

Wie lässt sich die maximale Verständlichkeit optimal bestimmen?

Thomas Brand¹⁾, Kirsten Wagener^{1, 2)}

¹⁾ Medizinische Physik, Universität Oldenburg, 26111 Oldenburg, ²⁾ Hörzentrum Oldenburg GmbH, Marie-Curie-Str. 2, 26129 Oldenburg

Einleitung

Der Diskriminationsverlust bzw. die maximal erreichbare Verständlichkeit ist eine wichtige Kenngröße für die Sprachaudiometrie in Ruhe. Das Standardverfahren zur Bestimmung der maximalen Diskrimination ist die Verwendung von jeweils 20 Wörtern des Freiburger Einsilbertests bei zwei bis vier verschiedenen Messpegeln.

Bei der Interpretation der Ergebnisse ist jedoch zu beachten, dass die Genauigkeit eines Sprachverständlichkeitstests durch die Binomialverteilung begrenzt wird. D.h., wenn die Schätzung der Sprachverständlichkeit P in Prozent auf einer Anzahl N getesteter Wörter beruht, gilt in erster Näherung für den Standardfehler SE_P der Sprachverständlichkeit:

$$SE_P \approx \sqrt{\frac{P \cdot (100\% - P)}{N}}$$

Eine gemessene Verständlichkeit von 50% entspricht daher tatsächlich einem 95% Konfidenzintervall von 25% bis 75%. Dieses Konfidenzintervall wird durch mögliche Schwankungen der Verständlichkeit der einzelnen Testwörter des Tests zusätzlich verbreitert. Vom Freiburger Einsilber ist z.B. bekannt, dass sich die verschiedenen Testlisten in ihrer Verständlichkeit zum Teil deutlich unterscheiden (v. Wedel, 1986). Leider werden diese Begrenzungen der Messgenauigkeit nicht immer beachtet und die Ergebnisse werden oft überinterpretiert, was z.B. zu falsch diagnostizierten Diskriminationsverlusten oder auch Helmkurven führen kann.

In dieser Studie werden Möglichkeiten zur Verbesserung der Messgenauigkeit untersucht.

Einsilber Reimtest als normale oder verkürzte Version in Ruhe

Ein Ansatz die Messgenauigkeit zu erhöhen besteht darin, einen Test zu verwenden, dessen Testlisten untereinander geringere Unterschiede aufweisen als die Testlisten des Freiburger Einsilbertests. Hierfür steht der Einsilber Reimtest nach v. Wallenberg und Kollmeier (1989) zur Verfügung. Dieser hat in seiner ursprünglichen Fassung Testlisten mit jeweils 72 Wörtern. Da dies für viele Anwendungen zu viel ist, wurde eine Version mit Testlisten mit jeweils 47 Wörtern und eine weitere Version mit Testlisten mit jeweils 25 Wörtern erstellt. Auch bei diesen verkürzten Versionen wurde darauf

geachtet, dass die Phonemverteilung möglichst der deutschen Sprache entspricht.

Da der Einsilber Reimtest ursprünglich für Sprachverständlichkeitsmessungen im Störgeräusch optimiert wurde, musste zunächst untersucht werden, ob der Test und seine beiden verkürzten Versionen auch für Messungen in Ruhe geeignet ist, d.h., ob die verschiedenen Testlisten auch in Ruhe gleich verständlich sind.

An den Messungen nahmen 30 normalhörende Versuchspersonen teil, die Reihenfolge der Testlisten und Messpegel wurde randomisiert. Pro Liste wurde bei drei festen Pegeln 10, 15 und 20 dB SPL (bzw. 9, 14 und 19 dB SPL bei den 25er Listen) gemessen. Jede Versuchsperson hat jede Liste nur einmal bei einem festen Pegel gemessen. Es gab kein Training. Die Auswertung erfolgte unter Berücksichtigung der Ratewahrscheinlichkeit (5 Antwortalternativen). Die Ergebnisse sind in Tabelle 1 dargestellt. Es zeigt sich, dass die verschiedenen Testlisten auch im Störgeräusch sehr ähnliche Diskriminationsfunktionen aufweisen. Wobei mit kürzer werdenden Testlisten die Standardabweichungen von Schwelle und Steigung ansteigen. Eine Verwendung des Tests in Ruhe erscheint damit sinnvoll.

Wörter pro Testliste	Sprachverständlichkeitsschwelle	Steigung
72	14,1 dB SPL \pm 0,3 dB	4,6 %/dB \pm 0,4 %/dB
47	16,5 dB SPL \pm 0,4 dB	5,1 %/dB \pm 0,6 %/dB
25	13,8 dB SPL \pm 0,7 dB	5,5 %/dB \pm 0,7 %/dB

Tabelle 1: Ergebnisse der Evaluationsmessungen des Einsilber Reimtests (Versionen mit 72, 47 und 25 Wörtern) in Ruhe. Die zweite Spalte zeigt die mittleren 60%-Sprachverständlichkeitsschwellen (Mittelpunkte der Diskriminationsfunktion) mit den Standardabweichungen zwischen den Listen. Die dritte Spalte zeigt die mittleren Steigungen der Diskriminationsfunktionen an der Stelle der 60%-Schwelle mit den Standardabweichungen zwischen den Listen.

Test-Retest Messungen

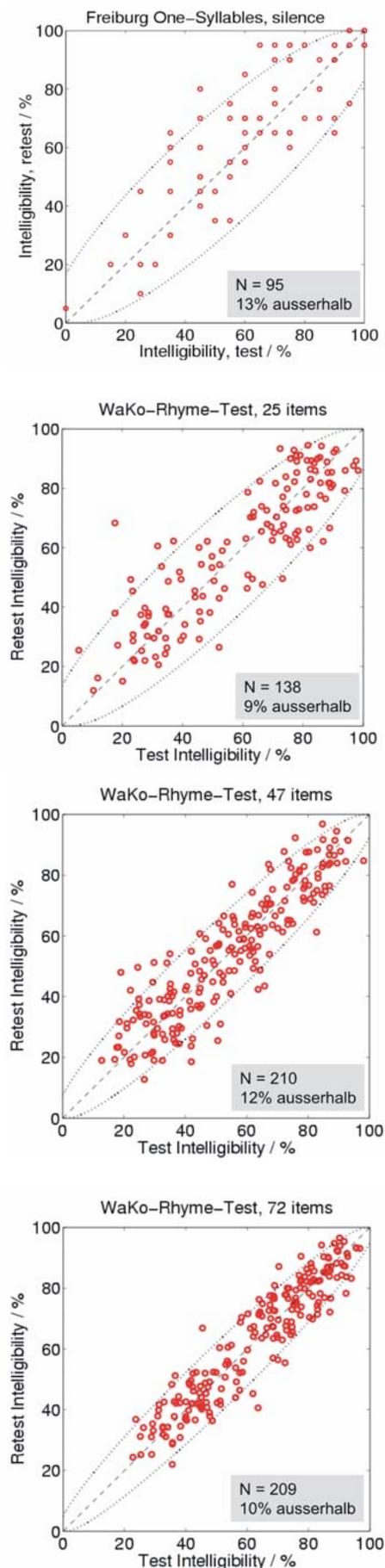
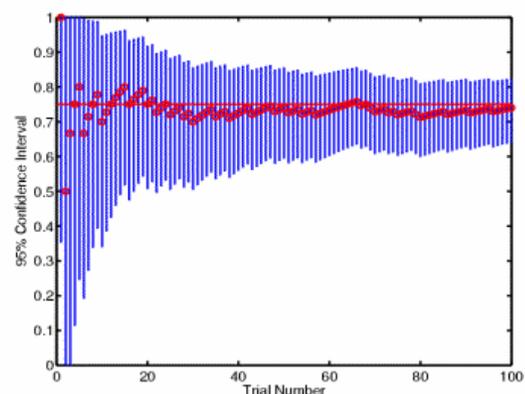


Abbildung 1. Test-Retest Messungen mit dem Freiburger Einsilbertest (20 Wörter) und den verschiedenen langen Testlisten des Einsilber Reimtests nach Wallenberg und Kollmeier.

Abbildung 1 zeigt, wie sich die Anzahl der Wörter pro Testliste (Items) in der Praxis auf die Test-Retest Reliabilität auswirkt. Über der Verständlichkeit der Test-Messung (Abszisse) ist jeweils die Verständlichkeit der Retest-Messung (Ordinate) dargestellt. Jeder Datenpunkt steht somit für eine Test- und eine Retest-Messung bei gleichem Sprachpegel mit einer Versuchsperson. Die gepunkteten Linien zeigen die durch die Binomialverteilung gegebenen theoretischen Konfidenzintervalle. Mit zunehmender Anzahl der Items werden diese Intervalle nach dem „Eins durch Wurzel N“-Gesetz schmaler. Sofern die Inhomogenitäten zwischen den Testlisten keine Rolle spielen, sollten jeweils 10% der Datenpunkte außerhalb der 95% Konfidenzintervalle liegen, da es sich um einen zweiseitigen Test handelt, d.h., die Retest-Messung kann sowohl nach unten als auch nach oben von der Test-Messung abweichen. Dies ist bei allen Test annähernd der Fall. D.h., der entscheidende Faktor für die Genauigkeit der Messung ist die Anzahl der Items. Inhomogenitäten zwischen den Testlisten spielen eine untergeordnete Rolle. Die Testlisten des Einsilber Reimtests sind somit in allen Versionen (25, 47 und 72 Wörter) homogen. Die bekannten Inhomogenitäten der Testlisten des Freiburger Einsilbers fallen hier wegen der Verwendung von lediglich 20 Items gegenüber der durch die Binomialverteilung gegebene statistische Ungenauigkeit nicht ins Gewicht, zumal bei diesen Messungen die bekanntermaßen vom Mittelwert abweichenden Testlisten nicht verwendet wurden.

Verlauf der Sprachverständlichkeitsmessung



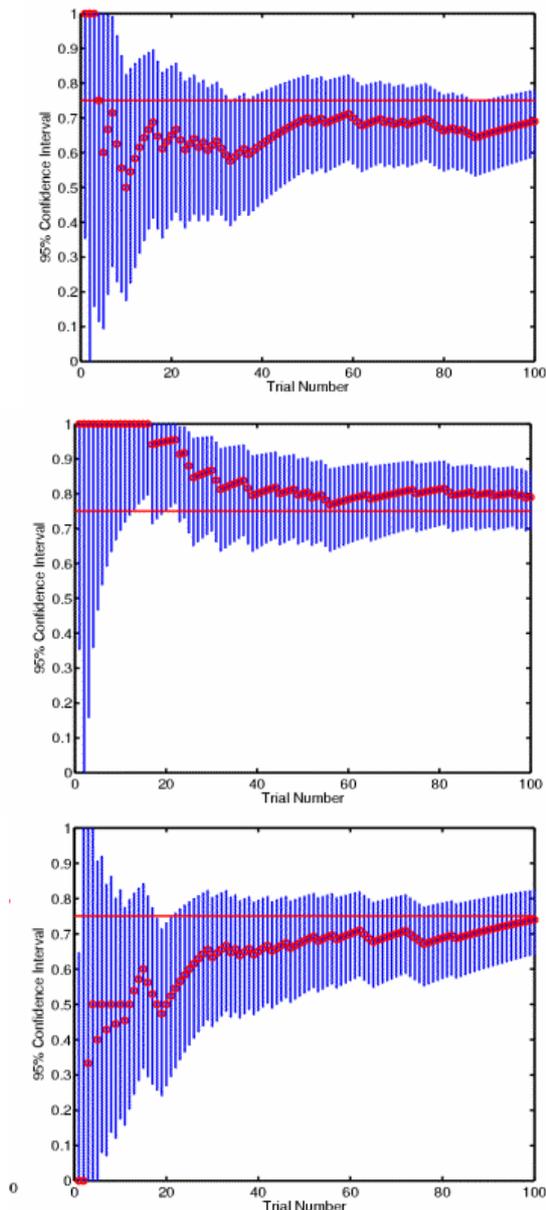


Abbildung 2: Vier Beispiele für den Verlauf einer Verständlichkeitsmessung (Monte-Carlo Simulationen).

Als ein Indikator, ob eine Sprachverständlichkeitsmessung störungsfrei abgelaufen ist, oder ob es z.B. Schwankungen der Aufmerksamkeit der Versuchsperson gab, könnte der Verlauf einer Sprachverständlichkeitsmessung dienen. Abbildung 2 zeigt vier verschiedene Beispiele für solche Verläufe. Eine Besonderheit der hier dargestellten Verläufe ist, dass der tatsächliche exakte Wert der Verständlichkeit (dargestellt als waagerechte rote Linie) bekannt ist, da es sich um Monte-Carlo Simulationen handelt. Die roten Kreise stellen den jeweiligen Schätzwert der Verständlichkeit dar (Anteil der bisher richtig verstandenen Antworten).

Die senkrechten blauen Linien stellen das 95% Konfidenzintervall dar. An den Beispielen lässt sich erkennen, dass der Verlauf der Schätzung nur sehr bedingt als Indikator für die Verlässlichkeit nutzbar ist. Während in der links oben dargestellten Simulation der Schätzwert bereits sehr früh mit dem tatsächlichen Wert übereinstimmt und dies auch im weiteren Verlauf so bleibt, gibt es beim unten rechts dargestellten Verlauf einen stetigen Anstieg des Schätzwertes. Eine solche Messung würde in der Praxis möglicherweise den Verdacht erregen, dass die Aufmerksamkeit der Versuchsperson während der Messung nicht konstant war. Tatsächlich lag der Simulation aber stets die gleiche Verständlichkeit zu Grunde. Es handelt sich also um einen rein statistischen Effekt.

Zusammenfassung

- Ein optimierter Einsilber Reimtest in drei Versionen (72, 47 und 25 Wörter) für Messungen in Ruhe steht zur Verfügung.
- Test-Retest Messungen zeigen, dass der Einsilber Reimtest genauere Messungen ermöglicht als der Freiburger Reimtest. Dieses Potential lässt sich allerdings nur nutzen, wenn genügend Wörter verwendet werden.
- Der Verlauf einer Messung ist kein verlässlicher Indikator für die Güte einer Messung. Auch Schätzungen, die über den gesamten Verlauf der Messung sehr konstant bleiben, können vom tatsächlichen Wert abweichen bzw. auch sich verändernde Schätzungen können letztlich dem tatsächlichen Wert entsprechen.

Dieser Beitrag wurde gefördert von DFG KO 942/13-3 und von dem BMBF Kompetenzzentrum „HörTech“.

Literatur

- v. Wedel H (1986) Untersuchungen zum Freiburger Sprachtest – Vergleichbarkeit der Gruppen in Hinblick auf Diagnose und Rehabilitation (Hörgeräteanpassung und Training). *Audiologische Akustik* 25, 60-73
- v. Wallenberg EL, Kollmeier B (1989) Sprachverständlichkeitsmessungen für die Audiologie mit einem Reimtest in deutscher Sprache: Erstellung und Evaluation von Testlisten, *Audiologische Akustik* 28(2), 50-65