

Cocktail-Partys und Hörgeräte: Biophysik des Gehörs

Physikalisch inspirierte Hörmodelle weisen den Weg zu intelligenten Hörgeräten

Birger Kollmeier

Von aktiven, nichtlinearen Prozessen im Ohr über numerische Hörmodelle bis hin zur Musikübertragung im Internet mit MP3 oder objektiver Beurteilung der Sprachgüte von Handys: Die Hör-Akustik hat unmittelbare Auswirkungen auf unser tägliches Leben – nicht nur wenn eine Hörstörung auftritt oder wenn man auf einer lebhaften Party nichts mehr versteht. Für die Physik ist die Analyse der effektiven Funktion des Gehörs als komplexes Gesamtsystem interessant. Die Umsetzung dieser Analyse in ein Hörmodell ermöglicht eine Vielzahl technischer Anwendungen.

Die Physik „wächst“ an den Rändern: Während vor 20 Jahren Physiker, die sich mit neuronalen Systemen oder Sinnesorganen beschäftigten, eher zu den Außenseitern zählten, ist die Beschäftigung mit komplexen, nichtlinearen biologischen Systemen zu einem auch für den wissenschaftlichen Nachwuchs attraktiven Feld geworden, das von der Interdisziplinarität und der Anwendungsbreite physikalischer Methoden lebt. Nicht erst seit Helmholtz's frühen Arbeiten über „Die Lehre von den Tonempfindungen“ [1] ist die Beschäftigung mit dem Hörsinn ein wichtiges Beispiel hierfür. Dieser Beitrag zeigt die Verbindung auf zwischen der medizinischen Sichtweise des Hörvorgangs (und seinen möglichen Störungen) und der physikalischen Sichtweise. Mögliche Anwendungen des quantitativen Zugangs zur Modellierung des Hörvorgangs werden exemplarisch für die Bereiche Telekommunikation, Audio-Signalverarbeitung und digitale Hörgeräte vorgestellt.

Eine wesentliche Motivation für die Beschäftigung mit dieser Thematik ist der „Cocktail-Party-Effekt“, der Partygästen die Konversation erleichtert und dessen Ausfall eines der schwerwiegendsten Probleme für Schwerhörige darstellt: In einer störrauschbehafteten Umgebung (z. B. einer lebhaften Party) können sich Normalhörende relativ gut auf einen Sprecher konzentrieren und die Störrausche unterdrücken. Schwerhörigen fällt dies jedoch besonders schwer, sodass sie derartige Situationen und größere Menschenansammlungen meiden, was häufig zur sozialen Isolation führt. Selbst die modernsten kommerziellen, digitalen Hörgeräte können zwar in akustisch „einfachen“ Situationen (mit nur einer stationären Störquelle, die aus einer ganz anderen Richtung als das Nutzsignal kommt) die



Störrausche wesentlich besser unterdrücken als es den Geräten noch vor etwa zehn Jahren möglich war, versagen aber nach wie vor in akustisch „schwierigen“ Situationen wie etwa einer Party. Um hier Lösungsmöglichkeiten aufzuzeigen, muss das zugrundeliegende „System Ohr“ erst einmal analysiert werden.

Aufbau und Funktionsweise des Gehörs

Das Außen- und Mittelohr (vgl. Abb. 2) leitet den Schall mit nur geringen Energieverlusten in das mit Flüssigkeit gefüllte Innenohr. Funktionsstörungen des Außen- und Mittelohres machen sich in einer *Schallleitungs-Schwerhörigkeit* bemerkbar, die durch HNO-ärztliche Eingriffe oder durch ein einfaches, lineares Hörgerät ausgeglichen werden können. Die eigentliche Umwandlung der Schallschwingungen in Nervenimpulse findet im Innenohr (Kochlea oder Hörschnecke) statt. Die Hörschnecke ist der Länge nach durch die Basilarmembran unterteilt, wobei die auf der Basilarmembran angeordneten inneren und äußeren Haarzellen die mechano-elektrischen Wandler darstellen. Störungen des Innenohrs oder der nachfolgenden neuronalen Strukturen führen zu der *Schallempfindungs-Schwerhörigkeit*, die wesentlich häufiger vorkommt (ca. 15 % unserer Bevölkerung), aber leider auch problematischer ist als die Schallleitungs-Schwerhörigkeit [2].

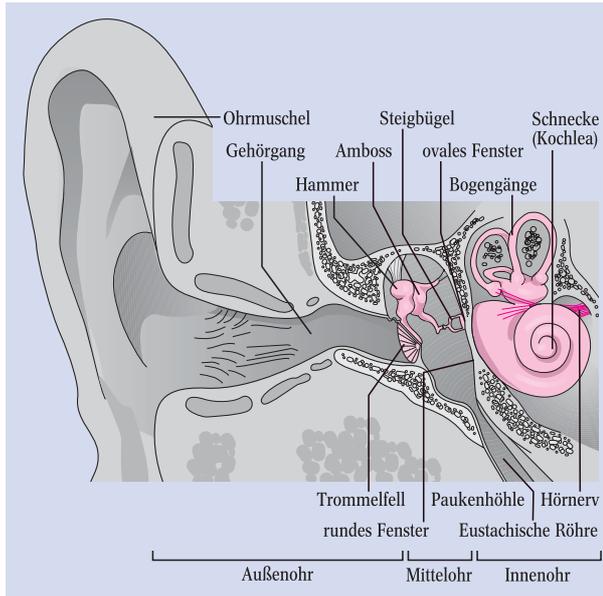
Das Problem liegt in der Nichtlinearität des Innenohrs (siehe Infokasten „Wie nichtlinear ist das Ohr?“), die sich auch objektiv mit einem empfindlichen Mikrofon ausmessen lässt. Abbildung 3 zeigt das gemittelte Spektrum des akustischen Signals, das im abgeschlossenen Gehörgang eines normalhörenden Probanden gemessen werden kann, wenn die mit f_1 und f_2 gekenn-

Abb. 1: Auf einer Cocktail-Party können Normalhörende die vielen Nebengeräusche unterdrücken und sich auf ein Gespräch konzentrieren. Schwerhörige haben dagegen große Probleme – eine besondere Herausforderung an die Kommunikationsakustik!

Prof. Dr. Dr. Birger Kollmeier, Medizinische Physik, Universität Oldenburg, 26111 Oldenburg; <http://medi.uni-oldenburg.de>

zeichneten reinen Sinustöne als Eingangssignale gegeben werden. Die Verzerrungsprodukte (z. B. bei $2f_1 - f_2$) werden vom Innenohr akustisch als eine Art Echo abgestrahlt („otoakustische Emissionen“, oto ist lateinisch für Ohr) und verschwinden bei Hörstörungen oder bei Unterbrechung der Sauerstoff-Zufuhr des Innenohrs. Die negative Dämpfung bei kleinen Auslenkungen der Basilarmembran und die Ankopplung an äußere Reize lässt sich in guter Näherung durch einen nichtlinearen van-der-Pol-Oszillator beschreiben [4].

Abb. 2: Das Ohr aus Sicht des Mediziners besteht (von links nach rechts) aus dem Außenohr (Ohrmuschel, Gehörgang, Trommelfell), dem Mittelohr (Paukenhöhle mit den Gehörknöchelchen Hammer, Amboss, Steigbügel), dem Innenohr (Gehörschnecke (Cochlea) mit der Basilarmembran) und dem Hörnerv sowie dem hier nicht skizzierten zentralen auditorischen System im Gehirn (aus [18]).



Weil die otoakustischen Emissionen in der Regel nur bei gesunden Ohren auftreten, werden sie auch als objektiver Hörtest bei Neugeborenen eingesetzt. Dazu beschallt man das Ohr mit einem leisen „Klick“ und misst Zeitverzögerung, Frequenzspektrum und Reproduzierbarkeit des Echos. Otoakustische Emissionen können auch ohne äußeren Reiz („spontan“) im abgeschlossenen Gehörgang mit niedrigen Schalldruckpegeln auftreten (Hörbeispiel 2)^{*)} und sind normalerweise unhörbar. Sie weisen keinerlei Beziehung zu dem subjektiv empfundenen „Fiepen“ im Ohr auf, das sporadisch oder nach einer Lärmbelastung bei Gesunden auftritt

oder als andauerndes Ohrensausen („Tinnitus“) Ausdruck einer Hörstörung ist.

Als wichtigste Quelle der Nichtlinearitäten im Innenohr und damit der otoakustischen Emissionen gelten die Haarzellen, die die Bewegung der Basilarmembran in eine elektrische Spannung umsetzen und dabei sogar Auslenkungen im Sub-Nanometer-Bereich (!) hörbar machen. Die inneren Haarzellen dienen dabei primär als „Mikrofone“, deren Ausgangsspannung über eine Erregung im Hörnerv zum Gehirn übertragen wird. Die äußeren Haarzellen fungieren zusätzlich als „Lautsprecher“, da sie durch Wechsellspannung zu aktiven Kontraktionen angeregt werden und quasi als Resonanzverstärker von Basilarmembran-Schwingungen die Empfindlichkeit und die Frequenzauflösung des Hörorgans steigern [5]. Die genauen mikromechanischen Mechanismen und die Kodierung akustischer Information im Nervensystem sind nach wie vor Gegenstand biophysikalischer Forschung. In *physiologischen* Messungen wird dazu die Reaktion des Körpers (z. B. einer Nervenzelle im Tierexperiment) auf einen definierten akustischen Reiz erfasst. Beim Menschen kann man dazu das Elektro-Enzephalogramm (EEG) bzw. das Magneto-Enzephalogramm (MEG) als Spannungs- bzw. Magnetfeldverlauf auf der Schädeloberfläche erfassen oder mit funktioneller Kernspintomographie die (zeitlich nur wenig aufgelöste) Aktivierung bestimmter Hirnareale. Bei *psychophysikalischen* Messungen wird dagegen die subjektive Reaktion einer Versuchsperson auf einen Reiz erfasst, wobei die Versuchsperson eine vorgegebene Aufgabe durchführen muss (z. B. Erkennen eines Tons in einem zweimal angebotenen Rauschen, von denen nur eines den Ton enthält). Durch die sich gegenseitig ergänzende Information aus beiden Bereichen besitzen wir relativ genaue Vorstellungen über die Verarbeitungsprinzipien des zentralen Hörsystems im Gehirn:

- ▶ **Frequenzabbildung:** Im Innenohr werden hohe Frequenzen an der Basis der Cochlea abgebildet, niedrige Frequenzen an der Spitze. Gemäß dieser Frequenz-Orts-Transformation werden auf allen Stationen der Hörbahn benachbarte Frequenzen an benachbarten Orten verarbeitet.
- ▶ **Einhüllenden-Extraktion und Dynamikkompression:**

Wie nichtlinear ist das Ohr?

Dass die Funktion des normalen Innenohrs hochgradig nichtlinear ist, lässt sich an folgendem einfachen Hörexperiment mit einem Paar Stereolautsprecher demonstrieren (Hörbeispiel 1)^{*)}: Auf einem Lautsprecher wird ein konstanter Sinuston von $f_1 = 2$ kHz abgespielt. Mit dem zweiten Lautsprecher wird dagegen ein Ton mit einer aufsteigenden Frequenz f_2 („Sweep“) zwischen 2 kHz und 2,8 kHz erzeugt. Werden nun beide Lautsprecher gleichzeitig angeschaltet (lineare Addition des Schallfeldes in der Luft), nimmt man subjektiv (je nach Lautstärke der beiden Stimuli und Funktionszustand des eigenen

Mittel- und Innenohres) einen abwärts verlaufenden Sweep („kubischer Differenzton“ $2f_1 - f_2$) und einen Aufwärts-Sweep mit niedrigerer Frequenz („quadratischer Differenzton“ $f_2 - f_1$) wahr. Diese so genannten Verzerrungsprodukte werden durch eine kompressive Nichtlinearität im Innenohr erzeugt: Für schmalbandige Signale ist die dort angeregte Schwingung proportional zu einer Potenzreihe des Eingangssignals, wobei die dritte Potenz zum kubischen Differenzton führt. Allerdings wird diese Nichtlinearität normalerweise in der nachfolgenden Signalverarbeitung und durch kognitive Prozesse

im Gehirn weitgehend ausgeglichen. Bei natürlichen (breitbandigen) Signalen werden die Verzerrungsprodukte nämlich von den stärkeren Komponenten des Eingangsschallsignals bei der jeweiligen Frequenz verdeckt („maskiert“), sodass sie unhörbar sind. Dies führt zusammen mit anderen Effekten zu einer „effektiven Linearität“ unserer Schallwahrnehmung. Sie ist eine wichtige Voraussetzung für das Trennen von überlagerten akustischen „Objekten“ und lässt uns zudem selbst kleinste Nichtlinearitäten im Schallsignal erkennen (z. B. den „Klirrfaktor“ der HiFi-Anlage kurz vor deren

Übersteuerung). Bei Innenohr-Störungen entfällt die periphere Dynamik-Kompression. Die auf eine funktionierende Kompression eingestellte Signalverarbeitung im Gehirn bleibt aber bestehen, sodass das Gesamtsystem „effektiv nichtlinear“ wird [3]. Dies liefert eine Erklärungsmöglichkeit für den gestörten Cocktail-Party-Effekt bei Schwerhörigen: Möglicherweise interagieren bei der gestörten Schallwahrnehmung die verschiedenen akustischen Objekte nichtlinear miteinander und können daher nicht mehr so getrennt werden wie bei Normalhörenden.

^{*)} Die in diesem Artikel erwähnten Hörbeispiele sind unter <http://medi.uni-oldenburg.de/demos/pj/> und www.physik.de/Phy/PJ/kollmeier.html abrufbar. Zum Anhören sind eine Soundkarte und Lautsprecherboxen oder Kopfhörer erforderlich.

In jedem Frequenzband, das zu einem bestimmten Ort auf der Basilarmembran gehört, überträgt der Hörnerve nicht mehr die vollständige Feinstruktur mit der Phaseninformation des Eingangssignals, sondern approximativ seine Einhüllende, d. h. die langsamer schwankende momentane Schallintensität. Dabei wird ein sehr großer Dynamikbereich von bis zu 120 dB Pegeldifferenz zwischen Hörschwelle und Unbehaglichkeitsschwelle erreicht, obwohl jedes einzelne beteiligte Neuron nur maximal etwa 40 dB überdeckt. Die Mechanismen der Adaptation (Anpassen der Empfindlichkeit eines Neurons an den gerade vorliegenden mittleren Eingangspegel) und der Umsetzung der Schallintensität in einen subjektiven Lautheitseindruck sind noch nicht abschließend geklärt.

► **Modulationsabbildung:** Innerhalb jedes Frequenzbandes wird die Signal-Einhüllende in verschiedene Amplituden-Modulationsfrequenzen aufgespalten. Damit werden über sämtliche Frequenzen hinweg die verschiedenen Rhythmen des Eingangssignals, aber auch periodische Zeitstrukturen (z. B. die Grundfrequenz bei Sprachlauten) systematisch auf benachbarte Nervenzellen im Gehirn abgebildet. Diese zweidimensionale Abbildung (Mittelfrequenz versus Modulationsfrequenz) wurde aufgrund von physiologischen Messungen gefunden [6] und anhand psychoakustischer Messungen und Modelle bestätigt [7].

► **Binaurale (räumliche) Abbildung:** Um den Ort bzw. die Einfallsrichtung einer Schallquelle zu ermitteln, wertet das Gehirn die zwischen beiden Ohren auftretende Laufzeit- und Pegeldifferenz aus und setzt sie in eine „innere Karte“ der akustischen Umwelt um. Dabei wird eine Winkelauflösung von bis zu 1 Grad und eine Zeitgenauigkeit von bis zu 20 μ s erreicht – eine der phantastischsten Leistungen unseres Gehörs! (Siehe Infokasten auf der folgenden Seite).

Modelle der „effektiven“ Signalverarbeitung

Aus Sicht der Physik möchte man vor allem die Gesetzmäßigkeiten und Mechanismen verstehen, die möglichst vielen der beobachtbaren Phänomene zugrunde liegen, während eine detaillierte Nachbildung jeder einzelnen Nervenzelle zunächst nicht im Vordergrund steht. Ein wichtiges Anliegen der physikalischen Hörforschung ist daher die Entwicklung von quantitativen Hörmodellen, die mit einer möglichst kleinen Zahl von Annahmen und einzustellenden Parametern eine möglichst große Variationsbreite von Experimenten quantitativ vorhersagen und anhand „kritischer“ Experimente bestätigt oder falsifiziert werden können. Obwohl dieser Modellansatz damit von vornherein starken Einschränkungen unterliegt, kann er unser derzeitiges Wissen über die „effektive“ Verarbeitung akustischer Information überprüfbar zusammenfassen und zu einem funktionalen Verständnis des Hörvorgangs führen.

Der prinzipielle Aufbau eines derartigen Modells der „effektiven Verarbeitung“ im auditorischen System ist in Abb. 4 dargestellt. Einige der Funktionsblöcke wurden direkt aus physiologischen Erkenntnissen abgeleitet. So wird z. B. die Funktion der Basilarmembran als „effektive“ Bank von Bandpassfiltern modelliert, die „effektive“ Funktion von Haarzellen und auditorischem Nerv als Kompression und Einhüllendenbildung, die Funktionsprinzipien des Hirnstamms als Modulations-Filterbank und binaurale Verarbeitung, und schließlich die Funktion des Kortex als Mustererken-

nung und Interpretation der „internen Repräsentation“ des Schalls. Andere Funktionsblöcke (z. B. das interne Rauschen) und die funktionelle Gestaltung und Parameterwahl sämtlicher Blöcke stammen dagegen aus psychoakustischen Experimenten, zu deren Vorhersage das Modell primär eingesetzt wird. Ebenso stammen die hypothetischen Funktionsstörungen, die bei einem sensorineuralen Hörverlust auftreten können (Sternchen in Abb. 4), aus audiologischen Messungen mit schwerhörigen Patienten.

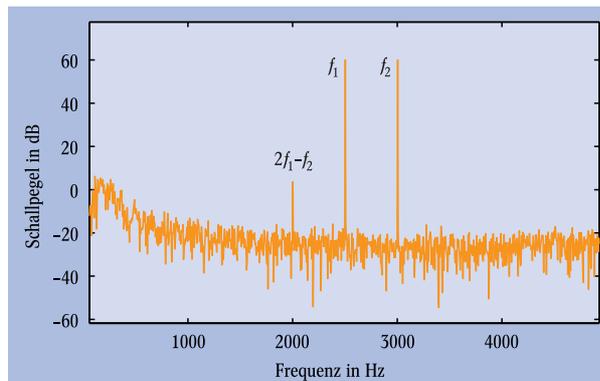


Abb. 3: Das „Echo“ des Innenohrs. Angeboten werden im abgeschlossenen Gehörgang zwei Sinustöne bei den Frequenzen f_1 und f_2 . Bei der Schallaufnahme mit einem empfindlichen Mikrofon erscheint im hier gezeigten gemittelten Spektrum bei gesundem Gehör zusätzlich das kubische Verzerrungsprodukt bei $2f_1-f_2$.

Um den prinzipiellen Rahmen des Modells in ein konkretes, numerisches Verarbeitungsmodell umzusetzen, sind umfangreiche theoretische und experimentelle Arbeiten notwendig. Typischerweise wird dem Modell am Eingang dasselbe Signal präsentiert wie der Versuchsperson, etwa ein Ton, versteckt im Rauschen. Vor dem letzten Block des Modells bildet sich dieses Eingangssignal zu jedem Zeitpunkt als mehrdimensionale Intensitätsverteilung von z. B. Frequenz, Modulationsfrequenz und binauraler Komponente ab. Aus der Ähnlichkeit des Musters mit einem zuvor gespeicherten Muster eines deutlich aus dem Rauschen herausragenden Tones berechnet nun der zentrale Mustererkenner

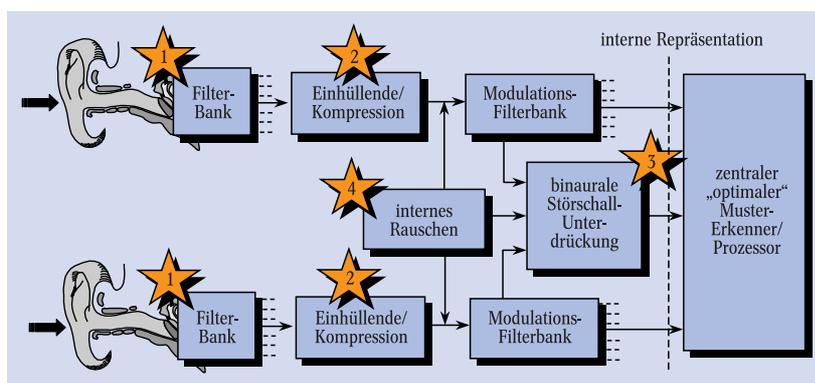


Abb. 4: Das Ohr aus Sicht des Physikers ist ein Modell der „effektiven“ Verarbeitung im auditorischen System. Es ist durch eine Reihe von parallelen, linearen und nichtlinearen Signalverarbeitungs-Operationen gekennzeichnet, die die Transformation von akustischen Signal in seine „interne Repräsentation“ im Gehirn beschreiben. Diese Repräsentation dient unserem Gehirn als Basis für höhere kognitive Leistungen wie das Verstehen von Sprache, wobei im Modell die Minderung dieser Leistungen allein durch die Unschärfe der internen Repräsentation angesetzt wird. Mögliche Störungen dieser Repräsentation bei Schwerhörigkeit sind durch Sternchen gekennzeichnet.

die Wahrscheinlichkeit, dass der Ton von der Versuchsperson wahrgenommen wird. Diese Wahrnehmbarkeit wird dann mit Experimenten verglichen, bei denen die Versuchsperson exakt dieselben Signale vorgespielt bekommt und dieselbe Erkennungsaufgabe durchführt wie der Computer. Anschließend kann das Modell durch die Vorhersage von Hörexperimenten mit vielen unterschiedlichen Signalen und Wahrnehmungsaufgaben überprüft und iterativ verbessert werden [7].

Der Teil des Modells vor dem Mustererkenner entspricht beim Menschen der Verarbeitung eines akustischen Signals zu einer „internen Repräsentation“, etwa einem neuronalen Erregungsmuster. Der kognitive Apparat, mit dem der Mensch solch ein Erregungsmuster interpretiert, wird durch den Mustererkenner modelliert. Am Ende dieses Prozesses steht eine Aussage wie „Die Versuchsperson hört den Ton.“ Sie ermöglicht den Vergleich von Theorie und Experiment. Dabei setzt jede Forschungsgruppe unterschiedliche Schwerpunkte im Detaillierungsgrad einzelner Funktionsblöcke des Modells und bei der Klasse von vorhersagbaren Experimenten:

► Das in München von Zwicker und Mitarbeitern entwickelte Lautheitsmodell [9] mit seinen Erweiterungen für die Beschreibung von Schwerhörigkeit und Schallfluktuationen beschreibt beispielsweise die primär von der Schallintensität, dem Schallspektrum und der Schalldauer abhängige Lautheitswahrnehmung, wobei die Blöcke „Modulationsfilterbank“, „Internes Rauschen“ und „binaurale Störschallunterdrückung“ aus Abb. 4 entfallen.

► Das Boston-Modell zur binauralen Informationsverarbeitung von Colburn und Mitarbeitern beschreibt dagegen die verschiedenen Leistungen des binauralen (zweiohrigen) Hörens unter expliziter Modellierung von Neuroneneigenschaften [10]. Andere binaurale Funktionsmodelle (z. B. das Bochumer Modell von Blauert und Mitarbeitern [11]) benutzen nachrichtentechnische Funktionselemente, um charakteristische Eigenschaften der binauralen Informationsverarbeitung im Gehirn funktionell zu modellieren.

► Das Modell der Cambridge-Arbeitsgruppe um Patterson und Meddis setzt dagegen Schwerpunkte bei der Tonhöhererkennung und dem Übergang von der Wahrnehmung aperiodischer Vorgänge in periodische Vorgänge mit zugehörigem „pitch“ [12].

► Das „Oldenburger Perzeptionsmodell“ legt einen besonderen Schwerpunkt auf die zeitlichen Eigenschaften der Signalverarbeitung im Gehör und bildet eine relativ große Zahl von psychoakustischen Effekten quantitativ nach. Es wurde zunächst am III. Physikalischen Institut in Göttingen und ab 1993 im Graduiertenkolleg „Psychoakustik“ der Universität Oldenburg von Dau et al. [7] als Modell der „effektiven“ Signalverarbeitung im auditorischen System entwickelt, und auf die Vorhersage von Sprachverständlichkeit bei Normal- und Schwerhörenden [13] sowie das binaurale Hören erweitert. Das Modell basiert auf einer geringen Zahl von Annahmen und Parametern, die in wenigen, „kritischen“ Experimenten festgelegt und für die quantitative Beschreibung anderer Experimente nicht mehr variiert werden. Eine Stärke des Oldenburger Modells ist die „optimale“ Mustererkennung: Der Algorithmus des Mustererkenners in Abb. 4 wird mithilfe empirischer Daten „geschult“. Auf diese Weise gehen Aspekte wie Aufmerksamkeit und Lernen nicht mehr explizit in die Modellierung ein, weil der „optimale Detektor“ bereits die „interne Repräsentation“ des Schallsignals perfekt interpretiert. Die gesamte Ungenauigkeit des Hörvorgangs wird auf die Nichtlinearität der Signalverarbeitung und das interne, neuronale Rauschen bei der Transformation vom akustischen Signal bis hin zu der internen Repräsentation reduziert. Dieses Vorgehen hat den entscheidenden Vorteil, dass man sich nur auf denjenigen Teil der Wahrnehmungsleistungen beschränkt, der reproduzierbaren psychophysikalischen Experimenten zugänglich ist und sich zudem möglicherweise physiologisch bestätigen lässt. Komplexe psychische Einflussfaktoren der menschlichen Wahrnehmung akustischer Ereignisse werden in der Modellierung dagegen nicht berücksichtigt.

Das Ohr als Spektralapparat oder als Zeit-Analysator?

Während frühere Hörtheorien und Erklärungen des Sprachverstehens vorwiegend von einer spektralen Sichtweise des Hörvorgangs ausgingen (z. B. Unterscheidung der Sprachvokale durch die Lage der spektralen Maxima/Formanten) und den zeitlichen Aspekt der Hörwahrnehmung als sekundär ansahen, verhält es sich bei modernen Hörtheorien genau umgekehrt: Das Ohr ist nicht nur das schnellste Sinnessystem des Menschen, es kann auch die einem akustischen Signal aufgeprägte zeitliche Information sehr genau verfolgen. So lassen sich zeitliche Lücken ab einer Dauer von ca. 5 ms in einem breitbandigen Signal sicher detektieren. Begrenzt wird die Zeitauflösung durch die Vor- und Nachverdeckung, d. h. ein Testsignal kann ab ca. 10 ms vor und bis zu 200 ms nach einem (lauteren) Maskierungssignal nicht mehr gehört werden. Das liegt an der Trägheit des zentralen Hörsystems bei der Anpassung an einen neuen Pegel.

Ein Beispiel gegen die rein spektrale Sichtweise ist die Verständlichkeit von „flat-spectrum speech“ [8], d. h. von gefilterter Sprache, deren Kurzzeitspektren ohne spektrale Information, also flach sind (Hörbeispiel 3). Ein anderes Beispiel für Sprachverstehen ohne intaktes Sprachspektrum ist Bandpassgefilterte Sprache, die in einer spektralen Lücke eines Rauschens dargeboten wird. Umgekehrt wird das Sprachverstehen durch eine Veränderung der

Zeitstruktur stärker beeinträchtigt als durch eine spektrale Verformung: So nimmt die Sprachverständlichkeit bei periodisch unterbrochener Sprache ab einer bestimmten Unterbrechungsrate sehr stark ab. Wenn in die zeitlichen Lücken jedoch anstelle der Originalsprache ein Rauschen mit festem Spektrum gefüllt wird, ist die Zeitstruktur nur noch relativ wenig gestört und unser Gehirn ist in der Lage, die Sprachinformation wieder zusammenzusetzen (Hörbeispiel 4): Das Abwechseln von Sprachsegmenten mit Rauschsegmenten hört sich wie stark verrauschte, kontinuierliche Sprache an, d. h. durch Zufügen (!) von Rauschen wird die Sprache „entstört“!

Die plausibelste Hörtheorie ist daher eine Kombination der spektralen und zeitlichen Analyse im Sinne einer „Demodulation“ der in jedem Frequenzband vorhandenen Zeit-Information durch das Innenohr mit anschließender Einhüllenden-Analyse in Modulationsfrequenzbändern im Gehirn. Besondere Bedeutung kommt dabei zeitlichen Merkmalen und Modulationsfrequenzen zu, die in mehreren Frequenzbändern gleichzeitig auftreten. Dadurch ist das Ohr in der Lage, akustische „Objekte“ (zum Beispiel Sprache) auch bei sehr ungünstigem Signal-zu-Rausch-Verhältnis noch sicher zu erkennen, bei dem das mittlere Sprachspektrum schon vollständig verdeckt ist.

Modellierung gestörter Hörfunktionen

Ein möglichst quantitatives Verständnis der gestörten Signalverarbeitung bei Schallempfindungs-Schwerhörigkeit ist sowohl für die Hördiagnostik als auch für die optimale Rehabilitation, z. B. mit „intelligenten“ Hörgeräten, unabdingbar. Wegen der Vielzahl der gestörten Einzelleistungen bei Schwerhörigkeit ist es eine besondere Herausforderung, die ursächlichen oder primären Defizite der Hör-Signalverarbeitung von den daraus ableitbaren Defiziten anderer Hörfunktionen zu trennen. Im Rahmen des in Abb. 4 dargestellten Modellschemas lassen sich nun die vier wichtigsten primären Komponenten von Hörstörungen (Sterne in Abb. 4) wie folgt charakterisieren:

1) *Abschwächungswirkung des Hörschadens (lineare Dämpfung)*: Eine Schallleitungs-Schwerhörigkeit oder

ein Ausfall der inneren Haarzellen führt vorwiegend zu einer Sensitivitäts-Verminderung, d. h. einer effektiven „Abschwächung“ des Schalls. Sie kann durch eine entsprechende lineare Verstärkung des Schalls kompensiert werden. Dies bewirkt aber meistens keine zufriedenstellende Wiederherstellung des Hörvermögens, sodass weitere Komponenten betrachtet werden müssen.

2) **Kompressionsverlust:** Ein Ausfall der äußeren Haarzellen führt zusätzlich zur Abschwächung zu einer „Verzerrung“: Bei niedrigen Pegeln entfällt die aktive Verstärkung, sodass sich der große Dynamikbereich der akustischen Eingangssignale nicht mehr vollständig im Gehirn abbilden lässt. Dies macht sich beim für die meisten Innenohr-Schwerhörigen typischen „Recruitment“-Phänomen bemerkbar (Hörbeispiel 5), bei dem nach dem subjektiven Eindruck „zu leise“ bei leichter Erhöhung des Schallpegels bereits der Eindruck „zu laut“ folgt. Diese gestörte Lautheitswahrnehmung kann durch eine Multiband-Dynamikkompression in modernen Hörgeräten nur teilweise kompensiert werden, da z. B. die Bandbreiten-Abhängigkeit der unterschiedlichen Lautheitswahrnehmung bei Normal- und Schwerhörigen nicht berücksichtigt wird. Die Entwicklung adäquater Lautheitsmodelle für Schwerhörige und ihre Integration in Hörgeräte ist daher Gegenstand laufender Forschung [14].

3) **Binauraler Hörverlust:** Normalhörende können die an beiden Ohren eintreffenden Signale im Gehirn vergleichen und durch binaurale (beidohrige) Signalverarbeitung den wahrnehmbaren Nachhall verringern und unerwünschte Schalleinfallrichtungen ausblenden. Bei Schwerhörigen kann aber – weitgehend unabhängig von den übrigen bisher genannten Faktoren der Hörstörung – genau diese binaurale Signalverarbeitung gestört sein. Dies bedingt u. a. die eingangs erwähnte Störung des „Cocktail-Party-Effektes“. In der derzeitigen Routine-Diagnostik und Hörgeräteversorgung mit unabhängigen Geräten auf beiden Seiten wird allerdings dieser Faktor noch nicht berücksichtigt. Erst „echt“ binaurale Hörgeräte versprechen Abhilfe.

4) **Zentrale Hörstörung:** Selbst bei nur gering gestörter Signalverarbeitung durch das Hörsystem kann bei Schwerhörigen die Auflösung der internen Repräsentation verringert sein, sodass die vom Gehör aufgenommenen und intern repräsentierten Schallsignale nicht mehr adäquat ausgewertet und interpretiert werden können. Dieser Effekt lässt sich ebenso wie andere Unzulänglichkeiten der zentralen Auswerteeinheit, z. B. geringe Aufmerksamkeit, mangelndes Training, am ehesten durch ein erhöhtes „internes Rauschen“ modellieren.

Sämtliche der hier genannten Komponenten tragen bei dem individuellen Patienten in unterschiedlichem Maße zu der Hörstörung und beispielsweise zur Verringerung des Sprachverstehens in Störgeräuschen bei. Daher ist es das Ziel aktueller Forschungsarbeiten, effiziente Messmethoden zu entwickeln, mit denen sich jede der Komponenten erfassen lässt, sowie das Modell so anzupassen, dass es das jeweilige Hörvermögen jedes individuellen Patienten korrekt beschreibt [15].

Anwendungen des Perzeptionsmodells

Unter der Voraussetzung, dass das oben beschriebene Perzeptionsmodell eine valide und objektive Beschreibung der Transformation des akustischen Signals in seine „interne Repräsentation“ im menschlichen Gehirn darstellt, erschließen sich eine Reihe von technischen Anwendungen:

► **Signalkodierung:** Die Bitraten-Reduktion bei der Speicherung von Sprach- und Audiodaten, z. B. mit dem MP3-Verfahren, kodiert das akustische Signal so, dass möglichst geringe Abweichungen zwischen Original und dekodiertem Signal auf der „perzeptiven“ Ebene am Ausgang des Hörmodells auftreten, obwohl diese Signale auf der akustischen Ebene am Eingang des Hörmodells sehr unterschiedlich sein können. Bei der standardisierten MP3-Kodierung wird im Wesentlichen das von Zwicker vor mehreren Jahrzehnten entwickelte Lautheits- und Maskierungsmodell verwendet, um unhörbare Signalbestandteile zu eliminieren und das Quantisierungsrauschen hinter den hörbaren Komponenten zu „verstecken“.

► **Signalqualitäts-Bewertung:** Der Unterschied am Ausgang des Hörmodells wird auch bei der objektiven Güte-Beurteilung von kodierten bzw. nichtlinear verarbeiteten Sprach- und Audiosignalen ausgewertet, die bis-

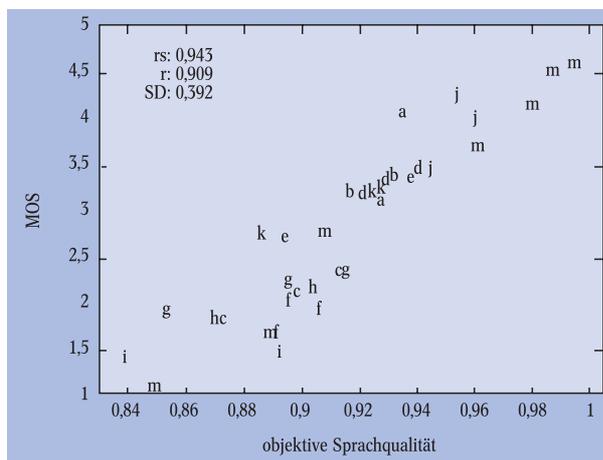


Abb. 5: Sprachqualitäts-Vorhersage (aus [16]): Die Ordinate zeigt das subjektive Qualitätsurteil einer Gruppe normalhörender Probanden (als mean opinion score, MOS) für eine Test-Datenbank von verschiedenen Mobilfunk-Verbindungen. Die Abszisse zeigt die Vorhersage dieser Daten auf der Basis des Oldenburger Perzeptionsmodells für verschiedene Kodierungsverfahren (Buchstaben). Die hohen Korrelations-Indices r und r_s bei niedriger Standard-Abweichung (SD) zeigen eine gute objektive Vorhersage der subjektiven Ergebnisse an.

her nur subjektiv mit aufwändigen Hörexperimenten ermittelt werden konnte. Beispielsweise lässt sich die wahrgenommene akustische Qualität einer Handy-Mobilfunkverbindung durch das Oldenburger Perzeptionsmodell mit dem Computer vorhersagen (objektiver Parameter q_c aufgetragen auf der Abszisse in Abb. 5). Der Vergleich mit dem subjektiven „Mean opinion score“ auf der Ordinate in Abb. 5 zeigt eine hohe Treffsicherheit der objektiven Vorhersage [16].

► **Sprach- und Mustererkennung:** Dem Computer werden „Ohren verliehen“, indem nicht eine technische Darstellung des Sprachsignals, sondern die „interne Repräsentation“ vom Ausgang des Gehörmodells als Eingangssignal für einen Spracherkennungs- bzw. Mustererkennungs-Algorithmus benutzt wird. Die Robustheit unseres Ohres gegenüber Störgeräuschen und Änderungen der Raumakustik soll damit auf den Computer übertragen werden: Abbildung 6 zeigt die Worterkennungsrate in Prozent als Funktion des Signal-zu-Rausch-Verhältnisses. Bei der konventionellen Sprachvorverarbeitung mit dem sog. MFCC-Verfahren (gestrichelte Linie) sinkt die Erkennungsrate mit zunehmendem Rauschen viel eher als bei der Sprachvorver-

arbeitung mit dem Oldenburger Perzeptionsmodell (durchgezogene Linie, aus [17]). Bei einer stark eingeschränkten Zahl von möglichen Alternativen des zu erkennenden Wortes oder Satzes erreicht der Computer sogar die Erkennungsleistung des menschlichen Gehörs – Anlass genug, das Gehörmodell in Zusammenarbeit mit der Informatik als Computer-Chip zu entwerfen („Silicon Ear“).

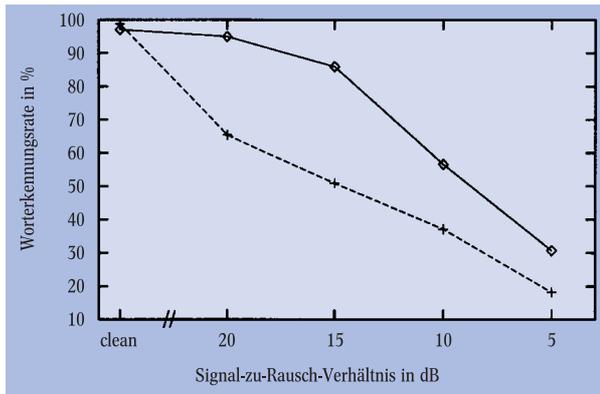


Abb. 6: Robuste Spracherkennung (aus [17]): Dargestellt ist die Worterkennungsrate eines künstlichen Spracherkenners unter Ruhebedingungen („clean“) und als Funktion des Signal-zu-Rausch-Verhältnisses für Sprache in Baustellenlärm für die konventionelle Sprachvorverarbeitung (MFCC, gestrichelte Linie) und für das Oldenburger Perzeptionsmodell (durchgezogene Linie).



Abb. 7: Digitales, binaurales Prototyp-Hörgerät, das im Rahmen eines BMBF-Verbundprojekts zusammen mit der FH Nürnberg und der Universität Gießen entwickelt wurde.

► „Intelligente“ Hörgeräte: Ziel der Signal-Manipulationen im Hörgerät sollte es sein, den Unterschied der „internen Repräsentation“ am Ausgang des Hörmodells zwischen Normal- und Schwerhörigen zu minimieren. Dies setzt jedoch neben dem Modell für das normale Gehör auch ein Modell des gestörten Hörvermögens voraus, das individuell angepasst wird. Obwohl inzwischen digitale Hörgeräte mit einem vergleichbaren Konzept auf dem Markt sind, ist dies noch ein Bereich aktueller Forschungs- und Entwicklungsarbeiten.

Intelligente Hörgeräte der Zukunft

Ein wichtiges Beispiel für die Umsetzung von Hörmodellen in die Praxis ist die Störgeräuschunterdrückung in Hörgeräten, die möglichst zu einer Wiederherstellung des „Cocktail-Party-Effektes“ bei Schwerhörigen führen soll und die Auswirkungen der reduzierten binauralen Interaktion und des erhöhten internen Rauschens zumindest teilweise kompensieren soll. Obwohl eine für alle möglichen akustischen Stör-schall-Nutzschall-Konfigurationen wirksame Störunterdrückung auch für andere Anwendungen in der Sprachkommunikation (z. B. automatische Spracher-

kennung) sehr erstrebenswert ist, zählt dies zu den großen, noch ungelösten Problemen der Akustik.

Einen vielversprechenden Ansatz bietet die Analyse mit dem so genannten Amplituden-Modulations-Spektrogramm (AMS), das in jedem Frequenzband die zeitlichen Fluktuationen in verschiedene Modulations-Frequenzen zerlegt. Es entspricht der bereits in Abb. 4 dargestellten Modellvorstellung, dass in jedem auditorischen Frequenzband die zeitliche Einhüllenden-Struktur durch eine Modulations-Filterbank ausgewertet wird. In dieser Darstellung ist Sprache durch in mehreren Frequenzbändern vorhandene kohärente Modulationen im Modulationsfrequenzbereich von 4 Hz (Silbenfrequenz) und mehreren hundert Hz (Sprachgrundfrequenz mit Harmonischen) gekennzeichnet. Störgeräusche weisen hingegen in der Regel weniger kohärente Modulationen und auch ein anderes Modulationsspektrum auf, sodass sich diese Unterschiede gut für die Störgeräuschunterdrückung ausnutzen lassen. Der Vorteil dieses Verfahrens ist seine Anwendbarkeit auch für monaurale (einkanalige) Mikrofon-signale und für fluktuierende Hintergrundgeräusche (Hörbeispiel 6). Der Nachteil des Verfahrens ist jedoch der hohe Rechenaufwand. Außerdem versagt es, wenn das Hintergrundgeräusch selbst Sprache (z. B. ein weiterer Sprecher) ist.

Daher bietet ein binaurales Verfahren Vorteile, bei dem die Signale an beiden Ohren aufgenommen werden und in einer zentralen Recheneinheit so gefiltert werden, dass die von vorn kommenden Signalanteile, d. h. der Nutzschall, verstärkt und der von anderen Richtungen kommende Störschall unterdrückt wird [18]. Damit wird das binaurale Hören gewissermaßen simuliert. Ein derartiges binaurales Hörgerät hat deutliche Vorteile gegenüber zwei unabhängigen Hörgeräten auf beiden Seiten und erst recht gegenüber einem monauralen Hörgerät. Ein binaurales Hörgerät ist zwar noch nicht kommerziell, zusammen mit der FH Nürnberg, der Uni Gießen und weiteren Partnern wurde aber bereits ein tragbarer Prototyp entwickelt (Abb. 7 und Hörbeispiel 7 für die Verarbeitungsleistung binauraler Hörgeräte-Algorithmen). Für ein kommerzielles binaurales Hörgerät ist jedoch eine drahtlose Verbindung zwischen den Geräten nötig, die wegen der erforderlichen Stromaufnahme noch eine große technologische Herausforderung darstellt.

Wie wird die Entwicklung der Hörgeräte weitergehen? Es ist abzusehen, dass das Hörgerät nur noch eine Option eines „Personal Communication Devices“ der Zukunft darstellen wird, in dem Mensch-Maschine- und Mensch-Mensch-Kommunikationsfunktionen wie MP3-Player, Handy, Laptop und eben Hörgerät in einem tragbaren und durch Sprachsteuerung bedienbaren Gerät verschmelzen. Zukünftige Hörgeräte werden zudem binaural sein und neben der modellgesteuerten Dynamikkompression eine an die jeweilige Kommunikationssituation optimal angepasste Stör-Reduktion mit automatischer Programmwahl aufweisen.

Eine Verbesserung dieser für den audiologischen Erfolg wichtigen Signalverarbeitungsleistungen von Hörgeräten ist u. a. Ziel des ab 1. Januar 2001 eingerichteten Oldenburger Kompetenzzentrums HörTech (www.hoertech.de), das zu den acht Gewinnern des BMBF-Wettbewerbs „Kompetenzzentren für die Medizintechnik“ zählt.

*

Gefördert von der DFG, dem BMBF und dem Land Niedersachsen. Herzlicher Dank allen Mitarbeiterinnen und Mitarbeitern der Medizinischen Physik, Universität Oldenburg.

Literatur

- [1] *H. v. Helmholtz*, Die Lehre von den Tonempfindungen als physiologische Grundlage fuer die Theorie der Musik, Vieweg, Braunschweig 1870
- [2] *G. Böhme* und *K. Welzl-Müller*, Audiometrie. Hörprüfungen im Erwachsenen- und Kindesalter, Verlag Hans Huber, Bern 1998.
- [3] *R.-P. Derleth*, *T. Dau* und *B. Kollmeier*, Hearing Research **159**, 132 (2001).
- [4] *S. Uppenkamp*, *J. Neumann*, und *B. Kollmeier*, Hearing Research **78**, 210 (1994).
- [5] *H. Zenner*, Hören, Thieme, Stuttgart 1994.
- [6] *C.E. Schreiner* und *G. Langner*, Nature **388**, 383 (1997)
- [7] *T. Dau*, *B. Kollmeier* und *A. Kohlrausch*, J. Acoust. Soc. Am. **102**, 2892 (1997).
- [8] *M. R. Schroeder*, Computer Speech: Recognition, Compression, Synthesis, Springer, Berlin 1999.
- [9] *E. Zwicker* und *H. Fastl*, Psychoacoustics – Facts and Models, Springer, Berlin 1990.
- [10] *H. S. Colburn*, in: Auditory Computation, Springer, New York 1996, S. 332.
- [11] *J. Blauert*, Spatial Hearing, MIT Press, Cambridge 1997.
- [12] *R. D. Patterson*, *M. Allerhand* und *C. Giguere*, J. Acoust. Soc. Am. **98**, 1890 (1995)
- [13] *I. Holube* und *B. Kollmeier*, J. Acoust. Soc. Am. **100**, 1703 (1996).
- [14] *V. Hohmann* und *B. Kollmeier*, in: Psychoacoustics, Speech and Hearing Aids, World Scientific, Singapore 1996, S. 193.
- [15] *J. Kießling*, *B. Kollmeier* und *G. Diller*, Versorgung und Rehabilitation mit Hörgeräten, Thieme Verlag, Stuttgart 1997.
- [16] *M. Hansen* und *B. Kollmeier*, J. Audio Eng. Soc. **48**, 395 (2000).
- [17] *J. Tchorz* und *B. Kollmeier*, J. Acoustical Soc. Am. **106**, 2040 (1999).
- [18] *T. Wittkop* et al., Acustica united with acta acustica **83**, 684 (1997)

Der Autor

Birger Kollmeier hat seit seiner Promotion in Physik und seinem Medizin-Studium in Göttingen immer den Verdacht erregt, weder Fisch noch Fleisch zu sein. Dazu trägt die Ausrichtung seiner Arbeitsgebiete Hör-, Sprach- und Hirnforschung wesentlich bei. Obwohl er sich mit seiner Habilitation in Physik in Göttingen und seiner Berufung an die Universität Oldenburg als Professor für Angewandte Physik/Experimentalphysik eindeutig festlegte, ist sein Motto der Brückenschlag zwischen den Fächern und die Anwendungsorientierung geblieben: Mit der Abteilung Medizinische Physik, dem Europäischen Graduiertenkolleg Neurosensorik, dem An-Institut Hörzentrum Oldenburg GmbH und dem Kompetenzzentrum HörTech hat er interdisziplinäre Strukturen aufgebaut, die er heute in Oldenburg leitet. Ausgezeichnet wurde er u. a. mit dem Lothar-Cremer-Preis der Dt. Gesellschaft für Akustik und dem Forschungspreis Technische Kommunikation der Alcatel-SEL-Stiftung. Sein Stolz gilt – neben den eigenen vier Kindern – den bisher 25 „Doktorkindern“, die u. a. bei ca. 70 % des Hörgeräte-Weltmarktes in der Entwicklung eine wichtige Rolle spielen.

