

Binaurale Sprachverständlichkeit mit frequenzabhängiger interauraler Phasendifferenz (IPD)

Rainer Beutelmann, Thomas Brand

Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, Deutschland
Email: rainer.beutelmann@uni-oldenburg.de, thomas.brand@uni-oldenburg.de

Einleitung

Binaurales Hören kann in vielen Situationen zu einer Verbesserung der Sprachverständlichkeitsschwelle um bis zu 12 dB gegenüber dem ungünstigsten monauralen Fall führen [3]. Deshalb ist es sinnvoll, die bereits existierenden monauralen Sprachverständlichkeitsmodelle [4,5] so zu erweitern, dass die Sprachverständlichkeit auch in Situationen mit binauralem Gewinn korrekt vorhergesagt werden kann. Die Modelle teilen die zu bewertenden Signale in mehrere Frequenzbänder auf. In (schmalbandigen) binauralen Detektionsexperimenten mit Sinustönen wurden allerdings zum Teil größere kritische Bandbreiten als für vergleichbare monaurale Situationen gefunden [6]. Diese Studie leistet einen Beitrag zur Beantwortung der Frage, welche Bandbreiten für die (breitbandigen) Signale eines binauralen Sprachverständlichkeitsmodells anzuwenden sind. Zu diesem Zweck wurde der binaurale Sprachverständlichkeitsgewinn mit interauralen Phasendifferenzen (IPDs) gemessen, die sich in benachbarten Frequenzbändern verschiedener Breite stark unterscheiden, und die Abhängigkeit des binauralen Gewinns von der Breite der Frequenzbänder mit den Vorhersagen des Modells verglichen.

Methoden

Die IPDs wurden durch ein FIR-Filter generiert, das gleichzeitig eine Bandbegrenzung der Signale auf den Bereich zwischen $f_l = 250$ Hz und $f_h = 4$ kHz vornahm. Die IPD-

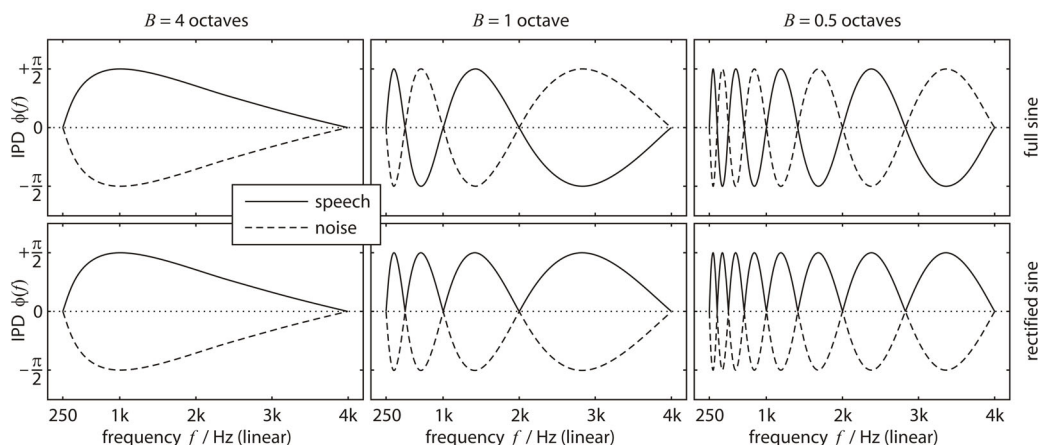
Funktion $\phi(f)$ war auf einer logarithmischen Frequenzachse sinusförmig moduliert, so dass der Bereich zwischen f_l und f_h immer durch eine ganze Anzahl Perioden aufgefüllt wurde. Der Parameter B gibt den Abstand zwischen zwei Nullstellen von

$$\phi(f) = \phi_0 \sin \left[4\pi \left(B \log \frac{f_l}{f_h} \right)^{-1} \log \frac{f}{f_l} \right] \quad (1)$$

in Oktaven an und wurde zwischen 0.125 und 4 variiert. Der Parameter ϕ_0 betrug immer $+\pi/2$ für das Sprachsignal und $-\pi/2$ für das Störsignal. Abb. 1 zeigt den Verlauf der IPD als Funktion der Frequenz beispielhaft für drei Werte von B . Für jeden Wert von B wurden insgesamt sechs Schwellen gemessen, jeweils drei mit voll durchmodulierter Sinusfunktion ("full sine") und drei mit dem Absolutwert der Sinusfunktion ("rectified sine"). Für beide IPD-Funktionen wurde jeweils eine Schwelle nur mit dem linken Ohr gemessen (monaural), um den Einfluss der rein monauralen Phasenverschiebungen abzuschätzen, eine Schwelle wurde binaural, aber mit gleichen IPDs für Sprache und Störgeräusch (kein Gewinn zu erwarten) als Referenz gemessen. Die dritte Schwelle enthielt die eigentliche binaurale Information mit entgegengesetzten IPD-Vorzeichen von Sprache und Störgeräusch.

Die Sprachverständlichkeitsmessungen wurden mit dem Oldenburger Satztest [7] durchgeführt. Die Sprachverständlichkeitsschwelle, das heißt das Signal-Rausch-Verhältnis,

Abbildung 1: Schematische Darstellung der interauralen Phasendifferenzen als Funktion der Frequenz, $\phi(f)$, beispielhaft für drei Halbperioden-Bandbreiten B . Durchgezogene Linien stellen die IPD des Sprachsignals dar, gestrichelte Linien die IPD des Störsignals. Die oberen drei Felder zeigen die IPD-Funktionen in den "full sine"-Bedingungen, in der sich positive und negative IPD-Vorzeichen abwechseln. Die unteren drei Felder zeigen die IPD-Funktionen in den "rectified sine"-Bedingungen, in denen Sprachsignale immer eine positive IPD und Störsignale immer eine negative IPD haben. Alle Signale wurden zwischen 250 Hz und 4 kHz bandpassgefiltert, IPDs außerhalb dieser Grenzen (wegen der endlichen Flankensteilheit des Bandpassfilters) waren immer Null.



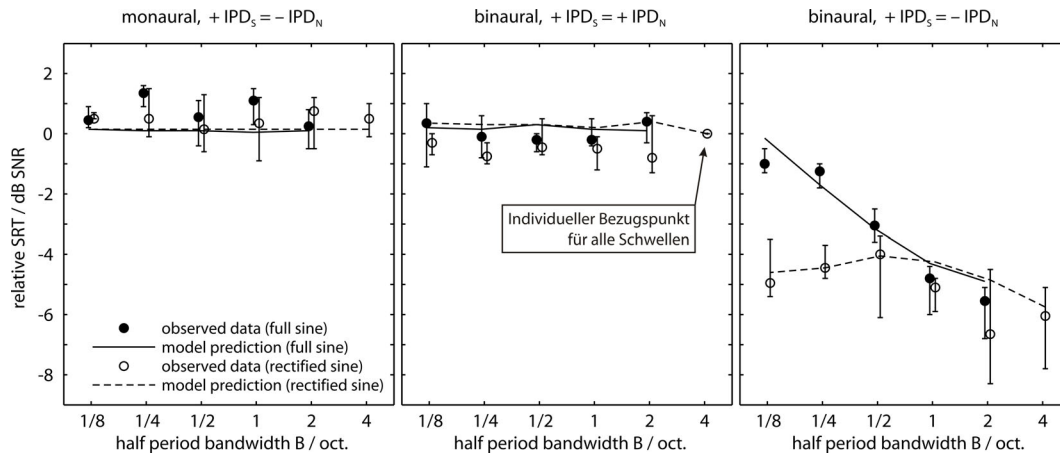


Abbildung 2: Gemessene Sprachverständlichkeitsschwellen (speech reception thresholds, SRT) von sechs normalhörenden Probanden (Kreise, Median mit oberen und unteren Quartilen) und Modellvorhersagen (Linien). Alle Schwellen sind relativ zur individuellen Schwelle bei $B=4$ Oktaven im mittleren Feld ($IPD_S=IPD_N$). Gefüllte Symbole und durchgezogene Linien repräsentieren die “full sine”-Bedingungen, offene Symbole und gestrichelte Linien stehen für die “rectified sine”-Bedingungen. Das linke Feld zeigt die Schwellen für monaurale Bedingungen, das mittlere Feld für die binauralen Bedingungen mit gleicher IPD für Sprache und Störgeräusch, und das rechte Feld für die binauralen Bedingungen mit entgegengesetzten IPD-Vorzeichen von Sprache und Störgeräusch.

bei dem der Proband im Mittel 50% der dargebotenen Wörter versteht, wurde adaptiv ermittelt, indem in Abhängigkeit von den Antworten des Probanden der Sprachpegel für den darauffolgenden Satz gesenkt oder angehoben wurde. Jede Schwelle wurde mit einer Liste von 20 Sätzen gemessen. Als Störgeräusch wurde aus dem Sprachmaterial des Tests generiertes, stationäres sprachsimulierendes Rauschen bei einem Pegel von 65 dB SPL verwendet. An der Studie nahmen sechs normalhörende Probanden im Alter von 24 bis 32 Jahren teil.

Das binaurale Sprachverständlichkeitsmodell [1] teilt die binauralen Eingangssignale, Sprache und Störgeräusch, mit einer Gammaton-Filterbank [8] in 30 Frequenzbänder mit jeweils 1 ERB [2] Breite auf, berechnet in jedem Frequenzband mit dem Equalization-Cancellation-Modell [9] die durch binaurale Verarbeitung mögliche Verbesserung des Signal-Rausch-Verhältnisses und daraus mit Hilfe des Speech Intelligibility Index [4] die Sprachverständlichkeit.

Ergebnisse

Messdaten und Modellvorhersage sind in Abb. 2 dargestellt. Der binaurale Gewinn ist in der binauralen “full-sine”-Bedingung mit entgegengesetztem IPD-Vorzeichen (rechtes Feld) stark von der IPD-Bandbreite B abhängig und für $B < 1/3$ Oktave nicht signifikant von Null verschieden. Für die “rectified sine”-Bedingungen ist diese Abhängigkeit wesentlich geringer und es ist sogar bei sehr kleinen Bandbreiten ein binauraler Gewinn möglich. Ein monauraler Effekt kann ausgeschlossen werden, da die Schwellen in den monauralen Bedingungen nicht signifikant voneinander abweichen.

Der Korrelationskoeffizient zwischen den Modellvorhersagen und dem Median der Messdaten über die Versuchspersonen liegt bei 0.97, der root-mean-square-Fehler bei 0.9 dB.

Referenzen

- 1 Beutelmann, R. und Brand, T. (2006): “Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners”, *J. Acoust. Soc. Am.*, 120, 331-342
- 2 Glasberg, B.R. und Moore, B.C.J. (1990): “Derivation of auditory filter shapes from notched noise data”, *Hear. Res.* 47, 103-138
- 3 Bronkhorst, A. W. (2000): “The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple Talker Conditions”, *ACUSTICA - acta acoustica*, 86, 117-128
- 4 ANSI (1997): “Methods for the Calculation of the Speech Intelligibility Index”, Standards Secretariat, Acoustical Society of America, S3.5-1997
- 5 IEC (1998): “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index”, International Standard, IEC 60268-16
- 6 Hall, J.W., Tyler, R.S. und Fernandes, M.A. (1983): “Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking”, *J. Acoust. Soc. Am.*, 73, 894-898
- 7 Wagener, K., Brand, T., Kühnel, V. und Kollmeier, B. (1999): “Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests”, *Zeitschrift für Audiologie*, 38, 4-14
- 8 Hohmann, V. (2002): “Frequency analysis and synthesis using a Gammatone filterbank”, *ACUSTICA - acta acoustica*, 88, 433-442
- 9 Durlach, N.I. (1963): “Equalization and Cancellation Theory of Binaural Masking-Level Differences”, *J. Acoust. Soc. Am.*, 35, 1206-1218