

Modellierung des Einflusses von Nachhall auf binaurale Sprachverständlichkeit in Ruhe und im Rauschen

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 vielen praktischen Anwendungen ist eine Vorhersage von Sprachverständlichkeit nützlich, so etwa in der Planung und Gestaltung von öffentlichen Räumen und Arbeitsplätzen oder bei der Entwicklung von Sprachsignalverarbeitungsalgorithmen. Dabei gilt es den Einfluss möglichst aller relevanten Schalleigenschaften wie Störgeräuschanteil oder Nachhall akkurat abzubilden. Kürzlich vorgestellte Sprachverständlichkeitsmodelle berücksichtigen darüber hinaus wichtige Faktoren wie binaurales Hören [1, 2, 3] oder die individuelle Hörschwelle und Maskiererfluktuationen [3]. In [4] wurde ein Erweiterungsansatz des Modells von Beutemann *et al.* [3] vorgestellt, der zudem den Einfluss von Nachhall auf das Sprachsignal berücksichtigen konnte und damit zu verbesserten Vorhersagen in stark verhalten Umgebungen führte. Dieser Beitrag stellt die weiter geführten Modellarbeiten vor. Der prinzipielle Ansatz besteht darin, in der Raumakustik etablierte Kenngrößen in das bestehende Modell zu integrieren. Die neuen Modellansätze werden anhand neu erhobener subjektiver Messdaten für Sprachverständlichkeit in Ruhe und im Störgeräusch evaluiert.

Experiment

An den Experimenten nahmen acht normalhörende Versuchspersonen teil, deren Reintonaudiogramme 15 dB HL nicht überschritten. Es wurden 50%-Sprachverständlichkeitsschwellen (speech reception thresholds, SRTs) mit der adaptiven Methode des Oldenburger Satztests [5] ermittelt, d.h. derjenige SNR, bei dem die Probanden die Hälfte der Wörter verstanden. Bei Messungen in Ruhe ist die SRT definiert als der Schalldruckpegel bei 50% Verständlichkeit. Der Störgeräuschpegel wurde stets auf 65 dB SPL (am rechten Ohr) gesetzt und der SNR durch Anpassung des Sprachpegels variiert. Die untersuchten akustischen Bedingungen wurden durch Faltung der Sprachsignale und Störgeräusche mit binauralen Raumimpulsantworten erzeugt, welche mit der CATT Acoustic Software simuliert wurden. Die Signalwiedergabe erfolgte über Sennheiser HDA200 Kopfhörer. Es wurde ein rechteckiger Raum simuliert (10m x 15m x 3m), auf dessen mittlerer Achse der Sprecher und der Zuhörer in unterschiedlichen Abständen platziert wurden (siehe Abb. 1). Die Absorptionskoeffizienten wurden für alle Raumflächen und Frequenzen auf 5% gesetzt, was zu einer

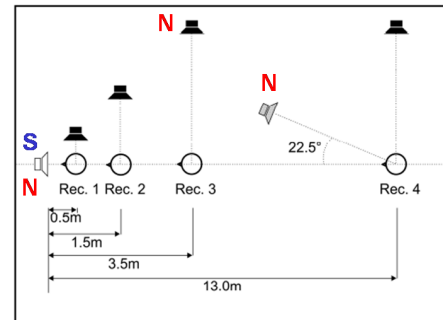


Abbildung 1: Simulierter Raum zur Messung und Vorhersage binauraler SRTs.

durchschnittlichen Nachhallzeit von etwa 2 s führte. Bei den Messungen mit Störgeräusch wurde die Sprache stets von vorne präsentiert (0°), das Störgeräusch aus drei verschiedenen Richtungen (0° , $22,5^\circ$ und 90°). Bei Messungen in Ruhe wurden die Sprachrichtungen 0° und 90° betrachtet. Die mittleren SRT-Werte sind als Symbole in Abb. 2 und 3 dargestellt. In Anwesenheit eines Störgeräusches zeigen sich zwei Haupteffekte. Einerseits verringert sich der binaurale Gewinn, also das Absinken der SRTs durch räumliche Trennung von Sprache und Störgeräusch mit zunehmendem Abstand zwischen Sprecher und Hörer. Dieser Effekt ist wesentlich durch die erhöhte Schwierigkeit zu begründen das stark verhaltene Störgeräusch von der Sprache zu trennen. Andererseits steigen die SRTs insgesamt mit Abstand an, auch in diotischen Hörbedingungen. Dies ist darauf zurückzuführen, dass der Nachhall das Sprachsignal selbst beeinträchtigt und dessen zeitliche Struktur verschmiert. Dieser Effekt wird umso größer, je geringer das Verhältnis zwischen Direkt- und Nachhallenergie ist (d.h. je größer der Abstand). Bei Messungen in Ruhe zeigt sich dieser zweite Effekt ebenfalls. Die zwei betrachteten Sprachrichtungen unterscheiden sich nur bei Position 4, was dadurch begründet ist, dass der Abstand für die 90° -Richtung geringer war als für 0° (siehe Abb. 1). Es stellt sich heraus, dass die individuellen Hörschwellen einen signifikanten Einfluss auf die Sprachverständlichkeit haben, was im Rauschen nicht beobachtet wurde. Dies gilt selbst dann, wenn alle Probanden klinisch normalhörig sind.

Modellierung

Das Modell von Beutemann *et al.* [3] basiert auf einer auditorischen Filterbank, gefolgt von einem equalization-cancellation (EC)-Mechanismus zur Modellierung

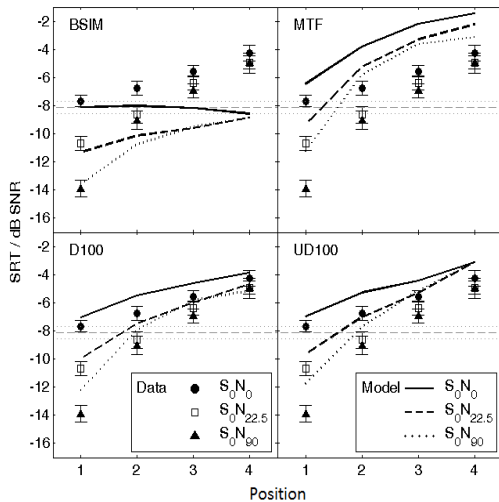


Abbildung 2: Messergebnisse (Symbole) und Modellvorhersagen (Linien) für binaurale SRTs in Rauschen in Abhängigkeit des Abstandes zwischen Sprecher und Hörer. Graue Linien kennzeichnen die Referenzbedingung des original Satztests (diotisch, kein Nachhall).

der binauralen Verarbeitung und schließlich dem Sprachverständlichkeitsindex (speech intelligibility index, SII). Individuelle Hörschwellen werden durch ein internes Rauschen modelliert. In [4] wurde eine Erweiterung des Modells vorgestellt, die den SII durch die Modulationsübertragungsfunktion so verändert, dass dieser Reduktionen von zeitlichen Modulationen ähnlich wie beim Sprachübertragungsindex (speech transmission index, STI) als verständlichkeitsmindernd bewertet (MTF). In dieser Studie wurde zwei weitere raumakustische Kenngrößen in das Modell integriert, nämlich zum einen das Maß der Deutlichkeit, das dem Verhältnis der Energie der frühen Reflektionen zur Gesamtenergie beschreibt (D100). Dieses Verhältnis wurde als Korrekturfaktor des SNR nach der binauralen Verarbeitungsstufe implementiert. Zum anderen wurde das Verhalten des binauralen Modells getestet, wenn die Eingangssprachsignale in einen frühen (nützlichen, engl. useful) und späten (schädlichen, engl. detrimental) Anteil aufgespalten werden und der späte Anteil dem Störgeräusch aufaddiert wird (UD100). Für beide Fälle wurden die ersten 100 ms als frühe Anteile verwendet. Die Modellvorhersagen sind den experimentellen Daten in Abb. 2 und 3 gegenüber gestellt. Alle drei Erweiterungsansätze zeigen höhere Korrelationen und durchschnittlich geringere Offsets als das Originalmodell (BSIM). Von den drei Modellerweiterungen stimmen die Vorhersagen basierend auf dem Deutlichkeitsmaß und der Trennung von frühen und späten Sprachanteilen am besten mit den Messdaten überein. Bei Messungen in Ruhe zeigt sich das gleiche Bild. Um eine hohe Korrelation basierend auf den individuellen Messdaten zu erzielen, ist es erforderlich, dass das Modell die individuellen Audiogramme berücksichtigt, d.h. auch das Modell reagiert bei Bedingungen in Ruhe sensitiv auf Ruhehörschwellen, was im Rauschen nicht der Fall war.

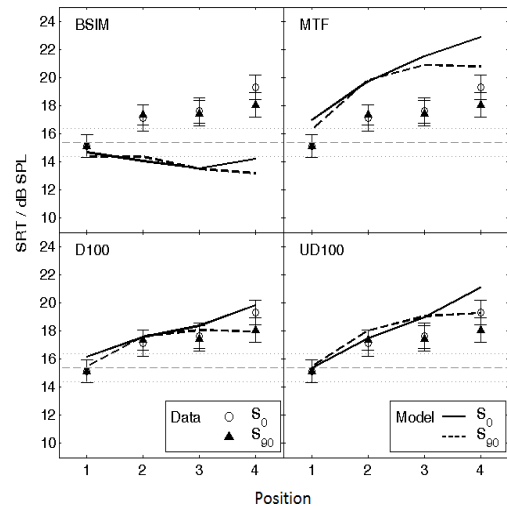


Abbildung 3: Darstellung wie in Abb. 2 für SRTs in Ruhe.

Diskussion

Das binaurale Sprachverständlichkeitsmodell in [3] sagt den beobachteten reduzierten binauralen Gewinn korrekt vorher, überschätzt jedoch die Sprachverständlichkeit in stark verhallten Bedingungen. Dies liegt daran, dass der verwendete SII auch die späten Reflektionen und den diffusen Nachhall des Sprachsignals als Nutzsignal ansieht. Korrigiert man diese Annahme mit aus der Raumakustik bekannten Maßen, verbessern sich die Vorhersagen in verhallten Umgebungen signifikant, während sie kaum verhallten Situationen mit denen des Originalmodells übereinstimmen. Liegen also Informationen über die Raumakustik vor (z.B. in Form einer gemessenen oder simulierten Impulsantwort), so können mit den Erweiterungen auch in stark verhallten Räumen genaue Vorhersagen getroffen werden. Dies hat beispielsweise praktische Relevanz für die Planung von großen öffentlichen Plätzen oder Räumen (Bahnhöfe etc.).

Literatur

- [1] van Wijngaarden, S.J. und Drullmann, R. (2008). Binaural intelligibility prediction based on the speech transmission index. *J. Acoust. Soc. Am.* 123, 4514-23.
- [2] Lavandier, M. und Culling, J.F. (2010). Prediction of binaural speech intelligibility against noise in rooms. *J. Acoust. Soc. Am.* 127, 387-399.
- [3] Beutelmann, R., Brand, T. und Kollmeier, B. (2010). Revision, extension, and evaluation of a binaural speech intelligibility model. *J. Acoust. Soc. Am.* 127, 2479-2497.
- [4] Rannies, J., Brand, T. und Kollmeier, B. (2010). Modellierung binauraler Sprachverständlichkeit in verhallter Umgebung. In *Fortschritte der Akustik - DAGA 2010*, 989-990, DEGA e.V., Berlin, 2010.
- [5] Wagener, K., Brand, T. und Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldeburger Satztests. *Z. Audiol.* 38, 86-95.