MODELLBASIERTE VERBESSERUNG VON SPRACHVERSTÄNDLICHKEIT IN HÖRGERÄTEN

Anton Schlesinger und Marinus M. Boone

Technische Universität Delft, Niederlande
a.schlesinger@tudelft.nl

Abstract: Dieser Artikel behandelt Algorithmen zur Verbesserung der Sprachverständlichkeit, welche auf physiologischen und psychoakustischen Modellen des Hörapparates beruhen. Neben den Methoden der Richtmikrofonierung ist der vorgestellte modellbasierte Ansatz der robusten und augenblicklichen Verbesserung des Signal-Rausch-Abstandes (SNR) fähig und eignet sich damit zum Einsatz in Hörgeräten. Ferner ermöglicht die Kombination beider Verfahren eine sich konstruktiv ergänzende Arbeitsweise bei der Trennung von Zielsprecher und Interferenz.
Eine aktuelle Fragestellung ist die Optimierung modellbasierter Algorithmen in exemplarisch kritischen Situationen. Die Verbesserung der Effizienz ist dabei im Wesentlichen funktional abhängig von algorithmischen Parametern. Die Suche nach einem optimalen Parameterset ist manuell aufwändig. Rationell ist die methodische Bewertung der komplexen Algorithmen allein mit objektiven Verfahren durchführbar. Da gegenwärtig keine objektiven Methoden für die Analyse binauraler und nicht-linearer Algorithmen verfügbar sind, wird an unserem Institut der *Speech Transmission Index* (STI), ein Maß zur Bewertung der Sprachverständlichkeit bei breitbandiger Interferenz, erweitert. Das neue, objektive Maß soll in einem genetischen Suchalgorithmus als Zielfunktion zur Optimierung von Hörgerätealgorithmen angewandt werden. Da der STI die individuelle Hörleistung einbeziehen kann, ist eine personenbezogene Adaption der modellbasierten Algorithmen möglich. Dieser Artikel skizziert den Optimierungsprozess anhand einer Demonstration. Aus den Ergebnissen werden die Anforderungen an eine verbesserte Bewertungsfunktion der Sprachverständlichkeit abgeleitet.

1 Einleitung

Der Erfolg von Hörgeräten ist abhängig von der Art des Hörverlustes und der akustischen Situation, ich welcher ein Hörgerät eingesetzt wird. Eine zentrale Hörstörungen, wie sie beispielsweise bei der Altersschwerhörigkeit auftritt, führt zu einer verminderten Sprachverständlichkeit in geräuschvollen Situationen. Während eine angehobene Hörschwelle und der damit einhergehende eingeschränkte Dynamikumfang allgemein durch eine Hörhilfe kompensiert werden können, wird die Verbesserung der Sprachverständlichkeit in geräuschvollen Situationen durch bisherige Hörgeräte nicht hinreichend erreicht.
Basierend auf kompakten Mikrophonarrays ist die Richtmikrophonierung ein verbreiteter Ansatz zur Verbesserung des SNR in Hörgeräten. Bei Implementierung eines *Delay-and-Sum*-Verfahrens oder eines Gradientenverfahrens mit zwei Mikrophonen ist beispielsweise eine theoretische Steigerung des SNR von 4 - 6 dB möglich. Beide Verfahren sind jedoch nur bedingt einsatzfähig. Das *Delay-and-Sum*-Verfahren baut allein eine schmalbandige Richtwirkung auf,

wenn die Wellenlänge klein im Verhältnis zum Mikrophonabstand wird. Auch eine dichtere räumliche Abtastung kann diesen Nachteil bei tiefen Frequenzen nicht kompensieren [8]. Das Gradientenverfahren ermöglicht eine breitbandige Richtwirkung, ist aber durch eine hohe Empfindlichkeit gegenüber internem Rauschen bei niedrigen Frequenzen instabil. Bei Gradientenverfahren höherer Ordnung nimmt die Empfindlichkeit weiter zu, was deren Implementierung unmöglich macht. In Hörversuchen ist der SNR-Gewinn dieser Verfahren in Hinter-dem-Ohr (HdO) Hörgeräten selten größer 3 dB. Innerhalb der zeitinvarianten Richtmikrophonierung wesentlich effektiver ist das *Minimum-Variance-Distortionless-Response* (MVDR) Beamforming. Dieses ist in der Lage, SNR-Verbesserungen von bis zu $20 \log N$ zu erreichen (N ist die Anzahl der Mikrophone). Das MVDR Beamforming ermöglicht ein optimales Filter im Sinne der Maximum-Likelihood-Schätzung. Angewandt in einem Hörgerät mit zwei Linienarrays von vier Mikrophonen in den Bügeln einer Brille, erzielt dieses Verfahren eine Verbesserung von mehr als 7 dB [2] unter realistischen Bedingungen und erfüllt damit die Forderung nach einer Verbesserung von mindestens 5 dB, um die Hörbehinderung aller Schwerheitsgrade um 50 % zu mindern [3]. Bei einer binauralen Versorgung mit einem zeitinvarianten Beamformer bleiben veränderte aber stabile binaurale Cues, die interauralen Zeit- und Pegelunterschiede (allg. *Interaural Time Differences*, ITDs und *Interaural Level Differences*, ILDs), erhalten. Dies ermöglicht zum einen die für den audiologischen Erfolg notwenige Nutzung durch das binaurale Resthörvermögen eines Schwerhörigen, sowie zum anderen die Nachschaltung eines binauralen Postfilter zur weiteren Verbesserung des SNR.

Die hier vorgestellten Filter basieren auf der auditorischen Szenenanalyse. Die tägliche Erfahrung kennt den Hörapparat als einen ausgezeichneten Sprachprozessor. Der Mensch ist in der Lage, in komplexen Störschallsituationen einem Sprecher inhaltlich zu folgen und bis zu sechs konkurrierende Sprecher zu identifizieren [1]. Psychoakustische und physiologische Modelle des Hörstamms haben zahlreiche Phänomene der akustischen Wahrnehmung erklären können. Abgeleitete Algorithmen bestätigten diese Modelle durch beachtliche Leistungen bei der Unterdrückung von Störgeräuschen [1]. In diesem Artikel werden drei dieser modellbasierten Algorithmen vorgestellt.

Die Kombination des modellbasierten Ansatzes mit einem Beamformer ermöglicht zudem, die Vorteile beider Verfahren zur Unterdrückung von Störquellen konstruktiv zu nutzen. Ein Nachteil einfacher modellbasierter Prozessoren ist zum Beispiel deren Lokalisationssymmetrie um die Ohrenachse, womit keine Unterscheidung zwischen frontalen und rückwärtigen Quellen möglich ist[1]. Eine örtliche Differenzierung von Quellen ist damit nur konzentrisch um die Ohrenachse auszuführen. Bei Überlagerung mit einem frontal gerichteten binauralen Beamformer kann dieses Symmetrieproblem überwunden werden. Eine weitere Verbesserung des gesamten SNR resultiert aus der Fähigkeit modellbasierter Verfahren, auf spektral dünn besetzte Sprachinformation zugreifen zu können. Diese Eigenschaft verzerrt jedoch das Zielsignal und bedingt das Auffinden einen Kompromisses zwischen Signalverzerrung und der Unterdrückung von Störquellen, um die Sprachvertändlichkeit zu fördern.

Die Optimierung der komplexen modellbasierter Verfahren hat dabei einen wesentlichen Einfluss auf deren Effizienz und wird im dritten Teil dieses Beitrags behandelt. Vorausgehend stellt sich die Frage nach repräsentativen akustischen Situationen, in welchen der Algorithmus die Signalqualität verbessern soll. Darüber hinaus bleibt unbeantwortet, ob ein fester Parametersatz des Algorithmus in unterschiedlichen Situationen hinreichend ist oder ob die algorithmischen Parameter in bestimmten Situationen jeweils angepasst werden müssen. Die Beantwortung dieser Fragen erfordert Vergleichsprüfungen. Die Beurteilung der Effizienz eines Algorithmus kann anhand subjektiver Hörtests vorgenommen werden. Diese sind wiederum in vielerlei Hinsicht aufwändig. Aus diesem Grund wurden objektive Methoden entwickelt, welche

[1] Der Hörapparat nutzt neben binauralen Cues die spektrale Information eines Signals für dessen Ortung.

den Bewertungsprozess wesentlich vereinfachen. Die bekannten objektiven Methoden erweisen sich jedoch als nicht geeignet zur Beurteilung der hier vorgestellten komplexen und binauralen Sprachprozessoren. In diesem Beitrag zeigen wir exemplarisch die Beschränkungen, die den STI als allgemeines Maß der Sprachverständlichkeit[2] in binauralen Situationen scheitern lassen. Dabei wurde zur Optimierung der algorithmischen Parameter ein genetischer Suchalgorithmus angewandt, in welchem eine monaurale Variante des STI als Zielfunktion arbeitete. Schließlich wurde der optimierte Algorithmus in einem subjektiven Hörtest evaluiert.
Die fehlende Übereinstimmung zwischen objektiven und subjektiven Ergebnissen wird im letzten Teil dieses Beitrages diskutiert. Zur Verbesserung der objektiven Bewertung der Sprachverständlichkeit wird eine Erweiterung des monauralen STI um ein Modell binauraler Interaktion vorgeschlagen.

2 Modellbasierte Sprachprozessoren

Die auditorische Szenenanalyse wird allgemein als zweifacher Prozess verstanden. Mit der Auswertung unterschiedlicher Signaleigenschaften, die vom Gehirn interpretiert werden, wird ein Merkmalsraum aufgebaut, in dem akustische Quellen, d.h. die gesamte akustische Szene, analysiert werden. Die Merkmalsextraktion beginnt in der Cochlea. Hier wird der Frequenzinhalt durch eine räumlich verteilte Stimulation der Basilarmembran in eine räumliche verteilte Nervenanregung transformiert, welche als tonotopische Abbildung bekannt ist. Diese Merkmalsabbildung setzt sich über alle Nervenkerne in das Sprachzentrum fort und bleibt also im gesamten Hörapparat erhalten. Auf den höheren Verarbeitungsstufen im Hirnstamm werden weitere Cues, wie bspw. impulsartige Einsätze, Periodizität oder Ortsinformationen, ausgewertet. Im zweiten Schritt werden die neuronalen Erregungsmuster an unterschiedlichen Stellen des Hörnervs zu kohärenten akustischen Objekten verknüpft. Durch eine bewusste Steuerung des Individuums ist es damit möglich, ein akustisches Objekt, beispielsweise einen Zielsprecher, zum Sprachzentrum in die Schläfenlappen weiterzuleiten. Man nimmt an, dass diese Verknüpfung durch ein zeitliches Angleichen der periodischen, neuronalen Erregungsmuster geschieht.
Die hier vorgestellten Sprachprozessoren basieren auf Modellen des Hirnstamms. Sie sind in der Lage, augenblicklich, d.h. in der Größenordnung von wenigen Millisekunden[3], den SNR anzuheben. Im Folgenden werden drei modellbasierte Sprachprozessoren vorgestellt. Allen gemein ist ein Modell binauraler Interaktion. In den Abbildungen 1 bis 3 sind jeweils der linke Kanal und die zentralen, binauralen Verarbeitungsstufen dargestellt. Die grundlegenden Besonderheiten des jeweiligen Prozessors sind rot unterlegt. Abbildung 1 zeigt einen Algorithmus nach Kollmeier und Peissig (KP) [7]. Die in der Cochlea stattfindende Frequenzkodierung wird durch eine FFT-basierte Analyse blockweise im linken und rechten Kanal des digitalisierten Signals vorgenommen. Von der Einhüllendeninformation des binauralen Signals werden ITD und ILD berechnet. Die Einhüllende von Sprache zeigt eine relative Robustheit gegenüber Störungen durch Nachhall oder breitbandige Interferenz. Durch eine Halbwellengleichrichtung und eine Tiefpassfilterung bei der Kodierung des akustischen Stimulus nutzt der Hörapprat diese physikalische Eigenschaft. Neben der Auswertung binauraler Cues wird auch die Kohärenz im binauralen Signal bestimmt. Die Analyse der Kohärenz ermöglicht es, zwischen direktem und diffusem Schall zu unterscheiden. Auf diese Weise kann im Signal der Nachhall identifiziert und unterdrückt werden. Psychoakustisch wird diese Fähigkeit dem Gesetz der ersten Wellenfront zugeschrieben. Mit Hilfe der Angabe von Referenzwerten für die interne Abbil-

[2] Allein der sprachbasierte STI ist gegenwärtig in der Lage, nichtlineare Sprachprozessoren, wie es die hier vorgestellten modellbasierten Systeme sind, auf ihre Sprachverständlichkeit hin zu bewerten.

[3] Hörgeräte fordern eine geringe Systemverzögerung, um Rückkopplungen, spektrale Färbungen sowie ein verzögertes Eigenhören zu vermeiden.

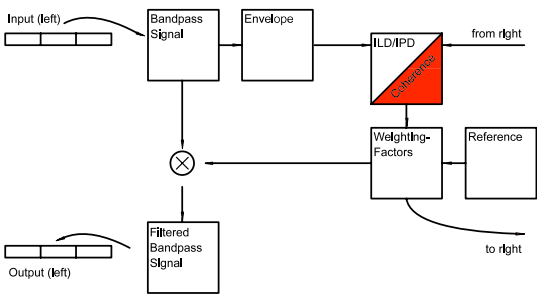

Abbildung 1: *Modellbasierter Algorithmus nach [7] zur Unterdrückung lateraler Quellen und Nachhall.*

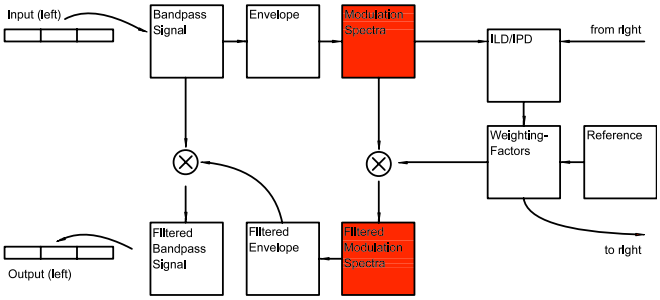

Abbildung 2: *Modellbasierter Algorithmus nach [6] zur Unterdrückung lateraler Quellen in Modulationsspektren.*

dung der akustischen Szene mittels ITD-, ILD- und der Kohärenzinformation, kann ein akustisches Objekt einer bestimmten Richtung in der Horizontalen des Gesichtsfeldes ausgewählt werden. Eine Gewichtungsfunktion unterdrückt adaptiv Quellen außerhalb der Vorzugsrichtung sowie inkohärente Signalanteile, d.h. den Nachhall. Nach der Modulation der Zeit-Frequenz-Darstellung des Ausgangssignals mit der Gewichtungsfunktion wird das verbesserte Signal in den Zeitbereich rücktransformiert. Der KP-Algorithmus zeigte eine hohe Effizienz in moderaten SNR-Situationen. Zudem besitzt er eine vergleichsweise geringe Systemverzögerung von 7 ms.

Der folgende Algorithmus von Kollmeier und Koch (KK) [6] in Abbildung 2 integriert neben dem allgemeinen Modell binauraler Interaktion ein physiologisches und psychoakustisches Modell der Modulationswahrnehmung. Demnach nutzt das Ohr die periodische Zeitstruktur des Signals für eine neuronale Ortskodierung von Modulationsfrequenzen (periodotopische Abbildung). Dies ermöglicht die Trennung von mehreren Sprechern, auch wenn diese die gleichen Mittenfrequenzen benutzen. Algorithmisch ist dies durch eine FFT-basierte Modulationsfrequenzanalyse in jedem Filterband gelöst. Anschließend werden in den Modulationsspektren binaurale Cues analysiert. Mittels einer Richtungsreferenz wird ähnlich dem KP-Prozessor eine Gewichtungsfunktion erstellt und mit den ursprünglichen Modulationsspektren multipliziert. Nach einer Rücktransformation wird mit dem verbesserten Signal das Ausgangssignal moduliert. Der KK-Algorithmus bietet eine robuste Verbesserung des SNR von etwa 2 dB bis hin zu sehr schlechten SNR-Bedingungen. Durch die Extraktion der Modulationsfrequenzen ist ein höhere Systemverzögerung von etwa 25 ms notwendig.

Der dritte Algorithmus (Abbildung 3) wurde von Albani und Kollegen (A) [4] entwickelt. Dieser integriert ein Lokalisationsmodell von akustischen Quellen. Das Lokalisationsmodell orientiert sich an der Mittenfrequenzinteraktion zur Auflösung räumlicher Mehrdeutigkeiten. So ist bekannt, dass auf den Radien eines geraden Kegelmantels konzentrisch um die Ohrenachse konstante ILD- und ITD-Werte auftreten (*cone of confusion*). Der Hörapparat integriert die neuronalen Aktivitätsantworten der Mittenfrequenzen allseitig, d.h. exzentrisch. Das Maximum der Integration entspricht dem Ort der Quelle. Dargestellt ist die physikalische Repräsentation einer Quelle mit ILD und ITD bei unterschiedlichen Mittenfrequenzen in Abbildung 4. Nach der Lokalisation von Quellen kann mit Referenzwerten, die bspw. prosodische Cues oder initiale binauale Cues enthalten, ein Zielsprecher verfolgt werden. Der Algorithmus soll in der Lage sein, mehrere Quellen gleichzeitig zu lokalisieren und ihnen folgen zu können [4]. Uns ist nicht bekannt, ob der Algorithmus auf die Verbesserung von Sprachverständlichkeit hin untersucht wurde.

Eine sich anschliessende Fragestellung ist, wie diese Algorithmen miteinander kombiniert werden können, um in einem breiten Spektrum akustischer Situationen die auditorische Szenenanalyse effektiv nachzubilden oder sogar in Kombination mit Richtmikrophonierung zu überbieten.

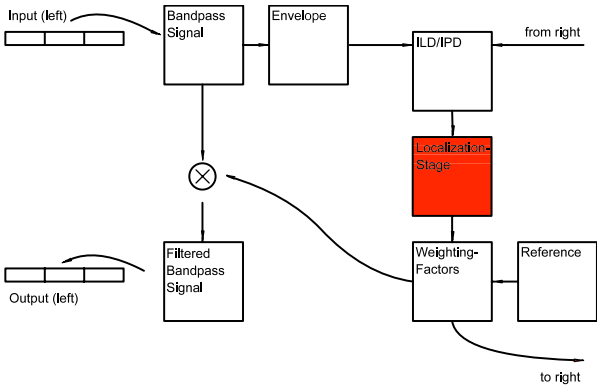

Abbildung 3: *Modellbasierter Algorithmus nach [4] zur Lokalisation und Unterdrückung von Störquellen.*

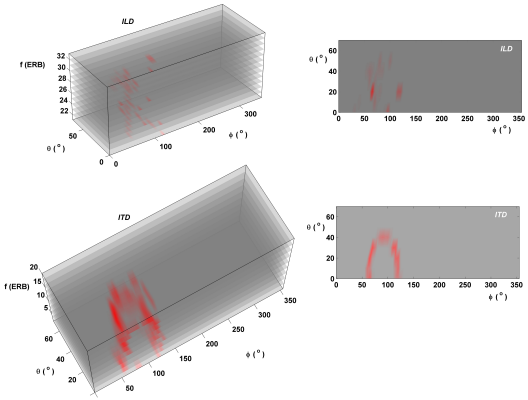

Abbildung 4: *Karten neuronaler Aktivität bei rosa Rauschen aus Richtung $\theta = 20°$ und $\phi = 60°$. Sichtbar ist für ITD and ILD ein Halbkreis pro Mittenfrequenz. Durch Frequenzbandsummation fällt das Maximum mit dem tatsächlichen Ort der Quelle zusammen (oben und unten rechts).*

Im folgenden Abschnitt wird eine solche Kombination vorgestellt.

3 Anwendungsbeispiel

Die diskutierten modellbasierten Sprachprozessoren wurden in den 1990er Jahren entwickelt. Ihr Einsatz in Hörgeräten ist durch einen hohen rechentechnischen Aufwand, sowie die aufwändige breitbandige Datenkommunikation zwischen den Ohren beschränkt. Diese Beschränkungen werden sich auf mittlere Sicht auflösen. Neben stets kleineren und effizienteren digitalen Signalprozessoren ermöglichen neue Transferprotokolle schon heute den breitbandigen, kabellosen Datentransfer.

Abbildung 5 zeigt die Kombination aus dem modellbasierten KP-Prozessor und der o.g. Hörbrille, welche zwei MVDR-Beamformer in den Bügeln der Brille besitzt und sich damit als binaurale Vorverarbeitungsstufe eignet. Auch sind die Anforderungen an die Integrationsdichte der Algorithmik sowie den Datentransfer im Falle der Unterbringung eines modellbasierten Algorithmus in einem Brillengestell geringer. Die Wahl fiel auf den KP-Prozessor, um einen maximalen Erwartungswert der Kombination zu ermitteln (der KP-Prozessor erzielte die höchste Verbesserung der Sprachverständlichkeit in moderaten SNR-Bedingungen). Die Hörbrille hebt den SNR als Vorverarbeitungsstufe um einen festen Wert und ermöglicht damit, bei Annahme einer entsprechenden SNR-Situation, den KP-Prozessor im Bereich seines Arbeitspunktes anzuwenden.

Das optimale Richtungsfilter des MVDR-Beamformer wurde für ein ideal diffuses Schallfeld berechnet und war fest implementiert [8]. Der KP-Prozessor umfasst in der vorgestellten Implementierung sieben algorithmische Parameter und wurde mit einem genetischen Suchalgorithmus entsprechend Abbildung 6 optimiert. Als Bewertungsfunktion der Sprachverständlichkeit diente eine sprachbasierte Einhüllenden-Regressionsvariante des STI von Goldsworthy und Greenberg [5]. Für die Optimierung wurde eine akustische Situation mit vier interferierenden Quellen auf 60°, 120°, 240° und 300° (Zielsprecher bei 0°) simuliert. Die Nachhallzeit des simulierten Raumes betrug 0,8 s. Für die genetische Optimierung wurde eine verknüpfte Auswahl von allgemeinen Sätzen aus der niederländischen Sprache verwendet. Das phonetisch ausgewogene Sprachbeispiel wurde von einer weiblichen Stimme gesprochen und hatte eine Länge von

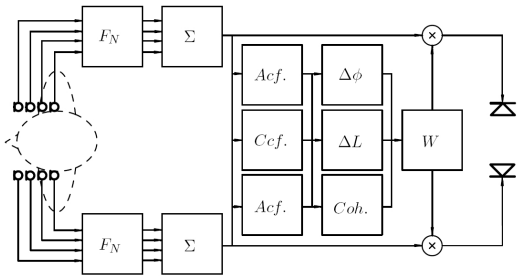

Abbildung 5: *Prototyp eines Hörgerätes zur Verbesserung der Sprachverständlichkeit in ak. schlechten Umgebungen. Der Prototyp besteht aus einem festen binauralen MVDR-Beamfomer (F_N: komplexe Filterkoeffizienten, $\sum$: Summation) und aus dem modellbasierten KP-Prozessor als Postfilter (Acf: Autokorrelationsfunktion, Ccf: Kreuzkorrelationsfunktion, $\Delta\phi$ und ΔL: ITD und ILD, Coh: Kohärenz und W: Gewichtungsfunktion).*

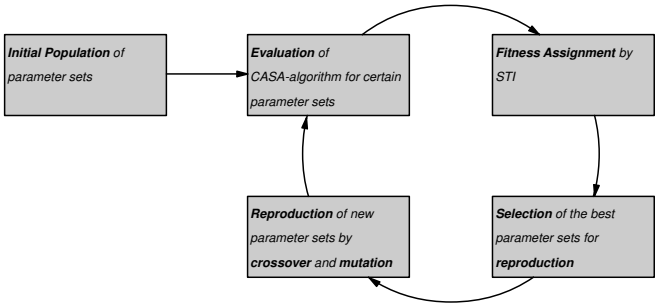

Abbildung 6: *Optimierungsroutine eines genetischen Suchalgorithmus mit dem STI (Speech Transmission Index) als Zielfunktion zur Verbesserung von Sprachprozessoren (hier dem KP-Algorithmus).*

45 s. Den interferierenden Quellen wurde das gemittelte Sprachspektrum desselben Sprachbeispiels von drei unterschiedlichen männlichen Stimmen und der weiblichen Zielstimme zugewiesen. Die Optimierung wurde bei vier SNR Verhältnissen von -12 bis -18 dB vorgenommen. Der genetische Optimierungszyklus wurde für jede SNR-Situation zweifach durchgeführt. Jeder Zyklus umfasste eine initiale Population von 4000 Sets von algorithmischen Parametern und 2000 Iterationen des Vererbungszyklus. Die Ergebnisse sind in Tabelle 1 notiert. Die Resultate

Tabelle 1: *STI-Ergebnisse (linkes Ohr / rechtes Ohr); die Kombination aus binauralem MVDR-Beamformer und modellbasiertem KP-Prozessor wurde zweifach mit dem genetischen Algorithmus (GA 1 und 2) optimiert.*

SNR	**-12 dB**	**-14 dB**	**-16 dB**	**-18 dB**
ohne Hilfe	0,22/0,29	0,17/0,23	0,12/0,16	0,05/0,08
binauraler MVDR-Beamformer (Hörbrille)	0,40/0,42	0,34/0,36	0,29/0,31	0,21/0,23
Kombination mit KP-Algorithmus (GA 1)	0,51/0,51	0,50/0,50	0,46/0,46	0,43/0,43
Kombination mit KP-Algorithmus (GA 2)	0,54/0,54	0,51/0,52	0,48/0,48	0,42/0,43

der Optimierung sprechen für die Kombination beider Verfahren zur Verbesserung des Sprachverständlichkeit. Zu beobachten ist eine stabile Verbesserung des binauralen Beamformers in allen SNR-Situationen. Eine Verringerung des SNR-Gewinns setzt für die Kombination mit dem KP-Algorithmus bei steigen SNR-Werten ein. So beträgt der Gewinn in der SNR-Situation -18 dB für die Kombination gegenüber der Hörbrille etwa 0,2, bei -12 dB noch etwa 0,1. In allen Situationen wird der binaurale Vorteil an beide Ohren weitergegeben. Eine detaillierte Auswertung der Untersuchung kann in [9] nachgelesen werden.
Im folgenden Schritt wurde die SNR-Situation von -12 dB in einem Hörtest evaluiert. Nach Verbesserung des SNR um etwa 7 dB mit dem vorgestellten binauralen MVDR-Beamformer, sollte die Verständlichkeit in der Größenordnung von 50 % liegen[4]. Als Parameterset der genetischen Optimierung des KP-Algorithmus wurde Lösung 2 (siehe Tabelle 1 GA 2) ausgewählt. Für die subjektive Bewertung wurde ein allgemeiner SRT-Test (*Speech Reception Threshold*

[4]Beim gesunden Gehör und monauraler Wiedergabe liegt eine Sprachverst. von 50% gemittelt bei -5 dB.

Test[5]) vorgenommen. Die vorgestellte Kombination zur Verbesserung des SNR wurde mit dem binauralen MVDR-Beamformer der Hörbrille verglichen und beide Methoden wurden jeweils um den SRT ohne Hörhilfe korrigiert. Daraus leitet sich der tatsächliche Gewinn in dB ab.
Am Hörtest nahmen 20 Personen teil. Von jedem Probanden wurde beidohrig ein Audiogramm erstellt. Als normalhörend wurde eingestuft, wessen Audiogramm keinen Hörverlust größer 20 dB zwischen 125 Hz und 4 kHz aufwies. Dies resultierte in 16 normalhörenden Probanden. Die Ergebnisse des Hörtests sind in Abbildung 7 dargestellt. Im Durchschnitt lag der SRT-Pegel

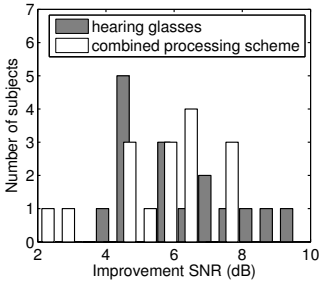

Abbildung 7: *Die Histogramme zeigten die Verteilung der individuellen SNR-Gewinne (SNR re SRT).*

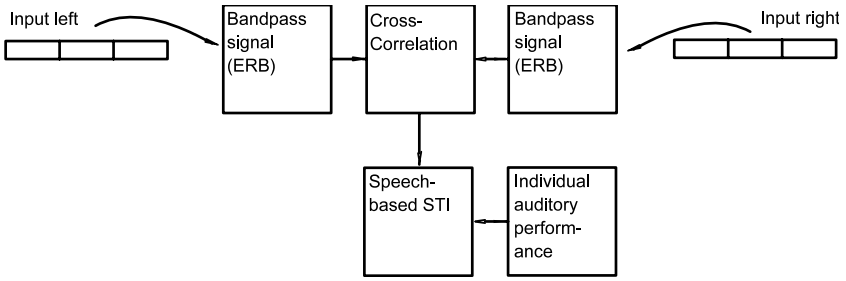

Abbildung 8: *Erweiterung des STI um ein Modell binauraler Interaktion mit auditorisch korrekten Filterbreiten (ERB: Equivalent Rectangular Bandwidth).*

ohne Hörhilfe bei −4,6 dB. Der Pegel verringerte sich beim Hören mit der Hörbrille auf −10,7 dB und beim Hören mit dem kombinierten Ansatz auf −10,4 dB. Dieses Resultat zeigt keinen signifikanten Unterschied zwischen den beiden Ansätzen zur Verbesserung des SNR.

4 Modellbasierte Bewertung der Sprachverständlichkeit

Ein audiologischer Erfolg des kombinierten Ansatzes gegenüber dem alleinigen binauralen MVDR-Beamformer der Hörbrille konnte dennoch für drei schlechthörende Probanden mit einem Hörverlust von mehr als 35 dB bei 4 bis 6 kHz erzielt werden. Die schlechthörenden Probanden zeigten jeweils eine Verringerung des SRT-Pegels von etwa 5 dB gegenüber der Hörbrille. Obwohl eine Generalisierung aufgrund der fehlenden statistische Grundlage unmöglich ist, findet dieses Resultat in der Literatur Bestätigung [5]. So zeigen sich bspw. bei einer verwandten Sprachverbesserungsmethode, der spektralen Subtraktion, dass für Normalhörende kein Vorteil erzielt werden kann, wenn die oberhalb der Hörschwelle liegende Feinstruktur eines Zielsprechers verstärkt wird. Der Schwerhörige benötigt jedoch eine Verstärkung der Feinstuktur, da seine verbreiteten auditiven Filter sowie eine angehobene Hörschwelle die informationstragenden Muster verschleifen und maskieren. Der breitbandige, monaurale STI ist für diese Differenzierung ungeeignet.
Wäre es möglich, die individuelle Sprachverständlichkeit hinreichend mit Modellen der Hörappartes zu simulieren, sollte es gelingen, modellbasierte Algorithmen in der vorgestellten, komplexen Implementierung eines Postfilters anzupassen und damit einen Gewinn für verschiedene Grade der Hörleistung zu erzielen. Abbildung 7 skizziert den von den Autoren vorgeschlagenen

[5]Mit dem SRT bestimmt man den individuellen SNR, bei welchem eine Sprachverständlichkeit von 50 % erzielt wird. Da der SRT-Test sich nicht für die Bewertung adaptiver Algorithmen eignet (der SRT ist dann nicht allein vom SNR abhängig, sondern auch von den algorithmischen Filtereigenschaften, welche ebenso eine Funktion des SNR sind), wurde ein diffuses Feld erzeugt und der Zielsprecher bei der Optimierung wie bei der Wiedergabe im SRT-Test innerhalb des Hallradius simuliert. Auf dieses Weise wurde das adaptive Filter des modellbasierten KP-Algorithmus in einem größeren SNR-Bereich stabilisiert .

und verfolgten Ansatz zur Erweiterung des STI um ein Modell auditorischer Filter und binauraler Interaktion unter Einbezug der individuellen Hörleistung mit der absoluten Hörschwelle. In diesem Ansatz werden die ersten neuronalen Verarbeitungsstufen des Hörapparates abgebildet. Das Design stellt ein Zusammenspiel erfolgreicher Modelle der auditorischen Verarbeitung dar. Die erste Annahme bezieht sich auf die periphere Verarbeitung mit auditorischen Filtern. In Verbindung mit der absoluten Hörschwelle sollte dieser Ansatz zur richtigen Bewertung von Verzerrungen (wie sie in modellbasierten Sprachprozessoren entstehen können) und schmalbandiger Interferenz fähig sein. Der Einbezug des binauralen Vorteils bei der Sprachverständlichkeit geschieht durch eine Kreuzkorrelation der binaural auditorischen gefilterten Signale. Und schließlich ist ein sprachbasierter STI auf Grundlage einer Einhüllenden-Regressionsmethode (s.o.) zur Bewertung nicht-linearer Sprachprozessoren geeignet. Die laufende Arbeit umfasst einen binauralen Sprachtest in verschiedenen räumlichen Umgebungen. Nach Abschluss des subjektiven Tests wird das Modell in Abbildung 7 mit den Ergebnissen abgeglichen.

5 Zusammenfassung

Dieser Beitrag stellte eine Auswahl modellbasierter Sprachprozessoren zur Verbesserung der Sprachverständlichkeit in schlechten akustischen Umgebungen vor. Eine Herausforderung bei der Anwendung modellbasierter Prozessoren stellt deren Optimierung dar. Der genetische Suchalgorithmus wurde als ein effizientes Optimierungswerkzeug vorgestellt. Die Verwendung des breitbandigen, monauralen STI zur objektiven Bewertung und Optimierung offenbarte dessen Unvermögen, nicht-lineare, binaurale Algorithmen zu beurteilen. Die Hoffnung ruht auf der Erweiterung des STI mit einem auditorischen Modell binauraler Interaktion, um Sprachprozessoren effizient zu bewerten und individuell optimieren zu können.

Literatur

[1] *Computational Auditory Scene Analysis*. A John Wiley & Sohns, Inc., Publication, 2006.

[2] M. M. Boone. Directivity measurements on a highly directive hearing aid: the hearing glasses. *AES 120th Convention*, 2006.

[3] A. J. Duquesnoy and R. Plomp. The effect of hearing impairment on the speech-reception threshold of hearing-impaired listeners in quiet and noise. *J. Acoust. Soc. Am.*, 1983.

[4] S. Albani et al. *Model of binaural localization resolving multiple sources and spatial ambiguities. In Psychoacoustics, speech and hearing aids*. World Scientific, 1996.

[5] R. L. Goldsworthy and J. E. Greenberg. Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *J. Acoust. Soc. Am.*, 2004.

[6] B. Kollmeier and R. Koch. Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction. *J. Acoust. Soc. Am.*, 1994.

[7] B. Kollmeier and J. Peissig. Real-time multiband dynamic compression and noise reduction for binaural hearing aids. *J. of Rehabilitation Research and Development*, 1993.

[8] Ivo Merks. *Binaural Application of Microphone Arrays for Improvement Speech Intelligibility in a noisy Environment*. PhD thesis, Technische Universiteit Delft, 2000.

[9] A. Schlesinger and M. M. Boone. Evolutionary optimization for hearing aids of computational auditory scene analysis. *AES 126th Convention, München*, 2009.