

Eignung des Göttinger und des Oldenburger Satztests für die Bestimmung von Verständlichkeitsfunktionen

Thomas Brand, Birger Kollmeier

AG Medizinische Physik, Carl-von-Ossietzky Universität Oldenburg, D-26111 Oldenburg

Einleitung

Die Sprachmaterialien des Göttinger Satztests [1] und des Oldenburger Satztests [2] sind auf eine maximale Steigung der Verständlichkeitsfunktion im Störgeräusch optimiert. Hierdurch wird eine hohe Effizienz bei der Bestimmung der Sprachverständlichkeitsschwelle ermöglicht. Durch die Schwelle allein wird die Verständlichkeitsfunktion jedoch nur unvollständig bestimmt. Daher wurde in dieser Studie untersucht, inwieweit sich neben der Sprachverständlichkeitsschwelle L_{50} auch die Steigung s_{50} der Verständlichkeitsfunktion bestimmen läßt. Um hierfür eine möglichst effektive Verteilung der Darbietungspegel zu erreichen, bieten sich adaptive Pegelsteuerungen an. Verschiedene adaptive Pegelsteuerungen für Satztestverfahren wurden mit Monte-Carlo Simulationen und mit Messungen an Normal- und Schwerhörenden untersucht. Dabei wurden sowohl Verfahren betrachtet, die allein für die Bestimmung der Sprachverständlichkeitsschwelle geeignet sind, als auch Verfahren, mit denen sich Schwelle und Steigung der Verständlichkeitsfunktion gleichzeitig bestimmen lassen.

Modellfunktion

An die Daten einer Sprachverständlichkeitsmessung wird die Modellfunktion

$$p(L, L_{50}, s_{50}) = \frac{1}{1 + \exp(4 \cdot s_{50} \cdot (L_{50} - L))} \quad (1)$$

mit einem Maximum-Likelihood Schätzer angepaßt. Hierbei bezeichnet p die Sprachverständlichkeit ($0 \leq p \leq 1$) beim Darbietungspegel L , L_{50} bezeichnet die Sprachverständlichkeitsschwelle (entspricht der Verständlichkeit $p = 0,5$) in dB und s_{50} bezeichnet die Steigung der Verständlichkeitsfunktion in dB^{-1} an der Stelle des L_{50} .

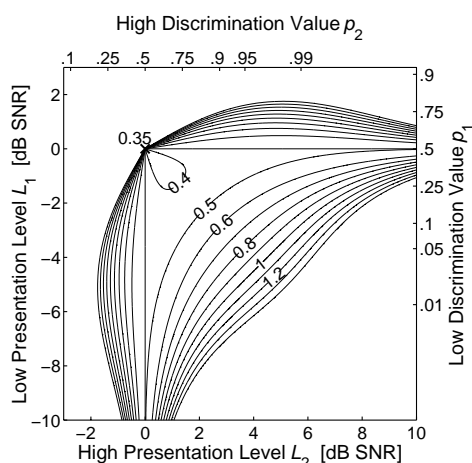


Abbildung 1: Standardabweichung der L_{50} -Schätzung in dB in Höhenliniendarstellung. Das Kreuz zeigt das Minimum (sweetpoint).

Statistische Grundlagen

Die Güte der Anpassung der Modellfunktion wird durch die Wahl der Darbietungspegel und die Anzahl der Datenpunkte (Stichprobenanzahl) bestimmt. Mit Hilfe der Binomialtheorie läßt sich die Abhängigkeit zwischen den Darbietungspegeln und den Standardabweichungen der Schätzwerte von L_{50} und s_{50} berechnen. Hierbei ist zu beachten, daß die Stichprobenanzahl $N \cdot j$ entspricht, wobei N der Anzahl der verwendeten Testsätze und j der Anzahl der statistisch unabhängigen verstandenen Wörter pro Satz entspricht [3]. Um sowohl L_{50} als auch s_{50} zu bestimmen, müssen Messungen bei mindestens zwei verschiedenen Meßpegeln durchgeführt werden. Abbildung 1 zeigt die Standardabweichung der L_{50} -Schätzung, wenn mit genau zwei Pegeln

gemessen wird. Die Diskriminationsfunktion entspricht Gleichung (1) mit $L_{50} = 0 \text{ dB S/N}$, $s_{50} = 0,2 \text{ dB}^{-1}$. $N_1 = N_2 = 10$ Sätze wurden an den beiden Meßpegeln gemessen. Die Anzahl der statistisch unabhängigen Wörter pro Satz wurde als $j = 2,5$ angenommen. Die Darbietungspegel sind sowohl als Pegel in dB S/N als auch als Verständlichkeiten dargestellt. Das Kreuz zeigt das Minimum der Standardabweichung (sweetpoint). Eine minimale Standardabweichung bei der Bestimmung des L_{50} wird erreicht, wenn die Darbietungen teilweise nah oberhalb und teilweise nah unterhalb des L_{50} erfolgen. Eine adaptive Pegelsteuerung, die optimale Effizienz bei der Schwellenbestimmung erreichen soll, sollte daher auf $p = 0,5$ konvergieren.

Abbildung 2 zeigt die Standardabweichung der s_{50} -Schätzung für die gleiche Diskriminationsfunktion wie in Abbildung 1. Das Kreuz zeigt das Paar der optimalen Darbietungspegel für die Steigungsschätzung (sweetpair) mit $p_1 = 0,08$ und $p_2 = 0,92$. Wenn nur die Steigung bestimmt werden soll, müssen die Darbietungspegel zum Teil bei sehr niedrigen Verständlichkeiten und zum Teil bei sehr hohen Verständlichkeiten positioniert werden. Dies hat jedoch negative Auswirkungen auf die Schätzung der Schwelle. Als Kompromiß für eine gleichzeitige Schwellen- und Steigungsschätzung sollten die Darbietungspegel teilweise bei $p = 0,2$ und teilweise bei $p = 0,8$ positioniert werden (pair of compromise).

Adaptive Pegelsteuerung

Die adaptive Pegelsteuerung startet bei einem Signal-Rauschverhältnis von 0 dB und ändert den Darbietungspegel von Satz zu Satz um:

$$\Delta L = \frac{f(i)(prev - tar)}{0,15 \text{ dB}^{-1}} \quad (2)$$

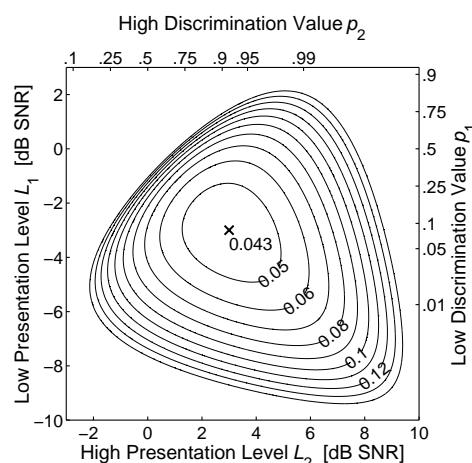


Abbildung 2: Standardabweichung der Steigungsschätzung in dB^{-1} als Höhenliniendarstellung. Das Kreuz zeigt das Minimum (sweetpair).

Hierbei bezeichnet tar die Sprachverständlichkeit, auf die die adaptive Pegelsteuerung konvergieren soll, $prev$ bezeichnet die Verständlichkeit des vorhergehenden Satzes und $f(i)$ ist ein Schrittwertenparameter, der die Konvergenz steuert (siehe nächsten Abschnitt). Dieser einfache Algorithmus wurde in zwei verschiedenen Pegelsteuerungen verwendet. Pegelsteuerung A1 konvergiert auf den sweetpoint ($tar = 0,5$), um eine maximale Effizienz bei der Steigungsschätzung zu erreichen. Pegelsteuerung A2 konvergiert in zwei ineinander verschachtelten adaptiven Zweigen auf das pair of compromise ($tar_1 = 0,2$ und $tar_2 = 0,8$), um neben der Sprachverständlichkeitsschwelle auch die Steigung der Sprachverständlichkeitsfunktion bestimmen zu können.

Optimierung der adaptiven Pegelsteuerungen

Die Auswirkung der adaptiven Pegelsteuerungen auf das Meßergebnis wurde zunächst mit Monte-Carlo Simulationen untersucht. Zuerst wurde der Schrittweitenparameter $f(i)$ in Gleichung (2) optimiert. Es zeigte sich, daß $f(i) = 1,5 \cdot 1,41^{-i}$ (i : Anzahl der Wendepunkte im Verlauf des Darbietungspegels) die schnellste Konvergenz des Verfahrens auf die Zielverständlichkeit ermöglicht (sowohl für A1 als auch für A2). Außerdem zeigte sich, daß dieser relativ einfache adaptive Algorithmus eine sehr hohe Effizienz erreicht. Die Standardabweichungen von L_{50} und s_{50} sind nur um 30% größer als die theoretisch optimal erreichbaren. Ein systematischer Fehler der Schwellenschätzung tritt nicht auf. Die Steigungen werden um jeweils 10% des tatsächlichen Wertes überschätzt.

Ein wichtiges Ergebnis der Simulationen war, daß sich die Effizienz der Algorithmen mit ansteigendem j (Anzahl der statistisch unabhängigen Wörter pro Satz) stark steigert. Aus diesem Grund war zu erwarten, daß der Oldenburger Satztest mit $j \approx 4$ eine etwa doppelt so hohe Effizienz erreicht als der Göttinger Satztest mit $j \approx 2$.

Messungen

An den Messungen nahmen 10 normalhörende und 10 schwerhörende Versuchspersonen teil. Jede Messung wurde adaptiv mit jeweils 30 Sätzen durchgeführt. Die Darbietung erfolgte monaural über Kopfhörer (Sennheiser HDA 200). Die Signale wurden freifeldentzerrt. Als Störgeräusch wurde ein unmoduliertes Rauschen mit dem gleichen Frequenzspektrum wie das jeweilige Sprachmaterial verwendet [1, 2]. Der Rauschpegel wurde für die normalhörenden Versuchspersonen auf 65 dB SPL eingestellt. Bei den schwerhörenden Versuchspersonen wurde eine kategoriale Lautheitsskalierung [4] durchgeführt und der Rauschpegel auf den individuellen Pegel mit der Lautheit ‚mittel‘ eingestellt. Bei jeder Versuchsperson wurde sowohl der Göttinger Satztest als auch der Oldenburger Satztest verwendet. Außerdem wurden jeweils beide Pegelsteuerungen A1 und A2 verwendet. Jede dieser insgesamt 4 Meßbedingungen pro Versuchsperson wurde jeweils dreimal durchgeführt. Für jede Versuchsperson ergab sich so ein Meßprogramm von etwa 3 bis 4 Stunden, das in zwei Sitzungen durchgeführt wurde. Zu Beginn jeder Sitzung wurden 1 bis 2 Übungsmessungen mit jeweils einer Testliste durchgeführt. Die Reihenfolge der verschiedenen Meßbedingungen war bei jeder Versuchsperson zufällig.

Abbildung 3 zeigt die mittleren intraindividuellen Standardabweichungen und die mittleren systematischen Fehler (Bias) der L_{50} - und s_{50} -Schätzungen für die beiden Versuchspersonengruppen und die beiden adaptiven Verfahren. Es sind sowohl die gemessenen Werte als auch die mit Monte-Carlo Simulationen vorhergesagten Werte dargestellt.

Das linke obere Diagramm in Abbildung 3 zeigt die Standardabweichungen der L_{50} Schätzung. Göttinger und Oldenburger Satztest liefern hier bei den Messungen vergleichbare Werte. Auffällig ist jedoch, daß der Oldenburger Satztest die vorhergesagte Verringerung der Standardabweichung nicht erreicht.

Das rechte obere Diagramm in Abbildung 3 zeigt den systematischen Fehler (Bias) der Schwellenschätzung. Wie vorhergesagt, tritt kein problematischer systematischer Fehler auf.

Das linke untere Diagramm in Abbildung 3 zeigt die Standardabweichung der Steigungsschätzung. Wie vorhergesagt schneidet Verfahren A1 hier deutlich schlechter ab als Verfahren A2. Die Erwartung, daß der Oldenburger Satztest mit Verfahren A2 etwa halb so große Standardabweichungen erzielt wie der Göttinger Satztest wurde jedoch nicht erfüllt.

Das rechte untere Diagramm in Abbildung 3 zeigt den systematischen Fehler der Steigungsschätzung. Auch hier bleibt der Oldenburger Satztest hinter der Erwartung zurück und die beiden Satzteste liefern etwa gleich große systematische Fehler.

Zusammenfassung und Diskussion

Göttinger Satztest und Oldenburger Satztest sind gleichermaßen für eine adaptive Meßdurchführung geeignet. Bei der Verwendung der hier vorgestellten adaptiven Pegelsteuerungen läßt sich die Sprachverständlichkeitsschwelle im Störgeräusch mit einer Standardabweichung von weniger als 1 dB bestimmen. Hierzu sind mindestens 20 Testsätze

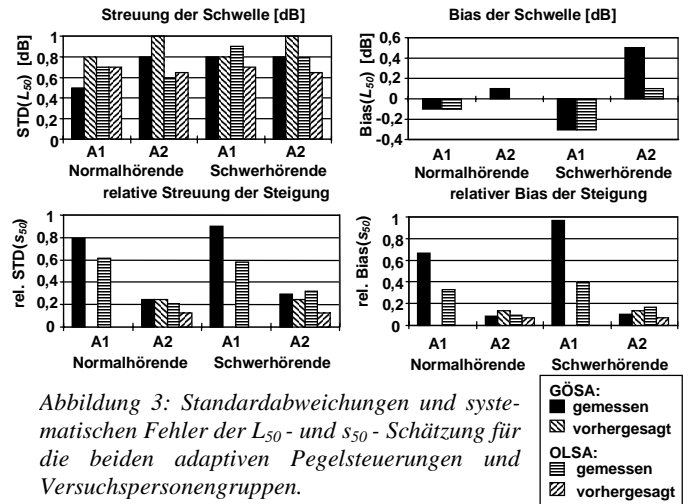


Abbildung 3: Standardabweichungen und systematischen Fehler der L_{50} - und s_{50} -Schätzung für die beiden adaptiven Pegelsteuerungen und Versuchspersonengruppen.

pro Messung erforderlich. Durch die adaptive Durchführung tritt kein systematischer Fehler auf. Bei der Verwendung von 30 Testsätzen ist die gleichzeitige Bestimmung der Steigung der Diskriminationsfunktion bedingt möglich. Hierzu muß die adaptive Pegelsteuerung verwendet werden, die auf die Verständlichkeiten 0,2 und 0,8 konvergiert (*pair of compromise*). Diese Pegelsteuerung erzielt eine Steigungsschätzung mit einer relativen Standardabweichung von 25% und einem relativen systematischen Fehler von 10%.

Weil beim Oldenburger Satztest die Sätze weniger vorhersagbar sind als beim Göttinger Satztest, ist bei ersterem die Anzahl der statistisch unabhängigen Wörter (Stichproben) pro Satz im Mittel etwa doppelt so hoch wie beim Göttinger Satztest. Wegen dieser erhöhten Anzahl von Stichproben wurde für den Oldenburger Satztest eine deutlich erhöhte Effizienz erwartet. Bei den Messungen zeigte sich allerdings, daß dieser Effekt durch die adaptive Pegelsteuerung teilweise gestört wird. Die Konsequenz ist, daß der Göttinger und der Oldenburger Satztest die gleiche Effizienz sowohl bei der Schwellenbestimmung als auch bei der gleichzeitigen Schwellen- und Steigungsbestimmung erreichen.

Der Göttinger Satztest empfiehlt sich für kurze Studien, bei denen nur wenige Testlisten erforderlich sind, da er über keinen Lerneffekt verfügt, solange keine Testlisten wiederholt werden. Der Oldenburger Satztest führt wegen seinem für alle Testlisten gleichem Wortmaterial zu einem gewissen Lerneffekt. Deshalb muß vor der ersten Messung mit jeder Versuchsperson mindestens eine Übungsliste durchgeführt werden. Der Vorteil des Oldenburger Satztests ist, daß nach erfolgtem Training der Test beliebig oft durchgeführt werden kann, weil die Testlisten nicht auswendig gelernt werden können. Der Oldenburger Satztest bietet sich daher für längere Studien an, bei denen eine große Anzahl von Testlisten erforderlich sind. Außerdem ist der Oldenburger Satztest etwas langsamer aufgesprochen als der Göttinger Satztest und eignet sich daher besser für hochgradig Schwerhörende (z.B. CI Patienten).

Diese Arbeit wurde gefördert von DFG KO 942/13-1 und dem EU Projekt NATASHA.

Literatur

- [1] Kollmeier B. und Wesselkamp, M. (1997). *Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment*. J. Acoust. Soc. Am., 102:2412-2421.
- [2] Wagener, K. und Brand, T. und Kollmeier, B. (1999). *Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests*. Zeitschrift für Audiologie, 38:86-95.
- [3] Boothroyd, A. und Nitttrouer, S. (1988). *Mathematical treatment of context effects in phoneme and word recognition*. J. Acoust. Soc. Am., 84:101-114.
- [4] Brand, T., Hohmann, V. Kollmeier, B. (1997): *Meßgenauigkeit der (adaptiven) kategorialen Hörflächenskalierung*. In: Fortschritte der Akustik - DAGA 97, DEGA e.V., Oldenburg