

VI. Sprachwahrnehmung und Sprachaudiometrie

VI.1 Grundlagen der Sprachwahrnehmung

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.

VI.1.1 Sprachproduktion

Im Gehirn des Sprechers werden die Sprachlaute als Steuersequenzen erzeugt, die die akustische Sprachproduktion steuern. Durch den Atem-Luftstrom wird dabei eine periodische Schwingung der Stimmlippen angeregt oder (bei Reibelauten) ein Strömungsgeräusch an der engsten Stelle des Vokaltraktes. Der dabei jeweils erzeugte Schall wird anschließend je nach Stellung der Artikulationsorgane (Zunge, Gaumen, Kiefer, Lippen) im Vokaltrakt verformt („gefiltert“). Man spricht daher von einer „akustischen Filterung“ des von der Schallquelle abgestrahlten Schalls durch die sich zeitlich ändernden Hohlräume des Vokaltraktes (vgl. Abb. 6.1). Der auf diese Weise geformte Schall wird vom Mund in die umgebende Luft abgestrahlt und breitet sich bis zum Ohr des Empfängers fort. Der Empfänger versucht nun, aus dem empfangenem akustischen Signal auf die ursprünglich gesendete Botschaft zu schließen. Dabei benutzt er seine Kenntnisse über die Filterwirkung des Vokaltraktes auf das jeweilige akustische Signal, so daß der Empfänger aus dem akustischen Signal auf die zugrundeliegende Stellung des Vokaltraktes und damit auf die vom Sender eigentlich gemeinten Sprachelemente zurückschließen kann.

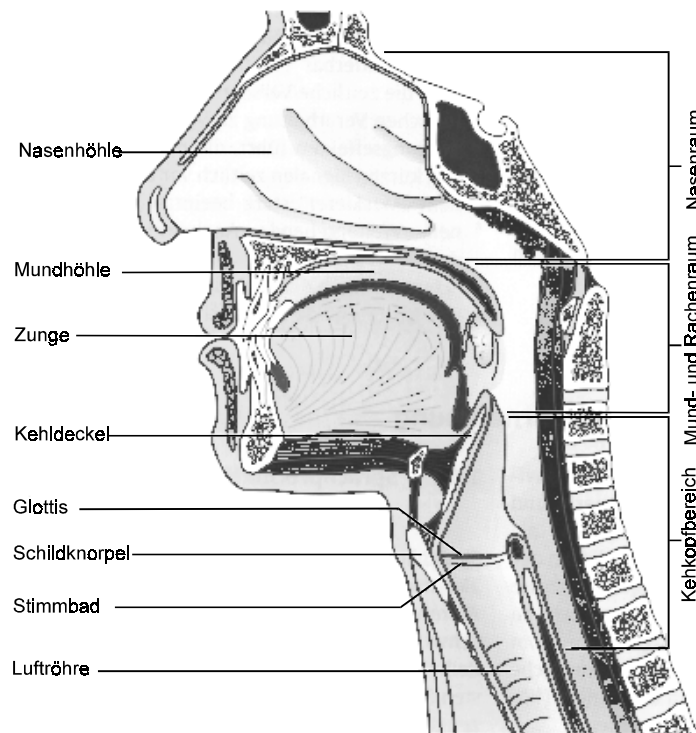


Abbildung 6.1: Vokaltrakt im Querschnitt (schematisch) (aus Kießling, J., Kollmeier, B., Diller, G.: Versorgung und Rehabilitation mit Hörgeräten. Thieme Verlag, Stuttgart 1997)

Die Beziehung zwischen den abstrakten, idealisierten sprachlichen Elementen (den **Phonemen**) und den zugehörigen akustischen Realisierungen (den **Lauten**) wird von der **Akustischen Phonetik** untersucht. Ein Phonem ist dabei die kleinste bedeutungsunterscheidende lautliche Einheit. Ein Beispiel dafür wäre die Umsetzung des Wortes „Sinn“ in [z], [I], [n]. Bei der akustischen Realisierung der Phoneme als Laute tritt eine große Variationsbreite auf, da z. B. der Dialekt des Sprechers oder der Kontext einen Einfluß auf die Art der akustischen Realisierung hat. Selbst bei demselben Sprecher und denselben Randbedingungen können unterschiedliche lautliche Realisierungen eines bestimmten Phonems auftreten. Zu einem (abstrakten) Phonem gehört also eine Vielzahl von akustisch realisierten Lauten. Diese Laute können aufgrund von akustischen Eigenschaften unterschieden werden (s. 6.2) oder aufgrund von artikulatorischen Merkmalen, d. h. Spracheigenschaften, die aufgrund der Produktion der jeweiligen Sprachelemente festgelegt werden können. Das Auffinden von derartigen Sprachmerkmalen und die Klassifikation eines jeden Phonems mit diesen Sprachmerkmalen ist insbesondere für die Beurteilung von **Phonem-Verwechslungen** wichtig. Systematische Verwechslungen bei der Sprachwahrnehmung treten beispielsweise aufgrund eines fehlerhaften Empfangs bestimmter

Sprachmerkmale auf.

Voraussetzung für die systematische Untersuchung derartiger Phonemverwechslungen ist ein geeignetes, die Sprache möglichst gut beschreibendes System von Sprachmerkmalen. Das Auffinden eines geeigneten Merkmalssystems erweist sich als äußerst schwierig. In der Literatur werden unterschiedliche Systeme verwendet, die vorwiegend auf der englischen Sprache basieren. Grundlagen für die am häufigsten verwendeten Typen von Merkmalssystemen ist das von Jakobsen et al. (1951) entwickelte „Distinctive-Feature-System“ (distinktive Sprachmerkmale für die englische Sprache). Dabei werden die Merkmale aus den Stellungen des Mund-Rachen-Raumes bei der Artikulation der jeweiligen Phoneme abgeleitet. Einem Merkmal entspricht dabei beispielsweise das Vorhandensein bzw. die Abwesenheit einer Anregung der Stimmbänder oder das Ausmaß der Deformation des Vokaltraktes aus seiner Ruhelage. Da es für jedes Phonem eine verhältnismäßig genau angebbare Kombination von Hals-, Zungen- und Lippenstellungen gibt und sich alle artikulierbaren Laute aus diesen Stellungen und der Atmung ableiten lassen, garantiert diese Definition die Eindeutigkeit und die Vollständigkeit der Darstellung. Eine weitere Eigenschaft der distinktiven Merkmale ist ihre Zweiwertigkeit (Binarität), d. h. bei einem Phonem ist ein Merkmal entweder vorhanden oder nicht, es gibt keine Abstufungen der Werte des Merkmals.

Die distinktiven Merkmale für die Konsonanten der deutschen Sprache sind in Tabelle 6.1 aufgelistet. Neben artikulatorischen Merkmalen, deren Bedeutung sofort ersichtlich ist (z. B. stimmhaft, nasal, scharf, vokalisch, konsonantisch) ergibt sich die Bedeutung anderer Merkmale (z. B. kompakt, dunkel, abrupt, gespannt) erst aus ihrer Zuordnung zu den jeweiligen Konsonanten. Man muß bei der Anwendung dieses Systems von Sprachmerkmalen jedoch beachten, daß es sich um rein artikulatorische Merkmale handelt, aus deren systematischer Verwechslung nur sehr schwer Rückschlüsse auf die akustische Grundlage dieser Verwechslung gezogen werden können. Da für diagnostische Zwecke (z. B. bei der Überprüfung von Hörgeräten oder Cochlea-Implantaten) jedoch die akustischen Übertragungseigenschaften und ihrer Optimierung im Vordergrund stehen, ist die Anwendbarkeit dieser phonetisch-artikulatorischen Analyseverfahren sehr begrenzt.

Tabelle 6.1: Distinktive Sprachmerkmale für die Konsonanten der deutschen Sprache

	voka- lisch	kom- pakt		nasal	ge- spannt		scharf		
		konso- nan- tisch	dunkel		abrupt	stimm- haft			
b (Bad)	-	+	-	+	-	+	-	+	-
d (Du)	-	+	-	-	-	+	-	+	-
f (Fee)	-	+	-	+	-	-	+	-	+
g (Gut)	-	+	+	+	-	+	-	+	-
h (Haar)	-	-	+	+	-	-	+	-	-
k (Kai)	-	+	+	+	-	+	+	-	-
l (Lag)	+	+	-	-	-	-	-	+	-
m (Mal)	-	+	-	+	+	+	-	+	-
n (Nun)	-	+	-	-	+	+	-	+	-
p (Pein)	-	+	-	+	-	+	+	-	-
r (Raus)	+	+	-	-	-	+	-	+	-
s (daS)	-	+	-	-	-	-	+	-	+
ʃ (SCHeu)	-	+	+	-	-	-	+	-	+
t (Tal)	-	+	-	-	-	+	+	-	-
v (Vase)	-	+	-	+	-	-	-	+	-
x (daCH)	-	+	+	+	-	-	-	-	-
z (Sinn)	-	+	-	-	-	-	-	+	+
j (Jod)	-	+	+	-	-	-	-	+	-

VI.1.2 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 6.2). 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 6.2 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).

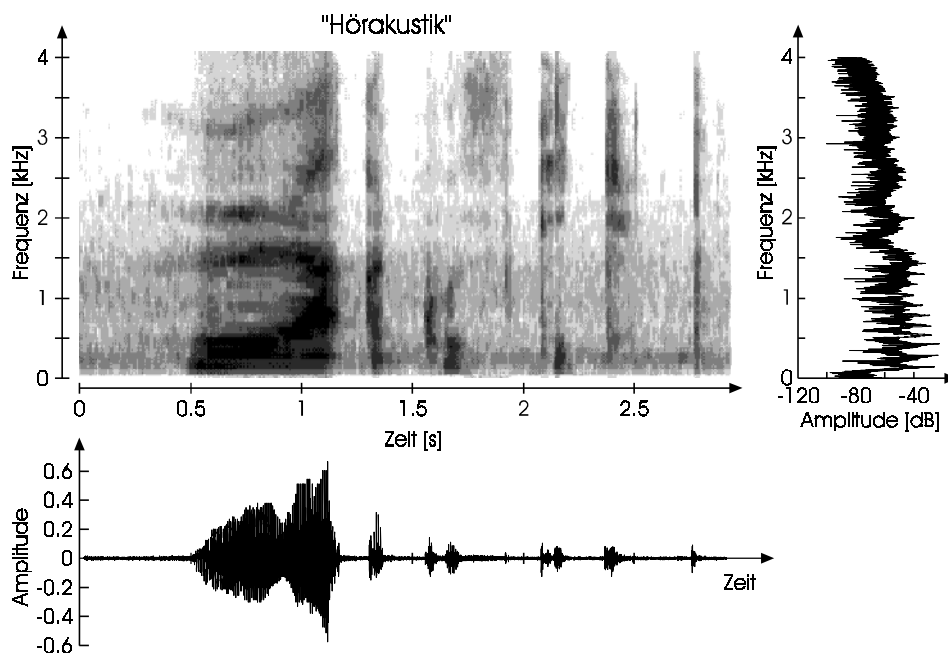


Abbildung 6.2: Zeit-Signal (unten), Spektrum (rechts) und Spektrogramm (links oben) des Wortes „Hörakustik“

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 6.3). 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.

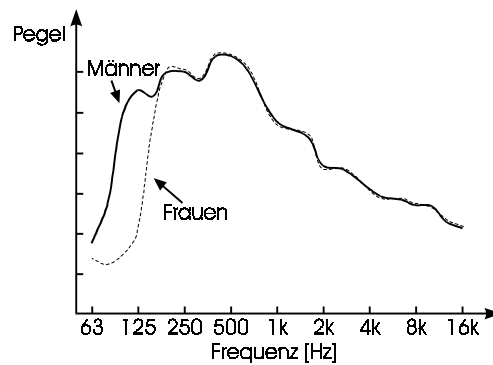


Abbildung 6.3: Mittleres Spektrum von männlichen bzw. weiblichen Sprechern

Aufgrund des Spektrums von gehaltenen Vokalen bzw. ausgeschnittenen Vokalen läßt sich ein Zusammenhang zwischen der Lage des ersten Formantens (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 6.4).

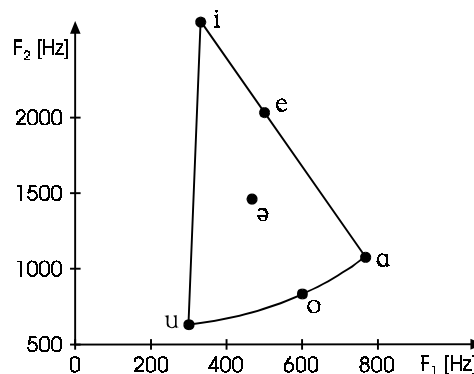


Abbildung 6.1.4: Vokaldreieck. Aufgetragen ist die Frequenz des ersten Formanten F_1 (Abzisse) und des zweiten Formanten F_2 (Ordinate) für unterschiedliche Vokale (schematisch)

Während die Vorne-Hinten-Bewegung der Zunge vorwiegend mit dem **zweiten Formanten** korreliert (d. h. „i“ wird vorne artikuliert und hat einen hohen zweiten Formanten, während das „u“ hinten artikuliert wird und einen niedrigen zweiten Formanten hat), ist die Lage des **ersten Formanten** vorwiegend mit der Mundöffnung während der Artikulation korreliert (geringe Öffnung beim i und u mit niedrigem ersten Formanten und relativ große Öffnung beim a mit hohem ersten Formanten). Dieses Vokaldreieck ist nur ein grober Anhaltspunkt für die tatsächlich realisierten Formant-Frequenzen bei verschiedenen Vokalen von verschiedenen Sprechern, da natürliche Sprache einer immensen Schwankung

unterliegt.

Um nun den zeitlichen Verlauf der Artikulation und die Entwicklung der Formanten über die Zeit verfolgen zu können, wird eine Darstellung des Frequenzgehaltes über der Zeit benötigt, die mit der als **Sonagramm** bezeichneten Kurzzeit-Spektralanalyse erfolgen kann. Dabei wird auf der x-Achse die Zeit und auf der y-Achse die Frequenz aufgetragen, während die bei einem bestimmten Zeitpunkt zu einer Frequenz auftretende Energie als Schwärzung über dem jeweiligen Punkt in der Zeit-Frequenz-Ebene aufgetragen wird. Mit dieser früher auch als „visible Speech“ bezeichneten Darstellungsweise ist es dem Spezialisten möglich, lediglich anhand des Spektrogramms Sprache zu erkennen, was anhand der Zeitfunktion oder des Spektrums allein nicht möglich ist. Als Beispiel ist in Abbildung 6.5 das Spektrogramm einer (künstlich erzeugten) Silbe „ba“ angegeben. Es kann deutlich die zeitliche Entwicklung des ersten und des zweiten Formanten beobachtet werden, sowie (aufgrund der hohen Zeitauflösung) eine mögliche Darstellung der Glottis-Periode (Periodizität im Zeitsignal).

Die Vorteile dieser Zeit-Frequenz-Analyse sind die realistische, freiwählbare Zeitauflösung, die gleichzeitig einstellbare Frequenzauflösung, die relativ gute Repräsentanz der Sprachinformation und die Eignung dieser Darstellungsweise für fließende Sprache. Aus diesem Grund ist das Spektrogramm das grundlegende Werkzeug in der Sprach- und Stimmanalyse. Die für das Spektrogramm durchgeführte Vorverarbeitung, die im wesentlichen dem Ohr nachempfunden ist, wird auch in technischen Systemen (beispielsweise zur Spracherkennung) eingesetzt.

Anhand des Spektrogramms lassen sich eine Reihe von akustisch definierten Sprachmerkmalen unterscheiden. Dazu zählen zum einen die bereits erwähnte **Lage der Formanten**, die zur Unterscheidung verschiedener Vokale (Vokal-Diskrimination) eingesetzt werden. Ein weiteres wichtiges Merkmal zur Unterscheidung zwischen stimmhaften und stimmlosen Konsonanten ist das Auftreten von **kurzen Sprachpausen** („Voice-onset-time“) von etwa 20 ms. Sie zeigen die Zeitverzögerung zwischen dem Beginn des stimmlosen Konsonanten (z. B. beim „pa“ den Zeitpunkt der Lippensprengung) und dem Einsatz der Stimmlippen-Schwingung an. Durch nachträgliches Einfügen von Pausen hinter dem Beginn eines Konsonanten kann so aus einem stimmhaften Konsonanten künstlich ein stimmloser Konsonant gemacht werden. Ein weiteres wichtiges Merkmal ist die zeitliche Entwicklung von Formanten (Formant-Übergänge), die beispielsweise zur Unterscheidung verschiedener (stimmhafter) Vokale dienen. Dies lässt sich im Spektrogramm anhand des Unterschiedes zwischen „ba“, „da“ und „ga“ demonstrieren (vgl. Abbildung 6.5). Die eigentliche Information des Konsonanten

liegt nicht zeitlich begrenzt am Anfang der Silbe, sondern bestimmt den **Übergang** der Formanten vom Anfang der Silbe bis zum hinteren Teil. Aus diesem Grunde kann ein geübter Hörer die drei Silben selbst dann noch unterscheiden, wenn der Anfang der Silben künstlich entfernt worden ist. Der im Spektrogramm ebenfalls sichtbare Zeitverlauf des gesamten Spektrums dient weiterhin zur Unterscheidung zwischen verschiedenen **Artikulations-Arten** von Konsonanten (z. B. zur Plosiv-Nasal-Plosiv-Diskrimination und zur Plosiv-Frikativ-Diskrimination) da einige Klassen von Formanten (z. B. Plosive) einen charakteristischen Zeitverlauf aufweisen. Die spektrale Grobstruktur ist auch notwendig, um verschiedene Konsonanten (z. B. Frikative, d. h. Reibelaute) akustisch voneinander unterscheiden zu können. Aus den genannten akustischen Parametern des Spektrogramms lassen sich **akustische Merkmalsysteme** bestimmen, die bei der Analyse von Phonem-Verwechslungen Rückschlüsse auf die zugrundeliegenden akustischen Übertragungsdefizite (allerdings nur im begrenztem Maße) zulassen. Zu derartigen akustischen Merkmalsystemen gehören beispielsweise die Lage des ersten und zweiten Formanten, die Vokaldauer, die Voice-onset-time und/oder andere aus dem Spektrogramm allgemein ablesbare Parameter.

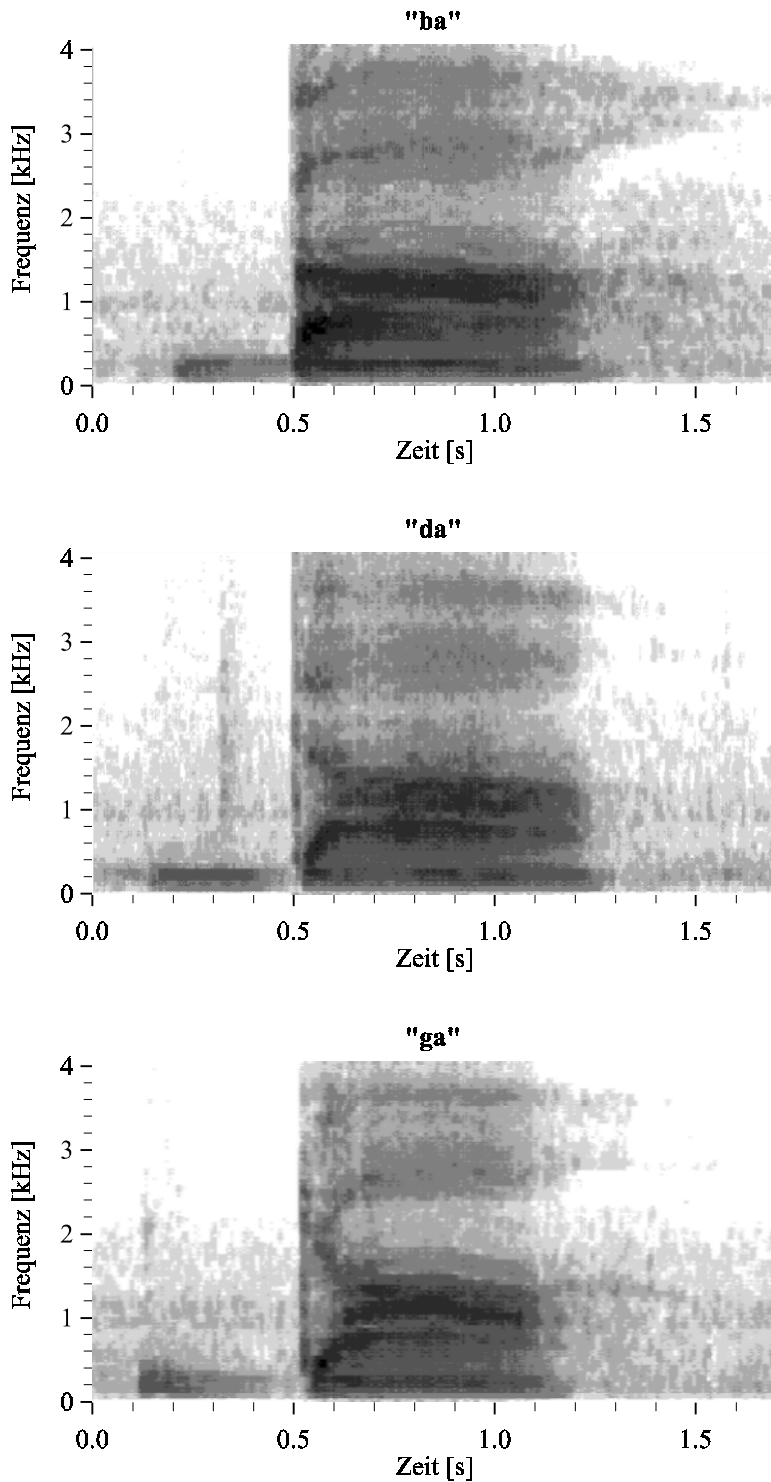


Abbildung 6.5: Spektrogramm „ba“, „da“, „ga“

Aufgrund des Erfolges, den die Spektrogramm-Darstellung bei der Visualisierung von Sprache und der Erklärung von Eigenschaften der Sprachwahrnehmung aufweist, erscheint es sinnvoll, die Arbeitsweise des Ohres in erster Näherung nach Art des Spektrogramms sich modellmäßig vorzustellen. Dafür sind die zeitliche Darstellung und die Darstellung im Frequenzbereich, sowie die entsprechende Darstellung der zu jeder

Kombination aus Frequenz und Zeit gehörenden Intensität von großer Bedeutung. Eine weitere Art der Darstellung, die eng mit der Spektrogrammdarstellung zusammenhängt, ist die Aufteilung des Zeitverlaufs in jedem Frequenzband nach unterschiedlichen **Modulationsfrequenzen** (Rhythmen). Es gibt physiologische Hinweise darauf, daß im auditorischem System eine Aufspaltung von jeder Mittenfrequenz in verschiedene Modulationsfrequenzen stattfindet, so daß die Untersuchung der Relevanz von Modulationen und Modulations-Übertragungsfunktionen für die menschliche Sprachwahrnehmung von großer Bedeutung ist. Wenn man die in natürlicher, fließender Sprache vorkommenden Modulationsfrequenzen analysiert, stellt man ein **Maximum des Modulationsspektrums bei etwa 4 Hz** fest, das ungefähr der Silbenfrequenz entspricht. Zu niedrigeren und höheren Modulationsfrequenzen nimmt diese Energie ab (vgl. Abbildung 6.6). Dabei ist das Vorhandensein dieser Modulationen essentiell für das Verstehen von Sprache: Wenn die tiefen Modulationsfrequenzen weggefiltert werden, wird mit zunehmender Eckfrequenz immer weniger von der Sprache verständlich. Dies tritt in ähnlicher Weise auf, wenn sämtliche hohen Modulationsfrequenzen in Sprache herausgefiltert werden und die Eckfrequenz langsam erniedrigt wird (vgl. Festen, 1996). Damit scheint die Modulationswahrnehmung neben der Zeit-Frequenz- und Intensitätsabbildung eine wesentliche Rolle für die Sprachwahrnehmung zu spielen. Sämtliche dieser für die Sprachwahrnehmung wichtigen Größen sind auch bei der Anwendung von Hörgeräten für schwerhörende Patienten von Bedeutung.

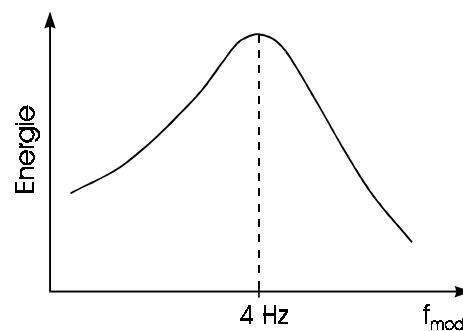


Abbildung 6.6: Modulationsspektrum von Sprache (Schema)

VI.1.3 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.

VI.1.3.1 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 Wörtertest**, 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).

VI.1.3.2 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 Transmission 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 Störrauschen 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 ($\text{SNR}(j)$) wird also auf einen Bereich von -15dB bis $+15\text{dB}$ 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:

$$\text{AI oder STI} = \sum_j W_j \cdot [\text{SNR}(j) + 15\text{dB}] \quad (6.1)$$

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 6.8). 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.

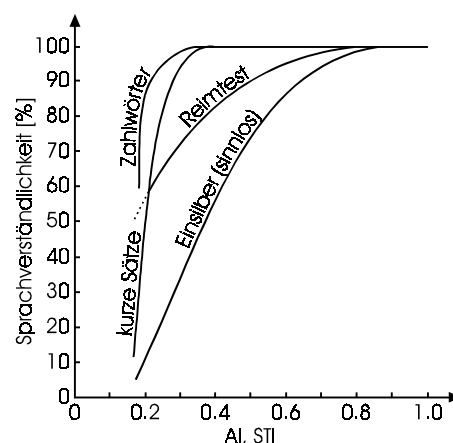


Abbildung 6.7: Sprachverständlichkeit für verschiedenes Sprachmaterial als Funktion des AI bzw. STI

Mit Hilfe der in Abbildung 6.7 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 Störgeräuschen 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).

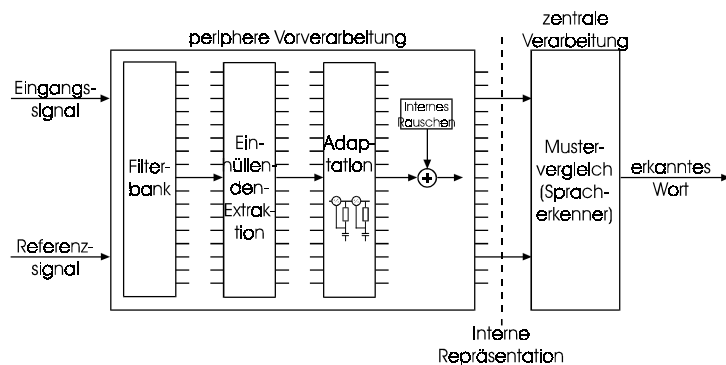


Abbildung 6.8: Perzeptionsmodell zum Sprachverstehen nach Holube und Kollmeier (1996)

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 6.8 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 (Abb. 5.10). 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 Rechenaufwandes für dieses Modell und die noch nicht in allen Einzelheiten geklärten Verarbeitungsprozesse sind jedoch noch weitere Forschungsarbeiten zu seiner Validierung notwendig.

VI.2 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. 6.1.3) 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.

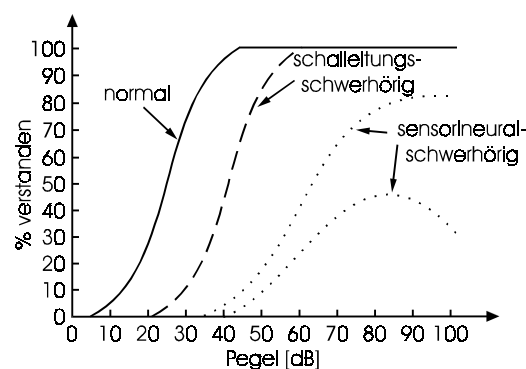


Abbildung 6.9: Diskriminationsfunktionen bei unterschiedlichen Pathologien

(schematisch)

Für den **Sprachtest in Ruhe** findet man bei den verschiedenen Schwerhörigkeitsformen prinzipiell die in Abbildung 6.9 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 Nutzsprach-Sprachschallsituationen**. 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.

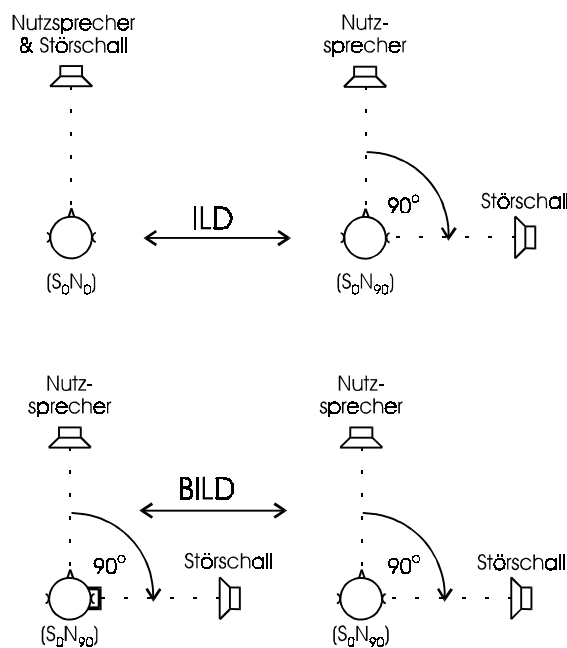


Abbildung 6.10: Testanordnung für ILD und BILD

Dieser Effekt kann mit der Intelligibility Level Difference (**ILD**) bzw. der binauralen Intelligibility Level Difference (**BILD**) quantitativ ausgemessen werden (vgl. Abb. 6.10): In der Referenzsituation wird Nutzsprache und Störgeräusch dem Probanden genau von vorne angeboten und die Sprachverständlichkeits-Schwelle bestimmt. Wenn der Störgeräusch nun von 90 Grad seitlich angeboten wird, sinkt die Sprachverständlichkeits-Schwelle bei Normalhörenden um bis zu 12 dB, während dieser Verständlichkeitsgewinn bei Schwerhörigen sehr individuell verschieden

ist und z. T. deutlich eingeschränkt sein kann. Um den monauralen Anteil (Kopfabstimmungseffekt an dem der Störgeräuschquelle abgewandten Ohr) vom binauralen Anteil (Gewinn durch binaurale Störgeräuschunterdrückung) trennen zu können, wird die BILD in der letztgenannten räumlichen Situation bestimmt. Dazu wird das „bessere“ (d. h. das der Störschallquelle abgewandte) Ohr verstopft und die resultierende Sprachverständlichkeitsschwelle mit der binauralen Sprachverständlichkeitsschwelle verglichen. Bei Normalhörenden wird durch Zuschalten des „schlechteren“ Ohrs ein Gewinn von maximal ca. 6 dB erzielt. Der Gesamt-Gewinn in der räumlichen Störschall-Nutzschallsituation wird also je zur Hälfte durch eine binaurale und eine monaurale Komponente hervorgerufen.

In der Praxis können diese Tests unter Kopfhörerbedingungen durch Verwendung der Kunstkopftechnik bzw. der virtuellen Akustik durchgeführt werden. In Freifeldbedingungen bei nicht-idealen akustischen Gegebenheiten (Hörprüfräumen mit einer gewissen Nachhallzeit) sinkt die maximal erzielbare ILD auf Werte um 6 dB und die BILD auf Werte um etwa 3 dB ab. Daher ist es notwendig, effiziente adaptive Testverfahren (wie beispielsweise den Göttinger Satztest mit adaptiver Pegelsteuerung einzusetzen), mit denen die ILD bzw. BILD für beide Ohren des Patienten in vertretbarem Zeitaufwand bestimmt werden können und so der Gewinn einer etwaigen zweiseitigen Hörgeräte-Versorgung quantifiziert werden kann.