

Psychoakustik und Sprachperzeption bei Normal- und Schwerhörigen

Birger Kollmeier / Oldenburg

1. Wahrnehmungs-Grundgrößen

Das menschliche Gehör ist in faszinierender Weise optimal an das Empfangen von akustischen Signalen aus der Umwelt angepaßt, insbesondere an das Verstehen von Sprache. Weil das akustische Sprachsignal sich zeitlich stark ändert und unterschiedliche Frequenzanteile aufweist, benötigt das Ohr für die Sprachwahrnehmung die Fähigkeit, zu jedem Zeitpunkt die Intensität wahrzunehmen, mit der jede Frequenz momentan vorliegt. Wesentliche Grundgrößen der Wahrnehmung sind daher die Umsetzung verschiedener Schallintensitäten in subjektiv empfundene **Lautheit**, die Umsetzung verschiedener Frequenzen in subjektiv empfundene **Tonhöhen**, die Umsetzung verschiedener Zeitdauern und Rhythmen in subjektiv empfundene **Zeitmuster**, die Umsetzung von akustischen Signalen in subjektiv empfundene **Klänge**, und das Trennen verschiedener Klänge (z. B. Nutzsignal vom störenden Hintergrundsignal). Diese Beziehungen zwischen dem akustischen Reiz und der subjektiven Wahrnehmung, die durch psychologische Meßmethoden erfaßt werden können, werden in der **Psychoakustik** untersucht.

1.1 Intensitätsabbildung (Lautheit) und Intensitätsauflösung

Als **Ruhehörschwelle** wird der Pegel eines soeben wahrnehmbaren Sinustons bei einer vorgegebenen Frequenz in Ruhe (d. h. ohne Vorliegen störender Signale) bezeichnet. Bei niedrigen und bei hohen Frequenzen muß ein wesentlich höherer Schalldruckpegel erzeugt werden als bei mittleren, damit der Ton von normalhörenden Menschen schwellenhaft wahrgenommen werden kann (vgl. Abbildung 1). Wenn ein Sinus-Ton im Pegel kontinuierlich erhöht wird, nimmt die Versuchsperson nach Überschreiten der Ruhehörschwelle einen immer lauterem Ton wahr. Der Grad, in dem diese Lautheitswahrnehmung mit zunehmendem Tonpegel ansteigt, hängt von der Frequenz ab (vgl. Abbildung.1): Ausgehend von einem Sinus-Ton bei 1000 Hz kann man bei jeder Frequenz denjenigen Tonpegel bestimmen, der zum gleichen Lautheitseindruck führt. Diese Kurve wird als **Isophone** oder Kurve gleicher Pegellautstärke bezeichnet. Während die 10 Phon-Isophone bei 1000 Hz genau 10 dB über der Ruhehörschwelle ist und zu niedrigen und hohen Frequenzen ähnlich steil ansteigt wie die Ruhehörschwelle, flachen die zu höheren Pegeln gehörenden Isophonen zunehmend ab, so daß die 120 Phon-Isophone, die ungefähr der Schmerzschwelle bei Normalhörenden für alle Frequenzen entspricht, etwa den gleichen Schalldruck für alle Frequenzen angibt.

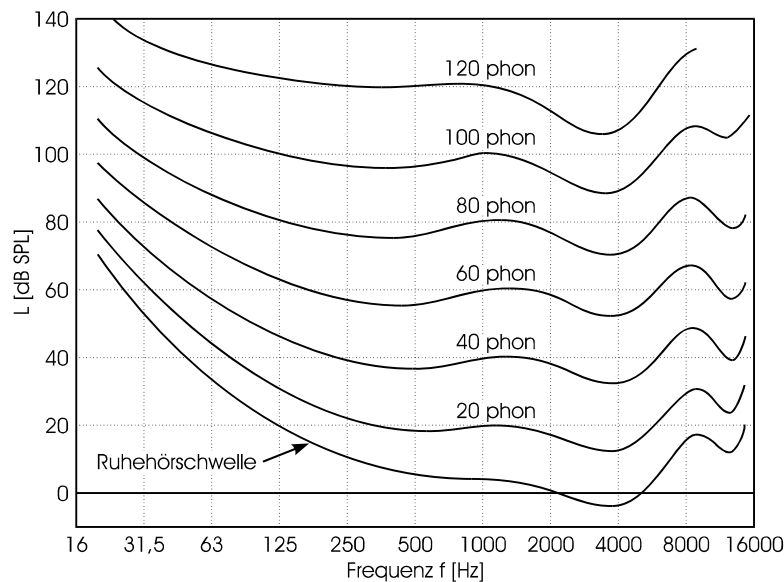


Abbildung 1: Ruhehörschwelle und Isophonen

Um den Lautheitsanstieg mit zunehmendem Pegel zu bestimmen, d. h. beispielsweise die Pegelerhöhung, die für eine Verdopplung der empfundenen Lautheit notwendig ist, bedient man sich verschiedener subjektiver Skalierungsmethoden. Bei der **freien Größenschätzung** nach Stevens und Mitarbeitern (Stevens, 1957) soll die Versuchsperson die jeweils empfundene Lautheit mit einer beliebigen Zahl kennzeichnen. Trägt man nun die von der Versuchsperson gegebenen Zahlenwerte über die angebotenen Pegel auf, erhält man die nach Stevens benannte **Potenzfunktion**

$$N[\text{sone}] = \left(\frac{I}{I_0} \right)^\alpha \quad (8.1)$$

Dabei bezeichnet N die Lautheit in der Einheit „sone“, I die Intensität des Testtons und I_0 die Intensität des Referenz-Sinustons bei 1 kHz mit einem Pegel von 40 dB SPL, der genau der Lautheit 1 sone entspricht. Der Exponent α nimmt im Mittel über viele Versuchspersonen und Experimente den Wert von etwa 0,3 an. Das bedeutet, daß bei einer Erhöhung der Schallintensität um 10 dB die Lautheit ungefähr um den Faktor 2 ansteigt (da 10 dB eine Verzehnfachung der Intensität bedeutet und $10^{0,3} \cong 2$ ist). Die Lautheit hängt damit in anderer Weise von der Schallintensität ab als der Schallpegel. Die Lautheit in sone darf auch nicht mit der Lautstärke (in Phon) verwechselt werden, da der erste Begriff die Größe einer subjektiven Empfindung kennzeichnet und der zweite Begriff einen Pegel bezeichnet (d. h. eine physikalische Reizgröße), die zu einer vorgegebenen Lautheitsempfindung führt (nämlich zu der Lautheit, die für alle Lautstärken auf einer Isophonen gleich ist). Daher wird der zweite Begriff in der Normung auch als „Pegellautstärke“ bezeichnet.

Eine andere Möglichkeit zur Bestimmung der subjektiven Lautheitsempfindung besteht bei der **kategorialen Skalierung** darin, daß der Versuchsperson eine gewisse Zahl von Lautheitskategorien (z. B. „sehr leise“, „leise“, „mittel“, „laut“ und „sehr laut“) als Antwortmöglichkeiten vorgegeben wird. Bei dem Verfahren der „Würzburger Hörfeldskalierung“ nach Heller (1985) wird jede dieser groben Kategorien in einem zweiten Schritt in 10 feinere Kategorien unterteilt. Bei der „Oldenburger Hörflächenskalierung“ (Hohmann und Kollmeier, 1995), die sich insbesondere für Messungen mit schwerhörigen Patienten eignet, werden neben den fünf Hauptkategorien vier Zwischenkategorien und die Grenzfälle „nicht gehört“ und „zu laut“ vorgegeben, so daß die Gesamtzahl der Antwortkategorien nicht 50, sondern 11 beträgt. Trägt man die Antwort der Versuchsperson in Kategorial-Einheiten über dem Pegel des Testschalls auf, erhält man im mittleren Pegelbereich eine Gerade, d. h. die empfundene Kategorial-Lautheit ist dort proportional zum Logarithmus der Intensität (vgl. Abbildung 2). Abweichungen zu der im logarithmischen Maßstab aufgetragenen Lautheit in sone treten bei sehr niedrigen und bei sehr hohen Pegeln auf:

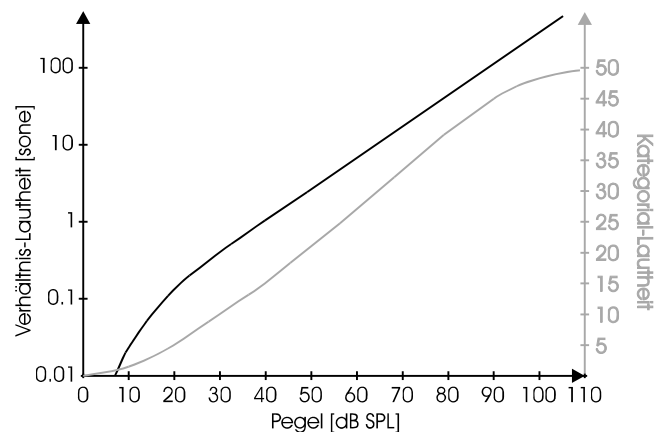


Abbildung 2: Sone-Skala und Kategorial-Skala über den Pegel

Bei sehr niedrigen Pegeln stehen für die relativ grobe Kategorial-Skalierung nicht genügend Antwortalternativen zur Verfügung, so daß die Kurve erst bei mittleren Pegeln ansteigt, während die Lautheit in sone gegen Null geht und ihr Logarithmus dementsprechend steil abfällt. Bei sehr hohen Pegeln ist der Antwortbereich der Kategorial-Lautheit ebenfalls begrenzt, so daß hier ein Sättigungseffekt auftritt, während der Logarithmus der Lautheit in sone weiter linear ansteigt. Dieser Effekt ist auf den „Logarithmic Response Bias“ (Poulton, 1989) zurückzuführen, d. h. die Versuchspersonen beurteilen bei der Erhöhung einer großen Zahl nicht mehr den absoluten Zuwachs dieser Zahl, sondern das Verhältnis der neuen Zahl zur alten Zahl. Damit tendieren die Versuchspersonen bei hohen Pegeln bei der absoluten Größenschätzung dazu, ihre Empfindungsgröße mit dem Logarithmus der angenommenen, großen Zahlen zu beschreiben. Dieser Effekt tritt bei der Kategorial-Skalierung mit einer wesentlich kleineren Zahl von Antwortalternativen nicht auf, so daß die ungefähre Proportionalität zwischen dem Logarithmus der Lautheit in sone und dem absoluten Wert der Kategorial-Skala plausibel erscheint (paralleler Verlauf der Kurven in Abbildung 2 über einen relativ großen Pegelbereich).

Eine logarithmische Beziehung zwischen der Reizgröße (in unserem Fall die Intensität I) und der Empfindungsgröße E folgt auch aus dem **Weber-Fechner'schen Gesetz**, das eines der ersten und wichtigsten Gesetze der Psychophysik darstellt. Weber und Fechner konnten nachweisen, daß der Reizunterschied ΔI , der zu einer eben wahrnehmbaren Änderung der Empfindung ΔE führt, proportional zu der absoluten Größe I des Reizes ist, so daß der sogenannte **Weber-Bruch** $\frac{\Delta I}{I}$ für alle Werte von I annähernd konstant sein muß und proportional der kleinsten wahrnehmbaren Empfindungsänderung ΔE ist:

$$\Delta E = k \cdot \frac{\Delta I}{I} \quad (8.2)$$

Durch Aneinanderreihen von eben merklichen Unterschieden ΔE (was mathematisch der Aufintegration von Gleichung 8.2 entspricht) läßt sich nun eine Beziehung zwischen der Empfindungsgröße E und der Reizgröße I herleiten:

$$E = k' \cdot \log_{10} I + k'' \quad (8.3)$$

k' und k'' sind Konstanten, die aus den Randbedingungen festgelegt werden. Diese logarithmische Abhängigkeit der Wahrnehmungsgröße von der Reizgröße entspricht gerade dem Zusammenhang zwischen Kategorial-Lautheit und Tonpegel, d. h. dem Logarithmus der Intensität. Dieses Gesetz entspricht auch der sone-Skala unter der Annahme, daß die empfundene Lautheit durch den Logarithmus der von der Versuchsperson angegebenen Zahl, und nicht durch die Zahl selber wiedergegeben wird.

Neben der bisher behandelten absoluten Wahrnehmung der Schallintensität interessiert auch die **differentielle Wahrnehmung**, d. h. die kleinste wahrnehmbare Änderung einer Schallintensität. Sie wird auch als JND („just noticeable difference“) bezeichnet und tritt als ΔI im Weber-Fechner'schen Gesetz (Gleichung 8.2) auf. Für die meisten Schalle beträgt die JND etwa 1 dB, d. h.

$$\frac{\Delta I + I}{I} = 10^{\frac{1\text{dB}}{10}} \Rightarrow \frac{\Delta I}{I} = 0,259 \quad (8.4)$$

In der Tat mußte die erste logarithmische Pegelskala, deren Einheit das Bel ist (nach Alexander Graham Bell, der unabhängig von Philip Reis das Telefon erfand und zu einem kommerziellen Erfolg führte, $1 \text{ Bel} = \log_{10} (I/I_0)$) in eine 10-fach feinere Skala, das **Dezibel (dB)** unterteilt werden. Damit entspricht eine Einheit dieser Skala gerade einem kleinsten hörbaren Pegelunterschied. Während bei breitbandigen Signalen über einen

weiten Pegelbereich Webers Gesetz gültig ist und der kleinste wahrnehmbare Pegelunterschied 1 dB beträgt, ist bei schmalbandigen Signalen (wie Sinustönen) dieser Unterschied bei kleinen Pegeln größer als 1 dB und nimmt mit zunehmendem Pegel stetig ab. Diese Abweichung wird als „near miss to Weber’s law“ bezeichnet und hängt mit der speziellen Anregung der Basilarmembran bei Sinustönen zusammen.

1.2 Tonhöhe und Frequenzauflösung

Wird die Frequenz eines Sinustons erhöht, führt dies zu der Wahrnehmung einer ansteigenden Tonhöhe. Wenn diese Tonhöhenwahrnehmung dem Weber-Fechner’schen Gesetz entsprechend proportional zum Logarithmus der Frequenz wäre, entspräche dies genau der musikalischen Tonhöhenempfindung: Eine Oktave (Frequenzverhältnis 2:1) würde als doppelt so hoch empfunden werden, unabhängig von der jeweiligen Frequenz. Tatsächlich folgt die Tonhöhenwahrnehmung dieser Gesetzmäßigkeit nur in einem begrenzten Bereich mittlerer Frequenzen, der bei musikalisch ausgebildeten Versuchspersonen sich noch etwas erweitern lässt. Lässt man Versuchspersonen jedoch unabhängig von musikalischen Intervallen die empfundene Tonhöhe im Vergleich zum Referenzton bei 1 kHz angeben, so erhält man die Mel-Skala als subjektive Tonhöhenempfindungsskala. Sie entspricht ungefähr der Frequenz-Orts-Transformation auf der Basilarmembran, bei der die verschiedenen Frequenzen zu einer maximalen Auslenkung an unterschiedlichen Orten führen (gepunktete Linie in Abb. 3). Die gleiche Beziehung zwischen subjektiver Tonhöhe und objektiver Frequenz liefert die in Abb. 3 dargestellte Bark-Skala (benannt nach dem deutschen Physiker und Akustiker Heinrich Barkhausen), die auf dem Konzept der Frequenzgruppe beruht (Zwicker und Mitarb. 1957).

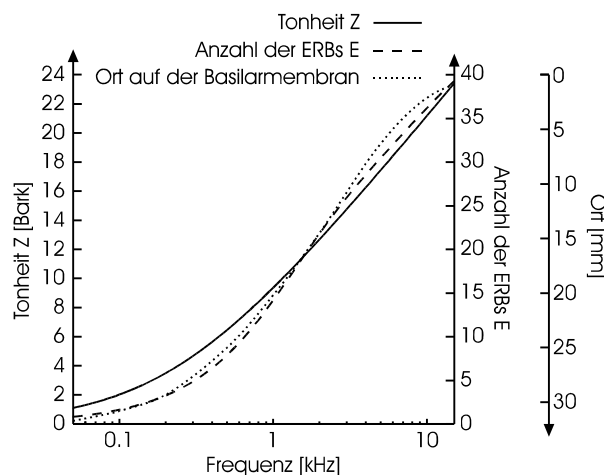


Abbildung 3: Mel-Skala, Bark-Skala, ERB-Skala

Sie basiert auf der Frequenzabhängigkeit der Lautheitswahrnehmung und der Maskierung: Bei der **Lautheitssummation** wird die empfundene Lautheit eines schmalbandigen Signals (z. B. eines Schmalbandrauschens) bei gleicher Leistung, aber verschiedener Bandbreite bestimmt (vgl. Abbildung 4). Wenn dieselbe Leistung auf

einen größeren Frequenzbereich verteilt wird, steigt die wahrgenommene Lautheit auf etwa das Doppelte an (entspricht ungefähr 10 dB Lautstärkengewinn). Bleibt die Leistung jedoch in einem Frequenzbereich, der kleiner als die **Frequenzgruppenbreite** ist, hängt die wahrgenommene Lautheit nicht von der Bandbreite ab, so daß aus dieser Bandbreitenabhängigkeit auf die Größe der Frequenzgruppe geschlossen werden kann. Die Frequenzgruppenbreite beträgt etwa 100 Hz unterhalb von 500 Hz und etwa 1/5 der Frequenz oberhalb von 500 Hz (vgl. Abbildung 3).

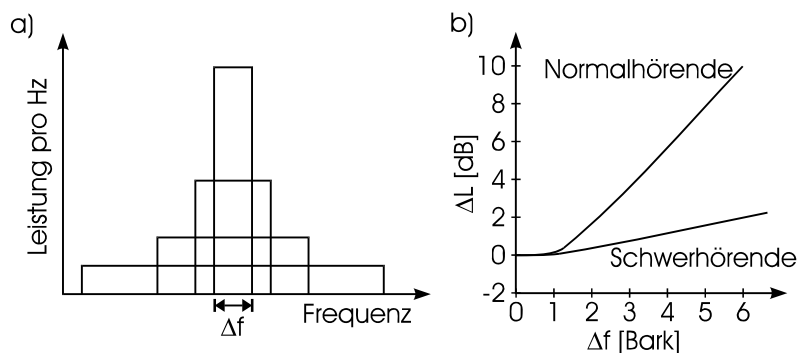


Abbildung 4: Bandbreitenabhängigkeit der Lautheit

Anschaulich versteht man unter der Frequenzgruppe diejenige Bandbreite im Gehör, innerhalb derer sämtliche Signale gemeinsam verarbeitet und zu einem „Erregungspegel“ zusammengefaßt werden, der der Lautheitswahrnehmung zugrunde liegt. Wenn das Signalspektrum in verschiedene Frequenzgruppen fällt, wird nicht mehr die Leistung aufsummiert, sondern die Teil-Lautheiten in den verschiedenen Frequenzgruppen.

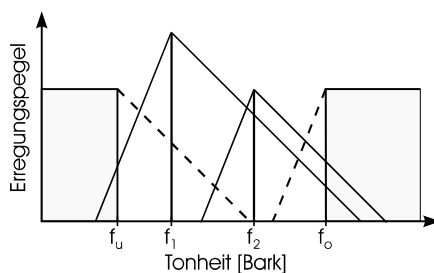


Abbildung 5: Erregungspegelmuster

Eine Vorstellung zum Zustandekommen der Frequenzgruppe und der Frequenzauflösung im auditorischen System ist die Erregungsverteilung auf der Basilarmembran (vgl. Abbildung 5). Bei einem Sinus-Ton mit f_1 wird die Basilarmembran nicht nur bei dieser Frequenz, sondern im schwächerem Maße auch bei den darüber- bzw. darunterliegenden Frequenzen angeregt. Wenn auf der x-Achse die Frequenz in Bark aufgetragen ist, kann diese Verbreiterung der Erregung als ein dreieckförmiges Muster dargestellt werden, dessen Flanken zu tiefen Frequenzen mit etwa 25 dB pro Bark ansteigen und zu hohen Frequenzen mit etwa 10 dB pro Bark abfallen. Bei hohen Pegeln werden diese Flanken flacher. Wird nun ein weiterer Sinus-Ton mit einer Frequenz f_2 mit einem Pegel angeboten, der unterhalb dieses

Erregungspegels bei der Frequenz f_2 liegt, wird dieser zusätzliche Ton vom ersten Ton vollständig **maskiert**, d. h. der zweite Ton wird im Beisein des ersten Tons nicht mehr gehört. Durch derartige Verdeckungsexperimente kann die Form und Steilheit der Erregungspegel-Verteilung über der Frequenz ausgemessen werden.

Um aus derartigen Experimenten auf die Frequenz-Gruppe zu schließen, wird meist ein **Notched-Noise** verwendet, bei dem ein Rauschen nur unterhalb einer Grenzfrequenz f_u und oberhalb einer Frequenz f_o vorliegt. Je schmaler die zwischen f_u und f_o liegende spektrale Lücke („notch“) ist, desto höher liegt die **Mithörschwelle** eines in diesem Frequenzbereich liegenden Tons, d. h. der Pegel des Tons muß relativ hoch sein, damit er gehört wird. Aus der Abnahme dieser Mithörschwelle mit zunehmender Breite der spektralen Lücke kann auf die Frequenzgruppenbreite geschlossen werden, die in Einheiten einer „equivalent rectangular bandwidth“(ERB) nach Moore und Glasberg (1987) gemessen wird. Darauf beruht die ebenfalls in Abb. 3 angegebene ERB-Skala, die sehr ähnlich der Bark-Skala ist.

Die Verteilung der Erregungspegel über der Frequenz kann für die Berechnung der Lautheit verwendet werden, wobei die sich überlappenden Bereiche des Erregungspegel-Verlaufs für verschiedene Frequenzen nur einmal berücksichtigt werden. Dazu wird bei jeder Frequenzgruppe der Erregungspegel gemäß dem Potenzgesetz (Gleichung 8.1) umgeformt. Die Gesamt-Lautheit wird dann durch Integration (Aufsummation) der spezifischen Lautheiten über sämtliche Frequenzen gebildet. Diese Berechnung entspricht dem nach Iso 532 B genormten **Lautheitsmodell nach Zwicker**, das für stationäre Schallsignale die subjektiv empfundene Lautheit in sone sehr gut berechnen kann. Dabei werden sowohl spektrale Maskierungseffekte als auch Lautheitssummutations-Effekte und Kombinationen dieser Effekte richtig vorhergesagt. Eine Anwendung auf zeitlich stark schwankende Schall (wie z. B. Sprache) ist jedoch problematisch, da das Modell für stationäre Signale entwickelt wurde.

Die Frequenzgruppe läßt sich auch mit einer Reihe weiterer Maskierungsexperimente bestimmen, bei denen sich der maskierende Schall (meistens ein Rauschen) stark im Frequenzbereich ändert, z. B. sinusförmige Änderung im Frequenzbereich („ripple-noise“), Tiefpaßrauschen oder Hochpaßrauschen. Für alle Fälle werden ähnliche Frequenzgruppenbreiten gemessen, wobei sich neben der Breite dieser Frequenzgruppen auch die effektive Form der Filter im Gehör angeben läßt, mit denen jeweils ein begrenzter Frequenzbereich analysiert und zu einer Frequenzgruppe zusammengefaßt wird. Die einfachste Form ist rechteckförmig, eine realistischere Filterform ist ein gerundetes Exponentialfilter dessen Flanken die Verhältnisse im Erregungspegelmuster widerspiegeln. Man darf sich eine Frequenzgruppe nicht als eine starre, bei einer festen Frequenz gelegene rechteckförmige Filterung des Schallsignals vorstellen, sondern eine kontinuierliche, bei jeder Mittenfrequenz auftretende Filterung mit einer abgerundeten, kontinuierlichen Filterform.

Im Gegensatz zu der relativ geringen Frequenzauflösung durch die Frequenzgruppenfilter bei der Maskierung steht die sehr hohe Frequenzauflösung bei der **Diskrimination** verschiedener Frequenzen, die nicht gleichzeitig, sondern in Folge angeboten werden. Dies bei der Musikwahrnehmung (z. B. beim Stimmen von Instrumenten) wichtige Phänomen der **Tonhöhenunterscheidung** ermöglicht einen kleinsten hörbaren Frequenzunterschied von etwa 3 Hz für Frequenzen unterhalb von 500 Hz und etwa 0,6 % für Frequenzen über 1000 Hz. Dies entspricht ungefähr 1/30 Bark, d.h. die Tonhöhenunterscheidung ist wesentlich feiner als die Maskierung im Frequenzbereich. Dies liegt an der massiven Parallelverarbeitung im Gehör: Wenn die Frequenz eines einzelnen Sinustons um wenige Herz verschoben wird, verschiebt sich das ganze Erregungsmuster auf der Basilarmembran, so daß im Mittel über sämtliche beteiligten Nervenfasern selbst ein kleiner Unterschied in der Verschiebung des Erregungsmusters feststellbar ist, obwohl das von einem einzelnen Ton hervorgerufene Erregungsmuster selbst relativ breit sein kann. Daher mißt die Frequenzauflösung (Diskrimination) eine andere Leistung des Gehörs als die Maskierung im Frequenzbereich, die auch als **spektrale Integration** bezeichnet werden kann.

1.3 Zeitliche Verarbeitung im Hörsystem

Neben der spektralen Verschmierung (Integration) und Frequenzauflösung spielt die zeitliche Verarbeitung von akustischen Signalen im auditorischen System eine wichtige Rolle, die ebenfalls durch eine zeitliche Verschmierung (Zeitintegration) und eine Zeitauflösung gekennzeichnet werden kann. Die **zeitliche Integration** bezeichnet dabei die Fähigkeit, einen lang andauernden, stationären Klang bei gleichem Pegel als lauter wahrzunehmen als einen kurzen Klang. Diese Eigenschaft kann mit einem Maskierungsexperiment demonstriert werden (vgl. Abbildung 6), bei der die Mithörschwelle für einen Signalton bestimmt wird, d. h. der Pegel, bei dem der Ton im Rauschen soeben noch hörbar ist. Sie nimmt mit zunehmender Dauer T des Testtons ab und erreicht für Werte von etwa 200 ms einen stabilen Wert, der durch weitere Verlängerungen des Tons nicht mehr verändert wird. Diese Eigenschaft kann durch eine Energieintegration über einen Bereich von 200 ms erklärt werden. Bei Innenohrschwerhörigen ist diese Energie-Summentation gestört, so daß die Schwelle bei zunehmender Dauer T nicht weiter abnimmt.

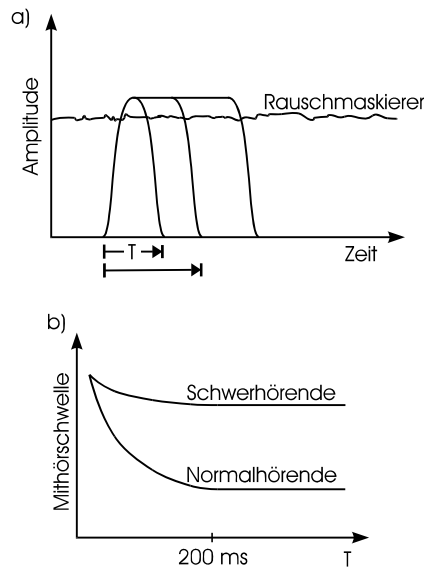


Abbildung 6: Zeitliche Integration (Schema, s. Text)

Die zeitliche Verschmierung der internen Repräsentation akustischer Signale lässt sich durch die psychoakustisch meßbaren Phänomene der **Nachverdeckung** und **Vorverdeckung** beschreiben. Bei der Nachverdeckung wird ein Testsignal (z. B. ein kurzer Testtonpuls) zeitlich nach dem Abschalten eines Maskierungssignals (z. B. Rauschen) angeboten und der Pegel des Tons wird solange variiert, bis er soeben nicht mehr hörbar ist (Mithörschwelle). Diese Mithörschwelle nimmt mit zunehmendem Abstand des Testtons vom Maskierende ab und erreicht bei etwa 200 ms die Ruhehörschwelle. Dabei ist der Verlauf dieser Nachverdeckungskurve abhängig von der Maskiererdauer, d. h. vom Adaptationszustand des auditorischen Systems (vgl. Abbildung 7). Bei der Vorverdeckung wird der Testton dagegen nicht nach dem Abschalten, sondern zeitlich vor dem Anschalten eines Maskierers angeboten, wobei sich eine ähnliche Verschmierung der Mithörschwelle als Funktion der Verzögerungszeit ergibt wie bei der Nachverdeckung. Allerdings ist der Zeitbereich, über den sich die Vorverdeckung erstreckt, wesentlich kürzer (etwa 10 ms) und es ergibt sich keine vergleichbar starke Abhängigkeit von der Dauer des Maskierers.

Einen kombinierten Effekt von Vor- und Nachverdeckung beobachtet man bei der **Lückendetektion** (Englisch: Gapdetection), bei der die kleinste in einem Rauschen wahrnehmbare Pausendauer gemessen wird (vgl. Abb. 8). Da aufgrund der zeitlichen Verschmierung im auditorischen System kleine Lücken im Rauschen nicht wahrgenommen werden können, liegt bei Normalhörenden bei der Verwendung von breitbandigen Rauschen die minimal detektierbare Lückendauer bei etwa 8 ms. Da sie ein einfach und schnell zu bestimmendes Maß für die zeitliche Verarbeitung ist, wird sie oft zur Charakterisierung der gestörten zeitlichen Verarbeitung bei Schwerhörenden eingesetzt, wobei aufgrund starker interindividueller Schwankungen Werte zwischen 8 und 100 ms auftreten können.

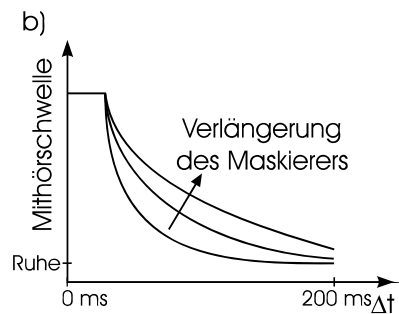
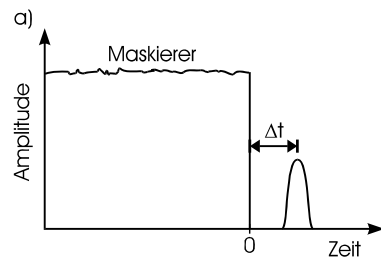


Abbildung 7: Nachverdeckung

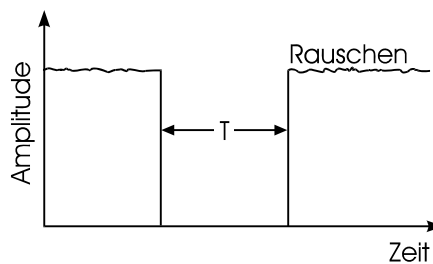


Abbildung 8: Gap-detection

Eine weitere Eigenschaft der zeitlichen Verarbeitung akustischer Signale ist die **Modulationswahrnehmung**, d. h. die Wahrnehmung von Schallen mit einer aufgeprägten (z. B. sinusförmig variierenden) Einhüllenden (vgl. Abb. 9). Als Trägersignale können Sinussignale oder Rauschen verwendet werden, während als Modulationssignale (d. h. Verformung der Einhüllenden) zumeist sinusförmige Signale mit einer bestimmten Modulationsfrequenz f_{mod} verwendet werden. Eine derartige Signalform ist sehr ähnlich der Form von Sprache, weil bei der Sprachartikulation ein Trägersignal (z. B. die Schwingung der Stimmritze) durch den Vokaltrakt in der Intensität zeitlich stark verändert wird. Im Modulationsfrequenzbereich zwischen 0 und etwa 10 Hz werden Modulationen als Schwankungen der Lautstärke wahrgenommen, die Empfindung wird als **Schwankungsstärke** bezeichnet. Bei Frequenzen zwischen etwa 10 und 20 Hz entsteht ein „knarrender“ Klangeindruck, der auch als R-Rauhigkeit bezeichnet wird, während Modulationen zwischen etwa 20 Hz und 80 Hz als **Rauhigkeit** des Klanges wahrgenommen werden können. Bei noch höheren Modulationsfrequenzen treten im Frequenzspektrum des akustischen Signals Seitenlinien auf, die aufgrund der spektralen Verfärbung als Klangänderungen wahrgenommen werden können.

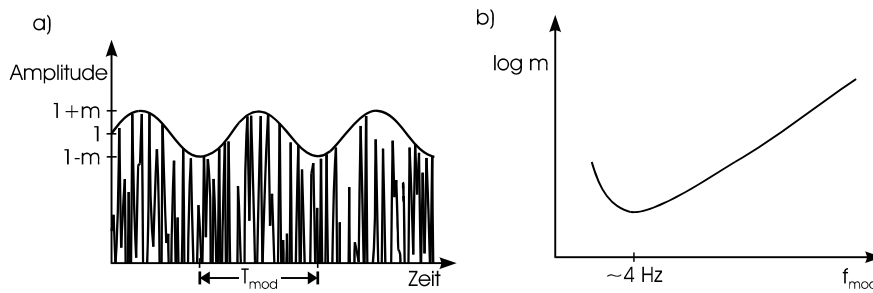


Abbildung 9: Moduliertes Signal, PMTF für breitbandige Träger, d. h. der kleinste detektierbare Modulationsgrad als Funktion der Modulationsfrequenz

Als Maß für die Zeitauflösung kann die **psychoakustische Modulationstransfer-Funktion** bestimmt werden, d. h. für jede Modulationsfrequenz derjenige Modulationsgrad m , der soeben noch wahrgenommen werden kann. Sie zeigt für Modulationen von etwa 4 Hz ein Maximum (dies entspricht in etwa der Sprachsilbenfrequenz und ist daher gut an die zu empfangene Sprache angepaßt) und für zunehmende Modulationsfrequenzen wird die Empfindlichkeit geringer. Diese Modulations-Tiefpaßcharakteristik des auditorischen Systems korrespondiert zu der zeitlichen Verschmierung, die in den Vor- und Nachverdeckungsexperimenten aufgezeigt wurde.

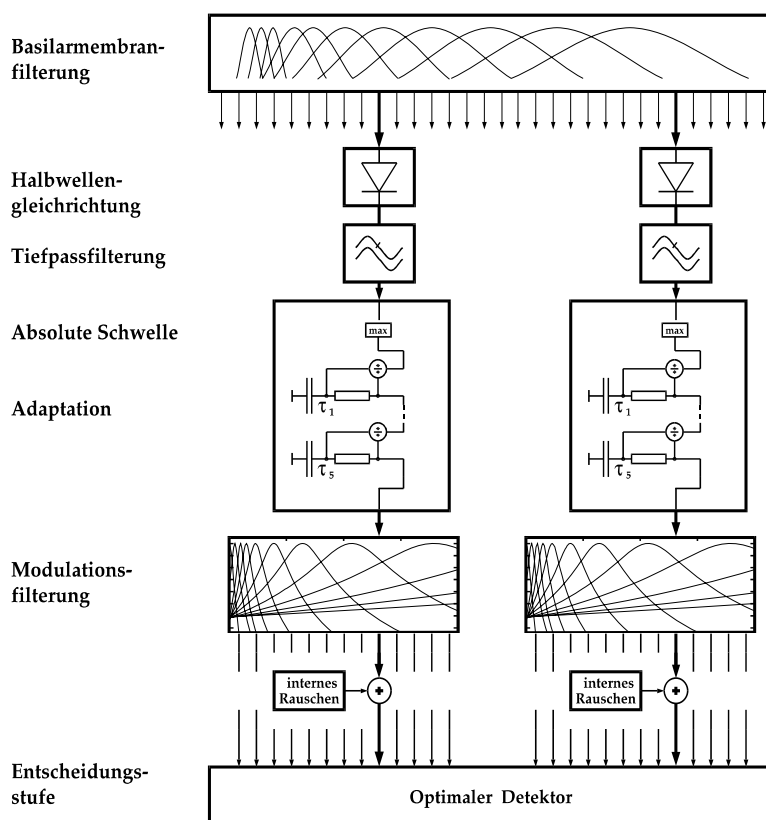


Abbildung 10: Verarbeitungsmodell nach Dau et al.

Um die verschiedenen zeitlichen Eigenschaften des Hörsystems (zeitliche Integration, Vor- und Nachverdeckung, Modulationswahrnehmung) mit möglichst wenig Annahmen beschreiben zu können, wurden Hörmodelle zur zeitlichen Verarbeitung aufgestellt. Bei dem in Abbildung 10 schematisch dargestellten Verarbeitungsmodell nach Dau et al. (1997) wird die Verarbeitung unterschiedlicher Frequenzen zunächst durch eine Filterbank beschrieben, die das Eingangssignal in verschiedene Frequenzgruppenbreite Bandpaß-Signale aufteilt. In jedem Bandpaßsignal wird eine Halbwellengleichrichtung mit Tiefpaßfilterung durchgeführt, das ungefähr der Aktion der Haarzellen entspricht und eine Extraktion der Einhüllenden bewirkt. Die zeitlichen Verarbeitungs-Eigenschaften werden durch die nachgeschalteten Adaptations-Schleifen nachgebildet, bei denen das Eingangssignal durch das tiefpaßgefilterte Ausgangssignal geteilt wird. Dadurch wird eine gewisse Adaptation an den Mittelwert des Eingangssignals ermöglicht, während schnelle Änderungen (Fluktuationen) ohne weitere Beeinflussung durchgelassen werden. Durch Hintereinanderschalten mehrerer dieser **Nachregelschleifen** mit unterschiedlichen Zeitkonstanten zwischen 1ms und 500 ms wird approximativ eine logarithmische Kompression für langsame Änderungen im Eingangssignal erreicht und eine unveränderte Durchlässigkeit für schnelle Änderungen im Eingangssignal. Mit dieser Modellstruktur lassen sich Vor- und Nachverdeckung sowie die Lückendetektion (Gap detection) bereits sehr gut beschreiben. Die zeitliche Integration von Testtönen und die Modulationswahrnehmung kann durch Einfügen einer Modulationsfilterbank erreicht werden, die für jede Mittenfrequenz die Einhüllenden-Fluktuationen in verschiedene Modulationsfrequenzen aufspaltet. Am Ausgang dieser Modulationsfilterbank steht ein bestimmtes, zeitlich variables Muster zur Verfügung, das die **interne Repräsentation** von akustischen Signalen darstellen soll. Mit dieser internen Repräsentation lassen sich verschiedene Experimente einfach dadurch beschreiben, daß eine bestimmte, minimale Änderung in dieser Repräsentation zur Wahrnehmung eines Unterschieds im Klang führt, der gerade zur Detektion des „gesuchten“ akustischen Stimulus führt.

1.4 Binaurale Interaktion

Neben den bisher beschriebenen Leistungen des Gehörs bei der Intensitäts-, Frequenz-, und Zeitverarbeitung im auditorischen System ist die binaurale Interaktion, d. h. die im Gehirn stattfindende vergleichende Verarbeitung der an beiden Ohren anliegenden Signale von besonderer Wichtigkeit für das Hören in natürlicher Umgebung. Es trägt signifikant zur **Hallunterdrückung**, zur Ortung (**Lokalisation**) von Schallquellen im Raum und zur **Unterdrückung** von „unerwünschten“ Störgeräuschquellen in realen akustischen Situationen bei.

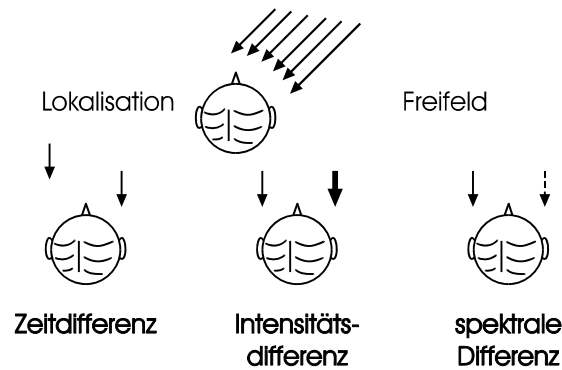


Abbildung 11: Physikalische Merkmale für die Lokalisation

Wenn eine Schallquelle so im Raum angeordnet ist, daß der von ihr ausgesendete Schall mit einem bestimmten Einfallswinkel auf den Kopf des Hörers fällt, bewirkt dieser (winkelabhängige) Schalleinfall eine **interaurale Zeitverzögerung** (d. h. der Schall erreicht das der Schallquelle zugewandte Ohr eher als das abgewandte Ohr), einen **interauralen Intensitätsunterschied** (d. h. das der Schallquelle zugewandte Ohr empfängt eine höhere Schallintensität), sowie eine von der Einfallsrichtung abhängige **spektrale Verfärbung** (d. h. der Frequenzgehalt des empfangenen Schalls wird je nach Einfallsrichtung unterschiedlich verändert vgl. Abb. 11.). Aufgrund dieser akustischen Merkmale ist der normalhörende Mensch bei breitbandigen Signalen in der Lage, den Ort einer Schallquelle aus der Vorne-Richtung mit einer Ungenauigkeit von etwa 1 Grad bzw. bei seitlichem Einfall oder Schalleinfall von oben mit einer Ungenauigkeit von etwa 5 Grad aufzulösen. Diese erstaunlich hohe Ortungsleistung wird allerdings bei schmalbandigen Signalen und bei Vorliegen von Nachhall bzw. Echo eingeschränkt.

Dieser Effekt ist darauf zurückzuführen, daß die bei einer bestimmten Frequenz anliegenden interauralen Intensitäts- und Zeitunterschiede nicht eindeutig zu einer bestimmten Einfallsrichtung gehören, sondern daß mehrere Schalleinfallsrichtungen bei derselben Frequenz zu denselben interauralen Intensitäts- und Zeitunterschieden führen können. Beispielsweise führen sämtliche Einfallsrichtungen in der Medianebene zu einer interauralen Zeit- und Laufzeitdifferenz von ungefähr 0, so daß bei schmalbandigen Signalen nicht sicher zwischen der Vorne-, Oben-, und Hinten-Einfallsrichtung unterschieden werden kann und es zu **Richtungsverwechslungen** kommt. Ein Schalleinfall von vorne links kann in ähnlicher Weise mit einem Schalleinfall von hinten links und Einfallsrichtungen verwechselt werden, die auf den sogenannten „**Cone of Confusion**“ liegen. Bei schmalbandigen akustischen Signalen werden zudem bestimmte Frequenzen zu bestimmten Schalleinfallsrichtungen zugeordnet („**richtungsbestimmende Bänder**“ nach Blauert, 1974), so daß hier die spektrale Information teilweise die Information aus den interauralen Zeit- und Intensitätsunterschieden überschreibt. Erst bei breitbandiger Schalldarbietung kann durch den Vergleich über mehrere Frequenzen hinweg eine eindeutige, sämtliche Mehrdeutigkeiten bei schmalbandiger Signaldarbietung vermeidende Lokalisation durchgeführt werden.

Die Rolle der interauralen Zeit- und Intensitätsdifferenzen für die Lokalisation und die binaurale Verarbeitung kann mit Kopfhörer-Experimenten erforscht werden, bei denen jeder dieser physikalischen Parameter einzeln variiert werden kann. So stellt sich

heraus, daß die interauralen Zeitdifferenzen insbesondere bei niedrigen Frequenzen (unter 1,5 kHz) eine dominierende Rolle für die Lokalisation spielen und der kleinste wahrnehmbare interaurale Zeitunterschied etwa 20 μs beträgt, während er oberhalb von 1,5 kHz etwa 50 μs beträgt. Die interauralen Intensitäts-Unterschieds-Schwellen betragen dagegen unterhalb von 1,5 kHz etwa 3 dB und oberhalb von 1,5 kHz etwa 1 dB, so daß sie besonders bei hohen Frequenzen für die Lokalisation dominant sind. Diese bereits von Lord Rayleigh (1877) formulierte **Duplex-Theorie** des binauralen Hörens hat ihre physikalische Begründung darin, daß bei niedrigen Frequenzen aufgrund der (relativ zum Kopfdurchmesser) großen Schallwellenlänge nur geringe Intensitätsunterschiede zwischen den beiden Ohren auftreten, so daß Intensitätsunterschiede erst bei hohen Frequenzen verlässlich ausgewertet werden können. Bei niedrigen Frequenzen können die Neuronen im auditorischen System der Schallwellenform noch sehr gut folgen und daher interaurale Zeit-Unterschiede gut detektieren, während bei hohen Frequenzen interaurale Zeit-Differenzen nur im Einhüllenden-Verlauf der Signale, nicht jedoch in der (von den Neuronen nicht mehr aufgelösten) Feinstruktur der übertragenden Signale erfaßt werden können.

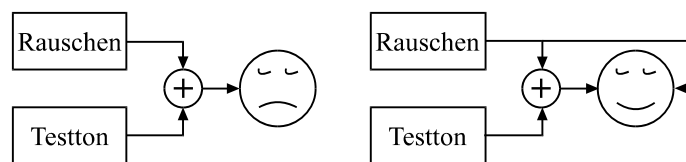


Abbildung 12: Binaurale Störgeräuschbefreiung

Zur Erfassung der **Störgeräuschbefreiung** im binauralen System kann ein psychoakustisches Experiment verwendet werden, bei dem die Detektion von Testtönen im Rauschen zwischen der binauralen und der monauralen Situation verglichen werden (vgl. Abb. 12). Zunächst wird der Versuchsperson ein Ton im Rauschen angeboten, dessen Pegel solange verringert wird, bis die Versuchsperson ihn nicht mehr hört (Mithörschwelle). Wenn anschließend dasselbe Rauschen am anderen Ohr gleichzeitig angeboten wird, wird die wahrgenommene Lautheit des Rauschens zwar höher, der Ton wird jedoch wieder hörbar, so daß eine Verringerung der Mithörschwelle gemessen werden kann, die als binaurale **Masking Level Difference** (MLD) bezeichnet wird. Ihre Größe hängt von der interauralen Phasenlage des Testtons und der zwischen den beiden Ohren bestehenden Beziehung zwischen dem Rauschen ab. Beispielsweise wird die MLD maximal (ca. 14-18 dB), wenn dasselbe Rauschen und ein um 180 Grad (Phase π) verschobener Sinuston bei etwa 500 Hz verwendet wird. Keine MLD tritt dagegen auf, wenn für das Rauschen und das Signal dieselbe Beziehung an beiden Ohren vorliegt (z. B. bei identischem Rauschen und identischem Ton an beiden Ohren oder bei Phaseninvertierung zwischen beiden Ohren).

In einer guten Näherung kann diese Störgeräuschunterdrückung des binauralen Systems durch eine (mit Restfehlern behaftete) **Subtraktion** der beiden Ohrsignale aufgefaßt werden, die im Gehirn nach einer Anpassung der vom rechten und linken Ohr stammenden Signale (z. B. durch entsprechende Zeitverzögerung und Amplituden-Anpassung) vorgenommen wird. Mit dieser als „Equalization and Cancellation Theory“ (EC-Theorie) nach Durlach (1972) bezeichneten Modellvorstellung kann bereits eine

Vielzahl der binauralen Maskierungsexperimente quantitativ gedeutet werden. Ein eher auf physiologischen Vorstellungen basierendes Modell der binauralen Signalverarbeitung wurde dagegen von Jeffress (1948) aufgestellt. Es sieht eine Art neuronales Kreuzkorrelations-Netzwerk vor, bei dem die an den beiden Ohren ankommenden Signale in einer Laufzeitkette sukzessiv zeitverzögert werden und durch Koinzidenz-Detektoren die Laufzeitunterschiede zwischen den beiden Ohren jeweils abgegriffen werden. Auf der Basis dieses Modells wurden in der neueren Literatur eine Reihe weiterer Modelle entwickelt, die die binaurale Signalverarbeitung einschließlich einer Reihe von Effekten richtig vorhersagen (z. B. die Lateralisation bzw. Lokalisation von Schallereignissen mit bestimmter interauraler Zeit- und Pegeldifferenz, das Gesetz der ersten Wellenfront, die MLD für verschiedene Signale und Maskierer, sowie zeitliche und spektrale Eigenschaften dieser Effekte). Da auf diesem Gebiet noch weiter intensiv geforscht wird, sei auf die weiterführende Literatur von Blauert (1983) und Colburn (1996) verwiesen.

2. Psychoakustik des pathologischen Gehörs

-Intensitätsabbildung

Bei sämtlichen Formen der Schwerhörigkeit tritt ein **Hörverlust** auf, d. h. eine Verringerung der Sensitivität des Hörorgans für Schall. Der Pegel eines akustischen Signals muß daher gegenüber Normalhörenden angehoben werden, um eine Empfindung hervorzurufen, d. h. die Hörschwelle ist angehoben. Die Frequenzabhängigkeit der Hörschwelle wird im **Tonschwellenaudiogramm** erfaßt. Bei weiterer Erhöhung des Pegels ergibt sich ein Anstieg der Lautheit, der bei Innenohrschwerhörigen in der Regel steiler ist als bei Normalhörenden (**Recruitment**). Dieser Effekt führt zu dem bei vielen Schwerhörenden beobachteten Phänomen, daß sie bei niedrigen Sprachlautstärken nichts verstehen. Wenn der Gesprächspartner aufgefordert wird, doch etwas lauter zu reden, beklagen sich die Schwerhörigen darüber, daß es schnell zu laut ist und die Sprecher nicht gleich schreien sollen. Dieses Phänomen kann mit **überschwelligen Tests** erfaßt werden. In der Routine-Audiometrie wird bei seitendifferentem Hörverlust dazu der Fowler-Test verwendet.

Den Lautheitsanstieg mit zunehmenden Pegel kann man für jedes Ohr separat mit der Hörlächenskalierung ermitteln, bei der sich bei Patienten mit Recruitment ein steilerer Anstieg der Kategoriellautheit ergibt. Die Steigung dieser Lautheitsfunktion ist jedoch individuell sehr unterschiedlich und kann möglicherweise von dem Verhältnis der Schädigung innerer und äußerer Haarzellen abhängen (vgl. 3.2). Daher ist der **Kompressionsverlust** des auditorischen Systems, der sich in der Steigung der Hörlächenskalierung ausdrückt, der als eine weitere, vom Hörverlust weitgehend unabhängige Komponente zu erfassen. Der Kompressionsverlust nimmt mit zunehmenden Hörverlust zwar tendenziell zu, läßt sich aufgrund der großen individuellen Schwankungen aber nicht daraus vorhersagen (vgl. Kießling, 1995, Launer et al, 1996). Bei Schalleitungs-Schwerhörigkeit ist die Pegel-Lautheitsfunktion der Hörlächenskalierung zu höheren Pegeln hin verschoben. Aufgrund des Adaptationseffektes ist diese Verschiebung jedoch nicht so groß wie der an der Schwelle meßbare Hörverlust. Außerdem kann auch die Steigung leicht variieren, so

daß die Hörflächenskalierung keine sichere Unterscheidungsmöglichkeit zwischen Schalleitungs- und Schallempfindungs-Schwerhörigkeit bietet.

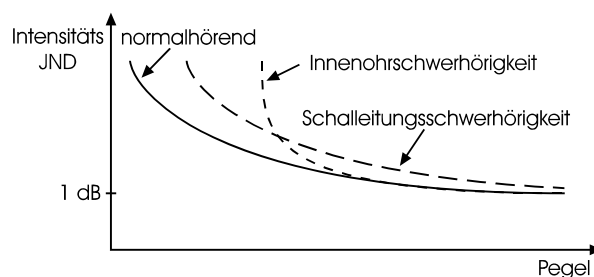


Abbildung 13: Intensitäts-JND für verschiedene Schwerhörigkeitsformen

Für die Schalleitungs-Schwerhörigkeit verschiebt sich die Pegelabhängigkeit der **Intensitätsauflösung** (d. h. der kleinste hörbare Intensitätsunterschied, Intensitäts-Just Noticeable Difference, JND) für Sinustöne zu höheren Pegeln (Abb 13). Bei Innenohrschwerhörigen mit Recruitment geschieht die Abnahme der Intensitäts-JND dagegen in einem wesentlich kleineren Pegelbereich. Sie haben bei derselben wahrgenommenen Lautheit ungefähr die gleiche bzw. eine geringgradig erhöhte Intensitäts-JND wie Normalhörende. Die früher verbreitete Annahme, daß Innenohrschwerhörige eine kleinere Intensitäts-JND besitzen, weil sie einen steileren Lautheitsanstieg haben, hat sich als falsch herausgestellt (Hohmann, 1993). Wenn man die Intensitäts-Diskrimination allerdings knapp oberhalb der Hörschwelle mißt (z. B. 20 dB wie beim SISI-Test in der überschwelligen Audiometrie), können Innenohrschwerhörige etwas niedrigere Diskriminations-Schwellen erreichen, weil sie dort eine größere Lautheit wahrnehmen als Normalhörende. Allerdings ist diese Diskriminationsschwelle sehr variabel, so daß der SISI-Test ein sehr indirektes Maß für ein Recruitment darstellt, was sich als sehr unzuverlässig herausgestellt hat. Bei breitbandigen Signalen (z. B. Rauschen oder Sprache) hängt die Intensitäts-Diskrimination sehr wenig vom Pegel ab. Innenohrschwerhörige zeigen daher in ihrem Dynamikbereich eine leicht erhöhte Intensitäts-JND gegenüber Normalhörenden, wobei jedoch die gleiche (geringere) Pegelabhängigkeit auftritt. Bei hochgradig Schwerhörigen mit einer sensorineuralen Schwerhörigkeit bzw. bei Patienten mit einer neuralen Schwerhörigkeit kann die Intensitäts-JND relativ stark erhöht sein, so daß es diesen Patienten schwerfällt, akustische Signale (wie z. B. Sprache) in ihrem auditorischen System richtig abzubilden und zu verstehen.

-Frequenz- und Zeitabbildung

Bei Schalleitungsschwerhörigkeit treten keine nennenswerten Abweichungen in den überschwelligen Funktionen gegenüber Normalhörenden auf, sofern der Darbietungspegel auf die individuelle Hörschwelle bezogen wird. Im folgenden sollen daher nur die bei sensorineuralen Schwerhörigkeiten auftretenden Veränderungen der Frequenz- und Zeitabbildung betrachtet werden:

Aufgrund der bei Schädigung der äußeren Haarzellen verringerten Frequenzselektivität des Gehörs nimmt auch die psychoakustisch meßbare Frequenzauflösung des Gehörs ab. Dies macht sich in einer **Verbreiterung der Frequenzgruppe** bemerkbar, d. h. bei Innenohrschwerhörenden wird ein größerer spektraler Bereich im Gehör zusammengefaßt, um die Lautheit und die Maskierung zu bestimmen (s. Abb. 14). Die Verbreiterung der Frequenzgruppe führt zu einem **Anstieg der Mithörschwelle** im Rauschen, d. h. ein in einem Rauschen versteckter Testton muß stärker im Pegel erhöht werden als bei Normalhörenden, um detektiert werden zu können. Dies liegt daran, daß mehr Energie von benachbarten Frequenzen zur Maskierung des Ziel-Testtons beiträgt. Allerdings hängt diese Verbreiterung vom Signalpegel ab: Auch bei Normalhörenden tritt eine Verbreiterung der Frequenzgruppe mit hohem Pegel auf. Daher sollte man die Frequenzgruppe von Schwerhörenden nicht bei gleichem Pegel über der individuellen Hörschwelle, sondern bei gleichem absoluten Pegel vergleichen. Dabei ist die Frequenzgruppe von Schwerhörenden nur geringgradig gegenüber der von Normalhörenden verbreitert. Bei einer entsprechenden Anhebung des Pegels über die Hörschwelle des Innenohrschwerhörigen (z. B. durch ein Hörgerät) wird daher der Effekt der verbreiterten Frequenzgruppe geringer.

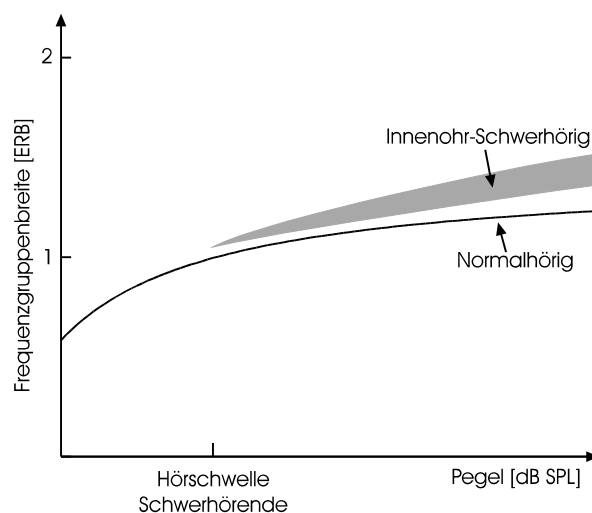


Abbildung 14: Frequenzgruppenbreite als Funktion des Pegels für Normal- und Innenohrschwerhörende

Die **Zeitauflösung** bei Innenohrschwerhörigen ist ebenfalls verringert, d. h. man mißt eine verlängerte Vor- und Nachverdeckung und eine Verlängerung der kleinsten detektierbaren Pausendauer im Rauschen. Allerdings ist hier genau wie bei der Frequenzauflösung die Pegelabhängigkeit zu berücksichtigen, da bei den Innenohrschwerhörenden ein wesentlich kleinerer Pegelbereich zur Verfügung steht als bei Normalhörenden. Wenn man die Experimente bei dem gleichen Pegel oberhalb der Ruhehörschwelle (bzw. bei der gleichen wahrgenommenen Lautheit) durchführt und auch den Effekt des Recruitments berücksichtigt, ist die Zeitauflösung nicht wesentlich gegenüber der von Normalhörenden gestört. Dies ist auch der Grund, wieso die psychoakustisch meßbare Modulations-Transferfunktion ungefähr den gleichen Tiefpaßverlauf wie bei Normalhörenden aufweist. Allerdings ist die Schwelle insgesamt

erhöht, so daß Schwerhörnde nicht so geringe Modulationstiefen detektieren können wie Normalhörende.

Im Gegensatz zu den relativ wenig geänderten Zeitkonstanten bei der Zeitauflösung steht das geänderte Verhalten bei der **zeitlichen Integration** bei Innenohrschwerhörigen (s.o.): Während bei Normalhörenden die Wahrnehmbarkeit und die Lautstärke eines Tones mit zunehmender Tondauer bis zu einer Maximaldauer von etwa 200 ms zunimmt, tritt dieser Effekt bei Innenohrschwerhörigen nicht in gleicher Weise auf. Dies ist ein Grund dafür, daß Signale, die im Rauschen versteckt sind (maskierte Signale), von Innenohrschwerhörigen erst bei wesentlich höheren Signal-Rauschverhältnissen detektiert werden können als von Normalhörenden.

-Weitere auditorische Funktionen

Generell zeigt sich bei Innenohrschwerhörigen eine allgemeine Verschlechterung sämtlicher psychoakustisch meßbarer auditorischer Funktionen im Vergleich zu Normalhörenden. Zu diesen eingeschränkten Funktionen gehören neben der oben erwähnten Intensitäts-, Frequenz- und Zeitauflösung auch andersartig gemessene Detektions- und Diskriminationsleistungen im Zeit- und Frequenzbereich (z. B. Tonhöhenunterscheidung), komplexere Erkennungs- und Unterscheidungsleistungen, sowie die binaurale Interaktion (d. h. die Extraktion von Unterschieden zwischen den Signalen, die an den beiden Ohren anliegen), die zur Ortung von Schallquellen und zur Trennung zwischen Nutzsignalen und Störsignalen beiträgt. Dabei stellt man fest, daß sämtliche dieser auditorischen Funktionen primär vom Hörverlust beeinflusst werden. Im Mittel über alle Patienten verschlechtern sich alle Funktionen mit zunehmendem Hörverlust. Abgesehen von diesem generellen Effekt gibt es jedoch nur sehr geringe gegenseitige Abhängigkeiten zwischen den Schädigungen in der Hörfunktion. Beispielsweise kann die Frequenzauflösung stark gestört sein und das binaurale Hören relativ intakt (oder umgekehrt). Dies führt zu der Vorstellung, daß sich Innenohr-Hörstörungen primär durch die **zwei Komponenten Hörverlust und Kompressionsverlust** charakterisieren lassen. Die darüber hinausgehenden Einschränkungen unterliegen einem diffusen, an unterschiedlichen Orten in verschiedener, nicht vorhersehbarer Weise wirkenden Schädigungsprozeß.

3. Sprachwahrnehmung bei Normal- und Schwerhörigen

Die sprachliche Kommunikation zwischen Menschen ist eine hochspezialisierte und faszinierende Leistung, die letztendlich die Grundlage unserer Kultur darstellt. Um die beim Verstehen von Sprache (und ihrer möglichen Störungen) beteiligten Prozesse verstehen zu können, muß der gesamte Übertragungsweg vom Sender (sprechender Mensch) über die akustische Darstellung von Sprache bis zum Empfänger (hörender Mensch) genau betrachtet werden.

3.1 Sprachakustik

Wenn das akustische Sprachsignal z. B. mit einem Mikrophon aufgenommen wird, läßt sich aus dem Zeitverlauf des Signals nur sehr wenig Information über das geäußerte Sprachelement ableiten. Beispielsweise kann gesehen werden, ob es sich um einen Vokal (periodisches Signal aufgrund der Stimmlippen-Schwingung) oder einen stimmlosen Konsonanten mit Rauschanteilen handelt (irreguläres Zeitsignal). Weiterhin lassen sich mit hoher Zeitauflösung zeitliche Übergänge feststellen (vgl. Abbildung 15). Eine Möglichkeit, mehr Informationen über die akustische Filterung des Sprachsignals im Vokaltrakt zu erhalten, bietet die Analyse des Frequenzgehalts des Schallsignal (**Leistungsspektrum**). Dabei wird für jede Frequenz über die gesamte Dauer des Signals gemittelt die Energie errechnet, die in das jeweilige Frequenzband fällt (vgl. Abbildung 15 rechts). Bei einem stationären Signal, das sich nicht zeitlich ändert (z. B. bei gehaltenen Vokalen) kann man in diesem (Leistungs-)Spektrum sehr gut die Grundfrequenz (Schwingungsfrequenz der Glottis) mit ihren Obertönen (ganzzahlige Vielfache der Grundfrequenz) erkennen, sowie die **Formanten**, d. h. die Frequenzbereiche, bei denen der Vokaltrakt eine besonders hohe Verstärkung des akustischen Signals bewirkt. Die Formanten stellen sich damit als Spitzen im Spektrum dar und sind charakteristisch für den jeweils artikulierten Vokal (s. unten).

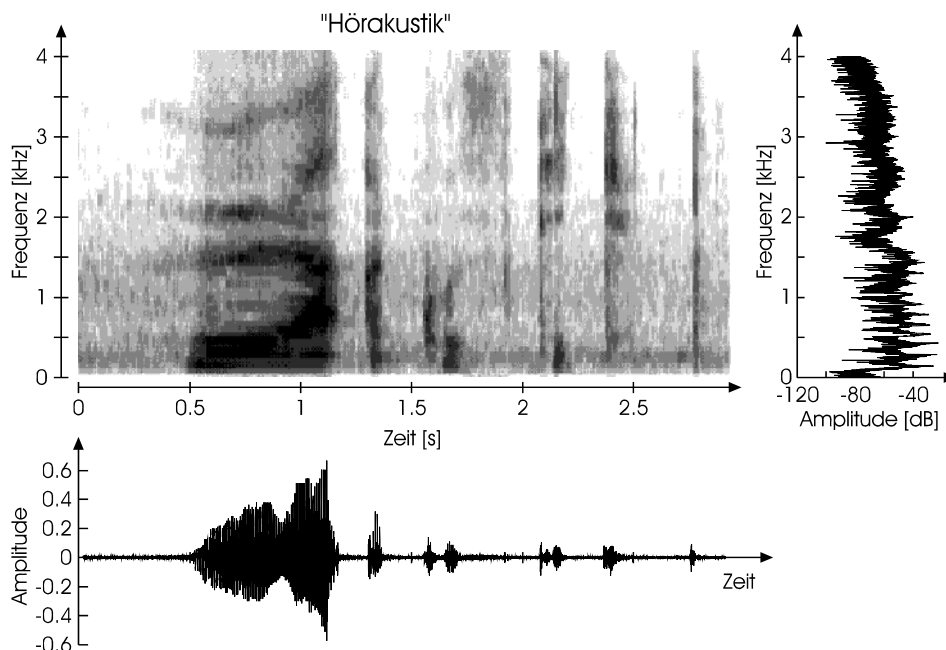


Abbildung 15: Zeit-Signal (unten), Spektrum (rechts) und Spektrogramm (links oben) des Wortes „Hörakustik“

Da das Spektrum über eine lange Zeit gebildet wird, ist es ungeeignet zur Beurteilung der sich zeitlich stark ändernden akustischen Anteile von fließender Sprache. Es hat allerdings eine große Bedeutung als **mittleres Sprachspektrum**, d. h. die Energieverteilung im Frequenzbereich, die im Mittel bei einem männlichen bzw. weiblichen Sprecher auftritt (vgl. Abbildung 16). Daraus läßt sich ablesen, daß im

Bereich der Grundfrequenzen zwischen etwa 300 und 800 Hz am meisten Sprachenergie vorliegt (Maximum des Spektrums), während zu niedrigen Frequenzen ein Abfall von etwa 20 dB pro Oktave erfolgt und zu hohen Frequenzen von etwa 10 dB pro Oktave. Die Langzeitspektren von männlichen und weiblichen bzw. kindlichen Sprechern unterscheiden sich durch die Lage des Maximums im Frequenzbereich, das bei Frauen und Kindern deutlich höher liegt als bei Männern. Mit zunehmender **Sprechanstrengung** verschiebt sich das spektrale Maximum zu hohen Frequenzen, so daß sich die Stimme aus dem Hintergrund (mit vorwiegend tieffrequenter Leistung) deutlicher abgrenzen läßt, ohne daß die akustische Gesamtleistung der Stimme in gleichem Maß ansteigt.

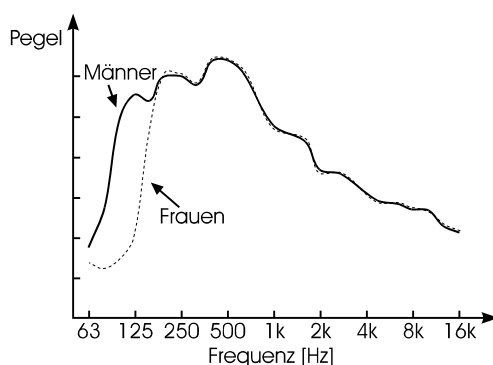


Abbildung 16: Mittleres Spektrum von männlichen bzw. weiblichen Sprechern

Aufgrund des Spektrums von gehaltenen Vokalen bzw. ausgeschnittenen Vokalen läßt sich ein Zusammenhang zwischen der Lage des ersten Formanten (d. h. spektrales Maximum im Bereich zwischen etwa 200 und 800 Hz), dem zweiten Formanten (zweites spektrales Maximum zwischen etwa 600 und 2.500 Hz), der Stellung der Artikulationsorgane und dem jeweils geäußerten Vokal aufstellen (**Vokaldreieck** bzw. Vokaltrapez, vgl. Abbildung 17).

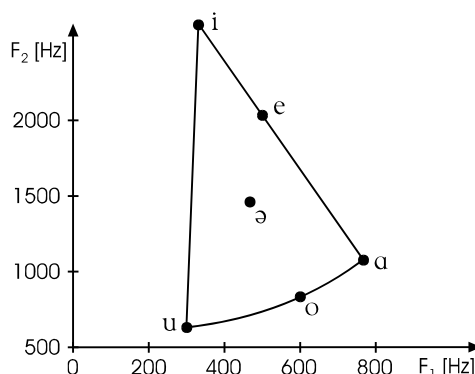


Abbildung 17: Vokaldreieck. Aufgetragen ist die Frequenz des ersten Formanten F_1 (Abzisse) und des zweiten Formanten F_2 (Ordinate) für unterschiedliche Vokale (schematisch)

3.2 Sprachverständlichkeit

Beim Verstehen von Sprache spielt einerseits die **Sprachverständlichkeit** eine Rolle, d. h. die Eigenschaft des Sprachmaterials, von einem durchschnittlichen, normalhörenden Probanden verstanden zu werden. Andererseits spielt die individuelle **Sprachperzeptionsleistung** eine Rolle, d. h. die von dem individuellen Patienten erbrachten Voraussetzungen zum Sprachverstehen, die mehr oder weniger stark gestört sein können. In der **Sprachaudiometrie** mißt man die Verständlichkeit von standardisiertem Sprachmaterial bei dem individuellen Patienten, um Rückschlüsse auf dessen Sprachperzeptionsleistungen führen zu können.

3.2.1 Methoden zur Bestimmung der Sprachverständlichkeit

Das für Verständlichkeitsmessungen verwendete Sprachmaterial sollte möglichst repräsentativ sein für die Sprache und für die zu betrachtende Kommunikationssituation. Dabei tritt als grundlegendes Problem die hohe Zahl von Variablen auf, für die eine möglichst vernünftige Festlegung erfolgen sollte. Zu diesen Variablen gehört die Art des Tests: „**offene**“ **Tests**, bei dem ein Test-Item (z. B. ein Wort oder Satz) dem Probanden dargeboten wird, das er möglichst korrekt wiederholen soll, und „**geschlossene**“ **Tests**, bei dem der Proband das richtige Test-Item aus einer Liste von möglichen Antworten bezeichnet. Weitere Variablen sind die Länge der Test-Items (z. B. einsilbige oder mehrsilbige Wörter oder Sätze), die Auswerte-Methode (Bewertung richtig erkannter Phoneme, Silben, Wörter oder Sätze), der Sprecher (männlicher oder weiblicher Sprecher, geschulter oder ungeschulter Sprecher oder etwa synthetische Sprache), sowie die Wahl eines Störgeräusches und einer räumlichen Anordnung von Nutzschall- und Störgeräuschquelle. Jedes der gebräuchlichen Sprachtestverfahren besitzt eine eigene Kombination dieser Variablen, so daß die Test-Ergebnisse sich nicht leicht miteinander vergleichen lassen.

Außerdem ist die Festlegung der Variablen abhängig vom Einsatzzweck des Tests. Beispielsweise steht bei der **Diagnostik** von Hörstörungen die analytische Fähigkeit des Tests im Vordergrund (d. h. die Möglichkeit, aus den auftretenden Phonem-Verwechslungen Rückschlüsse auf das gestörte Hörsystem zu ziehen), während bei der **Begutachtung** eher eine hohe Reproduzierbarkeit des Tests und eine hohe Repräsentanz des Sprachmaterials für alltägliche Kommunikation im Vordergrund steht. Bei der **Hörgeräteanpassung** ist dagegen eine hohe Sensitivität gegenüber kleinen Änderungen der Einstellparameter des Hörgerätes wichtig.

Der derzeit in der Standard-Audiometrie am häufigsten eingesetzte Sprachtest, der **Freiburger Wörkertest**, ist ein „offener“ Test mit Einsilbern bzw. mit Mehrsilbern (Zahlen). Als Alternative dazu wurde in jüngerer Zeit das **Reimtestverfahren** mit

geschlossenen Antwortalternativen eingeführt, die sich in nur einem Phonem unterscheiden und sich daher reimen (z. B. die Alternativen „Sinn“, „Hin“, „bin“, „Kinn“, „Zinn“). Dieses Verfahren vermeidet Fehlerquellen bei der Testbewertung durch den Testleiter und ist automatisierbar. Die Einsilber-Tests haben den Vorteil einer geringen Redundanz (d. h. aus einem richtig erkannten Teil des Wortes kann nicht auf den anderen, unverständlichen Teil des Wortes geschlossen werden) und bieten eine hohe analytische Aussagekraft bei den auftretenden Phonem-Verwechslungen. Da im Deutschen **Zweisilber** jedoch häufiger als Einsilber sind, sind Zweisilber-Tests eher repräsentativ für die deutsche Sprache, so daß auch ein Reimtest-Verfahren für Zweisilber entwickelt wurde. Sprachtests mit **Sätzen** bieten dagegen eine sehr realitätsnahe Kommunikations-Situation. Ihre Diskriminationsfunktion (d. h. der Prozentsatz richtig erkannter Wörter als Funktion des Sprachpegels) ist sehr steil, so daß sie eine hohe Sensitivität gegenüber Änderungen im Sprachpegel oder im Signal-Rauschabstand aufweisen. Für die deutsche Sprache gibt es den standardisierten Marburger Satztest und den neueren Göttinger Satztest. Ein Überblick über moderne Verfahren der deutschsprachigen Sprachaudiometrie findet sich bei Kollmeier (1992).

3.2.2 Methoden zur Berechnung und Modellierung der Sprachverständlichkeit

Um das Verstehen von Sprache quantitativ zu erfassen und die Sprachverständlichkeit für eine vorgegebene akustische Situation oder einen bestimmten Hörverlust vorhersagen zu können, wurden verschiedene Methoden zur Berechnung bzw. Vorhersage der Sprachverständlichkeit aus dem zugrundeliegenden Sprachmaterial und den akustischen Gegebenheiten (z. B. Nachhall, Störgeräusch, angeborene Ruhehörschwelle) entwickelt. Die klassische Methode der Sprachverständlichkeitsvorhersage ist der **Artikulations-Index (AI)** und der **Speech Transmision Index (STI)**. Sie beruhen auf der Annahme, daß die gesamte Sprach-Information auf die verschiedenen Frequenzbänder des akustischen Signals verteilt ist, und daß jedes Band einen gewissen Beitrag zur Gesamt-Sprachverständlichkeit liefert. Die Breite jedes dieser Bänder orientiert sich dabei an der Frequenz-Gruppenbreite (Bark-Skala oder in erster Näherung Terz-Bandbreite). In jedem dieser Frequenzbänder kann nun ein „effektiver“ Signal-Rauschabstand ermittelt werden, d. h. das Verhältnis zwischen der Nutzenergie des zu verstehenden Sprachsignals und den Störanteilen. Diese werden durch Nachhall oder durch ein Störgeräusch verursacht, oder dadurch, daß die Energie in diesem Band unterhalb der Ruhehörschwelle des jeweiligen Patienten liegt, was durch ein angenommenes „internes Rauschen“ nachgebildet wird. Wenn der Signal-Rauschabstand in dem Band größer als 15 dB ist, geht man davon aus, daß dieses Band vollständig zur Verständlichkeit beiträgt, während bei einem Signal-Rauschabstand von kleiner als -15 dB sämtliche Sprachinformation in diesem Band maskiert ist, so daß das Band nicht zur Gesamt-Sprachverständlichkeit beiträgt. Der Signal-Rauschabstand im Band j ($SNR(j)$) wird also auf einen Bereich von -15dB bis +15dB begrenzt. Zur Ermittlung des Artikulationsindex wird nun eine gewichtete Mittelung über die Signal-Rauschabstände in den einzelnen Bändern durchgeführt, bei der jedes Band j mit einem Gewichtungsfaktor W_j multipliziert wird:

$$\text{AI oder STI} = \sum_j W_j \cdot [\text{SNR}(j) + 15\text{dB}] \quad (1)$$

Die Gewichtungsfaktoren W_j sind dabei so gewählt, daß der STI oder AI nur Werte zwischen 0 (entsprechend einer Sprachverständlichkeit von 0, wenn in jedem Band nur Rauschen und keine signifikanten Sprachanteile vorliegen) und 1 variiert (vollständige Sprachverständlichkeit, da in jedem Frequenzband die Sprache von dem Rauschen kaum beeinflußt wird). Damit stellt der AI bzw. STI ein Maß für die Sprachverständlichkeit dar, das direkt mit der meßbaren mittleren Sprachverständlichkeit zusammenhängt. Diese Beziehung hängt allerdings von dem verwendeten Sprachmaterial und Test ab (vgl. Abbildung 18). Bei Sprachtests mit geringer Redundanz (z. B. bei sinnlosen oder sinnbehafteten Einsilbern) wächst die Sprachverständlichkeit mit zunehmendem AI bzw. STI nur langsam an, während die Sprachverständlichkeit bei Sätzen mit hoher Redundanz schon bei relativ kleinem AI sehr hoch ist. Dies liegt an der Möglichkeit, den Satz aufgrund des Satzzusammenhanges schon dann vollständig richtig zu verstehen, wenn nur einige Teile des Satzes verstanden werden.

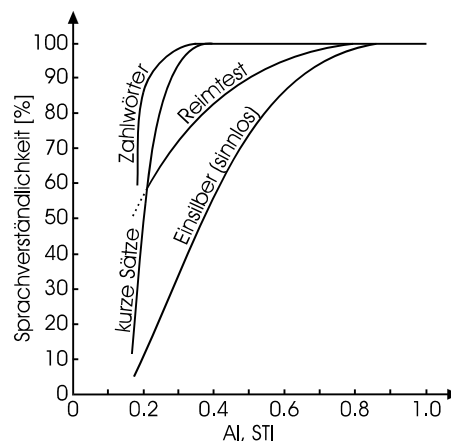


Abbildung 18: Sprachverständlichkeit für verschiedenes Sprachmaterial als Funktion des AI bzw. STI

Mit Hilfe der in Abbildung 18 angegebenen Kurven ist es daher prinzipiell möglich, die Verständlichkeit für verschiedene Sprachmaterialien bei vorgegebenen akustischen Bedingungen ineinander umzurechnen. Es muß jedoch betont werden, daß der Artikulations Index und der Speech Transmission Index nur für den Mittelwert von Sprachverständlichkeiten über ein normalhörendes Versuchspersonenkollektiv und eine große Anzahl von Tests gelten. Für den individuellen Patienten und für die Verständlichkeit eines einzelnen Wortes oder eines einzelnen Satzes treten große Schwankungen in der Verständlichkeit auf, so daß diese Berechnungsmethode für diesen Fall nicht angewendet werden kann (vgl. Kollmeier, 1990).

Der STI unterscheidet sich vom AI durch die Vorgehensweise zur Berechnung des „effektiven“ Signal-Rauschabstandes SNR (j) in jedem Band j. Während beim AI das Leistungsspektrum des

Sprachmaterials und das Leistungsspektrum des Störsignals getrennt voneinander bekannt sein müssen und ins Verhältnis gesetzt werden, wird beim STI der Signal-Rauschabstand anhand des Sprachsignals und der Mischung aus Sprachsignal und Störsignal mit Hilfe der **Modulations Transfer Funktion (MTF)** bestimmt. Dazu werden in jedem Frequenzbereich die im Original-Signal auftretenden Modulationen gemessen und die in der Mischung verbleibenden Modulationen werden ins Verhältnis zu den ursprünglichen Modulationen gesetzt. Wenn die ursprünglichen Modulationen vollständig erhalten sind (Modulations-Transfer-Funktion von 1), spricht dies für ein sehr hohes Signal-Rauschverhältnis, während die Anwesenheit von Störgeräusch oder von Nachhall die im gemischten Signal verbleibenden Modulationen stark verringert. Das Konzept des STI geht also davon aus, daß die Erhaltung der Modulationen im gemischten Signal für die Sprachverständlichkeit von entscheidender Bedeutung ist und errechnet den Signal-Rauschabstand aus diesem Erhalt der Modulationen. Dadurch können der Einfluß von Nachhall und der Einfluß von Störgeräusch in gleicher Weise behandelt werden, so daß sich der STI besonders für den Einsatz in der Raumakustik bewährt hat. Die zeitlichen Eigenschaften eines Sprachübertragungssystems (z. B. des gestörten Gehörs) werden ebenfalls besser berücksichtigt als beim AI (vgl. Houtgast und Steneken, 1973).

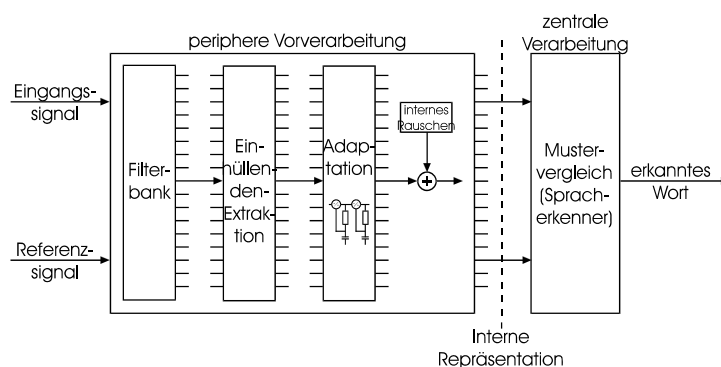


Abbildung 19: Perzeptionsmodell zum Sprachverstehen nach Holube und Kollmeier (1996)

Neben den auf den Eigenschaften des Sprachmaterials beruhenden Berechnungsverfahren des AI und STI gibt es Modelle zur Sprachverständlichkeitsvorhersage, die sich eher an den Eigenschaften des Gehörs orientieren und die Verarbeitungsschritte bei der Sprachwahrnehmung explizit nachbilden. Als Beispiel für ein derartiges **Perzeptionsmodell** soll das in Abbildung 19 skizzierte Modell nach Holube und Kollmeier (1996) erläutert werden: Das gestörte Sprachsignal wird zunächst einer peripheren Vorverarbeitungs-Stufe zugeführt, die die „effektive“ Signalverarbeitung im auditorischen System möglichst gut nachbildet. Die Elemente dieser Vorverarbeitungsstufe entsprechen dabei genau den anhand von psychoakustischen und physiologischen Experimenten gewonnenen Verarbeitungseinheiten. Zunächst wird eine Aufspaltung des Signals in verschiedene Frequenzbereiche vorgenommen (entspricht der Frequenz-Orts-Transformation der Basilarmembran), anschließend wird eine Extraktion der Einhüllenden und eine Adaptation mit zeitlicher Kontrastierung in jedem Frequenzkanal vorgenommen (dies entspricht ungefähr der Aktion der Haarzellen und des Hörnervs). Die statistischen Übertragungsfehler durch das Nervensystem und die Begrenzung der Empfindlichkeit durch die Ruhehörschwelle wird durch ein zusätzliches „internes“ Rauschen

nachgebildet. Am Ausgang der peripheren Vorverarbeitungsstufe liegt dann zu jedem Zeitpunkt für jede Frequenz ein komprimierter und zeitlich kontrastierter Intensitätswert vor, der ungefähr der **internen Repräsentation** der akustischen Signale entspricht und eine ähnliche Darstellung der Sprachinformation wie das Spektrogramm ermöglicht.

Dieselben Vorverarbeitungsschritte werden auch für ein Referenzsignal (ungestörtes Sprachsignal) durchgeführt. In einer zentralen Verarbeitungsstufe (Mustererkenner) kann dann überprüft werden, inwiefern die interne Repräsentation des Eingangssignals mit derjenigen eines vorgegebenen Vergleichsignals übereinstimmt. Diese zentrale Mustererkennung entspricht ungefähr dem Prozeß der Sprachwahrnehmung, bei dem die aktuelle akustische Wahrnehmung mit der möglichen Wahrnehmung von allen möglichen bekannten Wörtern verglichen wird. Als „erkanntes Wort“ wird dasjenige ausgewählt, dessen interne Repräsentation am ähnlichsten mit der Repräsentation des Eingangssignals ist. Dieser zentrale Verarbeitungs-Vorgang kann daher mit einem Spracherkennungsalgorithmus durchgeführt werden und bildet das „Weltwissen“ des Hörers nach.

Mit einer derartigen Modellstruktur kann untersucht werden, inwiefern sich Verarbeitungsfehler in der peripheren Vorverarbeitung (z. B. Änderung der spektralen und zeitlichen Auflösung oder eine Anhebung der Ruhehörschwelle) auf die veränderte Erkennungsleistung beim Sprachverstehen auswirkt. Der Vorteil eines derartigen Modells besteht darin, daß die Sprachverständlichkeitsvorhersage nicht global, im Mittel über viele Versuchspersonen und viele Wörter erfolgen muß, sondern daß für jede individuelle Versuchsperson und jedes zu erkennende Wort eine eigene Verständlichkeitsberechnung durchgeführt werden kann. Die Vorhersagen mit diesem Modell stimmen für die Sprachverständlichkeit in Ruhe und insbesondere für verschiedene Störgeräusche sehr gut mit den gemessenen Verständlichkeitswerten bei Normal- und Schwerhörenden überein (vgl. Holube und Kollmeier, 1996). Aufgrund des hohen Rechenzeitaufwandes für dieses Modell und die noch nicht in allen Einzelheiten geklärten Verarbeitungsprozesse sind jedoch noch weitere Forschungsarbeiten zu seiner Validierung notwendig.

3.3 Sprachperzeption bei pathologischem Gehör

Bei Schwerhörigkeit ist das Verstehen von Sprache in Ruhe und unter Störgeräuscheinfluß reduziert. In der **Sprachaudiometrie** wird diese Reduktion quantitativ erfaßt, in dem die Sprachverständlichkeit (d. h. der Prozentsatz korrekt verstandener Test-Items einer vorgegebenen Testliste von Wörtern, Zahlen oder Sätzen) für verschiedene Sprachpegel und für verschiedene Testbedingungen bestimmt wird (z. B. in Ruhe oder unter Störgeräusch mit unterschiedlicher räumlicher Anordnung von Störschall und Nutzschall). Das Ziel dieser Messungen ist entweder die differenzierte **Diagnostik** von Hörstörungen, die **Begutachtung** eines Hörschadens oder die **Anpassung von Hörhilfen** (d. h. das Ausmessen der Sprachverständlichkeit

ohne Hörhilfe und den Gewinn an Sprachverständlichkeit, der durch die Hörhilfe erzielt wird). Für die Bestimmung der Sprachverständlichkeit gibt es mehrere Testverfahren (vgl. 1.2) von denen der Freiburger Einsilber- bzw. Zahlentest und der Marburger Satztest standardisiert sind. Neuere und weniger fehleranfällige Verfahren wie der Einsilber- oder Zweisilber-Reimtest und der Göttinger Satztest befinden sich im Erprobungsstadium.

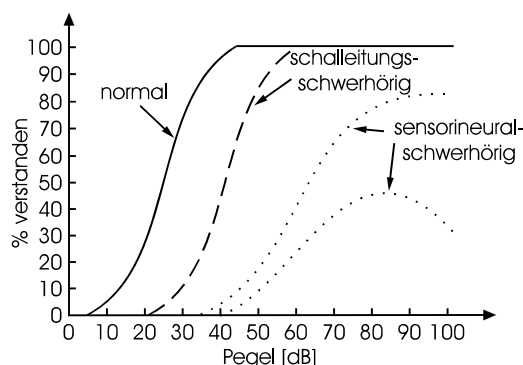


Abbildung 20: Diskriminationsfunktionen bei unterschiedlichen Pathologien (schematisch)

Für den **Sprachtest in Ruhe** findet man bei den verschiedenen Schwerhörigkeitsformen prinzipiell die in Abbildung 20 dargestellten **Diskriminationsfunktionen** (d. h. Prozent Sprachverständlichkeit als Funktion des Sprachpegels). Bei Normalhörenden hängt die Steigung der Diskriminationsfunktion vom verwendeten Sprachmaterial ab: Die Diskriminationsfunktion ist relativ flach bei Einsilbern mit offenem Antwortformat, etwas steiler beim Reimtestverfahren und extrem steil (ca. 20 % Verständlichkeit pro dB) bei Sätzen wie z. B. beim Göttinger Satztest. Bei Schalleitungs-Schwerhörigkeiten ist die Diskriminationsfunktion zu hohen Pegeln hin verschoben. Bei Schallempfindungs-Schwerhörigkeiten ist die Diskriminationsfunktion ebenfalls verschoben, was auf die „Abschwächungswirkung“ des Hörverlusts zurückgeführt wird. Die darüber hinausgehende „Verzerrungswirkung“ des Hörverlusts („**Fehlhörigkeit**“) macht sich dagegen in einer **Abflachung** (Verringerung der Steigung) der Diskriminationsfunktion und in einem **Diskriminationsverlust** bemerkbar, d. h. die Diskriminationskurve erreicht bei hohen Pegeln nicht 100 % sondern bleibt um einen bestimmten Wert, der als Diskriminationsverlust bezeichnet wird, unterhalb von 100 %. Bei sehr hohen Pegeln kann bei einem Patienten sogar eine Verschlechterung der Sprachverständlichkeit mit zunehmendem Pegel beobachtet werden („roll-over“-Phänomen oder R-Kurve), das im Zeitalter verbesserter audiologischer Technik (verzerrungsfreie Verstärker und Kopfhörer, CD-Wiedergabe des Sprachmaterials) allerdings selten geworden ist. Die beiden erstgenannten Effekte lassen sich bei Innenohrschwerhörigen darauf zurückführen, daß sie bestimmte Wörter erst bei sehr hohen Pegeln oder überhaupt nicht verstehen können, während andere Wörter schon bei relativ niedrigen Pegeln verständlich sind. Der Übergangsbereich der Diskriminationsfunktion zwischen „unverständlich“ und „vollständig verständlich“ überdeckt daher einen größeren Pegelbereich mehr als bei Normalhörenden bzw. Schalleitungsschwerhörigen.

Ziel der Sprachaudiometrie in Ruhe ist es nun, die wesentlichen Eigenschaften der Diskriminationsfunktion des individuellen Patienten zu erfassen, d. h. die **Verständlichkeitsschwelle** (der zu 50 % Sprachverständlichkeit gehörende Sprachpegel), die **Steigung** der Diskriminationsfunktion und den **Diskriminationsverlust** bzw. das Vorliegen einer R-Kurve. In der Routine-Audiometrie wird daher für die einsilbigen Wörter des Freiburger Sprachtest die Sprachverständlichkeit bei 65 dB bestimmt und bei um jeweils 15 dB erhöhtem Sprachpegel bis entweder die Erkennungsrate 100 % erreicht wird oder die Unbehaglichkeits-Schwelle überschritten wird. Daraus ergibt sich die **Diskriminationsfähigkeit bei 65 dB** (in %), der Pegel bei dem die höchste Erkennungsrate erzielt wird (**dB Opt.**), die Differenz zu 100 % bei diesem Pegel (**Diskriminationsverlust** in Prozent) und die Gesamtwortverständlichkeit (**GWV** in %), die sich aus der Summe der Erkennungsraten in Prozent bei 60, 80 und 100 dB ergibt. Bei den mehrsilbigen Zahlwörtern des Freiburger Tests wird die Erkennungsrate für einen oder zwei verschiedene Pegel bestimmt. Unter Berücksichtigung der Steigung der Diskriminations-Normkurve für Normalhörende wird anschließend durch Interpolation derjenige Pegel bestimmt, bei dem 50 % der Zahlen verstanden werden. Die Differenz zu dem Wert für Normalhörende wird als **Hörverlust (HV) für Zahlwörter** in dB angegeben. Neuere sprachaudiometrische Testverfahren ermitteln die Sprachverständlichkeitsschwelle, den Diskriminationsverlust und die Steigung der Diskriminationsfunktion durch **adaptive Messungen**, bei denen der Sprachpegel in Abhängigkeit von den Antworten des Probanden variiert wird. Dadurch wird mit wenigen Versuchsschritten der Pegel maximalen Sprachverstehens ermittelt und anschließend die Sprachverständlichkeitsschwelle. Durch das Anpassen einer Diskriminationsfunktion an die Antworten des Probanden läßt sich anschließend die Steigung der Diskriminationsfunktion angeben. Diese Verfahren werden vom Computer gesteuert bzw. ausgewertet, so daß sie bei zunehmender Verbreitung der computergesteuerten Audiometrie zunehmende Bedeutung für die Praxis gewinnen werden (vgl. Kollmeier, 1996).

Bei der **Sprachaudiometrie unter Störgeräusch** interessiert vorwiegend die Sprachverständlichkeits-Schwelle, d. h. derjenige Signal-Rauschabstand (Sprachpegel in Relation zum Störgeräuschpegel), bei dem eine Sprachverständlichkeit von 50 % erzielt werden kann. Dieser Wert ist bei sensorineural Schwerhörigen in der Regel deutlich erhöht gegenüber Normalhörenden und Schalleitungs-Schwerhörigen. Dies ist ebenfalls Ausdruck der „Verzerrungswirkung“ des Hörverlusts und entspricht den Beschwerden der Schwerhörigen, daß sie unter Umgebungsgeräusch mehr Schwierigkeiten haben, einer Unterhaltung zu folgen, als in Ruhe („Cocktail-Party-Effekt“ oder **Gesellschaftsschwerhörigkeit**). Da bei zusammenhängenden Sprachmaterialien wie einer Unterhaltung oder Sätzen eine Zunahme der Sprachverständlichkeit um etwa 20 % pro Verbesserung des Signal-Rauschabstands um 1 dB erfolgt, ist eine hohe Meßgenauigkeit zur Erfassung dieser Komponente des Hörverlusts notwendig. Einen besonders großen Unterschied in der Sprachverständlichkeitsschwelle zwischen Normal- und Innenohrschwerhörigen beobachtet man bei der Verwendung von

fluktuierenden Störgeräuschen, welche zeitliche Einhüllenden-Schwankungen in ähnlicher Weise aufweisen, wie ein einzelner, störender Sprecher (Fastl, 1987). Während Normalhörende in den „Lücken“ des Störgeräuschs noch einen Teil der Sprache gut verstehen können und daher durch ein derartig fluktuierendes Störgeräusch relativ wenig gestört sind, ist bei Innenohrschwerhörigen diese Fähigkeit stark gestört.

Ein weiteres Handicap beim Sprachverstehen im Störgeräusch haben Innenohrschwerhörige gegenüber Normalhörenden in **räumlichen Nutzschall-Störschallsituationen**. Während Normalhörende sowohl durch monaurale Verarbeitung als auch durch binaurale Signalverarbeitung (d. h. durch den Vergleich zwischen den an beiden Ohren jeweils anliegenden Signalen) eine gewisse Störgeräuschunterdrückung durchführen können und sich auf den Nutzsprecher konzentrieren können, ist dieser Effekt bei Schwerhörigen in sehr unterschiedlichem Maße gestört.

Literatur.

- Blauert, J. (1983): Spatial Hearing: The psychophysics of human sound localization, (Cambridge, Massachusetts, MIT Press).
- Böhme, G., Welzl-Müller, K.: Audiometrie - Hörprüfungen im Erwachsenen- und Kindesalter. Huber, Berlin 1993.
- Dau, T., Kollmeier, B., Kohlrausch, A. (1996). Modeling modulation perception: modulation low-pass filter or modulation filterbank. In: B. Kollmeier (Ed.), Psychoacoustics, Speech and Hearing Aids. World Scientific, Singapur, pp. 45-48.
- Durlach, N.I. (1972). Binaural signal detection: Equalization and cancellation theory. In: J.V. Tobias (Ed.) Foundations of Modern Auditory Theory, Academic Press, New York, Vol. 2, 363-462.
- Fastl, H. (1987). Ein Störgeräusch für die Sprachaudiometrie. Audiologische Akustik 26, 2-13.
- Festen, J.M. (1996): Temporal resolution and the importance of temporal envelope cues for speech perception. In: B. Kollmeier (Ed.), Psychoacoustics, Speech and Hearing Aids. World Scientific, Singapur, pp.95-102.
- Gelfand, S.A.: Hearing, an Introduction. M. Decker, 1990.
- Hellbrück, J.: Hören - Physiologie, Psychologie und Pathologie. Hogrefe Verlag, Göttingen 1993.
- Heller, O. (1985): Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). Psychologische Beiträge 27, 478-493.
- Helmholtz, H.L.F.: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik, Vieweg, Braunschweig, 1863.
- Hohmann, V. (1993). Dynamikkompression für Hörgeräte - Psychoakustische Grundlagen und Algorithmen. Fortschritt-Berichte VDI, Reihe 17 Nr. 93, Düsseldorf. Zugl.: Dissertation, Universität Göttingen.
- Hohmann, V., Kollmeier, B. (1995). Weiterentwicklung und klinischer Einsatz der Hörfeldskalierung. Audiol. Akustik 34 (2), 48-59.
- Holube, I., Kollmeier, B. (1996). Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. J. Acoust. Soc. Am. (im Druck)
- Houtgast, T., Steeneken, H.J.M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 28, 66-73.
- Jakobsen, R., Halle, M., Fant, G. (1951). Preliminaries to Speech Analysis: The distinctive Features and their Correlate. MIT Press, Cambridge, MA.
- Jeffress, L.A. (1948): A place theory of sound localization. J. Comp. Physiol. Psych. 61, 468-486.
- Kießling, J. (1995). Zum überschwelligem Lautheitsanstieg bei Schallempfindungsschwerhörigen - Konsequenzen für die Hörgeräte-Entwicklung und - Anpassung. Audiol. Akustik 34 (2), 82-89.
- Kießling, J., Kollmeier, B., Diller, G. Versorgung und Rehabilitation mit Hörgeräten. Thieme-Verlag
- Launer, S., Hohmann, V., Kollmeier, B. (1996). Modeling loudness growth and loudness summation in hearing-impaired listeners. In: W.
- Jestaedt (Ed.), Modeling Sensorineural Hearing Loss. Lawrence Erlbaum & Assoc., Hillsdale, NJ (im Druck).

- Kollmeier, B. (1990): Meßmethodik, Modellierung und Verbesserung der Verständlichkeit von Sprache. Habilitationsschrift, Universität Göttingen.
- Kollmeier, B. (Ed.): Moderne Verfahren der Sprachaudiometrie. Median-Verlag, Heidelberg, 1992
- Kollmeier, B. (1996): Computer-controlled speech audiometric techniques for the assessment of hearing loss and the evaluation of hearing aids. In: B. Kollmeier (Ed.), Psychoacoustics, Speech and Hearing Aids. World Scientific, Singapur, pp.57-68.
- Kollmeier, B.(2001)Die Oldenburger Hörforschung. Einblicke - Forschungsmagazin der Carl von Ossietzky Universität Oldenburg, 33, p. 3-6.
- Kollmeier, B.: Skript zur Vorlesung "Physiologische, psychologische und audiologische Akustik, Universität Oldenburg http://www.physik.uni-oldenburg.de/Docs/medi/html/teaching_download.html.
- Kollmeier, B. (2002) Cocktail-Partys und Hörgeräte: Biophysik des Gehörs. Physik Journal, 4, p. 39-45
- Moore, B.C.J.: An Introduction to the Psychology of Hearing, Academic Press, New York, 1989.
- Moore, B.C.J.: Perceptual Consequences of Cochlear Damage. Oxford University Press, Oxford 1995.
- Moore, B.C.J., Glasberg, B.R. (1987). Formulae describing frequency selectivity in the perception of loudness and level and their use in calculating excitation patterns. Hear. Res. 28, 209-225.
- Pickles, J.A.: Hearing, an Introduction to the Physiology of Hearing, Academic Press 1982.
- Poulton, E.C. (1989). Bias in Quantifying Judgements. Lawrence Erlbaum Assoc., Hillsdale, N.J.
- Rayleigh, Lord. Theory of Sound. London, Macmillan & Co. Ltd., 1877, Nachdruck 1929.
- Stevens, S.S. (1957). On the psychophysical law. Psychol. Rev. 64, 153 - 181.
- Zwicker, E., Fastl, H. Psychoacoustics - Facts and Models. Springer Verlag, Berlin 1990.

