

Modellierung binauraler Sprachverständlichkeit in verhallter Umgebung

Jan Rennies¹, Thomas Brand², Birger Kollmeier^{1,2}

¹ Fraunhofer IDMT, Hör-, Sprach- und Audiotechnologie, Oldenburg, Email: jan.rennies@idmt.fraunhofer.de

² Carl-von-Ossietzky Universität Oldenburg, Medizinische Physik

Einleitung

In räumlichen Situationen sind binaurale Sprachverständlichkeitsschwellen (speech reception thresholds, SRTs) oftmals niedriger, wenn Sprache und Störgeräusch aus unterschiedlichen Richtungen angeboten werden. In verhallten Umgebungen ist dieser Effekt häufig verringert, da Reflektionen und Nachhall das Störgeräusch diffuser erscheinen lassen und somit eine räumliche Trennung von Stör- und Nutzschall erschwert wird. Ein auf [1] aufbauendes Modell konnte den binauralen Gewinn in verhallten Umgebungen vorhersagen [2, 3]. In stark verhallten Umgebungen überschätzt das Modell allerdings die Sprachverständlichkeit, da es den schädlichen Einfluss des Nachhalls auf die Sprache selbst nicht berücksichtigt (z.B. [4]). Diese Studie stellt eine Erweiterung des Modells vor, die diesen Nachteil überwindet, indem zusätzlich Informationen über die Modulationsübertragungsfunktion (modulation transfer function, MTF) genutzt werden. Als Grundlage zur Verifikation der Modellerweiterung dient ein Experiment, in dem der Abstand von Schallquellen und Empfänger variiert wurde, um dadurch den Einfluss von Nachhall auf binaurale Sprachverständlichkeit systematisch zu untersuchen.

Experiment

Mit Hilfe des Oldenburger Satztests [5] wurden SRTs in unterschiedlichen binauralen Konditionen gemessen, die durch Faltung der Stimuli (Sätze und sprachsimulierendes Rauschen) mit simulierten Impulsantworten virtuell erzeugt wurden. In einem simulierten Raum der Dimensionen 10m x 15m x 3m wurde Sprache stets von vorne (0°) und die Störgeräuschquelle aus drei verschiedenen Richtungen dargeboten (0 , 45 und 90°). Die Nachhallzeit des Raumes betrug etwa 1,5s. Zusätzlich wurden acht verschiedene Abstände zwischen den Signalquellen und dem Hörer simuliert. Abbildung 1 zeigt beispielhaft die feste Position des Sprechers (x) sowie die acht Positionen der Hörer (Kreise auf der Mittelachse) und der Störgeräuschquellen (Kreise in der linken Raumlälfte) für die S_0N_0 -Bedingung (Sprache und Störgeräusch von vorne). Sechs normal hörende Versuchspersonen nahmen am Experiment teil. Im linken Teilbild von Abbildung 2 sind die mittleren Daten und Standardfehler dargestellt. Es ist zu beobachten, dass die unterschiedlichen Konditionen die SRTs beeinflussen. Der binaurale Gewinn, d.h. die Differenz der Schwellen in S_0N_0 -Bedingung und S_0N_{45} - bzw. S_0N_{90} -Bedingung

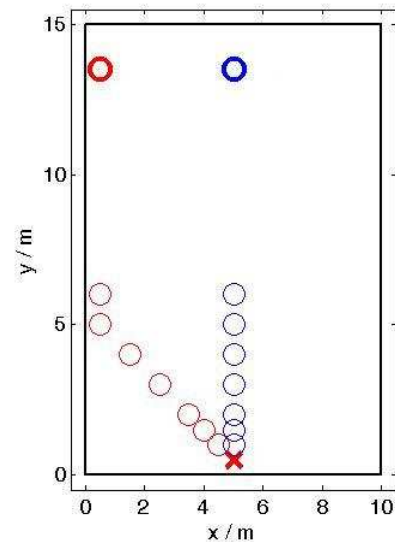


Abbildung 1: Simulierter Raum zur Messung und Vorhersage binauraler SRTs. Das X kennzeichnet die Position des Sprechers für alle Konditionen. Kreise auf der Mittelachse zeigen die acht Positionen des Hörers. Kreise in der linken Raumlälfte veranschaulichen die Positionen der Störerschallquellen für jeden Hörer in der S_0N_{90} -Kondition.

ist für geringe Abstände zwischen Quellen und Hörer deutlich ausgeprägt. Der binaurale Gewinn verschwindet vollständig bei größeren Abständen. Für alle Konditionen ist der binaurale Gewinn für Störgeräuschrichtungen von 45° und 90° sehr ähnlich. Zusätzlich fällt auf, dass auch in den diotischen S_0N_0 -Bedingungen die SRTs mit dem Abstand ansteigen.

Modellierung

Im in [2, 3] beschriebenen Modell erfolgt eine frequenzbandweise binaurale Vorverarbeitung von Sprach- und Störsignalen nach dem Equalization-Cancellation (EC) Prinzip [6]. Wenn Sprach- und Störgeräusch aus unterschiedlichen Richtungen eintreffen, führt die EC-Stufe durch Ausnutzen interauraler Phasen- und Pegeldifferenzen zu einem gegenüber diotischen Situationen erhöhten SNR. Nach der EC-Stufe wird die Sprachverständlichkeit mit Hilfe des Sprachverständlichkeitsindex (SII) berechnet. Der SII basiert dabei im Wesentlichen auf einem gewichteten SNR in den unterschiedlichen Frequenzbändern. In dieser Studie wird die SII-Stufe des Modells erweitert, indem zusätzlich Information über die MTF ausgenutzt werden. Dieses Prinzip wurde bereits

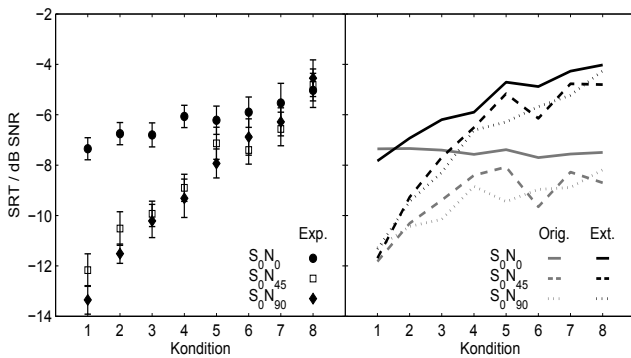


Abbildung 2: Messergebnisse der binauralen SRTs (links) und Modellvorhersagen (rechts) mit dem Originalmodell (grau) und dem erweiterten Modell dieser Studie (schwarz) in Abhängigkeit des Abstandes zwischen Sprecher und Hörer.

in vielen Studien erfolgreich eingesetzt, um den negativen Einfluss von Nachhall auf Sprachverständlichkeit zu beschreiben (z.B. [7]) und ist auch die Grundlage zur Bewertung der Sprachverständlichkeit mit dem Sprachübertragungsindex (STI). Die Information der MTF wird dem Modell in Form der Raumimpulsantwort, die die Übertragung der Sprache von der Quelle zum Hörer beschreibt, bereitgestellt. Diese Form der Erweiterung des SII ist als Zusatz auch in der ANSI-Norm [8] beschrieben.

Das rechte Teilbild von Abb. 2 zeigt die Simulationsergebnisse für die im Experiment verwendeten binauralen Konditionen. Die grauen Linien stellen die Vorhersagen des Originalmodells nach [3] dar, die schwarzen Linien zeigen Vorhersagen des erweiterten Modells dieser Studie. Der Vergleich der Modellvorhersagen zeigt, dass beide Modelle die Abnahme des binauralen Gewinns mit steigendem Abstand nachbilden können. Für alle S_0N_0 -Situationen sagt das Originalmodell die gleiche Schwelle vorher und kann somit den im Experiment gefundenen Anstieg nicht vorhersagen. Die in dieser Studie vorgestellte Erweiterung des Modells sagt auch diesen Effekt vorher. Ein Vergleich mit den experimentellen Daten zeigt, dass beide Modelle insgesamt den binauralen Gewinn unterschätzen. Zudem überschätzt das erweiterte Modell den schädlichen Einfluss des Nachhalls auf die Sprachverständlichkeit: der Unterschied zwischen den S_0N_0 -Schwellen in Kondition 1 und 8 beträgt etwa 2,5 dB, während das erweiterte Modell etwa 4 dB vorhersagt.

Diskussion

Die Daten dieser Studie zeigen, dass die Sprachverständlichkeit in verhallten Räumen vom Abstand zwischen Sprecher und Hörer abhängt. Wie erwartet verringert sich der binaurale Gewinn, wenn bei größeren Abständen das Verhältnis zwischen Direktschall- und Nachhallenergie abnimmt. Dies verringert die Lokalisierbarkeit der Störgeräuschquelle, so dass der Cocktail-Party-Effekt reduziert ist. Auch das Sprachsignal selbst wird vom Nachhall beeinflusst. Die SRTs in S_0N_0 -Bedingungen steigen mit zunehmendem

Einfluss des Nachhalls um etwa 2,5 dB an. Das Originalmodell sagt diesen Einfluss nicht vorher, da es die Standardimplementierung des SII verwendet. Dieser berechnet die Sprachverständlichkeit auf Grundlage der spektralen Eigenschaften von Sprach- und Störsignal. Wird zusätzlich der Effekt zeitlicher Modulationen im Modell berücksichtigt, ist es möglich auch den negativen Einfluss des Nachhalls auf die Sprache selbst vorherzusagen, wie er auch durch mehrere raumakustische Maße, wie z.B. C_{50} , D_{80} oder die Direct-to-reverberant Ratio, vorhergesagt wird. In der derzeitigen Implementation des Modells wird der Einfluss des Nachhalls allerdings etwas überschätzt. Ein mögliche Ursache dafür könnte sein, dass der im STI (bzw. im erweiterten SII) verwendete Ansatz der MTF jede Form von Nachhall und Echo als schädlich bewertet. Aus der Raumakustik bekannt ist jedoch der Effekt, dass frühe Reflektionen positiven Einfluss auf die Sprachverständlichkeit haben können, da sie effektiv den Sprachpegel erhöhen (vgl. C_{50} , D_{80} , etc.). Da dieser Effekt in der MTF nicht berücksichtigt wird, ist es möglich, dass die Sprachverständlichkeit unterschätzt wird. Dieser Effekt soll in Zukunft durch Messungen von SRTs in Ruhe weiter untersucht werden. Es ist zu erwarten, dass das erweiterte Modell im Einklang mit den raumakustischen Maßen ebenfalls steigende Schwellen für zunehmend verhallte Situationen vorhersagt.

Literatur

- [1] vom Hövel H. (1984). Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung, Fakultät für Elektrotechnik, RWTH Aachen
- [2] 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
- [3] Beutelmann, R., Brand, T. und Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model, *J.Acoust.Soc.Am*, *in press*
- [4] Duquesnoy, A.J. und Plomp, R. (1980). Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis, *J.Acoust.Soc.Am*, 62, 537-544
- [5] Wagener, K., Brand, T. und Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldeburger Satztests, *Zeitschrift für Audiologie*, 38, 86-95.
- [6] Durlach, N.I. (1963). Equalization and Cancellation Theory of Binaural Masking-Level Differences, *J.Acoust.Soc.Am.*, 35, 1206-1218
- [7] Houtgast, T. and Steeneken, H.J.M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, *J.Acoust.Soc.Am*, 77, 1069-1077
- [8] ANSI (1997). Methods for the Calculation of the Speech Intelligibility Index, Standards Secretariat, Acoustical Society of America, American National Standard, S3.5-1997