

Modell der effektiven Signalverarbeitung im Gehör

Von Torsten Dau

In den letzten Jahren ist in Oldenburg ein Signalverarbeitungsmodell entwickelt worden, das eine Vielzahl von experimentell ermittelten Leistungen des menschlichen Gehörs quantitativ beschreiben kann. Das Modell enthält dabei sowohl physiologisches Wissen über die neuronale Verarbeitung im Gehör als auch physikalische Prinzipien bei der Signalerkennung. Besonders interessant ist die Fähigkeit des Modells, die Verarbeitung von zeitlich in der Amplitude schwankenden (modulierten) Signalen "gehörgerecht" nachzubilden. Sämtliche in der Natur vorkommenden Signale, insbesondere Sprache, sind durch solche Modulationen gekennzeichnet. Aufbauend auf dem Modell können die Sprachverständlichkeit in Ruhe und unter Störgeräusch sowie die Sprachübertragungsqualität von Kodiersystemen (z.B. in Mobiltelefonen) erfolgreich vorhergesagt werden.



Vorbereitungen auf einen psychoakustischen Test, bei dem einer Versuchsperson Schallsignale über Kopfhörer dargeboten werden und die Antworten durch Antippen eines berührungssensitiven Bildschirms an den Computer gemeldet werden.

A quantitative model of the effective signal processing in the auditory system

Over the last few years a signal processing model has been developed in Oldenburg that is able to quantitatively reproduce a wide range of data from experiments in human hearing. The model incorporates physiological aspects of neural processing in the auditory system as well as the physical principles involved in signal detection. The ability of the model to simulate the processing of temporally fluctuating (modulated) sounds is particularly interesting. Many naturally occurring sounds, in particular speech, are characterized by such fluctuations in level. Using the model, it is possible to successfully predict speech intelligibility in quiet and in noisy conditions, as well as the speech transmission quality of various coding systems, such as those used in mobile telephones.

Unser Gehör ist in beeindruckender Weise an das Erfassen von akustischen Signalen aus der Umwelt angepaßt, insbesondere an das Verstehen von Sprache. Akustische Sprachsignale weisen sehr unterschiedliche Frequenzanteile auf und haben vor allem die Eigenschaft, sich zeitlich stark zu ändern. Das Gehör benötigt also für die Sprachwahrnehmung die Fähigkeit, sozusagen zu jedem Zeitpunkt die Intensität wahrzunehmen, mit der jede im Schall vorkommende Frequenz momentan vorliegt. Aber wie macht es das? Was bedeutet es, wenn wir sagen, daß das Gehör einen Schall oder ein bestimmtes Schallmerkmal "wahrnimmt"? Und vor allem: Wie können wir die Verarbeitung des Schalls von den Luftdruckschwankungen, die am Außenohr anliegen, über die Umsetzung dieser Schwingungen in neuronale Erregungsmuster bis hin zur Wahrnehmung am Ende der Verarbeitungskette modellhaft beschreiben, ohne gleich an den komplizierten anatomischen und physiologischen Details der einzelnen Stufen zu "scheitern"? Unser Ansatz wird sein, uns dem "System" physikalisch bzw. nachrichtentechnisch zu nähern und zu versuchen, die Signalverarbeitung im Gehör durch eine Reihe von Funktions-

elementen modellhaft zu beschreiben. Darum werden wir es in diesem Beitrag mit der "effektiven" Signalverarbeitung beim Hören zu tun haben.

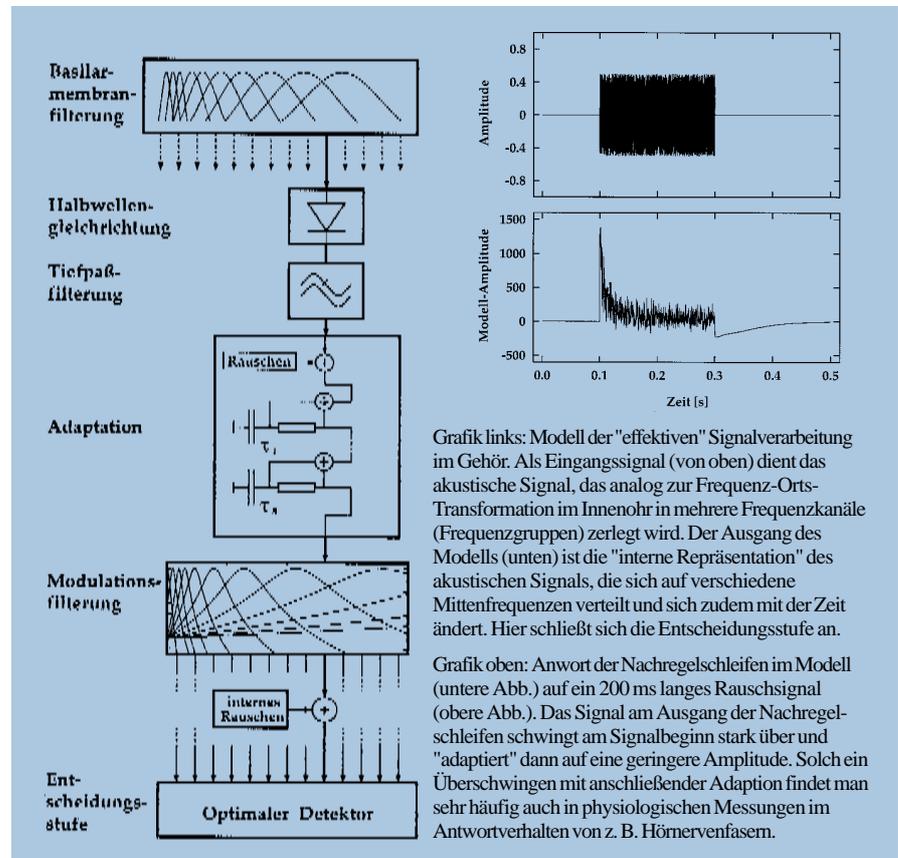
Was ist Psychoakustik?

Grundsätzlich ist diese Art der Modellierung in dem Bereich der Psychoakustik angesiedelt, der die Zusammenhänge zwischen den akustischen Reizen und den durch sie beim Menschen hervorgerufenen Empfindungen beschreibt. Die speziellen Meßmethoden zur Erfassung der Empfindungsgrößen können dabei sehr unterschiedlich sein, je nachdem, ob es sich z.B. um die Erkennung (Detektion) eines Signals in Anwesenheit eines Störschalls (Maskierer), um die Identifikation von Schall oder auch um die Bewertung oder Skalierung eines bestimmten akustischen Reizes oder Reizattributs handelt (wie z.B. die wahrgenommene Lautheit, Rauigkeit oder Klangfarbe). Der quantitative funktionale Zusammenhang zwischen Reiz und Empfindung, nach dem wir suchen, reicht in der Literatur von einfachen Beschreibungen bis hin zu komplexen Modellen der auditorischen Signalverarbeitung, in denen

Kenntnisse über die neuronale Verarbeitung von Schall im Nervensystem sowie eine Bewertungsstufe (z.B. ein "Detektor") am Ende dieser Vorverarbeitung enthalten sind. Im folgenden schauen wir uns ein Modell an, das entwickelt wurde, um Leistungen und Grenzen des Gehörs bei der sogenannten "differenziellen" Verarbeitung zu beschreiben. Wie gut sind wir z.B. in der Lage, die Frequenzen zweier Töne bzw. die durch sie hervorgerufenen Tonhöhen gerade voneinander zu unterscheiden? Wie gut können wir einem Schall zeitlich folgen? Wie gut sind also die spektralen und zeitlichen Auflösungsgrenzen unserer auditorischen Wahrnehmung? Bereits diese grundlegenden Aspekte des Hörens sind sehr entscheidend für spätere Anwendungen wie z.B. die Entwicklung von modernen Hörgeräten.

Vom Gehör zum Modell

Starten wir bei den ersten, sogenannten "peripheren", Verarbeitungsstufen wie dem Außen-, Mittel-, und Innenohr und arbeiten uns dann zu "zentraleren" Stufen der Hörbahn vor, die im sogenannten Hirnstamm und schließlich in der Hirnrinde, dem auditorischen Kortex, zu finden sind. Das Außenohr dient vor allem zur richtungsabhängigen Verfärbung (Filterung) des auf das Ohr einfallenden akustischen Signals. Diese je nach Einfallsrichtung unterschiedliche Klangfärbung kann bereits zur Ortung von Schallquellen verwendet werden. Der Schall wird dann durch das Mittelohr an das flüssigkeitsgefüllte Innenohr weitergeleitet. Das Mittelohr ist dabei so aufgebaut, daß es eine fast verlustfreie Energieübertragung zwischen der akustischen Wellenfortbewegung in Luft (im Außenohr) und der Wellenausbreitung in den mit Flüssigkeit gefüllten Kammern im Innenohr ermöglicht. Im Innenohr wird der Schall in verschiedene Frequenzanteile zerlegt und es findet eine *Frequenz-Orts-Transformation* statt: unterschiedliche Frequenzen werden an unterschiedlichen Orten abgebildet. Dieses Ordnungsprinzip wird *Tonotopie* genannt und setzt sich auch auf den weiteren "Stationen" der Hörbahn bei der Reizfortleitung ins Gehirn fort. Aus physikalischer Sicht entspricht die Abbildung im Innenohr einer *Filterbank*: der Schall wird spektral in verschiedene bandpaßgefilterte Signale, sogenannte Frequenzgruppen, zerlegt. Anschließend werden die mechanischen Schwingungen in den einzelnen Frequenzgruppen durch sogenannte Haarzellen in Nerveneregerungen umgewandelt. Bei tiefen Frequenzen des Schalls können diese Nerveneregerungen dem genauen Verlauf des Schalls folgen, während sie dies für hohe Frequenzen nur mit einer gewissen Trägheit können. Physi-



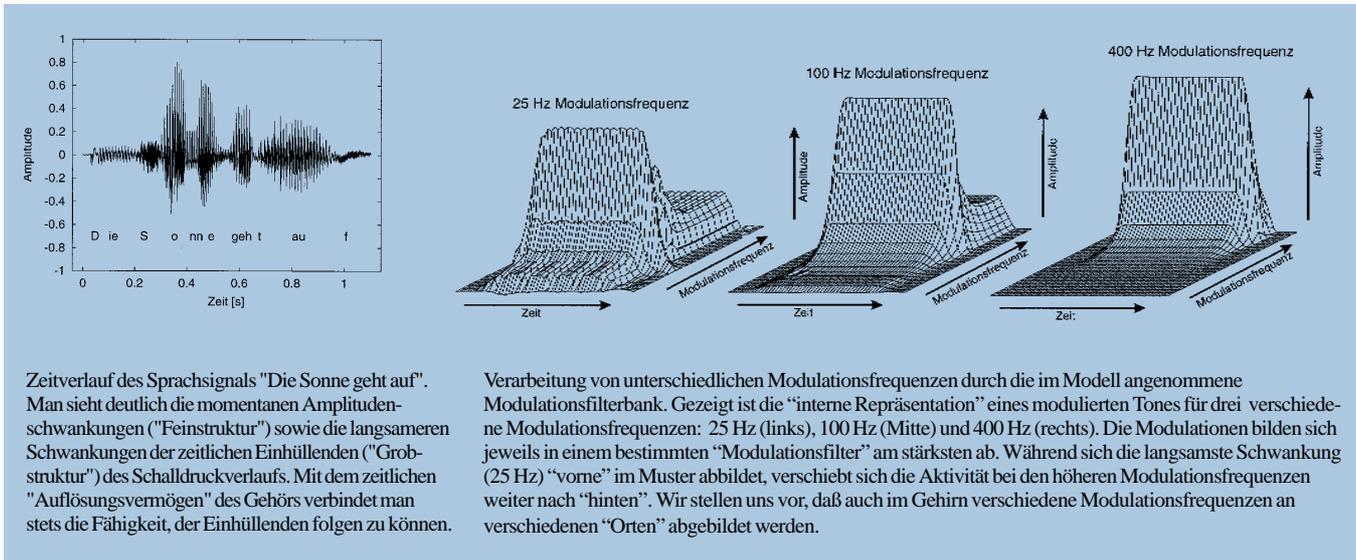
Grafik links: Modell der "effektiven" Signalverarbeitung im Gehör. Als Eingangssignal (von oben) dient das akustische Signal, das analog zur Frequenz-Orts-Transformation im Innenohr in mehrere Frequenzkanäle (Frequenzgruppen) zerlegt wird. Der Ausgang des Modells (unten) ist die "interne Repräsentation" des akustischen Signals, die sich auf verschiedene Mittenfrequenzen verteilt und sich zudem mit der Zeit ändert. Hier schließt sich die Entscheidungsstufe an.

Grafik oben: Antwort der Nachregelschleifen im Modell (untere Abb.) auf ein 200 ms langes Rauschsignal (obere Abb.). Das Signal am Ausgang der Nachregelschleifen schwingt am Signalbeginn stark über und "adaptiert" dann auf eine geringere Amplitude. Solch ein Überspringen mit anschließender Adaption findet man sehr häufig auch in physiologischen Messungen im Antwortverhalten von z. B. Hörnervenfasern.

kalisch können wir diese Transformation durch eine *Einhüllendenextraktion* beschreiben, die sich durch Halbwellengleichrichtung und anschließende Tiefpaßfilterung realisieren läßt. Im anschließenden Hörnerv wird nun die akustische Information durch Erhöhung der neuronalen Aktivität der verschiedenen Nervenfasern kodiert, so daß zu jedem Zeitpunkt die Schallintensität für unterschiedliche Frequenzen verschlüsselt wird. Zudem finden wir im Antwortverhalten von Hörnervenfasern stark nichtlineare, sogenannte *adaptive* Effekte. Im Hörsystem werden plötzliche Änderungen im Schall wie z.B. Ein- und Ausschaltvorgänge neuronal stärker bewertet als statische, unveränderliche Anteile im Signal. Ein solches Verhalten ist typisch für die Verarbeitung von zeitlicher Information und zeigt sich in allen Stufen entlang der Hörbahn bis hin zum Kortex. Physikalisch können wir uns ein solches adaptives Verhalten durch das Hintereinanderschalten von sogenannten *Nachregelschleifen* mit unterschiedlichen "Zeitkonstanten" vorstellen, bei denen jeweils das Eingangssignal durch das tiefpaßgefilterte Ausgangssignal geteilt wird. Hierdurch wird eine gewisse Adaption an den Mittelwert des Eingangssignals ermöglicht, während schnelle Änderungen unbeeinflusst durchge-

lassen werden. Das Antwortverhalten der im Modell enthaltenen Adaptationsstufe ist tatsächlich gemessenen neuronalen Antwortmustern von Hörnervenfasern sehr ähnlich, so daß z.B. Signalbeginn und Signalende besonders betont werden. Allerdings können wir diese im Modell angenommenen Adaptationsstufen nicht direkt einzelnen lokalen Strukturen zuschreiben, wie dies noch bei den ersten Stufen des Modells der Fall war.

Die neuronalen Reize werden vom Hörnerv in den sogenannten *Hirnstamm* weitergeleitet, in dem bereits komplexe Funktionen ausgewertet werden. Beispielsweise erfolgt hier, in der sogenannten *Oberen Olive*, ein erster interauraler Vergleich, d.h. eine Auswertung der zwischen den beiden Ohren auftretenden Zeit- und Intensitätsunterschiede zur Lokalisation von Schallquellen. Weiterhin erfolgt in der vielleicht wichtigsten "Schaltstelle" im Hirnstamm, dem *Inferior Colliculus*, eine Auswertung von *Modulationsfrequenzen*. Modulationen bezeichnen die Schwankungen der zeitlichen Einhüllenden von Signalen. Alle für uns wichtigen Kommunikationssignale wie z.B. Sprache und Musik weisen Einhüllenden-schwankungen bzw. Modulationen auf. Deshalb ist es besonders interessant zu verstehen, wie solche Modulationen in unserem Gehirn abgebildet und weitergeleitet wer-



Zeitverlauf des Sprachsignals "Die Sonne geht auf". Man sieht deutlich die momentanen Amplitudenschwankungen ("Feinstruktur") sowie die langsameren Schwankungen der zeitlichen Einhüllenden ("Grobstruktur") des Schalldruckverlaufs. Mit dem zeitlichen "Auflösungsvermögen" des Gehörs verbindet man stets die Fähigkeit, der Einhüllenden folgen zu können.

Verarbeitung von unterschiedlichen Modulationsfrequenzen durch die im Modell angenommene Modulationsfilterbank. Gezeigt ist die "interne Repräsentation" eines modulierten Tones für drei verschiedene Modulationsfrequenzen: 25 Hz (links), 100 Hz (Mitte) und 400 Hz (rechts). Die Modulationen bilden sich jeweils in einem bestimmten "Modulationsfilter" am stärksten ab. Während sich die langsamste Schwankung (25 Hz) "vorne" im Muster abbildet, verschiebt sich die Aktivität bei den höheren Modulationsfrequenzen weiter nach "hinten". Wir stellen uns vor, daß auch im Gehirn verschiedene Modulationsfrequenzen an verschiedenen "Orten" abgebildet werden.

den. Im Frequenzbereich zwischen 0 und etwa 10 Hz werden Modulationen als Lautstärkeschwankungen wahrgenommen. Bei Frequenzen zwischen 10 und 80 Hz entsteht eine "rauhe" Wahrnehmung. Bei noch höheren Modulationsfrequenzen werden durch die gleichzeitige spektrale Verfärbung des Schalles komplexere Klangänderungen wahrgenommen.

Erst seit kurzem ist bekannt, daß im *Inferior Colliculus* der Zeitverlauf der Nervenerregungen in verschiedene Modulationsfrequenzbereiche aufgespalten wird. Man findet hier Neuronen, die auf bestimmte Modulationsfrequenzen abgestimmt sind, während sie auf andere Modulationsfrequenzen gar nicht reagieren. Neben dem schon im Innenohr gebildeten Ordnungsprinzip der Tonotopie (Frequenz-Orts-Abbildung) zeigt sich auf dieser höheren Stufe zusätzlich das Prinzip der *Periodotopie*, d.h. verschiedene Modulationsfrequenzen werden an verschiedenen Orten abgebildet. Interessanterweise bilden sich dabei die beiden "Achsen" Frequenz und Modulationsfrequenz unabhängig voneinander im Gehirn ab. Physikalisch entspricht dies einer *Modulationsfilterbank*, die die einzelnen vorverarbeiteten Signale in *Modulationsfrequenzgruppen* zerlegt, so daß sich am Ausgang der bisherigen Verarbeitungsstufen im Modell ein zweidimensionales Muster ergibt (Frequenz x Modulationsfrequenz). Diese Modellstufe ist fundamental für die gesamte Modellierung der Signalverarbeitung, denn sie ermöglicht eine realistische Nachbildung vieler unterschiedlicher akustischer Phänomene, bei denen die zeitlichen Aspekte des Hörens eine Rolle spielen.

Im Modell wird am Ausgang der Modulationsfilterbank noch ein "internes Rauschen" des neuronalen Systems angenommen, das

die neuronalen Verarbeitungsfehler repräsentiert. Das zeitliche Muster am Ausgang der Vorverarbeitung im Modell stellt die sogenannte *interne Repräsentation* des ursprünglichen akustischen Eingangssignals dar. Einer solchen internen Repräsentation liegt somit die Modellvorstellung zugrunde, daß wir die wesentlichen Vorverarbeitungsschritte des Hörsystems mit technischen Schaltkreisen effektiv nachbilden können. Wir nehmen also an, daß auf diese Weise eine Art Abbildung vom "Zustand des Gehirns" erstellt wird. Auf diesem Zustand bauen nun verschiedene Leistungen des Gehörs auf. Er gilt sozusagen als Eingangsgröße für den folgenden Mustererkenner (Detektor), durch den verschiedene Signale erkannt bzw. unterschieden werden können. Dem Mustererkenner liegt die Idee zugrunde, daß eine Änderung im Eingangssignal gerade wahrnehmbar wird, wenn die Änderung in der zugehörigen internen Repräsentation des Signals gerade so groß ist, daß sie aus dem internen Rauschen herausragt.

Was bringt das Modell?

Aufbauend auf dem aktuellen Modell kann bisher bereits die *Sprachverständlichkeit* in Ruhe und unter Störgeräusch bei normal- und schwerhörenden Versuchspersonen sehr gut vorhergesagt werden. Zudem kann die Vorverarbeitung im Modell für die Vorhersage der *Sprachübertragungsqualität* von Kodiersystemen (die z.B. in Mobiltelefonen zur Reduktion der zu übertragenen Datenmenge verwendet werden) und zur robusten *Spracherkennung* in verschiedenen Störgeräuschen erfolgreich verwendet werden. Natürlich hat unser Gehör eine Reihe von Eigenschaften, die durch das bisherige Modell noch nicht

erfaßt werden können. Beispielsweise wissen wir, daß im Gehirn sehr viele Verschaltungen und Wechselwirkungen zwischen den neuronalen Aktivitäten in den verschiedenen Frequenzbereichen stattfinden. Solch eine Informationsverarbeitung über Frequenzgruppen hinweg spielt bei der Wahrnehmung in komplexer akustischer Umgebung wie z.B. in typischen "Cocktail-Party"-Situationen eine wesentliche Rolle. Inwieweit sich das Modell auch im Bereich der digitalen Kodierung von Audiosignalen und in der Hörgerätetechnologie bewähren kann, werden wir in naher Zukunft erfahren können.

Der Autor



Dr. Torsten Dau (33), wiss. Assistent am Fachbereich Physik, AG Medizinische Physik. Maschinenbaustudium 1987-1989 in Hannover (Abschluß Vordiplom). Physikstudium 1987-

1992 in Göttingen. Promotion in Physik 1996 im Graduiertenkolleg "Psychoakustik" in Oldenburg. Forschungsaufenthalte in Cambridge (England) 1994 und 1996. Seit 1996 Mitarbeiter im Sonderforschungsbereich "Neurokognition". Auszeichnung als Nachwuchswissenschaftler im Bereich der Akustik mit dem Lothar-Cremer-Preis 1998. Forschungsschwerpunkte: Psychoakustik, digitale Signalverarbeitung, Neuronale Korrelate von Wahrnehmungsgrößen mittels akustisch evoked Potentiale (EEG).