

# Optimierung eines Modells der binauralen Sprachverständlichkeit

Rainer Beutelmann<sup>1</sup>, Thomas Brand<sup>2</sup>

Medizinische Physik, Universität Oldenburg, 26129 Oldenburg, Deutschland

<sup>1</sup>Email: rainer.beutelmann@uni-oldenburg.de <sup>2</sup>Email: thomas.brand@uni-oldenburg.de

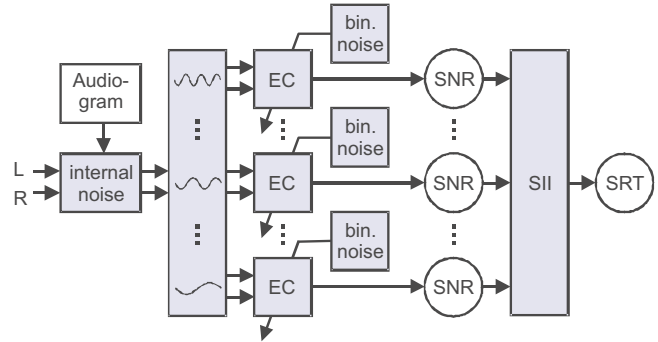
## Einleitung

Sprache als Kommunikationsmittel ist ein wichtiger Bestandteil unseres sozialen Lebens. Häufig findet sie in Umgebungen statt, die durch Störgeräusche, andere Sprecher und oft ungünstige Raumakustik ein Gespräch erschweren. Ein entscheidender Faktor in solchen Situationen ist das binaurale Hören. Durch Auswertung unterschiedlicher interauraler Zeit- und Pegeldifferenzen räumlich getrennter Nutz- oder Störschallquellen kann die Sprachverständlichkeitsschwelle (Signal-Rausch-Verhältnis, das zu einer Verständlichkeit von 50% führt) um bis zu 12 dB verbessert werden. Dieser Gewinn ist bei Schwerhörigen je nach Hörverlust stark eingeschränkt. Mit einem quantitativen Modell der binauralen Sprachverständlichkeit kann dieses Phänomen untersucht und außerdem Hilfe für die Diagnose von Hörstörungen, die Entwicklung von Hörgerätealgorithmen und das Fitting von Hörgeräten gegeben werden.

Das hier vorgestellte Modell führt die Idee von v. Hövel [4] fort und basiert auf dem Equalization-Cancellation-Prinzip von Durlach [3]. Es konnte Messdaten von Normal- und Schwerhörigen mit einer Korrelation von 0,9 vorhersagen [2]. Allerdings waren insbesondere die vorhergesagten Sprachverständlichkeitsschwellen von Messungen in reflexionsarmer Umgebung sowie die der meisten schwerhörigen Probanden noch zu niedrig. Mithilfe einer durch analytische Vorberechnung beschleunigten Version des Modells wurden in dieser Studie die Sprachverständlichkeitsschwellen von 238 schwerhörigen Patienten aus einer diagnostischen Datenbank vorhergesagt und durch eine Faktorenanalyse Zusammenhänge zwischen den Messdaten, den Hörverlusten und den Vorhersagefehlern ermittelt.

## Modell

Abb. 1 zeigt eine schematische Übersicht des Modells. Es benötigt als Eingangsgrößen die getrennten Sprach- und Störsignale sowie die Hörschwellen des linken und rechten Ohrs. Aus den Hörschwellen werden jeweils ein zusätzliches internes Rauschen (rechts und links unkorreliert) generiert, um die Hörschwellen zu simulieren und zum Störgeräusch addiert. Die Signale werden in 30 ERB-breite Frequenzbänder zwischen 140 Hz und 9 kHz aufgeteilt. In jedem Frequenzband wird, unabhängig von den anderen Frequenzbändern, mithilfe eines Equalization-Cancellation-Prozesses (EC) das bestmögliche Signal-Rausch-Verhältnis (signal-to-noise-ratio, SNR) bestimmt. Der EC-Prozess (Gleichung 1) verschiebt den rechten und linken Kanal mit der Delay-Zeit  $\tau$  gegeneinander und passt die Amplituden mit dem



**Abbildung 1:** Schematische Darstellung des Modells. Das Modell verarbeitet die Signale als getrennt in Sprache und Störgeräusch vorliegende Wellenformen (L bzw. R). EC: Equalization-Cancellation-Prozess. SNR: Signal-Rausch-Verhältnis. SII: Speech Intelligibility Index. SRT: Sprachverständlichkeitsschwelle.

Gain-Faktor  $\alpha$  so an, dass in der Differenz der Kanäle das Störgeräusch ( $N_L, N_R$ ) möglichst ausgelöscht wird, die Sprache ( $S_L, S_R$ ) jedoch erhalten bleibt.

$$\frac{S}{N} = \frac{\int |\alpha S_L(\omega)e^{i\omega\tau} - S_R(\omega)|^2 d\omega}{\int |\alpha N_L(\omega)e^{i\omega\tau} - N_R(\omega)|^2 d\omega} \quad (1)$$

Da der EC-Prozess den SNR auch verschlechtern kann, wird im Zweifelsfall der SNR des günstigeren monauralen Kanals verwendet. Die SNRs der einzelnen Frequenzbänder werden an den Speech Intelligibility Index (SII, [1]) weitergereicht, über den die Sprachverständlichkeitsschwelle (SRT) berechnet wird.

Mit den optimalen Gain- und Delay-Parametern sind die vorhergesagten Schwellen viel zu niedrig. Deshalb wurden von Durlach [3] künstliche binaurale Verarbeitungsfehler ( $\varepsilon, \delta$ ) eingefügt, d.h. statistische Abweichungen von den optimalen Parametern, die den Einfluss der EC-Parameter abschwächen. Sie sind normalverteilt und ihre Varianzen sind von den zugehörigen EC-Parametern abhängig.

$$\sigma_\varepsilon = \sigma_{\varepsilon 0} \left( 1 + \left( \frac{|\alpha|}{\alpha_0} \right)^p \right), \quad \sigma_\delta = \sigma_{\delta 0} \left( 1 + \frac{|\tau|}{\tau_0} \right) \quad (2)$$

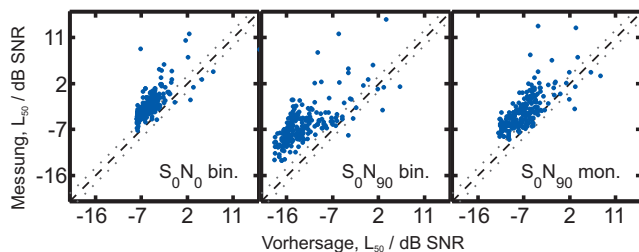
Die Parameter  $\sigma_{\varepsilon 0}, \alpha_0, p, \sigma_{\delta 0}$  und  $\tau_0$  wurden von v. Hövel [4] an Daten psychoakustische Messungen (Detektion von Sinustönen in dichotischem Rauschen) angepasst. Die Fehler werden gleichermaßen auf Sprache und Störgeräusch, aber im Gegensatz zu  $\alpha$  und  $\tau$  auf beide Kanäle aufgeprägt.

Die ersten Versionen des Modells benutzten eine Simplex-Suche über beide Parameter,  $\alpha$  und  $\tau$ , um die optimalen Parameter in jedem Frequenzband zu berechnen und er-

mittelte den Einfluss der binauralen Verarbeitungsfehler mithilfe eines Monte-Carlo-Verfahrens. Es ist jedoch möglich, Gleichung 1 weiter aufzulösen und das optimale  $\alpha$  in Abhängigkeit von  $\tau$  analytisch zu berechnen. Damit reduziert sich das Problem auf eine eindimensionale Suche.

## Ergebnisse

Mit dem Modell wurden Sprachverständlichkeitsschwellen von 238 Patienten aus einer diagnostischen Datenbank vorhergesagt. Die Sprachverständlichkeitsmessungen waren Teil eines größeren Diagnoseprogramms. Das Alter der Patienten lag zwischen 18 und 87 Jahren, die Hörverluste (Pure Tone Average PTA, über 1, 2 und 4 kHz) betragen 0–105 dB HL, im Mittel 41 dB HL am linken Ohr und 2–102 dB HL, im Mittel 40 dB HL am rechten Ohr. Es lagen sowohl symmetrische als auch asymmetrische Hörverluste mit flachem und schrägem Verlauf sowie mit Hochtonsteilabfall vor. Die Ergebnisse sind in Abb. 2 dargestellt. Die



**Abbildung 2:** Vorhergesagte und gemessene Sprachverständlichkeitsschwellen von 238 schwerhörenden Versuchspersonen. Die Störquelle befand sich in den dichotischen Situationen am schlechteren Ohr des Probanden. „bin.“: binaurale Signaldarbietung „mon.“: der Störquelle zugewandter Kopfhörerkanal wurde abgeschaltet.

Korrelationen zwischen Vorhersagen und Messwerten betragen 0,73 für die  $S_0N_{90}$ -Situationen und 0,77 für die  $S_0N_0$ -Situation. Die vorhergesagten SRTs sind in fast allen Fällen zu niedrig, d.h. es liegen Störungen der Sprachverständlichkeit vor, die nicht durch das Modell, das nur auf dem Audiogramm basiert, vorhergesagt werden können. In der binauralen  $S_0N_{90}$ -Situation sind alle vorhergesagten SRTs deutlich zu niedrig, was ein Hinweis darauf ist, dass die Parameter der binauralen Verarbeitungsfehler besser angepasst werden müssen. Da die Varianz der Vorhersagefehler auch in der monauralen Situation nicht geringer wird, obwohl das binaurale Modell hier keinen Einfluss hat, müssen Faktoren in Betracht gezogen werden, die entweder im monauralen SII zu suchen sind, oder die mit dem Modellansatz dieser Studie nicht vorhergesagt werden können.

Durch eine Faktorenanalyse über 12 ausgewählte Variablen (siehe Tab. 1) konnte gezeigt werden, dass die Vorhersagefehler bis auf eine signifikante negative Korrelation mit der Hörschwelle am besseren Ohr bei 500 Hz einen eigenen unabhängigen Faktor darstellen. Das bedeutet, dass sie nicht durch eine der anderen Variablen zu erklären sind. Die Vorhersagefehler korrelieren aber untereinander sehr stark. Da sie auch mit den

gemessenen Schwellen nicht korrelieren, liegt der Schluss nahe, dass sie durch eine „zentralere“ Hörstörung erklärt werden kann, also nicht nur durch den Hörverlust im Innenohr.

	Faktor		
	1	2	3
SRT Messg. ( $S_0N_0$ )	0,937	0,145	0,047
SRT Messg. ( $S_0N_{90}$ mon.)	0,917	0,241	-0,102
SRT Messg. ( $S_0N_{90}$ bin.)	0,910	0,140	0,138
PTA besseres Ohr	0,841	-0,234	-0,260
HL besseres Ohr 4 kHz	0,753	0,008	-0,186
HL besseres Ohr 500 Hz	0,656	-0,441	-0,108
Vorh.-fehler ( $S_0N_{90}$ mon.)	0,054	0,923	0,022
Vorh.-fehler ( $S_0N_{90}$ bin.)	-0,078	0,869	-0,294
Vorh.-fehler ( $S_0N_0$ )	0,156	0,869	0,130
Diff. PTA	-0,074	-0,003	0,933
Diff. HL 500 Hz	0,053	-0,030	0,859
Diff. HL 4 k	-0,267	-0,026	0,846

**Tabelle 1:** Ergebnisse der Faktorenanalyse. Faktorladungen zu den Variablen in der ersten Spalte. PTA: Pure Tone Average über 1, 2 und 4 kHz, HL: Hörschwelle. Differenz der Hörschwellen immer besseres minus schlechteres Ohr. 80% der Varianz konnte durch die 12 Variablen erklärt werden.

## Zusammenfassung

Das hier vorgestellte Modell der binauralen Sprachverständlichkeit, basierend auf dem Equalization-Cancellation-Prinzip ist in der Lage, Sprachverständlichkeitsschwellen in Abhängigkeit der räumlichen Anordnung von Sprache und Störgeräusch für Normal- und Schwerhörende vorherzusagen [2]. Bei der Anwendung in der klinischen Diagnostik bleibt jedoch eine signifikante Varianz, die sich offensichtlich nicht mit einem Modell erklären lässt, das nur auf der individuellen Hörschwelle basiert.

Diese Studie wurde gefördert durch das Kompetenzzentrum HörTech und das EU-Projekt HearCom.

## Literatur

- [1] ANSI (1997). „Methods for the Calculation of the Speech Intelligibility Index,“ Standards Secretariat, Acoustical Society of America, American National Standard, S3.5–1997
- [2] Beutelmann R, Brand T, Kollmeier B (2004). „Sprachverständlichkeit in räumlichen Störgeräuschsituationen,“ ZfA, Suppl. 7, auf CD
- [3] Durlach N.I. (1963). „Equalization and Cancellation Theory of Binaural Masking-Level Differences,“ J. Acoust. Soc. Am., 35, 1206–1218
- [4] vom Hövel H. (1984). „Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung,“ Fakultät für Elektrotechnik, RTWH Aachen