rechts) rekonstruiert das verbesserte Sprachsignal mit Hilfe der erlernten latenten Repräsentation und lateraler Verbindungen.

im Labor repliziert werden. Dies ist wichtig, um die zugrunde liegenden auditiven Mechanismen und ihre Relevanz in realen Situationen zu verstehen sowie immer anspruchsvollere Algorithmen für Hörgeräte zu entwickeln und zu bewerten. Eine Reihe von „klassischen“ Stimuli und Paradigmen haben sich in der Psychoakustik zu De-facto-Standards entwickelt, die einfach sind und leicht in Laboren reproduziert werden können. Obwohl sie idealerweise Vergleiche zwischen Laboren und reproduzierbare Forschung ermöglichen, fehlt ihnen die akustische Stimuluskomplexität und die Verfügbarkeit von visuellen Informationen, wie sie in der Kommunikation und Hörsituationen des täglichen Lebens beobachtet werden.
Kürzlich wurde ein erweiterbarer Satz komplexer audiovisueller Szenen für die Hörforschung bereitgestellt und etabliert, der ökologisch valide Tests in realistischen Szenen ermöglicht und gleichzeitig die Reproduzierbarkeit und Vergleichbarkeit wis-

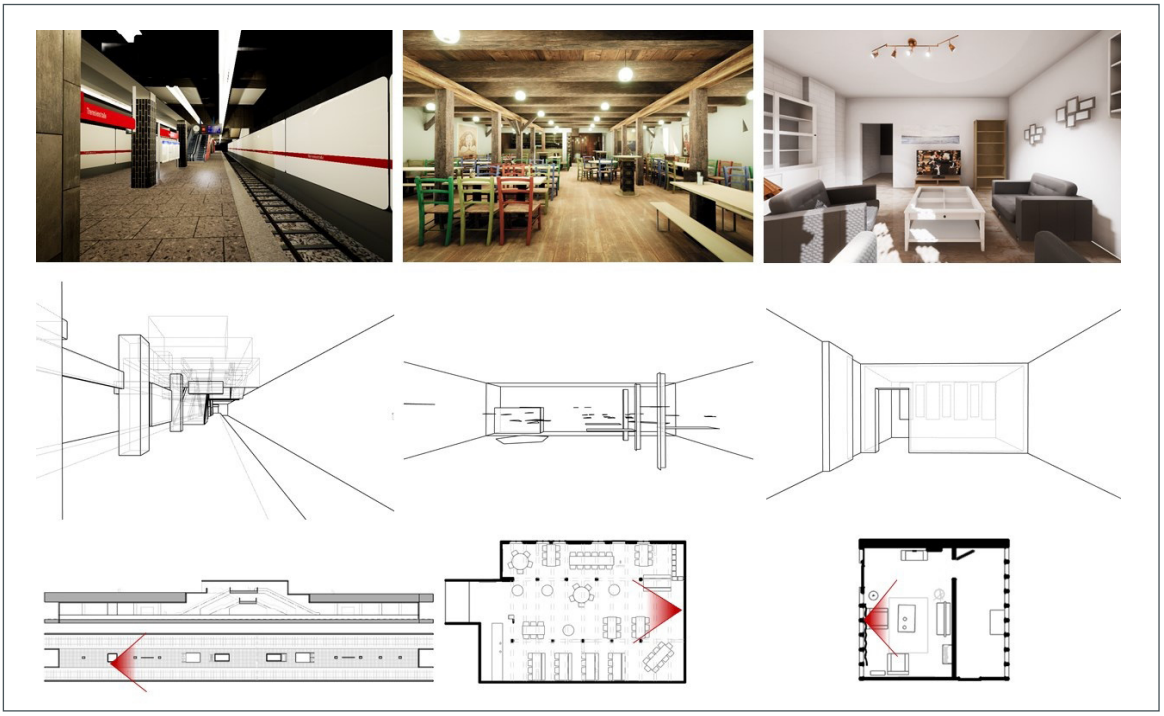

Abb. 6: Open-source audio-visuelle Umgebungen und Szenen: U-Bahn-Station (links), OL's Pub (Mitte), Living Room Lab (rechts). Die obere Zeile zeigt ein visuelles Rendering, die mittlere Zeile das akustische Geometriemodell und die untere Zeile eine Aufsicht mit Verteilung der Quellen und Empfänger inkl. Orientierung (rot). Aus [28].

senschaftlicher Ergebnisse unterstützt. Drei Umgebungen (siehe Abbildung 6) mit jeweils zwei Szenen mit unterschiedlichen akustischen Simulations- und Höranforderungen wurden auf Basis von Messungen in realen Umgebungen entwickelt [28]:

1) U-Bahn Station Theresienstraße in München mit sehr großem Volumen und gekoppelten akustischen Räumen,
2) OL's pub in Oldenburg mit vielen Oberflächen und lokalen und entfernten Quellen,
3) Living Room Lab (LIROLA) an der Universität Oldenburg mit einem gekoppelten Küchenraum für Kommunikationssituationen ohne Direktschall.

Die als „open-source" auf Zenodo zur Verfügung gestellten Szenen bestehen zum einen aus der Definition der visuellen und der akustischen Umgebung und zum anderen aus der Definition von jeweils zwei Kommunikationsszenen in diesen Umgebungen. Die Szenendefinitionen schreiben Quellen und Hörerpositionen und Orientierungen in audiologischen Standardsituationen und in für den jeweiligen Raum relevanten Situationen vor. Für diese liegen zur Verifikation von raumakustischen Simulationen auch Impulsantwortmessungen vor.

In [34] wurde die Sprachverständlichkeit für Normalhörende im realen LIROLA mit verschiedenen simulierten akustischen Reproduktionen verglichen: Binaurale Kopfhörerwiedergabe, ein 4-Kanal horizontales Lautsprecherarray in einer Hörkabine (Positionen bei 45°, 135°, 225°, 315°, Radius 1 m) sowie ein sphärisches 86-Kanal Lautsprecherarray in einem reflexionsarmen Raum. Als Referenz dienten Lautsprecher in vier unterschiedlichen Positionen im LIROLA, wobei die Versuchsperson auf dem Sofa saß. In allen Messräumen wurden die Sprachverständlichkeitsmessungen auch über Kopfhörer mit der simulierten Szene durchgeführt. Die Simulationen der Raumimpulsantworten und das Lautsprecherrendering erfolgten mit dem Raumakustiksimulator RAZR [35,36]. Die Ergebnisse verdeutlichten, dass die Sprachverständlichkeit, wie erwartet, stark von der Position des Zielsprechers abhängt – so war Sprache aus der Küche signifikant schlechter verständlich als aus dem Wohnzimmer. Besonders für Hörgeschädigte sind diese Ergebnisse relevant, da Sprache aus angrenzenden Räumen oft als sehr herausfordernd empfunden wird. Interessanterweise zeigten sich kaum Unterschiede zwischen den verschiedenen Mess-Setups, was in der klinischen Praxis ermutigend ist, da oft keine spezialisierten Messräume, wie das LIROLA oder ein reflexionsarmer Raum, verfügbar sind. Allerdings basierten die Messungen auf relativ einfachen, rein akustischen Szenarien. Zukünftige Studien sollen die Grenzen der ökologischen Validität in simulierten Umgebungen mit komplexeren Szenarien weiter erforschen.

Akustisches Ohrpassstück (Hearpiece)

Zukünftige Hörsysteme, die im Ohr getragen werden – wie In-Ear-Kopfhörer, Hörgeräte und Hearables –

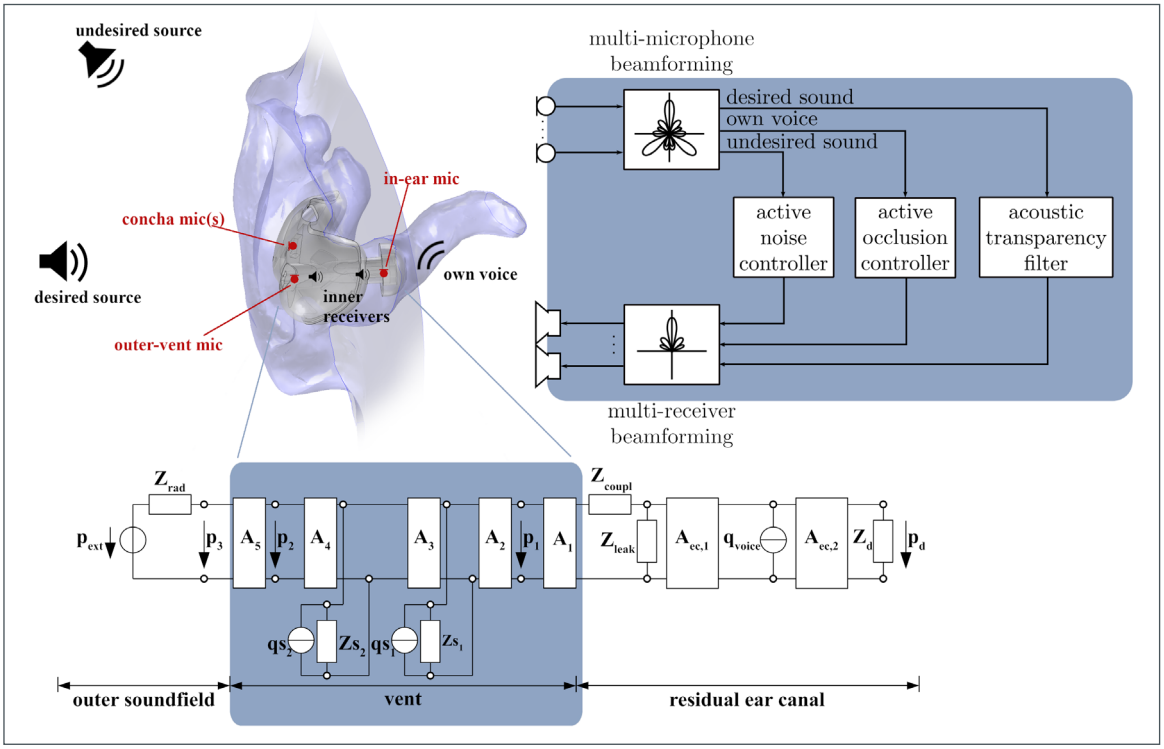

Abb. 7: Im-Ohr-Hörsystem mit mehreren internen Lautsprechern ("receivers") und Mikrofonen (in rot) mit externen und internen Schallquellen. Rechts oben: multiple-input-multiple-output (MIMO) Signalverarbeitungsschema, links unten: elektroakustisches Modell

werden nicht nur über Miniaturlautsprecher verfügen, sondern auch Miniaturmikrofone beinhalten. Dies führt zu der wichtigen Frage, wie alle vorhandenen Mikrofone und Lautsprecher optimal genutzt werden können. Abbildung 7 veranschaulicht das betrachtete Szenario, das externe und interne Schallquellen umfasst, insbesondere die Miniaturlautsprecher („receiver") des Hörsystems und die Körperschallkomponente der eigenen Stimme („own voice"). Das Ziel besteht darin, kombinierte Lösungen für Herausforderungen, wie individualisierte Schalldruckentzerrung am Trommelfell, Unterdrückung von Rückkopplungen sowie aktive Reduktion von Störgeräuschen und Okklusionseffekten zu entwickeln. Um diese Ziele zu erreichen, sind sowohl Multiple-Input-Multiple-Output-Signalverarbeitungsalgorithmen (dargestellt im Kasten oben rechts) als auch elektroakustische Modelle (im Kasten unten links) erforderlich.

Zur Schalldruckentzerrung wurden in [37] robuste Transparenzfilter mit ein und zwei Lautsprechern des Im-Ohr-Hörsystems vorgeschlagen, die zunächst auf gemessenen akustischen Übertragungsfunktionen bei 20 Versuchspersonen beruhten. In Hörversuchen wurden diese Transparenzfilter als mindestens gleichwertig mit dem Transparenzmodus kommerzieller In-Ear-Hörsysteme bewertet [14]. Parallel dazu wurde ein elektroakustisches Modell für einen Ohrpassstück-Prototyp, das sogenannte Hearpiece [38], entwickelt. Dieses Modell ermöglicht es nicht nur, das elektroakustische Verhalten des Prototyps zu beschreiben, sondern auch die akustische Eingangsimpedanz des individuellen Gehörgangs im eingesetzten Zustand zu messen. Diese Messung kann verwendet werden, um ein Scheibchenmodell des individuellen Gehörgangs anzupassen und somit den Schalldruck am individuellen Trommelfell vorherzusagen [39]. Dies erlaubt die Realisierung von individualisierten Transparenzfiltern, die im Vergleich zu Filtern, die auf mittleren Übertragungsfunktionen basieren, zu einer verbesserten Qualitätsbeurteilung führen.

Im Bereich der Rückkopplungsunterdrückung wurden in [40] neuartige Verfahren vorgeschlagen, die auf Beamforming basieren und die Mikrofonsignale so verarbeiten, dass eine räumliche ‚Null' in Richtung des Lautsprechers erzeugt wird. Durch diesen Ansatz lassen sich auf robuste Weise erhebliche Verbesserungen in der Unterdrückung von Rückkopplungen erzielen, ohne die Sprachqualität zu beeinträchtigen. Zur aktiven Unterdrückung externer Störquellen wurde in [41] ein Verfahren mit einem virtuellen Mikrofon am Trommelfell und mehreren Kontroll-Lautsprechern vorgeschlagen. Durch diesen Ansatz wird nicht nur die Bandbreite der aktiven Störgeräuschunterdrückung erweitert, sondern auch die Tiefe der Unterdrückung erhöht. Des Weiteren wurde in [42] ein neuartiges Verfahren entwickelt, das es erlaubt, externe Störquellen aus allen Richtungen außer aus einer bestimmten Richtung aktiv zu unterdrücken.

Das Im-Ohr-Mikrofon des Ohrpassstücks kann des Weiteren genutzt werden, um die eigene Stimme des Nutzers aufzunehmen, während dieser in einer lauten Umgebung spricht (z. B. zur Übertragung zu einem Mobiltelefon oder einem anderen Hörsystem). Da Aufnahmen der eigenen Stimme im Ohr jedoch stark bandbegrenzt sind, ist ein System zur Rekonstruktion der Breitbandsprache aus dem Im-Ohr-Mikrofonsignal erforderlich. In [43] wurde hierfür ein Deep Learning-basiertes Verfahren entwickelt, das Bandbreitenerweiterung, Entzerrung und Geräuschreduktion kombiniert. Für das Training dieses Systems, das vorzugsweise sehr viele Im-Ohr-Aufnahmen benötigt, wurde in [44] ein phonemabhängiges Modell zur Simulation dieser Aufnahmen vorgeschlagen.

Zum besseren Verständnis des Einflusses individueller anatomischer Gegebenheiten wurde eine Datenbank mit dreidimensionalen Anatomien von Oberkörper, Kopf, Ohrmuschel (Pinna) und Gehörgang bis zum Trommelfell erhoben und auf Zenodo zur Verfügung gestellt [45]. Dies soll die Entwicklung verbesserter Prototypen von aktiven, offenen Im-Ohr-Hörsystemen unterstützen.

Folgerungen

Die systematische Einbeziehung des Modells der Kommunikationsschleife hat zu neuartigen Forschungsmethoden mit alltagsrelevanten Ergebnissen geführt, wie etwa Echtzeit-Modellen für Sprachverstehen und Höranstrengung in realen Situationen, immersiven Closed-Loop-Hörgeräten mit Algorithmen des maschinellen Lernens, akustisch transparenten Hearables mit Hörunterstützung sowie interaktiven audiovisuellen Umgebungen mit hoher ökologischer Validität für Forschung und Anwendungen. Durch Einbeziehung weiterer experimenteller Methoden etwa aus der Kognitionsforschung sowie die Erweiterung auf weitere Forschungsfelder, wie z. B. die Mensch-Maschine-Kommunikation, eröffnen die vorgestellten Closed-Loop-Methoden viele spannende Möglichkeiten für die weitere Forschung.

Danksagung

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 352015383 – SFB 1330.

Literatur

[1] Hauth, C. F.; Berning, S. C.; Kollmeier, B.; Brand, T.: Modeling binaural unmasking of speech using a blind binaural processing stage. Trends in Hearing 24, 2020. *https://doi.org/10.1177/2331216520975630*

[2] Huber, R.; Krüger, M.; Meyer, B. T.: Single-ended prediction of listening effort using deep neural networks. Hearing Research 359, pp. 40–49, 2018. *https://doi.org/10.1016/j.heares.2017.12.014*

[3] Kayser, H.; Herzke, T.; Maanen, P.; Zimmermann, M.; Grimm, G.; Hohmann, V.: Open community platform for hearing aid algorithm research: open Master Hearing Aid (openMHA). SoftwareX 17, 100953, 2022. *https://doi.org/10.1016/j.softx.2021.100953*

[4] Rennies, J.; Röttges, S.; Huber, R.; Hauth, C. F.; Brand, T.: A joint framework for blind prediction of binaural speech intelligibility and perceived listening effort. Hearing Research 426, 108598, 2022. *https://doi.org/10.1016/j.heares.2022.108598*

[5] Biberger, T.; Schepker, H.; Denk, F.; Ewert, S. D.: Instrumental quality predictions and analysis of auditory cues for algorithms in modern headphone technology. Trends in Hearing 25, 2021. *https://doi:10.1177/23312165211001219*

[6] Biberger, T.; Fleßner, J.-H.; Huber, R.; Ewert, S. D.: An objective audio quality measure based on power and envelope power cues. Journal of the Audio Engineering Society 66(7/8), pp. 578–593, 2018. *https://doi.org/10.17743/jaes.2018.0031*

[7] Moore, B. C. J.; Tan, C.-T.: Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. Journal of the Audio Engineering Society 52(9), pp. 900–914, 2004. *https://www.aes.org/e-lib/browse.cfm?elib=13018*

[8] Kates, J. M.; Arehart, K. H.: The Hearing-Aid Speech Quality Index (HASQI) version 2. Journal of the Audio Engineering Society 62(3), pp. 99–117, 2014. *https://doi.org/10.17743/jaes.2014.0006*

[9] Fleßner, J.-H.; Biberger, T.; Ewert, S. D.: Subjective and objective assessment of monaural and binaural aspects of audio quality. IEEE Transactions on Audio, Speech and Language Processing 27(7), pp. 1 112–1 125, 2019. *https://doi.org/10.1109/TASLP.2019.2904850*

[10] Fleßner, J.-H.; Huber, R.; Ewert, S. D.: Assessment and prediction of binaural aspects of audio quality. Journal of the Audio Engineering Society 65(11), pp. 929–942, 2017. *https://doi.org/10.17743/jaes.2017.0037*

[11] Gößling, N.; Marquardt, D.;. Doclo S.: Perceptual evaluation of binaural MVDR-based algorithms to preserve the interaural coherence of diffuse noise fields. Trends in Hearing 24, 2020. *https://doi: 10.1177/2331216520919573*

[12] Nordholm, S.; Schepker, H.; Tran, L. T. T.; Doclo, S.: Stability-controlled hybrid adaptive feedback cancellation scheme for hearing aids. Journal of the Acoustical Society of America 143(1), pp. 150–166, 2018. *https://doi.org/10.1121/1.5020269*

[13] Schepker, H.; Denk, F.; Kollmeier, B.; Doclo, S.: Subjective sound quality evaluation of an acoustically transparent hearing device. In: Proc. 2nd AES Conference on Headphone Technology, San Francisco, USA, 2019. *https://www.aes.org/e-lib/browse.cfm?elib=20517*

[14] Schepker, H.; Denk, F.; Kollmeier, B.; Doclo, S.: Acoustic transparency in hearables – Perceptual sound quality evaluations. Journal of the Audio Engineering Society 68(7/8), pp. 495–507, 2020. *https://doi.org/10.17743/jaes.2020.0045*

[15] Schädler, M. R.: Interactive spatial speech recognition maps based on simulated speech recognition experiments. Acta Acustica united with Acustica 6:31, 2022. *https://doi.org/10.1051/aacus/2022028*

[16] Grimm, G.; Luberadzka, J.; Hohmann, V.: A toolbox for rendering virtual acoustic environ-ments in the context of audiology. Acta Acustica united with Acustica 105(3), pp. 566–578, 2019. *https://doi.org/10.3813/AAA.919337*

[17] Schädler, M. R.; Warzybok, A.; Ewert, S. D.; Kollmeier, B.: A simulation framework for auditory discrimination experiments: Revealing the importance of across-frequency processing in speech perception. The Journal of the Acoustical Society of America 139(5), pp. 2 708–2 722, 2016. *https://doi.org/10.1121/1.4948772*

[18] Schädler, M. R.; Hülsmeier, D.; Warzybok, A.; Kollmeier,

B.: Individual aided speech-recognition performance and predictions of benefit for listeners with impaired hearing employing FADE. Trends in Hearing 24, 2020. https://doi.org/10.1177/2331216520938929

[19] Kayser, H.; Herzke, T.; Maanen, P.; Pavlovic, C.; Hohmann, V.: Open master hearing aid (openMHA) – an integrated platform for hearing aid research. The Journal of the Acoustical Society of America 146(4), pp. 2 879–2 879, 2019. https://doi.org/10.1121/1.5136988

[20] Grimm, G; Kayser, H.; Hendrikse, M. M. E.; Hohmann, V.: A gaze-based attention model for spatially-aware hearing aids. In: Speech Communication; 13. ITG Symposium 2018, pp. 231–235, 2018. [Online]. https://ieeexplore.ieee.org/document/8578029

[21] Hartwig, M.; Hohmann, V.; Grimm, G.: Speaking with avatars – influence of social interaction on movement behavior in interactive hearing experiments. IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 94–98, 2021. https://doi.org/10.1109/VRW52623.2021.00025

[22] Gogate, M.; Dashtipour, K.; Adeel, A.; Hussain, A.: CochleaNet: A robust language-independent audio-visual model for real-time speech enhancement. Information Fusion 63, pp. 273–285, 2020. https://doi.org/10.1016/j.inffus.2020.04.001

[23] Tammen, M.; Doclo, S.: Deep multi-frame MVDR filtering for binaural noise reduction. In: 2022 International Workshop on Acoustic Signal Enhancement (IWAENC), Bamberg, Germany, pp. 1–5. IEEE, 2022. https://doi.org/10.1109/IWAENC53105.2022.9914742

[24] Nustede, E. J.; Anemüller, J.: Towards speech enhancement using a variational U-Net architecture. In: 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, pp. 481–485, IEEE. https://doi.org/10.23919/EUSIPCO54536.2021.9616114

[25] Dasenbrock, S.; Blum, S.; Debener, S.; Hohmann, V.; Kayser, H.: A step towards neuro-steered hearing aids: Integrated portable setup for time – synchronized acoustic stimuli presentation and EEG recording. Current Directions in Biomedical Engineering 7(2), pp. 855–858, 2021. https://doi.org/10.1515/cdbme-2021-2218

[26] Dasenbrock, S.; Blum, S.; Maanen, P.; Debener, S.; Hohmann, V.; Kayser, H.: Synchronization of ear-EEG and audio streams in a portable research hearing device. Frontiers in Neuroscience 16:904003, 2022. https://doi.org/10.3389/fnins.2022.904003

[27] Grimm, G.; Hendrikse, M.; Hohmann, V.: Pub environment. Zenodo, 2021. https://doi.org/10.5281/ZENODO.5886987

[28] van de Par, S.; Ewert, S. D.; Hladek, L.; Kirsch, C.; Schütze, J.; Llorca-Bofí, J.; Grimm, G.; Hendrikse, M. M. E.; Kollmeier, B.; Seeber, B. U.: Auditory-visual scenes for hearing research. Acta Acustica 6:55, 2022. https://doi.org/10.1051/aacus/2022032

[29] Grimm, G.; Kayser, H.; Kothe, A.; Hohmann, V.: Evaluation of behavior-controlled hearing devices in the lab using interactive turn-taking conversations. In: Proc. 10th Convention of the European Acoustics Association (Forum Acusticum), Turin, Italy, pp. 2 773–2 777, 2023. https://doi.org/10.61782/fa.2023.0127

[30] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015e. Lecture Notes in Computer Science 9351, pp. 234–241. Springer, Cham, 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[31] Nustede, E. J., Anemüller, J.: On the generalization ability of complex-valued variational U-networks for single-channel speech enhancement. IEEE/ACM Transaction on Audio, Speech, and Language Processing 32, pp. 3 838–3 849, 2024. https://doi.org/10.1109/TASLP.2024.3444492

[32] Kingma, D. P.; Welling, M.: Auto-encoding variational Bayes. In: International Conference on Learning Representations (ICLR). 2014. https://doi.org/10.48550/arXiv.1312.6114

[33] Nustede, E. J.; Anemueller, J.: Exploring visualization techniques for interpretable learning in speech enhancement deep neural networks. In: Speech Communication; 15th ITG Conference, Aachen, pp. 220–224, 2023. https://doi.org/10.30420/456164043

[34] Schütze, J.; Ewert, S. D.; Kirsch, C.; Kollmeier, B.: Speech intelligibility and hearing aid benefit in a living room: Comparison of real and simulated acoustics. The Journal of the Acoustical Society of America 154, A115, 2023. https://doi.org/10.1121/10.0022969

[35] Wendt, T.; van de Par, S.; Ewert, S. D.: A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. Journal of the Audio Engineering Society 62(11), pp. 748–766, 2014. https://doi.org/10.17743/jaes.2014.0042

[36] Kirsch, C.; Poppitz, J.; Wendt, T.; van de Par, S.; Ewert, S. D.: Spatial resolution of late reverberation in virtual acoustic environments. Trends in Hearing 25: 23312165211054924, 2021. https://doi.org/10.1177/23312165211054924

[37] Schepker, H.; Denk, F.; Kollmeier, B.; Doclo, S.: Robust single- and multi-loudspeaker least-squares-based equalization for hearing devices. EURASIP Journal on Audio, Speech, and Music Processing 2022:15, pp 1–14, 2022. https://doi.org/10.1186/s13636-022-00247-6

[38] Denk, F.; Kollmeier, B.: The Hearpiece database of individual transfer functions of an in-the-ear earpiece for hearing device research. Acta Acustica 5:2, pp. 1–16, 2021. https://doi.org/10.1051/aacus/2020028

[39] Vogl, S.; Blau, M.: Individualized prediction of the sound pressure at the eardrum for an earpiece with integrated receivers and microphones. The Journal of the Acoustical Society of America, 145(2), pp. 917–930, 2019. https://doi.org/10.1121/1.5089219

[40] Schepker, H.; Nordholm, S.; Doclo, S.: Acoustic feedback suppression for multi-microphone hearing devices using a soft-constrained null-steering beamformer. IEEE/ACM Transactions on Audio, Speech, and Language Processing 28, pp. 929–940, 2020. https://doi.org/10.1109/TASLP.2020.2975390

[41] Benois, P. R.; Roden, R.; Blau, M.; Doclo, S.: Optimization of a fixed virtual sensing feedback ANC controller for in-ear headphones with multiple loudspeakers. In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8 717–8 721, 2022. https://doi.org/10.1109/ICASSP43922.2022.9746327

[42] Xiao, T; Doclo, S.: Effect of target signals and delays on spatially selective active noise control for open-fitting hearables, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1 056–1 060, 2024. https://doi.org/10.1109/ICASSP48485.2024.10445843

[43] Ohlenbusch, M.; Rollwage, C; Doclo, S.: Multi-microphone noise data augmentation for DNN-based own voice reconstruction for hearables in noisy environments, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 416–420, 2024. https://doi.org/10.1109/ICASSP48485.2024.10447066

[44] Ohlenbusch, M.; Rollwage, C.; Doclo, S.: Modeling of speech-dependent own voice transfer characteristics for hearables with an in-ear microphone. Acta Acustica 8:28, 2024. https://doi.org/10.1051/aacus/2024032

[45] Roden, R.; Blau, M.: The IHA database of human geometries including torso, head and complete outer ears for acoustic research, Zenodo, 2021. https://doi.org/10.5281/zenodo.5528766

Editor:innen:
Volker Hohmann, Anna Warzybok-Oetjen, Simon Doclo, Karin Klink
Department für Medizinische Physik und Akustik, Fakultät VI für Medizin und Gesundheitswissenschaften, Carl von Ossietzky Universität Oldenburg