

Modelle der auditorischen Verarbeitung

Torsten Dau

Medizinische Physik, Carl von Ossietzky Universität-Oldenburg

Einleitung

Reale Systeme sind im häufig so komplex, dass nur Untersuchungen am weniger komplexen Modell ein Verständnis der Funktion des Systems, also z.B. des auditorischen Systems, ermöglichen. Ein Modell stellt stets eine generalisierte Beschreibung des realen Systems dar. Vorteile der Modellierung bestehen z.B. darin, dass Eingriffe am realen System nicht mehr notwendig sind, die stets auch zu Beeinflussungen führen. Modelle erlauben uns, Hypothesen über die zugrunde liegenden Mechanismen zu überprüfen und weiterzuentwickeln, sowie Vorhersagen des Systems in bisher noch nicht experimentell getesteten Situationen vorzunehmen. Je größer die Anzahl an korrekt vorhergesagten experimentellen Konstellationen ist, desto eher gehen wir davon aus, ein Problem „richtig verstanden“ zu haben. Modellierung ist immer dann besonders wertvoll, wenn verschiedene – bisher als unabhängig angenommene – Phänomene auf die gleichen grundlegenden Prozesse zurückgeführt werden können. Zugleich machen erst Modellrechnungen deutlich, wie komplex und vielschichtig z.B. die neuronalen Mechanismen im auditorischen System selbst bei vermeintlich „einfachen“ Wahrnehmungsprozessen sind. Je nach „Paradigma“ gibt es sehr verschiedene Modellklassen, die jeweils sehr unterschiedliche Aspekte der neuronalen Verarbeitung auf verschiedensten Ebenen der (Signal-)Verarbeitung nachbilden: z.B. mathematische (statistische) Modelle, biophysikalische Modelle, physiologische Modelle, und psychoakustische Modelle. Zudem unterscheiden sich die einzelnen Modelle durch ihren Grad an Komplexität (Black-Box-Modelle versus differenzierte Modelle), den Grad der Quantifizierung (begriffliche versus numerische Modelle) und die Berücksichtigung von Zufallsprozessen (deterministische versus statistische Modelle). Im Folgenden werden einige Beispiele unterschiedlicher Komplexität aus dem Bereich der psychoakustischen und physiologischen Modellierung vorgestellt. Nur grundlegende Aspekte können hier dargestellt werden. Die meisten Modellbeispiele haben gemeinsam, dass psychoakustische Daten mit Hilfe einer physiologisch motivierten „Signalverarbeitung“ im Gehör nachgebildet werden.

Beispiele für auditorische Modelle

(i) Jeffress' Lokalisationsmodell (1948) als Beispiel für ein Black-Box Modell

Unser Hörsystem ist enorm leistungsfähig bei der Lokalisation von Schall. Quellen aus seitlichen Einfallrichtungen, die sich lediglich um 5 Grad Azimut unterscheiden, können vom Gehör aufgelöst werden. Bei Schalleinfallrichtung von vorne liegt die Auflösungs-

leistung sogar bei nur etwa 1 Grad. Bei tieffrequenten Schallquellen erfolgt die Lokalisation insbesondere aufgrund von Laufzeitunterschieden zwischen den beiden Ohren, die an der Auflösungsgrenze nur etwa 10 μ s betragen. Die Frage ist, wie unser Gehirn eine solche hohe Zeitauflösung bewerkstelligt und welches die zugrunde liegenden neuronalen Mechanismen sind. Jeffress (1948) postulierte, dass es spezifische Neuronen („Koinzidenzdetektoren“) geben müsse, die ihre Antwortaktivität (Feuerrate) erhöhen, wenn der Schall aus einer bestimmten Richtung kommt. Der Koinzidenzdetektor „feuert“, wenn er gleichzeitigen neuronalen Input von beiden „Seiten“ bekommt, wobei die externen Laufzeitunterschiede gerade durch interne neuronale Laufzeitdifferenzen kompensiert werden (siehe Abbildung 1).

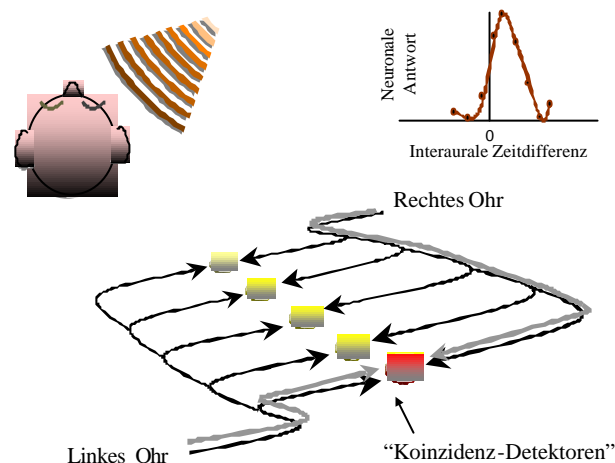


Abb. 1: Schematische Darstellung des Jeffress Modells (1948). Für eine bestimmte Schalleinfallrichtung erhöht eine der dargestellten Neurone (Koinzidenzdetektoren) seine Feuerrate, da die neuronale Erregung von beiden Seiten synchron erfolgt. Der neuronale Laufzeitunterschied „kompensiert“ dabei gerade den externen interauralen Laufzeitunterschied (ITD) zwischen den beiden Ohren. Verschiedene ITDs werden nach diesem Schema an verschiedenen Orten abgebildet.

Die Umsetzung zeitlicher Information in einen solchen Raten-Code wurde postuliert, obwohl es zu jener Zeit keine direkten Hinweise auf ein entsprechendes physiologisches Korrelat und auch nicht die technischen Möglichkeiten physiologischer Einzelzellaufnahmen im Hirnstamm gab. Das Jeffress Modell stellt demnach ein Black-Box Modell ohne Spezifikationen dar, das einen funktionellen Mechanismus für die Lokalisation externer Schallquellen postuliert. Später wurde das Konzept zunächst durch binaurale Detektionsexperimente zur „Binaural masking level difference“ (BMLD) bestätigt (McFadden et al., 1972). Erst

1990 wurden dann Neuronen in der Medialen Oberen Olive (MSO) im Hirnstamm nachgewiesen, die bei Anregung sowohl mit tieffrequenten Tönen als auch mit komplexeren Signalen ein Antwortverhalten zeigen, das genau der prinzipiellen Wirkungsweise von Koinzidenzdetektoren entspricht (Yin und Chan, 1990). In einer verallgemeinerten Beschreibung lassen sich solche Zellen als Kreuzkorrelationsoperatoren der beiden (neuronalen) Eingänge auffassen. Alle modernen Lokalisationsmodelle enthalten eine solche Korrelationsstufe am Ausgang der beiden monaural vorverarbeiteten neuronalen Eingänge (z.B. Colburn, 1995; Lindemann, 1986; Stern und Trahiotis, 1995; Breebart et al., 2001). Das Black Box Modell von Jeffress hat enorm zum Verständnis der Mechanismen bei der Lokalisation von Schall beigetragen.

(ii) Modellierung von sensorineuraler Schwerhörigkeit: Von der Physiologie zum psychoakustischen Modell

Die wesentlichen Kennzeichen einer (Innenohr) Schwerhörigkeit sind (i) erhöhte Hörschwellen, (ii) eine „Aufsteilung“ der Lautheitsfunktion (Recruitment-Phänomen), (iii) eine verringerte Frequenzauflösung (verbreiterte auditorische Filter), (iv) ein eingeschränktes binaurales Hörvermögen, sowie (v) eine verschlechterte Zeitauflösung (z.B. bei der Lückendetektion und Nachverdeckung). Eine wichtige Frage ist, ob die verschiedenen Kennzeichen der Schwerhörigkeit durch unterschiedliche und ggf. unabhängige Aspekte zustande kommen oder ob sie eine Folge von gekoppelten Prozessen darstellen. In physiologischen Arbeiten an der Basilarmembran (im Tierexperiment) wurden die nichtlinearen Eigenschaften des Übertragungsverhaltens der Basilarmembran (BM) im Detail untersucht (Ruggero et al., 1997). Besonders wichtig ist hierbei die kompressive Übertragungsfunktion der BM (an der charakteristischen Frequenz) bei Anregung mit Tönen mittleren Pegels, sowie die nahezu lineare Übertragung bei sehr kleinen bzw. sehr hohen Pegeln. Im geschädigten System fällt durch den Ausfall der aktiven Mechanismen in den äußeren Haarzellen die Kompression aus; das System wird linearisiert. Als Folge dieser Linearisierung wird ein wesentlich geringerer Dynamikbereich des äußeren Schalls auf den gleichen internen (neuronalen) Ausgangsbereich übertragen als im ungeschädigten System. Die Kompression ist allerdings nur bei Anregung an der charakteristischen Frequenz zu beobachten (Ruggero et al., 1997), woraus sich unmittelbar die pegelabhängigen frequenzselektiven Eigenschaften des ungestörten peripheren Systems ableiten. Im gestörten System ist die Frequenzselektivität geringer ausgeprägt und kaum pegelabhängig.

Erhöhte „Schwellen“, verringerte Dynamik und verringerte Frequenzauflösung lassen sich also bereits

auf lokaler (peripherer) Ebene im physiologischen Experiment nachweisen.

Interessant ist es nun, die Konsequenzen dieser cochleären Prozesse für die *Zeitwahrnehmung* zu untersuchen. Die Nachverdeckung ist eines der klassischen psychoakustischen Experimente zur Zeitauflösung. Es zeigt sich, dass Schwerhörige bei der Detektion eines Signals (z.B. eines Tonpulses) nach Abschalten eines Maskierers (z.B. eines Rauschens) länger benötigen, um wieder in den „Grundzustand“ zu adaptieren, sodass die Ruhehörschwelle also erst bei einem größeren zeitlichen Abstand zwischen Maskierer und Signal erreicht wird. Die veränderte Zeitauflösung ließe sich im Rahmen eines Detektionsmodells (siehe Abbildung 2) am einfachsten durch eine *veränderte Zeitkonstante* bei der Signalintegration erfassen. Die Unterschiede wären somit auf unterschiedlich „zentrale“ Prozesse bei der Integration von Signalinformation zurückgeführt. Es zeigt sich hingegen durch entsprechende Modellrechnungen, dass eine solche Unterscheidung im zentralen Prozessor nicht nötig ist, um die Daten zu erklären. Stattdessen können die Unterschiede zwischen Normal- und Schwerhörigen bei der Nachverdeckung auf die Unterschiede in der cochleären Kompression zurückgeführt werden, während der zentrale Prozessor unverändert bleibt. Um z.B. eine Änderung von 1 dB in der internen Repräsentation des Maskierers zu erhalten, ist bei einem Kompressionsverhältnis von 5:1 im normalhörenden System eine Änderung von 5 dB im Signalpegel nötig, während im (linearisierten) schwerhörenden System eine Änderung von 1 dB ausreicht, um die gleiche Änderung vorherzurufen. Abbildung 2 (untere Teilbilder) zeigt gemessene und mit einem solchen Modell simulierte Nachverdeckungskurven (Oxenham und Moore, 1997) für Normalhörende (links) und Schwerhörige (rechts).

Zusammenfassend lässt sich also zeigen, dass die Linearisierung des peripheren Systems zu einer *effektiv* größeren Zeitkonstante im gestörten System führt (Oxenham und Moore, 1997; Derleth et al., 2001). Das Modell zur Zeitverarbeitung im auditorischen System ist motiviert durch ein detaillierteres Verständnis der Prozesse in der Cochlea. In diesem Beispiel ging demnach der Entwicklung eines Wahrnehmungsmodell die Erfassung physiologischer Prozesse voraus.

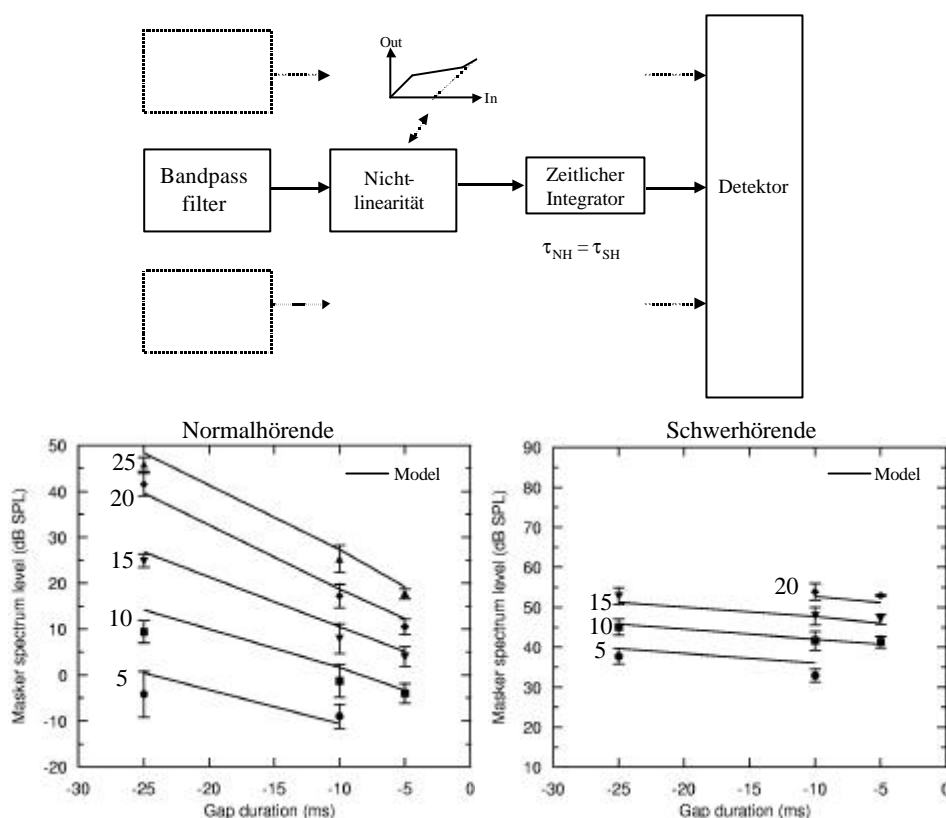


Abb. 2: Oben: Schematisches psychoakustisches Detektionsmodell. Nach Bandpassfilterung erfolgt im Modell für Normalhörende eine Kompression bei mittleren Pegeln, während im Modell für Schwerhörnde eine lineare Kennlinie angenommen wird. Der (zentrale) Integrationsprozess wird in beiden Modellen als gleich angenommen (mit gleichen Zeitkonstanten τ). In der anschließenden Detektorstufe werden Maskierer und Summe aus Maskierer und Signal miteinander verglichen. Das Entscheidungskriterium ist für beide Modelle gleich. Unten: Gemessene (Symbole) und mit obigem Modell simulierte (Linien) Nachverdeckungskurven für Normalhörende (links) und Schwerhörnde (rechts) für verschiedene zeitliche Abstände zwischen Maskierer und Signal (nach Oxenham und Moore, 1997). Parameter ist der Signalpegel (in dB SL). Der Maskierpegel wurde in Messung und Simulation variiert

(iii) Modellierung von Hirnstammpotenzialen: Von der Cochlea zum Fernfeldpotenzial

Akustisch evozierte Potenziale werden z.B. in der klinischen Diagnostik beim objektiven Nachweis des frequenzspezifischen Hörvermögens bei Neugeborenen und Kleinkindern eingesetzt. Evozierte Potenziale repräsentieren die an der Kopfoberfläche gemessene (gemittelte) elektrische Potenzialverteilung als Antwort auf einen (häufig wiederholten) akustischen Reiz. Je größer die Zahl gleichzeitig angeregter aktiver Neurone ist, desto größer ist das zugehörige generierte Potential. Der Klickreiz wird routinemäßig in der Diagnostik eingesetzt, da hiermit eine hohe Synchronizität erzielt werden kann. Allerdings repräsentieren klick-evozierte Potenziale nur neuronale Aktivitäten aus medialen und basalen Abschnitten der Cochlea, während Aktivität aus den apikalen Bereichen (tiefe Frequenzen) durch die Laufzeitunterschiede auf der Basilarmembran kaum zum Potenzial beitragen. Im Gegensatz hierzu generiert ein Chirpreiz, der die Laufzeitunterschiede auf der BM zwischen den verschiedenen Frequenzkomponenten gerade kompensiert, signifikant höhere Potenzialamplituden (Dau et al., 2000; Junius et al., 2000; Wegner und Dau, 2002). Ein Verständnis der dispersiven Pro-

zesse in der Cochlea ist also wichtig für die Interpretation von Hirnstammpotenzialen.

Um die von einem *beliebigen* Reiz hervorgerufenen Antworten interpretieren zu können, benötigt man demnach ein Modell, das eine möglichst realistische Nachbildung der cochleären Prozesse beinhaltet. Hierzu gehört insbesondere die Nachbildung der pegelabhängigen frequenzselektiven Eigenschaften der Cochlea, die ja eine unmittelbare Folge der pegelabhängigen kompressiven Nichtlinearität repräsentieren (s.o.). Ein Modell der Entstehungsmechanismen von evozierten Potenzialen ist in Abbildung 3 dargestellt.

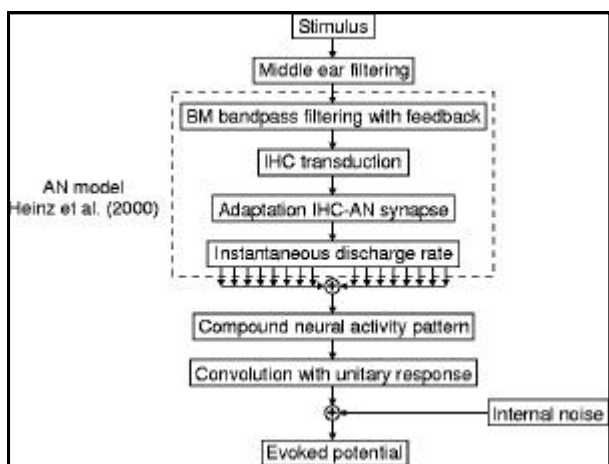


Abb. 3: Modell zur Generierung von Hirnstammpotentialen (Dau, 2002). Nach Aussen- und Mittelohrfilterung erfolgt die cochleäre Filterung nach Heinz et al. (2001). Die neuronale Aktivität am auditorischen Nerven wird über alle Frequenzkanäle aufsummiert. Es folgt die Faltung mit einer an physiologische Daten (von Katzen) angelehnte „Impuls-Antwort“, die die Übertragungseigenschaften des Kopfes beim Übergang vom „lokalen Potenzial“ der beteiligten Generatoren im Hirnstamm hin zum „Fernfeldpotenzial“ an den Oberflächen Elektroden berücksichtigt.

Für die cochleäre Vorverarbeitung wurde das von Heinz et al. (2001) entwickelte Modell des auditorischen Nerven (AN) gewählt. Die neuronale Aktivität im AN wird anschließend über alle Frequenzkanäle summiert und dann mit einer effektiven „Impulsantwort“ gefaltet, die die Übertragungseigenschaften des Kopfes von der lokalen Aktivität an den Generatoren der Wellen IV bis zum Ort der Elektroden repräsentiert (Melcher und Kiang, 1996). Im Rahmen des Modells (Dau, 2002) können viele Potenzialmuster als Funktion des Pegels befriedigend nachgebildet werden (siehe Abbildung 4). Insbesondere zeigt sich durch Modellrechnungen, dass die durch tieffrequente Töne hervorgerufenen Frequenzfolgepotenziale (FFR) eindeutig durch (die synchrone) mediale und basale Aktivität generiert werden, und keineswegs durch neuronale Aktivität aus den apikalen Bereichen (Janssen et al., 1991; Dau, 2002), die sich durch die Laufzeitunterschiede auf der BM im Fernfeld destruktiv überlagert. Ein auf der realistischen Nachbildung der peripheren Prozesse basierendes Modell kann deshalb ein nützliches Instrument im Bereich der klinischen Diagnostik darstellen. Unterschiedliche Formen von Schwerhörigkeit können im Modell berücksichtigt und deren Auswirkung auf die Ausbildung der Potenziale getestet werden.

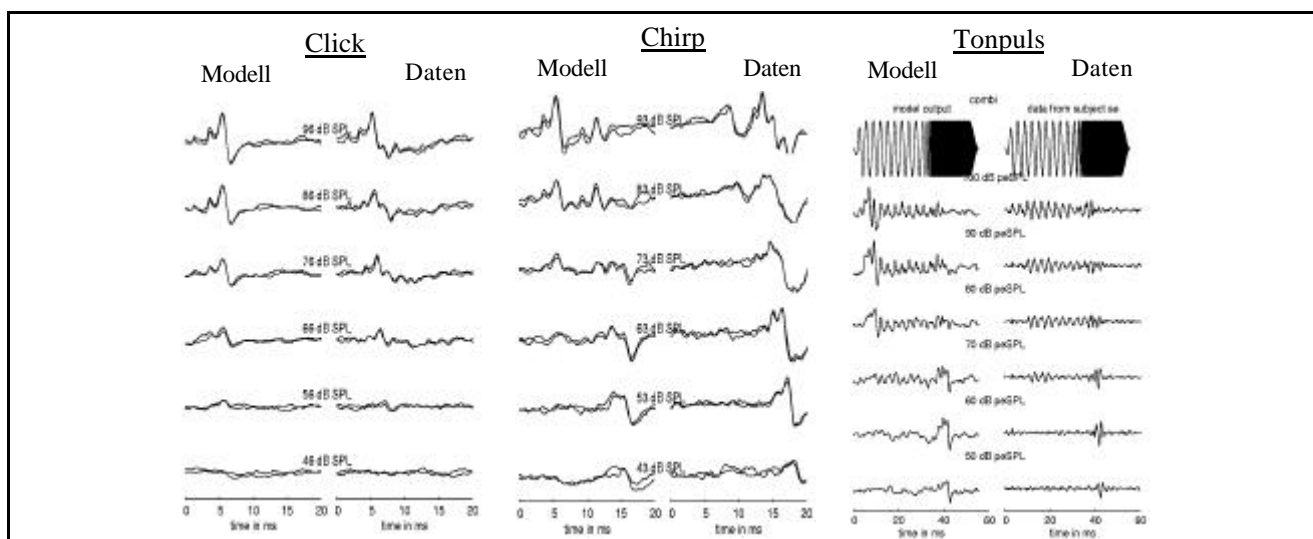


Abb. 4: Gemessene und mit obigem Modell simulierte Hirnstammpotentialen für drei verschiedene Reize (links: Klick; mitte: Chirpreiz nach Dau et al. (2000); rechts: „Kombinationsstimulus“ nach Junius et al. (2000), bestehend aus 0.3-kHz Ton, 0.3-4-kHz Chirp und 4 kHz Ton). Das Modell bildet viele der typischen Merkmale der gemessenen Potenziale nach. Insbesondere zeigt sich das richtige pegelabhängige Verhalten, was beim Kombinationsstimulus (rechts) besonders gut zu sehen ist.

(iv) Verarbeitung von Amplitudenmodulationen im Gehör: Modulationstiefpass oder Modulationsfilterbank? Neben der Signalanalyse in der Cochlea (Frequenz-Orts-Transformation, Wanderwellenkonzept) spielt die Analyse der zeitlichen Einhüllenden der Signale eine wesentliche Rolle für die Wahrnehmung. Wichtige Informationen in Kommunikationssignalen Modelle der auditorischen Verarbeitung

wie z.B. Sprache oder Musik sind in der Einhüllenden enthalten. In den inneren Haarzellen werden die mechanischen Schwingungen der Basilarmembran in Rezeptorpotenziale umgesetzt, die anschließend im auditorischen Nerven in Form von Aktionspotentialen ans weitergeleitet werden. Bei der Haarzellentransformation zeigt sich, dass vornehmlich Auslenkungen der

BM in eine Richtung, also nur „Halbwellen“, übertragen werden. Zudem folgt das Rezeptorpotenzial nur bis zu Anregungsfrequenzen von etwa 1 kHz ohne Abschwächung (z.B. Pickles, 1988). Diese Prozesse werden in gehörgerechten Modellen durch eine Halbwellengleichrichtung mit anschließender 1-kHz Tiefpassfilterung nachgebildet, was effektiv bei hohen Frequenzen) einer Einhüllendenbildung entspricht.

Eine wichtige Frage ist, wie das Gehör die im Schallsignal enthaltenen Fluktuationen der Einhüllenden – also Amplitudenmodulationen – verarbeitet. Ein grundlegendes psychoakustisches Experiment hierzu besteht darin, die Detektierbarkeit einer auf einen breitbandigen Rauschträger aufgeprägten Amplitudenmodulation als Funktion der Modulationsfrequenz zu ermitteln. Die entsprechende Modulationstransferfunktion (TMTF) hat typischerweise Tiefpasscharakter (Viemeister, 1979): Langsame Modulationen können scheinbar leicht detektiert werden, während schnelle Modulationen nur bei einem Modulationsgrad von nahezu 100% wahrgenommen werden. Das Gehör scheint demnach zu träge zu sein, schnellen Fluktuationen zu folgen. Ein aus diesen Messungen abgeleitetes klassisches psychoakustisches Detektionsmodell nimmt entsprechend einen Modulationstiefpass zwischen peripherer Bandpass-Zerlegung und Detektionsprozess an, um die Daten nachzubilden (Viemeister, 1979). Es zeigt sich allerdings, dass ein solches Modell ausschließlich für Breitbandrauschträger „funktioniert“.

Für Sinusträger oder schmalbandiges Rauschen ändert sich der gemessene Schwellenverlauf grundlegend, während das obige Modell, unabhängig vom verwendeten Träger, stets die Tiefpass-Charakteristik beibehält. Es musste demnach nach einer andere Verarbeitungsstrategie „gesucht“ werden.

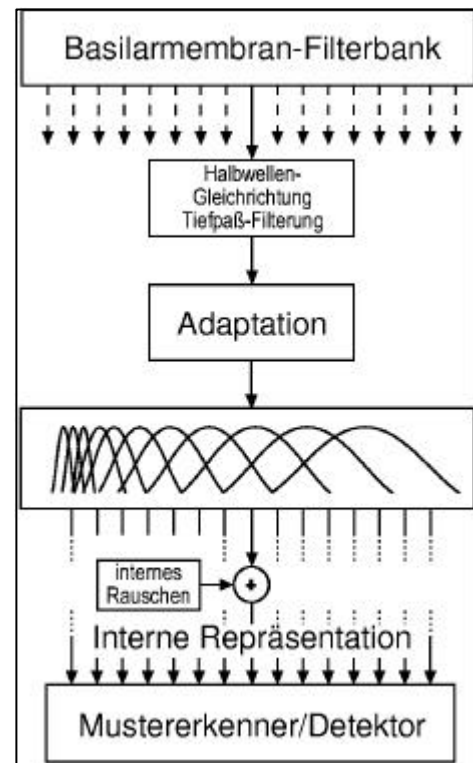
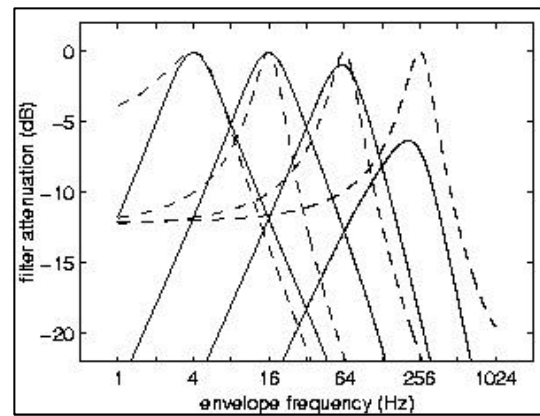
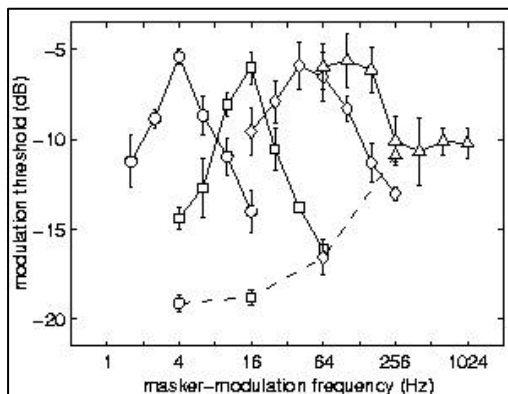


Abb. 5: Oben links: Gemessene Modulationsmaskierungsdaten aus Ewert und Dau (2000). Für vier verschiedene Signal-Modulationsfrequenzen (4, 16, 64, 256 Hz) sind die durch Maskierung mit Schmalbandrauschen erzielten Schwellen als Funktion des spektralen Abstands zwischen Signal- und Maskierermodulation gezeigt. Unten links: An die Maskierungsdaten angepasste Modulationsfilter (2. Ordnung, Güte 1). Ein Modulationstiefpass könnte die Daten nicht nachbilden. Rechts: Gesamt-Perzeptionsmodell (Dau et al. 1997a,b), in dem am Ausgang jedes peripheren Filters nach Einhüllendenextraktion und Adaptationsstufe eine Modulationsfilterbank angenommen wird. Nach Addition von internem Rauschen liegt die interne Repräsentation vor, die von einem optimalen Detektor ausgewertet wird.

In Modulationsmaskierungsexperimenten wurden Maskierungsmuster (Ewert und Dau, 2000) gefunden, die den aus klassischen Arbeiten bekannten Maskierungsmustern zur Charakterisierung von peripheren Filtern ähnlich sind (Abbildung 5). Basierend auf diesen

und anderen Daten wurde ein Modell entwickelt, das eine Modulationsfilterbank enthält (Abbildung 5). Dieses Modell, das eine Modulationsfrequenzerlegung innerhalb jeder Frequenzgruppe annimmt, kann die gemessenen TMTFs unabhängig von der Trägerbandbreite korrekt nachbilden (Dau et al., 1997a,b). Die oben beschriebenen „klassischen“ Daten sind demnach nicht durch einen Trägheitsprozess des Gehörs zu erklären, sondern lassen sich durch Berücksichtigung der physikalischen Eigenschaften des Trägers und die Annahme der modulationsfrequenzselektiven Verarbeitung berechnen. Die Tiefpasscharakteristik der Daten resultiert dabei insbesondere aus der Annahme einer logarithmischen Skalierung der Modulationsfilterbreite mit der Frequenz. Eine solche Abbildung von Modulationen zusätzlich zur Abbildung der (Audio-) Frequenzachse wurde auch in verschiedenen physiologischen Arbeiten im Hirnstamm (Langner und Schreiner, 1988) und im Kortex (Schulze et al., 2001) gefunden, sodass es hier konzeptionelle Parallelen zwischen lokalen Verarbeitungsprinzipien einerseits und globaler „effektiver“ Verarbeitung im Wahrnehmungsmodell andererseits gibt.

(v) Modellierung auditorischer Objektbildung: Anspruch und Wirklichkeit

Das obige Perzeptionsmodell ist in der Lage, eine Reihe von Experimenten zur zeitlichen und spektralen Maskierung mit einer sehr geringen Anzahl von Parametern erfolgreich nachzubilden. In der bisherigen Modellversion wird die Verarbeitung bei verschiedenen Frequenzen (und auch Modulationsfrequenzen) als

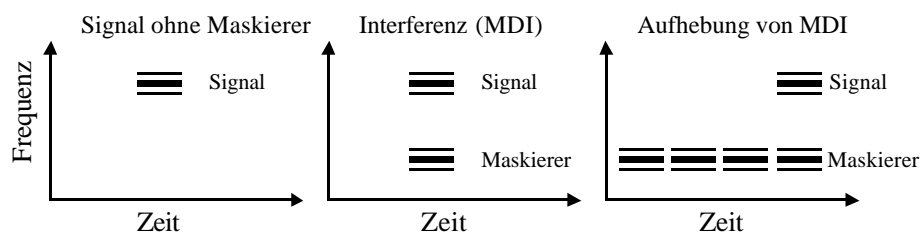


Abb. 6: Schematische Darstellung von Signalkonfigurationen, die in Experimenten zur „Modulation Detection Interference“ (MDI) verwendet wurden. Links: Referenzsituation, in der lediglich der modulierte Signalton angeboten wird. Mitte: Klassische MDI-Situation, in der bei gleichzeitiger Präsentation einer ähnlichen Modulation auf einem anderen Träger eine stark erhöhte Schwelle bei der Detektion der Signalmodulation gefunden wird (Yost et al., 1989). Rechts: Vollständige Aufhebung von MDI durch Präsentation von Pulsen vor dem Maskierer. MDI kann nur im Zusammenhang mit der Bildung auditorischer Objekte erklärt werden.

Es zeigt sich demnach, dass die Bildung von MDI (sowie deren Aufhebung) eng an Prozesse der auditorischen Objektbildung gekoppelt ist. Die Stärke der Interferenz hängt von objektbildenden Parametern im Signal ab, wie z.B. der Synchronizität, Harmonizität, Tonhöhe, und räumliche Position der beteiligten Komponenten. Dies bedeutet, dass eine „Bottom-Up“ Modellierung, wie sie bisher hier vorgestellt wurde, nicht ausreichen kann, um die Daten nachbilden zu können. Auch einfache Modifikationen des obigen Perzeptionsmodells, z.B. durch eine „fest-verdrahtete“ frequenzübergreifende Integration von Modulationen,

unabhängig angenommen, d.h. es gibt keine Wechselwirkung über Frequenzgruppen hinweg. Es zeigt sich allerdings, dass in vielen Hörsituationen eine unabhängige Verarbeitung unrealistisch ist. Werden beispielsweise zwei amplitudenmodulierte Töne sehr unterschiedlicher Trägerfrequenz (von z.B. 1 und 4 kHz) dargeboten, kann unser Gehör die beiden Modulationen nicht voneinander „trennen“, obwohl sich die beiden

Schalle nach dem klassischen Frequenzgruppenkonzept nicht beeinflussen sollten. Dieses spezielle Beispiel für eine *Interferenz* in der Wahrnehmung wird „Modulation Detection Interference“ (MDI) genannt (Yost et al., 1989). Es zeigt sich allerdings, dass MDI nur dann auftritt, wenn die beiden Signale *gleichzeitig* ein- und ausgeschaltet werden (Hall und Grose, 1991). Bei einer Verzögerung von z.B. 200 ms verschwindet der Effekt vollständig, was darauf zurückgeführt wird, dass nun die beiden Schalle vom Gehör nicht mehr zu einem gemeinsamen Objekt zugehörig aufgefasst werden. Genauso verschwindet die Interferenz, wenn vor einen der beiden Töne eine Serie von weiteren Tonpulsen vorgeschaltet werden (siehe Abbildung 6). In dieser Kondition des sogenannten „sequential streaming“ wird die gesamte zeitliche Sequenz (bei der einen Trägerfrequenz) zu einem Objekt zusammengruppiert: der zweite Träger und die auf ihn aufgeprägte Modulation werden nun unabhängig verarbeitet und wahrgenommen, als wäre gar kein weiterer Schall vorhanden (Oxenham und Dau, 2001).

sind unzureichend. Stattdessen müssen „Top-Down“ Mechanismen in die Modellierung einbezogen werden, die vom aktuellen akustischen Kontext abhängen. Eine solche Modellierung steht noch am Anfang und stellt eine große Herausforderung für die Zukunft dar.

(vi) Technische Anwendungen des Perzeptionsmodells

Trotz der obigen Einschränkungen der Modellierung in komplexen akustischen Szenen ist das Perzeptionsmodell interessant für technische Anwendungen und wurde bereits in einigen Bereichen eingesetzt. Bei der Kodierung von Sprach- und Audiodaten (z.B. mit MP3) geht es darum, Abweichungen zwischen Original und dekodiertem Signal auf *perzeptiver* Ebene möglichst gering zu halten. Eine entsprechende Auswertung dieser Unterschiede auf der Ebene der internen Repräsentation der Signale am Ausgang des Verarbeitungsmodells könnte vielversprechend sein. Im Bereich der objektiven Signalqualitätsbeurteilung wurde das Modell bereits eingesetzt, wobei Unterschiede zwischen Original und kodiertem Sprach- oder Audiosignal am Ausgang des Modells ausgewertet wurden (Hansen und Kollmeier, 1999). In der Sprach- und Mustererkennung wurde der Ausgang des Modells als Eingang für Mustererkennungsalgorithmen verwendet (Tchorz und Kollmeier, 1999). Zudem könnte das Modell zukünftig für Signalmanipulationen in Hörgerätealgorithmen mit dem Ziel eingesetzt werden, Unterschiede in der internen Repräsentation zwischen Normalhörenden und individuell Schwerhörenden zu minimieren. Für eine solche Anwendung ist allerdings noch viel Entwicklungsarbeit notwendig.

Zusammenfassung

Es wurden einige Beispielmodelle aus dem Bereiche der psychoakustischen und physiologischen Modellbildung beschrieben. In der Vergangenheit hat sich oft gezeigt, dass ein tieferes Verständnis physiologischer Prozesse zu einem besseren Verständnis von Wahrnehmungsprozessen und entsprechenden Modellen geführt hat (Beispiel Modellierung von Schwerhörigkeit). Umgekehrt führen psychophysikalische Daten und daraus abgeleitete Konzepte/Modelle zur Perzeption (z.B. Lokalisationsmodell, Modulationsverarbeitung), die nur die „effektive“ Verarbeitung nachbilden, häufig zu neuen Fragestellungen über die neuronale Kodierung mit Methoden aus der Physiologie. Während die Modellierung von relativ grundlegenden Wahrnehmungsphänomenen durch die Kombination von Information aus den verschiedenen Paradigmen recht erfolgreich ist und entsprechende Modelle auch bereits in einigen technischen Anwendungen bereits eingesetzt werden, sind die in komplexen akustischen Szenen beteiligten objektbildenden Prozesse noch nicht gut verstanden. Die Integration von Top-Down Prozessen in die Modellierung stellt hierbei eine große Herausforderung für zukünftige Entwicklungen dar.

Literatur

- Breebaart, J., van de Par S., and Kohlrausch A. (2001). „Binaural processing model based on contralateral inhibition. I. Model structure“. *J. Acoust. Soc. Am.* 110, 1074-88.
- Colburn, H.S. (1995). „Computational models of binaural processing“, in *Auditory computation*, edited by H. Hawkins and T. McMullin (Springer-Verlag, New York).
- Dau, T., Kollmeier, B, and Kohlrausch, A.(1997a). „Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers“, *Journal of the Acoustical Society of America* 102, 2892-2905.
- Dau, T., Kollmeier, B, and Kohlrausch, A. (1997b). „Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection“, *Journal of the Acoustical Society of America* 102, 2906-2919.
- Dau, T., Wegner, O., Mellert, V., and Kollmeier, B. (2000). „Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion“, *Journal of the Acoustical Society of America* 107, 1530-1540.
- Dau, T. (2002). „The importance of cochlear processing for the formation of auditory brainstem potentials“, submitted to the *Journal of the Acoustical Society of America*.
- Derleth, R.P., Dau, T., and Kollmeier, B. (2001). „Modelling temporal and compressive properties of the normal and impaired auditory system“, *Hearing Research* 159, 132-149.
- Ewert, S.D. and Dau, T. (2000). „Characterizing frequency selectivity for envelope fluctuations“, *J. Acoust. Soc. Am.* 108, 1181-96.
- Hall, J. W., and Grose, J. H. (1991). "Some effects of auditory grouping factors on modulation detection interference (MDI)," *J. Acoust. Soc. Am.* 90, 3028-3035.
- Hansen, M. und Kollmeier, B. (1999). „Continuous assessment of time-varying speech quality.“ *J. Acoust. Soc. Am.* 106, 2888-99.
- Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (2001). „Auditory nerve model for predicting performance limits of normal and impaired listeners,“ *Acoustics Research Letters Online* 2, 91–96.
- Janssen, T., Steinhoff, H.-J., and Böhnke, F. (1991). „Zum Entstehungsmechanismus der Frequenzfolgepotentiale.“ *Otorhinolaryngol. Nova.* 1, 16—25.
- Jeffress, L.A. (1948). „A place theory of sound localization“. *J.Comp. Physiol. Psychol.* 41, 35-39.

- Junius, D., Wegner, O. and Dau, T. (2000). „Human auditory brainstem potentials using optimized stimuli to compensate for basilar-membrane dispersion“, in *23th meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, Florida.
- Langner, G. and Schreiner, C. (1988). „Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanism.“ *J. Neurophysiol.* 60., 1799-1822.
- Lindemann, W. (1986). „Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wave front,“ *J. Acoust. Soc. Am.* 80, 1623-1630.
- McFadden, D., Jeffress, L.A., and Ermy, H.L. (1972). „Differences in interaural phase and level in detection and lateralization: 1000 and 2000 Hz,“ *J. Acoust. Soc. Am.* 52, 1197-1206.
- Melcher, J. R., Kiang, N. Y. (1996). „Generators of the brainstem auditory evoked potential in cat. III: Identified cell populations,“ *Hear. Res.* 93, 52-71.
- Oxenham, A. J., and Moore, B. C. J. (1997). “Modeling the effects of peripheral nonlinearity in listeners with normal and impaired hearing,” in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, New Jersey).
- Oxenham, A., and Dau, T. (2001). „Modulation detection interference: Effects of concurrent and sequential streaming“, *J. Acoust. Soc. Am.* 110, 401-408.
- Pickles, J.O. (1988). „An introduction to the Physiology of Hearing“. Academic Press (London).
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (1997). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," *J. Acoust. Soc. Am.* 101, 2151-2163.
- Schulze, H., Hess, A., Ohl, F.W., and Scheich, H. (2002). „Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the mongolian gerbil.“ *Europ. J. Neurosci*, in press.
- Stern, R.M., and Trahiotis, C. (1995). „Models of binaural interaction“, in *Hearing*, edited by B. C. J. Moore (Academic, New York).
- Tchorz, J. and Kollmeier, B. (1999). „A model of auditory perception as front end for automatic speech recognition.“ *J. Acoust. Soc. Am.* 106, 2040-50.
- Viemeister, N.F. (1979). „Temporal modulation transfer functions based upon modulation thresholds.“ *J. Acoust. Soc. Am.* 66:1364-80.
- Wegner, O., and Dau, T. (2002). „Frequency specificity of chirp-evoked auditory brainstem responses“, *Journal of the Acoustical Society of America*, 111, 1318-1329.
- Yin, T. C., Chan, J. C. (1990). „Interaural time sensitivity in medial superior olive of cat,“ *J. Neurophysiol.* 64, 465-488.
- Yost, W. A., Sheft, S., and Opie, J. (1989). "Modulation interference in detection and discrimination of amplitude modulation," *J. Acoust. Soc. Am.* 86, 2138-2147