Vergleich verschiedener Methoden zur Vorhersage der Sprachverständlichkeit in fluktuierendem Störgeräusch

Ralf Meyer, Thomas Brand, Birger Kollmeier

Carl-von-Ossietzky Universität Oldenburg, Medizinische Physik

Mit Hilfe einer im Kompetenzzentrum „Hörtech“ erhobenen Datenbasis von Sprachverständlichkeitsdaten in verschiedenen Hörsituationen wurden vier unterschiedliche Modelle zur Sprachverständlichkeitsvorhersage im modulierten Störgeräusch verglichen. Drei der Modelle basieren auf dem „Speech Intelligibility Index“ (SII), das vierte Modell (Wagener 2003) basiert auf einem Modell von Plomp (1979). Der Vergleich wurde für eine Hörsituation mit fluktuierendem Störgeräusch durchgeführt, dabei wurde ein Störsprecher mit dem ICRA5-250 Rauschen simuliert.

Einleitung

Fluktuierende Störgeräusche treten in verschiedenen Situationen im Alltag auf, z. B. bei einem Gespräch in einer Personengruppe. Diese Situationen werden häufig mit dem Begriff „Cocktail-Party-Effekt“ beschrieben. An der Vorhersage der Sprachverständlichkeit in solchen Situationen gibt es verschiedene Interessen. Zum einen gibt es ein wissenschaftliches Interesse an der Verarbeitung der Sprache allgemein. Zum anderen ist die Diagnostik interessiert an einer sinnvollen Kategorisierung der Ergebnisse von Sprachtests in auffällige und unaufällige Befunde. Hinzu kommt, dass Messungen in fluktuierenden Störgeräuschen besser zwischen verschiedenen Hörverlusten differenzieren (z.B. Wagener 2005a). Bisher ist allerdings kein Modell bekannt, welches die Sprachverständlichkeit in diesen Situationen zufriedenstellend beschreibt, und dazu auch Schwerhörigkeit angemessen berücksichtigt.

Modelle

Die im folgenden vorgestellten Modelle dienen dazu, die Schwelle im Störgeräusch zu berechnen, bei der fünfzig Prozent der Sprache verstanden wird. Diese Schwelle wird auch als L50 oder SRT („Speech Reception Threshold“) bezeichnet. Die vier Modelle lassen sich in zwei Gruppen einteilen. Die erste Gruppe, besteht aus drei SII-Modellen, die zweite Gruppe besteht aus dem Modell von Wagener.

1. **Der Speech Intelligibility Index (SII)** (ANSI S3.5-1997)

   ist ein ANSI-Standard und wurde aus dem Articulation Index (AI) entwickelt. Er ist ein Maß für den Anteil der Sprachinformation, die dem Hörer zugänglich ist. Die Berechnung basiert auf dem Ton-Audiogramm und den Langzeitspektren der Sprache und des Störgeräusches. Aus diesen Eingangsgrößen wird dann in verschiedenen Frequenzbändern eine Verständlichkeit berechnet. Der SII wird dann aus der Summe der Verständlichkeiten über alle Frequenzbänder berechnet, wobei jedes Frequenzband entsprechend einer von der Art des Sprachmaterials abhängigen Gewichtungsfunktion berücksichtigt wird. Da der SII mit den Langzeitspektren der Signale arbeitet, ist er unempfindlich gegenüber zeitlichen Fluktuationen der Signale.

2. **Das Kurzzeit-SII Modell nach Brand et. al (2002)**

   ist eine einfache Erweiterung des SII für fluktuierende Störgeräusche. Das Modell berücksichtigt dabei die Modulationen des Störgeräusches, allerdings nur breitbandig, d. h. die Frequenzabhängigkeit der Modulationen wird vernachlässigt.

3. **Das Kurzzeit-SII Modell nach Rhebergen et. al (2005)**

   berücksichtigt zusätzich auch die Frequenzabhängigkeit der Modulationen des Störgeräusches. Die Frequenzabhängigkeit der Modulationen der Sprache wird jedoch wie bei Brand nicht berücksichtigt.

4. **Das Kurzzeit-Plomp Modell nach Wagener (2003)**

   basiert nicht auf der Hörschwelle und den Signalspektren, sondern auf je einer SRT Messung in Ruhe und im nicht moduliertem Störgeräusch. Außerdem berücksichtigt es Kontexteffekte, indem es eine effektive Wortlänge modelliert. Diese effektive Wortlänge wurde für jede Versuchsperson angepasst, bei Versuchspersonen mit Hörverlusten war die effektive Wortlänge größer als bei Normalhörenden.

Während bei Wagener der SRT direkt berechnet wird, wird bei den drei SII-basierten Modellen der SRT erst durch eine variable Transformation aus den ermittelten SII-Werten der SRT bestimmt. Bei den beiden Kurzzeit-SII Modellen wird der SII für modulierte Störgeräusche angepasst, indem in kurzen Zeitfenstern ein SII bestimmt wird, wobei diese Zeitfenster bei Rhebergen et. al frequenzabhängig sind. Am Ende wird dann über alle Kurzzeit-SII-Werte gemittelt. Kontexteffekte werden in den SII-Modellen nicht berücksichtigt. Durch die Modellierung von Kontexteffekten in dem Modell von Wagener werden zum Teil auch kognitive Leistungen in das Modell eingebaut, die in den SII-Modellen so nicht berücksichtigt werden.

Ergebnisse

Alle vier Modelle wurden auf die vorhandene Hör-Tech-Datenbasis angewendet (Brand 2002). Der Datensatz enthält unter anderem Ergebnisse zweier verschiedener Sprachtests (Göttinger Satztest (GÖSA) und Oldenburger Satztest (OLSA)) jeweils in Ruhe, in unmoduliertem Störgeräusch und mit dem ICRA5-250 Störgeräusch (Simulation von einem mänlichen „Stör-Sprecher“ mit einer maximalen Sprachpausenlänge von 250ms). Insgesamt sind Daten für ca. 200 Ohren (ca. 100 Versuchspersonen) vorhanden, sowohl von Normalhörenden, als auch von Versuchspersonen mit unterschiedlichen Hörverlusten. Mit den verschiedenen Modellen wurde versucht den SRT für das ICRA5-250 Störgeräusch vorher zusagen.

Die Darstellung der Ergebnisse in den folgenden Abbildungen ist identisch, auf der Abzisse ist jeweils der vorhergesagte SRT und auf der Ordinate der beobachtete SRT aus dem Datensatz dargestellt. Zusätzlich ist jeweils die Diagonale eingezeichnet, welche einer perfekten Vorhersage der Beobachtung entspräche.

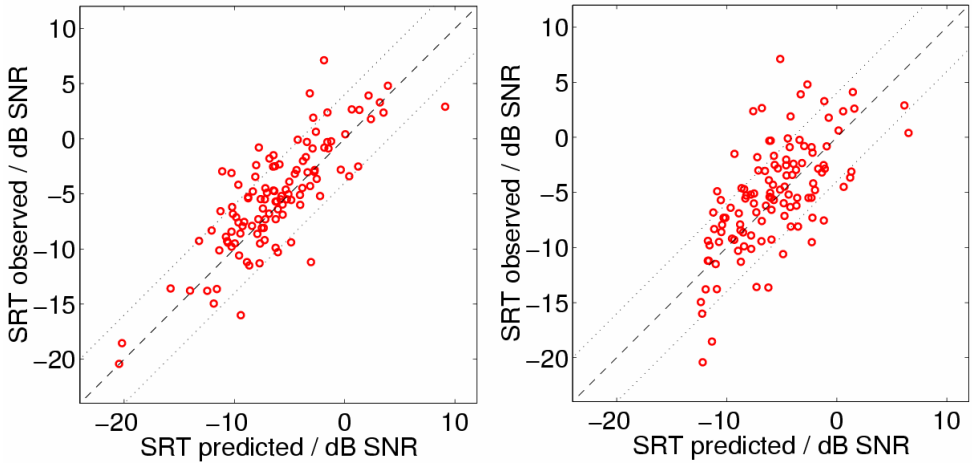

Abbildung 1: Vorhergesagter SRT (Abzisse) über SRT aus der Datenbasis (Ordinate), ICRA5-250 Störgeräusch, OLSA, Einheit: dB SNR. ***Links:*** *Ergebnisse für den SII nach einer lin. Transformation. Korrelationskoeffizient r=0.72.* ***Rechts:*** *Ergebnisse für den Kurzzeit-SII nach Brand et. al, Korrelationskoeffizient r=0.69*

In Abbildung 1 auf der linken Seite ist der Zusammenhang zwischen den Vorhersagen durch den Standard-SII nach einer linearen Transformation und den beobachteten Daten dargestellt. Die Daten zeigen den jeweiligen SRT für den OLSA mit dem ICRA5-250 Störgeräusch. Ohne eine Transformation der Vorhersagen würden diese nur in einem kleinen Bereich zwischen -5 und 2 dB liegen, während sich die Daten über einen Bereich von -20 bis 10 dB erstrecken. Durch die lineare Transformation der Vorhersagen ergibt sich die Darstellung in Abbildung 1. Jetzt decken die Vorhersagen einen ähnlichen Wertebereich ab wie die Daten. Die Korrelation zwischen Vorhersagen und Beobachtung liegt bei r=0.72. Obwohl der Standard-SII nicht für fluktuierende Störgeräusche entwickelt wurde, ergibt sich bereits eine relativ hohe Korrelation zwischen Daten und Vorhersagen.

In Abbildung 1 auf der rechten Seite und Abbildung 2 auf der linken Seite sind die Ergebnisse für die Vorhersagen durch die beiden Kurzzeit-SII Modelle nach Brand et. al (Abbildung 1 rechte Seite) und Rhebergen et. al (Abbildung 2 linke Seite) dargestellt, jeweils für den OLSA. Bei Brand lagen die Korrelationen bei r=0.67 für den GOESA und r=0.69 für den OLSA. Für Rhebergen ergeben sich Korrelationen von r=0.67 für den GOESA und r=0.71 für den OLSA.

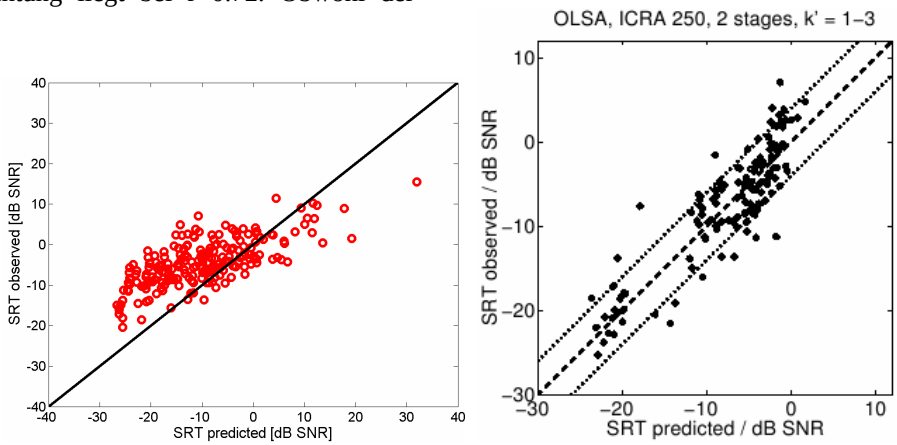

Abbildung 2: Vorhergesagter SRT (Abzisse) über SRT aus der Datenbasis (Ordinate), ICRA5-250 Störgeräusch, OLSA, Einheit: dB SNR. ***Links:*** *Ergebnisse für den Kurzzeit-SII nach Rhebergen et. al. Korrelationskoeffizient r=0.71.* ***Rechts:*** *Ergebnisse für das Kurzzeit-Plomp Modell nach Wagener, Korrelationskoeffizient r=0.89*

Die Korrelationen aller drei SII-basierten Modelle unterscheiden sich nicht wesentlich, die Anpassungen des Standard-SII bringen also keine großen Verbesserungen in den Korrelationen. Aber die Vorhersagen der beiden Kurzzeit-SII Modelle sagen die Daten besser voraus als der Standard-SII ohne eine lineare Transformation. Auch scheint die Korrelation beim OLSA geringfügig größer zu sein als beim GOESA (0.67 (Brand: 0.67) im Gegensatz zu 0.71 (Brand: 0.69)). Das Modell von Rhebergen (Abbildung 4) scheint außerdem die Versuchspersonen zu überschätzen, die meisten der Datenpunkte liegen bei beiden Satztests über der Diagonalen.

In Abbildung 2 sind die Ergebnisse des Kurzzeit-Plomp Modells von Wagener für den OLSA und ICRA5-250 Rauschen auf der rechten Seite dargestellt. Die Korrelation zwischen Vorhersagen und Daten liegt bei r=0.89. Man erkennt, dass die Vorhersagen deutlich näher an den Daten liegen als bei den drei anderen Modellen, alle Datenpunkte gruppieren sich entlang der Diagonalen. Dieser Ansatz liefert im Vergleich zu den anderen drei Modellen die höchste Korrelation zwischen Vorhersagen und Datenbasis.

Zusammenfassung und Ausblick

Sowohl der Standard-SII, als auch seine Erweiterungen auf fluktuierende Störgeräusche erreichen beim OLSA und GOESA mit ICRA5-250 Störgeräusch eine Korrelation im Bereich von r=0.7, wobei der Standard-SII nicht wesentlich schlechter abschneidet als seine Erweiterungen. Der Standard-SII ist durch eine nachträgliche lineare Transformation in der Lage auch die untersuchte Hörsituation erstaunlich gut zu beschreiben, obwohl er nicht für fluktuierende Störgeräusche konzipiert wurde.

Die größte Korrelation wird von dem Kurzzeit-Plomp Modell unter Berücksichtigung von Kontexteffekten mit r=0.89 erreicht. Dieses Ergebnis zeigt, dass Kontexteffekte eine besondere Rolle zu spielen scheinen, gerade bei fluktuierenden Störgeräuschen.

Bei den beiden Erweiterungen zum Standard-SII gilt es zu berücksichtigen, dass bisher nur die Modulationen des Störgeräusches in das Modell einfließen. Die Sprache geht in die Modelle bisher als zeitinvariantes Spektrum ein, so dass deren Modulationen der nicht berücksichtigt werden.

Um die Vorhersagen bei den SII-basierten Modellen weiter zu verbessern, gibt es mehrere Möglichkeiten:

Man kann zusätzlich zu den Modulationen des Störgeräusches die Modulationen der Sprache im SII-Modell berücksichtigen (frequenzunabhängig und frequenzabhängig).

Bei dem Modell von Rhebergen gibt es noch Parameter die noch nicht an die Daten angepasst wurden (Überschätzen der Leistungen Versuchspersonen).

Desweiteren scheint es sinnvoll und notwendig, Kontexteffekte bzw. die Redundanz der Sprache in ein Modell zur Vorhersage der Sprachverständlichkeit in fluktuierendem Störgeräusch einzubauen, wie die Ergebnisse von Wagener zeigen.

Unter Umständen lohnt es sich auch andere Ansätze zur Modellierung von Sprachverständlichkeit zu untersuchen, wie z. B. das SRS-Modell von Hannes Müsch (Müsch 2001).

Danksagung

Die hier vorgestellten Simulationen wurden durchgeführt im Rahmen des Kompetenzzentrums HörTech und gefördert vom europäischen Projekt HearCom.

Literatur

Brand, T. (2002). Vorhersage der Sprachverständlichkeit in Ruhe und im Störgeräusch aufgrund des Reintonaudiogramms. DGA 2002

Müsch, H. & Buus, S. (2001). Using statistical decision theory to predict speech intelligibility. I. Model structure J Acoust Soc Am, 109, 2896-909

Plomp R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. J. Acoust Soc Am, 63(2)

Rhebergen, K. & Versfeld, N. (2005a). 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, 2181-92

Wagener, K.C. (2003). Factors influencing sentence intelligibility in noise (Dissertation). Carl v. Ossietzky Universität Oldenburg.

Wagener, K.C. & Brand, T. (2005b). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: influence of measurement procedure and masking parameters. Int J Audiol, 44, 144-156