

# Sprachaudiometrie in der Forschung

Thomas Brand, Tim Jürgens

Medizinische Physik, Universität Oldenburg

## Einleitung

In der audiologischen Praxis in Deutschland werden meist Sprachverständlichkeitstests in Ruhe eingesetzt (Freiburger Einsilber und Zahlen). Für die Forschung ist die Sprachverständlichkeit in Ruhe jedoch nicht besonders interessant, weil hierfür die Abhängigkeit von Sprachpegel und Hörverlust weitestgehend geklärt ist. Interessanter ist hier die Sprachverständlichkeit im Störgeräusch, die besser diejenigen Situationen widerspiegelt, in denen Hörverluste zu einer pathologischen Beeinträchtigung der Kommunikationsfähigkeit führen. Neben den häufig verwendeten unmodulierten Störgeräuschen mit sprachähnlichem Frequenzspektrum sind besonders Störgeräusche mit Modulationen von Interesse, da diese Störgeräusche eine sehr gute diagnostische Differenzierung zwischen verschiedenen Hörstörungen ermöglichen (Wagner et al., 2006). Ein weiterer wichtiger Forschungszweig sind räumliche Störschall-Nutzschall-Anordnungen, die eine Erfassung der Leistung des binauralen Gehörs bei der Störgeräuschbefreiung in „Cocktail-Party-Situationen“ ermöglichen (siehe Überblick bei Bronkhorst, 2000).

Für die verschiedenen Messbedingungen (Ruhe, unmodulierte und modulierte Störgeräusche sowie räumliche Bedingungen) gibt es Modellansätze, die aus dem Tonaudiogramm die zu erwartende Sprachverständlichkeit vorhersagen (z.B. Brand und Kollmeier, 2002; Beutelmann und Brand, 2006; Meyer et al. 2007). Die Anwendungsmöglichkeit dieser Vorhersagemodelle in der audiologischen Praxis besteht darin, dass ausgehend vom Tonaudiogramm individuell für die Patientin oder den Patienten eine Vorhersage des zu erwartenden Ergebnisses angegeben werden kann. Diese Vorhersage ist auch für relativ komplexe Hörsituationen möglich, für die die Audiologin oder der Audiologe das zu erwartende Ergebnis nicht im Vorhinein einschätzen kann. Weicht dann das Messergebnis signifikant von der Prognose ab, so ist dies ein Hinweis darauf, dass die Beeinträchtigung des Sprachverstehens nicht durch die Hörbarkeit allein (ausgedrückt durch das Tonaudiogramm) erklärt werden kann, sondern dass es noch weitere Komponenten gibt, die die Sprachverständlichkeit beeinträchtigen. Als solche weiteren Komponenten kommen z.B. Störungen der zentralen auditorischen Verarbeitung, der Sprachverarbeitung oder der binauralen Verarbeitung in Frage.

Aus diesen Betrachtungen ergeben sich folgende Anforderungen bzw. Möglichkeiten für die moderne Sprachaudiometrie:

1. Die Sprachaudiometrie sollte Situationen erfassen, in denen schwerhörnde Patienten typischerweise ein Handicap durch ihre Schwerhörigkeit haben, also auch Situationen im Störgeräusch.

2. Sprachaudiometrische Daten können nur sinnvoll interpretiert werden, wenn sie mit Referenzwerten für die jeweilige Messsituation verglichen werden können. Dies gilt insbesondere für Daten von individuellen schwerhörnden Patienten.

3. Modellbasierte Vorhersagen der Sprachverständlichkeit können als solche individuelle Referenzdaten verwendet werden.

Die verschiedenen für die Sprachverständlichkeitsvorhersage zur Verfügung stehenden Modelle können in zwei Gruppen unterteilt werden. Erstens *makroskopische Modelle* bei denen Sprache, Störgeräusch und Versuchsperson nur grob beschrieben werden, z.B. durch Langzeit-Frequenz-Spektrum von Sprache und Störgeräusch sowie Reintonaudiogramm der Versuchsperson. Einer der heute am weitesten verbreiteten Vertreter dieser makroskopischen Modelle ist der Speech-Intelligibility-Index (SII, ANSI S3.5-1997), der auch die Grundlage der hier beschriebenen makroskopischen Modelle darstellt.

Der zweite in diesem Beitrag beschriebene Ansatz zur Sprachverständlichkeitsvorhersage ist die *mikroskopische Modellierung*. Hier werden die jeweiligen Zeitsignale von Sprache und Störgeräusch durch ein auditorisches Modell verarbeitet und auf diese Weise eine möglichst gehörgerechte *interne Repräsentation* der Signale berechnet. Diese interne Repräsentation wird anschließend von einem automatischen Spracherkenner weiterverarbeitet, der aus einem vorgegebenen Vokabular dasjenige Wort auswählt, dessen interne Repräsentation am besten mit der des dargebotenen Testworts übereinstimmt (Holube und Kollmeier, 1996, Jürgens et al., 2007). Diese mikroskopische Modellierung ermöglicht es im Gegensatz zur makroskopischen Modellierung, nicht nur die Verständlichkeitsfunktion, sondern auch die Erkennungs- und Verwechslungsraten einzelner Wörter (oder Phoneme, falls Phonemtests verwendet werden) zu berechnen. Wie in diesem Beitrag dargestellt, ist es so möglich,

Phonemverwechslungsmatrizen vorherzusagen, die mit Phonemverwechslungsmatrizen verglichen werden können, wie sie z.B. mit einem Logatomtest mit Schwerhörigen gemessen werden können.

## Teil I: Makroskopische Modellierung

### Sprachverstehen in Ruhe

Der Speech-Intelligibility-Index (SII, ANSI S3.5-1997) sagt die Sprachverständlichkeit in Abhängigkeit vom Langzeit-Frequenzspektrum der Sprache und des Störgeräusches sowie des Reintonaudiogramms der Versuchsperson vorher. Die Anwendung dieses Verfahrens auf audiologische Messungen mit dem Freiburger Einsilber- und Zahlentest sowie dem Göttinger und Oldenburger Satztest in Ruhe und in verschiedenen Störgeräuschen wurde von Brand und Kollmeier (2002) untersucht. Es zeigte sich, dass z.B. die 50%-Sprachverständlichkeitsschwelle in Ruhe gemessen mit dem Freiburger Zahlentest, Göttinger Satztest oder Oldenburger Satztest mit einer sehr hohen Korrelation ( $p > 0,9$ ) vorhergesagt werden kann. Eine ähnlich hohe Korrelation ist z.B. auch zwischen der Ruhehörschwelle bei 500 Hz und der 50%-Sprachverständlichkeitsschwelle in Ruhe bekannt. Das bedeutet, dass diese Art von Sprachverständlichkeitsmessungen keinen nennenswerten Erkenntnisgewinn bringen, da ihr Ergebnis bereits aus dem Tonaudiogramm vorhersagbar ist. Andererseits sind hier Inkonsistenzen zwischen Vorhersage und Messung besonders signifikant und können auf Fehler der Messwerterhebung, Störungen der höheren auditiven Sprachverarbeitung oder eine vom Patienten simulierte Schwerhörigkeit hinweisen.

Bei der Vorhersage der Verständlichkeitsfunktion (Verständlichkeit in Abhängigkeit vom Sprachpegel) ist keine hohe Korrelation zwischen Vorhersagen und Messwerten festzustellen. Dies kann zu einem großen Teil durch die geringe Messgenauigkeit von Verständlichkeitsmessungen erklärt werden. Durch die Begrenzung der für die klinische Sprachaudiometrie zur Verfügung stehenden Zeit können nur eine begrenzte Anzahl von Wörtern pro Pegel gemessen werden (z.B. 20 Wörter pro Sprachpegel beim Freiburger Einsilber). Da die Daten binomialverteilt sind, ergeben sich sehr weite Konfidenzintervalle der resultierenden Verständlichkeitsschätzungen. Aus diesem Grund beschränkt sich dieser Beitrag im Folgenden auf Messung und Vorhersage der 50% Sprachverständlichkeitsschwelle, die mit hoher Genauigkeit messbar und vorhersagbar ist.

### Sprachverstehen im Störgeräusch

Testverfahren wie der Göttinger Satztest und der Oldenburger Satztest ermöglichen eine sehr genaue Bestimmung der 50%-Sprachverständlichkeits-

schwelle im Störgeräusch, da diese Tests über eine sehr steile Verständlichkeitsfunktion verfügen, d.h. die Verständlichkeit steigt mit steigendem Sprachpegel steil an. Bei der Verwendung von 20 Sätzen pro Messung kann die 50%-Schwelle in der Regel mit einer Standardabweichung von weniger als einem dB gemessen werden. Bei Messungen in unmodulierten Störgeräuschen reicht allerdings selbst diese Genauigkeit oft kaum aus, um deutliche Unterschiede zwischen verschiedenen Graden von Schwerhörigkeit signifikant nachzuweisen. Die Sprachverständlichkeitsschwellen von Normal- und Schwerhörigen werden im unmodulierten Störgeräusch durch die Maskierungswirkung angeglichen. Folglich liegen typische Werte der 50%-Schwelle in einem relativ engen Bereich zwischen -6 dB S/N bei Normalhörenden und 15 dB S/N bei höhergradig Schwerhörigen. Diese vergleichsweise geringe Streuung der gemessenen 50%-Schwellen würde selbst bei einer sehr präzisen Vorhersagemethode keine besonders hohe Korrelation zwischen vorhergesagten und gemessenen 50%-Schwellen im unmodulierten Störgeräusch erlauben. Brand und Kollmeier (2002) erzielten für diese Situation lediglich eine Korrelation von 0,57. Kurioserweise stellen also die Standard-Störgeräusche von Göttinger und Oldenburger Satztest für die Vorhersage eine besonders schwierige Situation dar, wobei bisher unklar ist, ob dies noch wesentlich verbessert werden kann.

Die Vorhersagbarkeit verbessert sich wieder, wenn ein weiteres psychoakustisches Merkmal eingeführt wird, das von der Versuchsperson genutzt werden kann, um eine Reduzierung der Maskierung durch das Störgeräusch zu erreichen. Eine solche Reduzierung kann z.B. entweder durch eine Modulation des Störgeräusches erreicht werden, die es ermöglicht in den Lücken des Störgeräusches ein verbessertes Sprachverstehen zu erreichen oder durch eine räumliche Trennung von Sprache und Störgeräusch, die eine räumliche Demaskierung erlaubt. Da die Stärke dieser Demaskierung bei verschiedenen Graden der Schwerhörigkeit unterschiedlich ausfällt, kommt es zu einer vergrößerten Varianz der inter-individuellen Messwerte. Diese vergrößerte Varianz der Messwerte ermöglicht bei annähernd gleich bleibender Güte der Vorhersage für diese komplexeren Messsituationen (verglichen mit der Güte der Vorhersagen für unmodulierte Störgeräusche) eine Verbesserung der Korrelation zwischen Vorhersagen und Messwerten.

### Sprachverstehen in moduliertem Störgeräusch

Der standardisierte SII nach ANSI S3.5-1997 ist nicht für modulierte Störgeräusche vorgesehen, da er lediglich auf Langzeit-Frequenz-Spektren basiert

und die zeitliche Verarbeitung (z.B. das „in-die-Lücken-des-Störgeräusches-Hören“) nicht berücksichtigt. Modulierte Störgeräusche sind jedoch diagnostisch besonders interessant, da sie deutlich größere Unterschiede zwischen verschiedenen Schwerhörnden aufzeigen als nicht modulierte Störgeräusche. Aus diesem Grund wird an Erweiterungen des SII für modulierte Störgeräusche gearbeitet. Diese Erweiterungen beruhen auf der Berechnung von Kurzzeit-Frequenzspektren, die iterativ vom SII verarbeitet werden. Aus den so berechneten Kurzzeit-SII Werten wird durch geeignete Mittelung die Vorhersage der Sprachverständlichkeit für die Gesamtaufnahme (z. B. Testliste mit Sätzen) berechnet. In der Untersuchung von Meyer et al. (2007, in diesem Band) wurden verschiedene Ansätze für solche Kurzzeit-SII Berechnungen verglichen. Wobei sich diese Ansätze zum Teil deutlich in ihrem rechnerischen Aufwand unterschieden. Es zeigte sich jedoch, dass der Nutzen von relativ aufwendigen Modellansätzen begrenzt ist und bereits vereinfachende Ansätze (Brand und Kollmeier, 2002) ähnliche Vorhersagen liefern wie der etwas aufwendigere, auch die Frequenzabhängigkeit der Modulationen des Störgeräusches berücksichtigende Ansatz von Rhebergen und Versfeld (2005). Eine weitere Steigerung des Rechenaufwands zur zusätzlichen Berücksichtigung der Frequenzabhängigkeit der Modulationen der Sprache (Meyer et al., 2007) brachte nur noch einen vergleichsweise geringen Nutzen.

### **Sprachverstehen in binauralen Störgeräusch-Situationen**

Die Erweiterung des SII für binaurale Störgeräusch-Sprache Situationen besteht aus einer Kombination eines Equalization/Cancellation (EC) Modells mit dem SII (Beutelmann und Brand, 2006). Das EC Modell berechnet aus binauralen Aufnahmen von Sprache und Störgeräusch eine monaurale Aufnahme mit verbessertem Signal-Rausch-Verhältnis. Diese Aufnahme wird vom SII weiterverarbeitet und die Sprachverständlichkeit vorhergesagt. Bei der Verbesserung des Signal-Rausch-Verhältnisses mit dem EC-Modell werden auf einem Ohrkanal Pegeländerungen und Verzögerungen durchgeführt (Equalization) und anschließend die beiden Ohrsignale voneinander abgezogen (Cancellation). Dies wird unabhängig in verschiedenen Frequenzkanälen durchgeführt. Eine genauere Darstellung dieses Modells und seiner Anwendung auf Normal- und Schwerhörnde findet sich auch bei Brand und Beutelmann (2006). Die Anwendung dieses Modells zeigte sehr gute Vorhersagen von Laborexperimenten sowohl mit Normalhörenden als auch mit Schwerhörnden. Die Korrelation über alle gemessenen Situationen war besser als 0,9. Etwas problematischer erwies sich

die Situation bei diagnostischen Messungen mit untrainierten Versuchspersonen (Brand und Beutelmann, 2006). Hier fiel die Korrelation  $p=0,77$  (binaural) bzw.  $p=0,73$  (monaural) deutlich geringer aus.

### **Kombination aus modulierten und räumlichen Situationen und Informationale Masking**

Bisher unveröffentlichte Messungen ergaben, dass eine Kombination des Modells für modulierte Störgeräusche mit dem binauralen Modell gut möglich ist. Die Lücken im Störgeräusch können auch bei binauralen Situationen von der Versuchsperson genutzt werden, was durch eine einfache Kombination der Modelle gut beschrieben werden kann. Hiermit steht ein Modell für eine Vielzahl von Störgeräuschkombinationen zur Verfügung.

Allerdings wird durch dieses Modell nicht die Effekte des so genannten *Informational Maskings* berücksichtigt. Hierbei werden Sprache oder sehr sprachähnliche Signale als Maskierer verwendet, wodurch ein zusätzlicher Maskierungseffekt auftritt, weil die Objektrennung zwischen Sprache und Maskierer erschwert wird.

## **Teil II: Mikroskopische Modellierung**

### **Auditorisches Perzeptionsmodell**

Das in diesem Ansatz verwendete mikroskopische Modell basiert auf dem auditorischen Perzeptionsmodell nach Dau et al. (1996, 1997), welches in Anlehnung an das menschliche auditorische System eine interne Repräsentation des dargebotenen akustischen Signals berechnet. Das Perzeptionsmodell besteht aus einer Gammaton-Filterbank, einem einfachen Haarzellenmodell, dynamischen Nachregelschleifen und einer Modulationsfilterbank mit vier Modulationsfrequenzkanälen. Das Ergebnis ist eine spektrogrammähnliche Repräsentation des dargebotenen Signals, wobei für die vier verschiedenen Modulations-Frequenz-Kanäle jeweils eine Darstellung vorliegt. Entsprechend dem von Holube und Kollmeier (1996) vorgeschlagenen Verfahren wird diese interne Repräsentation als Eingangsgröße für einen einfachen automatischen Spracherkenner verwendet. Dieser Spracherkenner berechnet perzeptive Abstände der internen Repräsentation des dargebotenen Wortes zu den zuvor berechneten internen Repräsentationen des Wortschatzes des Modells. Das Wort mit dem geringsten Abstand wird als Antwort ausgewählt. Um zeitliche Unterschiede zwischen dem Testwort und den Wörtern aus dem Vokabular auszugleichen wird ein Dynamic-Time-Warping (DTW) Algorithmus verwendet.

In dieser Studie wurden zwei Konfigurationen des DTW Erkennens verwendet:

**Konfiguration A:** Fünf interne Repräsentationen pro Wort wurden als mögliche Zielmuster („Templates“) verwendet. Keine der zugehörigen originalen Aufnahmen war identisch mit dem zu erkennenden Testwort. Das Wort mit dem geringsten mittleren Abstand seiner fünf Templates zu der internen Repräsentation des Testworts wurde von dem Modell als erkannt ausgewählt. Dieses Vorgehen bildet die realistische Situation eines Spracherkenners nach, dem ebenfalls die tatsächliche Aufnahme des zu erkennenden Wortes nicht bekannt ist.

**Konfiguration B:** Es wurde für jedes Wort nur eine interne Repräsentation als Template verwendet, wobei das zu erkennende Testwort jeweils mit einem der Templates übereinstimmt. D.h. die resultierende interne Repräsentation des Testworts unterscheidet sich nur in dem hinzugefügten Rauschen von der internen Repräsentation des richtigen Wortes. Diese Vorgehensweise vernachlässigt die natürliche Variabilität von Sprache und nimmt eine perfekte Übereinstimmung zwischen Template und Testwort an. Für einen automatischen Spracherkenner ist diese Situation nicht realistisch. Für die Modellierung ist sie dennoch interessant, weil sie den Grenzfall beschreibt, in dem das Sprachmodell des menschlichen Zuhörers als perfekt angenommen wird.

### Oldenburger Logatom Korpus (OLLO)

Für die Experimente mit dem mikroskopischen Modell wurde der Oldenburger Logatom Korpus (OLLO) verwendet (Wesker et al., 2005). Dieser Sprachkorpus besteht aus 70 Vokal-Konsonant-Vokal (VCV) und 80 Konsonant-Vokal-Konsonant (CVC) Logatomen. Der Sprachkorpus wurde von 50 Sprechern und Sprecherinnen mit insgesamt fünf Dialekten bzw. Akzenten aufgenommen. Jedes Logatom wurde 18 mal von jedem Sprecher oder Sprecherin aufgesprochen, wobei verschiedene Artikulationsarten aufgenommen wurden („langsam“, „normal“, „schnell“, „laut“, „leise“, „fragend“). Der OLLO Logatomkorpus ist frei erhältlich unter: <http://sirius.physik.uni-oldenburg.de>. Die Verwendung dieses Sprachkorpus erlaubt systematische Untersuchungen von einzelnen Phonemerkennungsraten und vermeidet, dass die Versuchspersonen Kontext-Wissen verwenden können. Hierdurch wird die Modellierung vereinfacht.

### Ergebnisse

Die Verständlichkeitsfunktionen aus den automatischen Erkennungsexperimenten sind in Abb. 1 dargestellt. Außerdem sind zum Vergleich die Verständlichkeitsfunktionen aus den

Hörexperimenten mit den normalhörenden Versuchspersonen dargestellt (gestrichelt „Fit to HSR“). Konfiguration A zeigt dieselben Erkennungsraten für die CVCs als auch für die VCVs. Die 50%-Schwelle der angepassten Verständlichkeitsfunktion ist 1,3 dB und damit 13 dB höher als bei den Hörexperimenten. Eine deutlich bessere Vorhersage der Daten aus den Hörexperimenten ist mit Modelvariante B möglich. Die Reihenfolge der Verständlichkeiten von CVC und VCV Logatomen wird richtig vorhergesagt. Auch bei den Hörexperimenten zeigte sich, dass die CVCs besser erkannt werden als die VCVs (hier nicht dargestellt). Die 50%-Schwelle ist mit -7,6 dB deutlich näher an der 50%-Schwelle aus den Hörexperimenten, es bleibt aber ein Unterschied von 4,6 dB. Aus diesem Grund werden im Folgenden nur Verwechslungsmatrizen für Modelvariante B bei einem Signal-Rausch-Verhältnis von -10 dB betrachtet. Bei den Hörexperimenten wird ein Signal-Rausch-Verhältnis von -15 dB betrachtet, weil bei diesem Wert in etwa die gleiche Verständlichkeit erzielt wird.

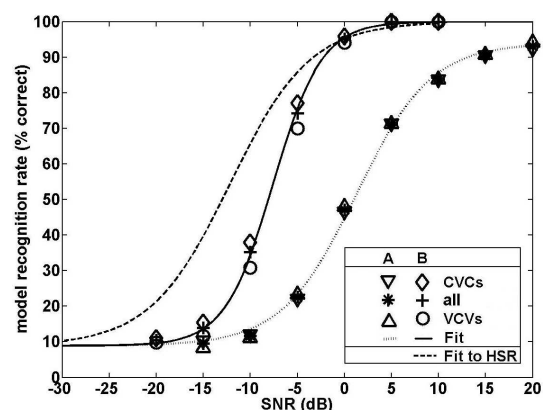


Abbildung 1: Vorhergesagte Verständlichkeitsfunktionen für Model-Konfigurationen A und B für Logatome in ICRA 1 Rauschen bei festem Signal-Rausch-Abstand. Zum Vergleich: an Messungen angepasste mittlere Verständlichkeitsfunktionen von Normalhörenden. (Aus Jürgens et al. 2007).

Abb. 2 zeigt die Phonem-Verwechslungsmatrix für die VCVs für 10 normalhörende Versuchspersonen. Der Signal-Rausch-Abstand entsprach -15 dB, was einer Verständlichkeit von 33% entspricht. (Die hier nicht dargestellten CVCs ergaben bei diesem Signal-Rausch-Abstand 52% Verständlichkeit.) Richtig erkannte Phoneme sind auf der Diagonalen dargestellt. Frikative wie „f“, „s“ und „ts“ werden am besten erkannt, stimmhafte Konsonanten wie „n“, „v“ und „b“ am schlechtesten. Auf der Diagonalen treten große Unterschiede auf, z.B. zwischen „n“ und „f“. Stimmlose Plosive („p“, „t“, „k“) werden deutlich besser erkannt als stimmhafte („b“, „d“, „g“). Es gibt kaum Verwechslungen zwischen hochfrequenten Konsonanten wie „f“, „ts“ mit tieffrequenten.

	d	t	g	k	f	s	b	p	v	ts	m	n	ʃ	l
d	16	12					8	14	12		10		12	
t		62												
g		16	12	24							14		8	
k		8	14	26	8							8		
f	8	8			30			10	12					
s						80				18				
b	8	18						10	10		10		8	
p					10			12	32					
v			12	10				14	12	8		12		16
ts						28				66				
m	10	10			8		12	8			12		10	
n	8	12	12	16						12			16	
ʃ													92	
l	18	12			10	8					8			18

Abbildung 2: Verwechslungs-Matrix (Antworthäufigkeiten in %) für Normalhörende bei -15 dB, gemessen mit in Logatome eingebetteten Konsonanten. Zeilen: präsentiertes Phonem. Spalten: erkanntes Phonem. Graustufen verdeutlichen unterschiedliche Antworthäufigkeiten. Antworthäufigkeiten unter 7 % sind nicht dargestellt. (Aus Jürgens et al. 2007).

Abb. 3 zeigt die Konsonanten-Verwechslungsmatrix aus den Modellrechnungen. Der Vergleich mit Abb. 2 ergibt, dass dieselben Konsonanten “ʃ”, “ts” und “s” am besten erkannt werden, allerdings nicht mit so hohen Erkennungsraten wie bei den Hörexperimenten. Jedoch werden einige Konsonanten wie “n”, “v” und “b” besser erkannt als bei den Hörexperimenten. Es gibt keine signifikanten Unterschiede zwischen den Erkennungsraten für stimmhafte und stimmlose Plosive. Insgesamt ist der “Kontrast” der Matrix zwischen den Diagonalelementen kleiner als bei den Hörexperimenten.

	d	t	g	k	f	s	b	p	v	ts	m	n	ʃ	l
d	22						8	16			16	20		
t		26					10				22	12		8
g			26				10	8			14	18		8
k			10	20			10				18	16		8
f					12		10	10			18	18		14
s						54					12			10
b							30				16	14		18
p							12	18	12		10	22		10
v	8								24		22	18		8
ts							8	8	38		16	14		
m							10				32	14		18
n							10	8	16		14	26		10
ʃ												8	72	
l							10	8			18	18		32

Abbildung 3: Verwechslungs-Matrix für Konsonanten-Erkennung mit Model Variante B bei -10 dB Signal-Rausch-Abstand. Darstellung wie in Abbildung 2. (Aus Jürgens et al. 2007).

## Diskussion und Ausblick

Die Anwendung von makroskopischen auf dem SII basierenden Modellen in der audiologischen Diagnostik ist bereit für die klinische Erprobung. Da für die Vorhersagen lediglich das Reintonaudiogramm benötigt wird, sind keine zusätzlichen Messungen erforderlich, um die Sprachverständlichkeit vorherzusagen. Besonders hilfreich erscheint die Anwendung dieser Vorhersagemethoden für Messungen mit (modulierten) Störgeräuschen und räumliche Situationen. Da solche Messungen in der klinischen Diagnostik noch nicht verbreitet sind, gibt es kaum audiologische Erfahrungen hinsichtlich der für eine gegebene Versuchsperson und Messsituation zu erwartenden Sprachverständlichkeit. Hier kann die Vorhersage helfen, da sie basierend auf dem Tonaudiogramm die Sprachverständlichkeit auch in komplexen Hörsituationen (konstante und modulierte Störgeräusche, binaurale Hörsituationen) vorhersagen kann und auf diese Weise die individuellen Referenzwerte liefert, die für die Interpretation der Ergebnisse erforderlich sind.

Die mikroskopische Modellierung der Sprachverständlichkeit ist gegenwärtig noch nicht für den Einsatz in der klinischen Diagnostik bereit, sondern in erster Linie für die Forschung interessant. Die mikroskopische Modellierung bietet prinzipiell deutlich weiter reichende Möglichkeiten als die makroskopische Modellierung. Neben der hier beschriebenen Vorhersage der Sprachverständlichkeit auf Phonemebene und die hierdurch mögliche Modellierung von Phonem-Verwechslungsmatrizen, können auch komplexere psychoakustische Effekte (wie Einschränkungen der Frequenzselektivität und zeitlichen Verarbeitung) und ihre Auswirkungen auf das Sprachverstehen modelliert werden. Hierzu bietet sich der Ansatz nach Derleth et al. (2001) an, bei dem das auditorische Perzeptionsmodell entsprechend der spezifischen psychoakustischen Leistungen von Schwerhörigen angepasst wurde. Die Kombination dieses Perzeptionsmodells für Schwerhörige mit der hier vorgestellten einfachen Spracherkennungsstufe steht noch aus.

Ein praktisches Problem bei der klinischen Auswertung von Phonemverwechslungsmatrizen besteht in der statistischen Unsicherheit, die entsteht, wenn diese Verwechslungsmatrizen im Einzelfall experimentell bestimmt werden müssen. Aufgrund der Binomialverteilung der Daten ist eine größere Anzahl von Messungen erforderlich, um

die Besetzung dieser Matrix mit hinreichender Genauigkeit zu bestimmen. Für die praktische Anwendung bedeutet dies, dass Phonemverwechslungen nur für eine sehr begrenzte Anzahl von Phonemen bestimmt werden können. Ziel der gegenwärtigen Untersuchungen ist es, durch ausgiebige Messungen bei einer ausgewählten Anzahl von Schwerhörenden und den Vergleich mit den Ergebnissen der Modellierung herauszufinden, welche Phonemverwechslungen als charakteristisch für verschiedene audiologische Befunde angesehen werden können und sich deshalb besonders für die audiologische Anwendung empfehlen.

## Danksagung

Dieser Beitrag wurde gefördert von der Audiologie Initiative Niedersachsen und vom SFB/TR 31 „Das aktive Gehör“.

## Literatur

- ANSI S3.5-1997 – Methods for Calculation of the Speech Intelligibility Index, 1997.
- Beutelmann R, Brand T (2006) Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 120, p. 331-342.
- Brand T, Beutelmann R (2006) Examination of an EC/SII based model predicting speech reception thresholds of hearing-impaired listeners in spatial noise situations. In: *Hearing aid fitting - Proceedings of 21st Danavox Symposium 2005*, editors: Rasmussen, Poulsen, Andersen, Larsen, ISBN 87-982422-0-2 Center Tryk Denmark 2006, p. 139-151.
- Brand T, Kollmeier B (2002) Vorhersage der Sprachverständlichkeit in Ruhe und Störgeräusch aufgrund des Reintonaudiogramms. In: *5. Jahrestagung der Deutschen Gesellschaft für Audiologie*, p. CD-ROM.
- Bronkhorst AW (2000) The Cocktail Party Phenomenon: A review of research on speech intelligibility in multiple talker conditions, *ACUSTICA - acta acoustica* 86, 117–128.
- Dau T, Püschel D, Kohlrausch A (1996) A quantitative model of the "effective" signal processing in the auditory system: I. Model structure. *J. Acoust. Soc. Am.* 99, p. 3615-3622.
- Dau T (1997) Modeling auditory processing of amplitude modulation: I Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.* 102, p. 2892-2905.
- Derleth, RP, Dau T, Kollmeier B (2001) Modelling temporal and compressive properties of the normal and impaired auditory system. *Hearing Research*, 159(1-2), p. 132-149.
- Holube, I, Kollmeier B (1996) Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. *J. Acoust. Soc. Am.* 100 p. 1703-1716
- Jürgens T, Brand T, Kollmeier B (2007) Modelling the human-machine gap in speech reception: Microscopic speech intelligibility prediction for normal-hearing subjects with an auditory model. In *Interspeech 2007, ISCA, Antwerp (im Druck)*.
- Meyer R, Kollmeier B, Brand T (2007) Predicting speech intelligibility in fluctuating noise. 8<sup>th</sup> EFAS Congress/10<sup>th</sup> Congress of the German Society of Audiology, Heidelberg 2007 (auf dieser CD).
- Rhebergen K, Versfeld N (2005) A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *J. Acoust. Soc. Am.* 117, 2181–2192.
- Wagener KC, Brand T, Kollmeier B (2006) The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45, p. 26-33.
- Wesker, T, et al. (2005) Oldenburg logatome speech corpus (OLLO) for speech recognition experiments with humans and machines. In *Interspeech 2005, ISCA, Bonn*. p. 1273-1276.