

Vergleich von fünf Sprachtests im sprachsimulierenden Störgeräusch

Comparison of five speech tests in speech-simulating noise

Abstract

There are several speech recognition tests available in Germany. They are often used with different background noises and with different calibration methods. Therefore, comparisons between the resulting speech recognition scores can only be made conditionally. Here, measurement conditions were adjusted to examine the differences in speech recognition using the Freiburg monosyllabic speech test, Oldenburg and Göttingen sentence test, Hochmair-Schulz-Moser sentence test, and monosyllabic rhyme test according to von Wallenberg and Kollmeier. First, all speech data sets were calibrated to the same average long-term level. Second, new speech-simulating noises (speech-adjusted noise, SAN) were generated from the speech materials. For this purpose, each set of speech material was superimposed 30 times. These SAN noises differ substantially from the spectrum of standard (STD) noises. The speech recognition thresholds in STD and SAN noises were measured with 22 young listeners without hearing deficits in the free field. Both speech and background noise were presented via one loudspeaker from the 0° direction. Each of the five speech tests was presented in background noise with at least three signal-to-noise ratios. Discrimination functions were fitted to the results and the thresholds for a speech recognition score of 50% (SRT_{50}) and the corresponding slopes were determined. The difference in SRT_{50} between STD and SAN noise was largest (4.0 dB) for the Hochmair-Schulz-Moser sentence test and was only 0.7 dB for the Göttingen sentence test. When using SAN noise and the same calibration method, the SRT_{50} values of the five speech sets deviated from each another by maximally 2.7 dB. The slopes of the discrimination functions differed by up to 15.1%-points/dB in STD noise. This deviation was reduced in SAN noise to 1.8%-points/dB within the sentence tests and to 0.2%-points/dB within the word tests. In addition, the slopes were significantly steeper for the sentence tests than for the word tests.

Keywords: speech test, speech in background noise, speech-simulating noise, calibration, OLSA, FBE, GÖSA, WAKO, HSM

Zusammenfassung

In Deutschland stehen mehrere Sprachtestverfahren zur Verfügung, die jeweils mit unterschiedlichen Störgeräuschen und teils auch verschiedenen Kalibrierungsverfahren verwendet werden. Dadurch können Vergleiche zwischen den Ergebnissen der jeweiligen Sprachtests nur bedingt erfolgen. Zur Untersuchung der Unterschiede zwischen dem Freiburger Einsilbertest, dem Oldenburger Satztest, dem Göttinger Satztest, dem Hochmair-Schulz-Moser-Satztest und dem Reimtest nach von Wallenberg und Kollmeier wurden deshalb die Messbedingungen angeglichen. Zum einen wurden alle Sprachmaterialien auf den gleichen mittleren Langzeitpegel kalibriert. Zum anderen wurden neue sprachsimulierende Rauschen (Speech-Adjusted Noise, SAN) durch 30fache Überlagerung des jeweiligen Sprachmaterials generiert. Diese SAN-Rauschen weichen zum Teil deutlich vom Spektrum der standardmäßig

Christina Zinner¹
Alexandra Winkler¹
Inga Holube¹

¹ Institut für Hörtechnik und Audiologie, Jade Hochschule, und Exzellenzcluster „Hearing4All“, Oldenburg, Deutschland

verwendeten Rauschen (STD-Rauschen) ab. Das Sprachverstehen in SAN- und in STD-Rauschen wurde mit 22 jungen Probanden ohne Hörbeeinträchtigungen im Freifeld ermittelt. Sowohl die Sprache als auch die Störgeräusche wurden über einen Lautsprecher aus der 0°-Richtung dargeboten. Jedes der fünf Sprachmaterialien wurde in den Störgeräuschen bei mindestens drei Signal-Rausch-Abständen präsentiert. An die Messergebnisse wurden Diskriminationsfunktionen angepasst und die Schwelle für 50%iges Sprachverstehen (SRT_{50}) mit der entsprechenden Steigung ermittelt. Die Differenz zwischen den SRT_{50} -Werten für die STD-Rauschen und die SAN-Rauschen war für den Hochmair-Schulz-Moser-Satztest mit 4,0 dB am größten und für den Göttinger Satztest mit 0,7 dB am geringsten. Bei Verwendung der SAN-Rauschen und gleichem Kalibrierverfahren wichen die SRT_{50} -Werte der fünf Sprachmaterialien um maximal 2,7 dB voneinander ab. Die Steigungen der Diskriminationsfunktionen unterschieden sich im STD-Rauschen um bis zu 15,1%-Punkte/dB. Diese Diskrepanz reduzierte sich im SAN-Rauschen innerhalb der Satztests auf 1,8%-Punkte/dB sowie innerhalb der Worttests auf 0,2%-Punkte/dB und war für die Satztestverfahren signifikant steiler als für die Wortteste.

Schlüsselwörter: Sprachtests, Sprache im Störgeräusch, sprachsimulierendes Störgeräusch, Kalibrierung, OLSA, FBE, GÖSA, WAKO, HSM

Einleitung

Für die Ermittlung des Sprachverstehens bei erwachsenen Personen im Rahmen der Hörrehabilitation stehen im deutschsprachigen Raum mehrere Testverfahren zur Verfügung. Neben dem Freiburger Einsilbertest (FBE, [1]), dem Oldenburger Satztest (OLSA, [2], [3], [4]) und dem Göttinger Satztest (GÖSA, [5]), werden auch der Hochmair-Schulz-Moser-Satztest (HSM, [6]) und der Reimtest nach von Wallenberg und Kollmeier (WAKO, [7]) verwendet. Jedoch unterscheiden sich diese Sprachtests in mehreren Aspekten, wie zum Beispiel im Sprachmaterial, der Art der Durchführung, im verwendeten Störgeräusch und der Kalibrierung [8].

Die Sätze des OLSA wurden jeweils nach gleicher Struktur aus fünf Wörtern (Name, Verb, Zahlwort, Adjektiv, Objekt) zusammengestellt, wobei für jede Wortart zehn Wörter zur Verfügung standen [2]. Diese Matrix-Struktur begrenzt den getesteten Wortschatz und führt zu Trainingseffekten [4], ermöglicht jedoch neben dem offenen Antwortformat (verbale Wiederholung der Wörter durch die Probanden) auch ein geschlossenes Antwortformat (Auswahl der Wörter aus der dargestellten Gesamtmatrix). Die Satztests GÖSA und HSM sind dagegen nur in der Anzahl der zur Verfügung stehenden Testlisten, jedoch nicht prinzipiell im Wortschatz begrenzt und können deshalb nur offen durchgeführt werden. Gleiches gilt für die einsilbigen Wörter des FBE. Der WAKO wird dagegen geschlossen durchgeführt. Die präsentierten einsilbigen Wörter müssen aus fünf dargestellten Antwortalternativen ausgewählt werden, so dass eine Ratewahrscheinlichkeit von 20% besteht.

Während es sich bei den nachzusprechenden Sätzen des OLSA um sinnleeres Sprachmaterial handelt, sind die

Sätze des GÖSA und HSM sinnbehaftet. Dadurch ergibt sich für die Sprachtests ein unterschiedlicher j -Faktor [9]. Dieser ist ein Indikator für die Anzahl statistisch unabhängiger Testitems innerhalb einer Darbietung, d.h. eines Wortes bei einem Worttest bzw. eines Satzes bei einem Satztest. Nach Wagener et al. [4] folgt aus einem höheren j -Faktor eine höhere Messgenauigkeit, da in einem bestimmten Zeitintervall der Messung mehr Informationen über das Sprachverstehen ermittelt werden können. Der OLSA erreicht einen j -Faktor von bis zu $j=4,29$ [4]. Nach Kollmeier und Wesselkamp [5] beträgt der maximale j -Faktor des GÖSA 2,38. Aufgrund des ähnlichen Testaufbaus ist davon auszugehen, dass der j -Faktor des HSM in der gleichen Größenordnung wie derjenige des GÖSA liegt. Die Einsilbertests FBE und WAKO besitzen aufgrund der Darbietung von nur einem Wort pro Darbietung den geringsten j -Faktor mit $j=1$. Somit unterscheidet sich die Genauigkeit, mit der das Sprachverstehen bei den Sprachtests ermittelt wird.

Auch das standardmäßig verwendete Störgeräusch der Sprachtests unterscheidet sich. Zur Messung des FBE und des HSM wird das Rauschen gemäß Comité Consultatif International Télégraphique et Téléphonique (sogenanntes CCITT-Rauschen, [10]) verwendet. Dagegen wurden für den OLSA [2] und den WAKO [11] Störgeräusche durch mehrfache Überlagerung des jeweiligen Sprachmaterials generiert. Für den GÖSA wird standardmäßig das Störgeräusch des WAKO verwendet. Die spektralen Unterschiede in den Störgeräuschen können das Sprachverstehen signifikant beeinflussen (siehe z.B. Winkler und Holube [12] für das Sprachverstehen von Einsilbern). Außerdem beschreiben Wagener et al. [2], dass mit den gleichen Langzeitspektren von Sprache und STD-Rauschen eine steilere Diskriminationsfunktion und

damit einhergehend eine höhere Effizienz der Messungen (d.h. ein geringerer Messaufwand bei gewünschter Messgenauigkeit einer adaptiven Messung) erreicht werden können. Eine Möglichkeit zur Simulation der Störgeräuscheffekte bietet der Speech Intelligibility Index (SII, [13]). Der SII schätzt anhand der Langzeitspektren von Sprache und Störgeräusch das Sprachverstehen bei verschiedenen SNR-Werten und ermöglicht somit eine objektive Bewertung der Verdeckungswirkung von Störgeräuschen. Verschiedene Studien nutzten den SII, um eine Vorhersage des Sprachverstehens für bestimmte Hörsituationen zu treffen (siehe z.B. [14] und [15]).

Ein weiterer Aspekt ist die unterschiedliche Kalibrierung der Sprachtests. Winkler und Holube [12] beschrieben den Einfluss der Kalibrierung auf das Messergebnis und die Vergleichbarkeit des FBE mit anderen Sprachtests. Bei der Kalibrierung des FBE wird nicht – wie in DIN EN ISO 8253-3 [16] gefordert – der äquivalente Dauerschalldruckpegel, sondern der Spitzenpegel des Sprachmaterials auf den Pegel des CCITT-Rauschens angepasst. Das genaue Vorgehen bei der Kalibrierung des HSM ist bei Schmidt et al. [6] nicht näher beschrieben. Jedoch kann nach Winkler et al. [17] angenommen werden, dass die Kalibrierung mit einer Impuls-Bewertung des Sprachmaterials durchgeführt wurde. Durch die Kalibriermethoden unterscheidet sich der äquivalente Dauerschalldruckpegel des FBE zum CCITT-Rauschen um ca. 6,5 dB [12] und des HSM zum CCITT-Rauschen um ca. 6,3 dB [18]. Beim OLSA, GÖSA und WAKO stimmen dagegen die äquivalenten Dauerschalldruckpegel von Sprachmaterial und Störgeräusch überein. Um eine Vergleichbarkeit zwischen den Sprachtests herstellen zu können, ist eine gleiche Kalibrierung oder eine nachträgliche Korrektur der Ergebnisse notwendig.

Durch die unterschiedlichen Testkonditionen ergeben sich Sprachtestergebnisse, die nur bedingt miteinander vergleichbar sind. Die vorliegende Studie beschäftigt sich deshalb mit der Frage, welche Unterschiede in den Ergebnissen der Sprachtests verbleiben, wenn die Störgeräusche dem Langzeitspektrum der entsprechenden Sprachmaterialien angepasst und die Kalibrierungen identisch sind.

Methodik

Probanden

An der Studie nahmen 22 otologisch normale Probanden (13 ♀, 9 ♂) nach DIN EN ISO 8253-1 [19] teil. Alle Probanden waren zwischen 18 und 25 Jahre alt (Median: 22,5 Jahre) und hatten keine Erfahrung mit Sprachtestverfahren. Vor den Messungen des Sprachverstehens wurde mit jedem Probanden eine seitengetrennte Tonaudiometrie mit dem Audiometer Unity 2 (Siemens Audiologische Technik GmbH, Erlangen) durchgeführt. Die Messung der Luftleitungshörschwelle erfolgte mit dem Kopfhörer HDA 200 (Sennheiser electronic GmbH & Co. KG, Wedemark), während für die Knochenleitungshörschwelle

der Wandler B-71 (Radioear, Middelfart, Dänemark) verwendet wurde. Das Hörvermögen der Probanden lag, wie von DIN EN ISO 8253-3 [16] gefordert, zwischen 250 Hz und 8 kHz bei maximal 10 dB HL bzw. betrug bei höchstens zwei Frequenzen maximal 15 dB HL. Für die Teilnahme an der Studie erhielt jeder Proband eine Aufwandsentschädigung von 10 € pro Stunde. Die Untersuchungen wurden von der Kommission für Forschungsfolgenabschätzung und Ethik der Carl von Ossietzky Universität Oldenburg (Drs. 34/2017) genehmigt.

Sprachtests

Das Sprachverstehen der Probanden wurde mit fünf verschiedenen Sprachtests ermittelt: FBE, OLSA, GÖSA, HSM und WAKO. Für den FBE wurde die CD mit der Sachnummer 7970155 HH 922 (Siemens Audiologische Technik GmbH, Erlangen) verwendet. Während FBE, OLSA, GÖSA und HSM offen mit jeweils 20 Darbietungen pro Testliste durchgeführt wurden, lag die Anzahl der Darbietungen beim WAKO bei 25. In den Satztests wurde jedes einzelne Wort als richtig oder falsch gewertet. Der WAKO wurde als geschlossener Test durchgeführt. Die Probanden wählten das gehörte Wort aus fünf Antwortalternativen aus. Die dadurch entstehende Ratewahrscheinlichkeit von 20% wurde in den Ergebnissen durch die Item-Ratekorrektur [20] ausgeglichen:

Gleichung 1:

$$X_{\text{korr}} = N_r - \frac{N_f}{4}$$

Dabei stellt N_r die Prozentzahl der richtigen Antworten, N_f die Prozentzahl der falschen Antworten und X_{korr} das korrigierte Messergebnis in Prozent dar.

Störgeräusch

Jeder Sprachtest wurde mit den standardmäßig verwendeten Störgeräuschen (STD-Rauschen) durchgeführt. Die spektralen Abweichungen, die die STD-Rauschen zu den jeweiligen Sprachmaterialien aufweisen, sind in Abbildung 1 dargestellt. Dazu wurden alle Signale auf den gleichen digitalen Root Mean Square (RMS) kalibriert. Zur Bestimmung der Leistungsdichtespektren der Sprachmaterialien wurden für jeden Sprachtest alle Sätze oder Wörter ohne Pause hintereinander zu einem Signal zusammengefasst und analysiert.

Das CCITT-Rauschen, das im FBE und HSM verwendet wird, wurde für die Telekommunikation entwickelt [10]. Die Terzpegel des CCITT-Rauschens können DIN 45626-1 [21] entnommen werden. Es ist als Kalibriersignal zum FBE auf jeder CD verfügbar. Für den GÖSA und den WAKO wird ebenfalls dasselbe Störgeräusch verwendet, das als GÖnoise bezeichnet wird. Das GÖnoise wurde durch mehrfache Überlagerung des WAKO-Sprachmaterials generiert [11]. Beim Vergleich der Leistungsdichtespektren von GÖSA-Sprachmaterial und GÖnoise zeigt sich

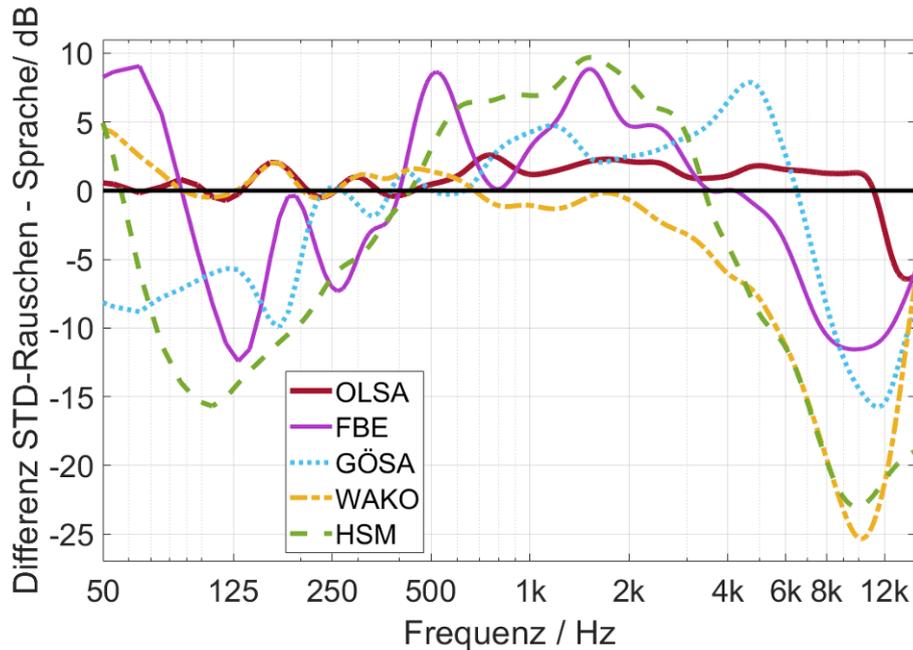


Abbildung 1: Frequenzabhängige Differenz der Leistungsdichtespektren zwischen den standardmäßig verwendeten Rauschen (STD-Rauschen) und den jeweiligen Sprachmaterialien

ein Unterschied von bis zu 16 dB bei ca. 11,5 kHz. Auch zwischen den Spektren des WAKO-Sprachmaterials mit dem GÖnoise treten Unterschiede auf. Diese sind bei Frequenzen oberhalb von 2 kHz beobachtbar und betragen bis zu ca. 25 dB bei ca. 10 kHz. Die genaue Ursache dieser Differenzen ist nicht näher beschrieben. Nach Brand (persönliche Mitteilung [22]) kann jedoch angenommen werden, dass sie auf eine nachträgliche Tiefpass-Filterung des GÖnoise zurückzuführen sind. In Analogie zum GÖnoise wurde das OLnoise aus dem Sprachmaterial des OLSA generiert. Dazu wurde das Sprachmaterial 30fach zufällig in der Zeit verschoben überlagert [2]. Nach Wagener et al. [2] treten beim OLnoise nur gering ausgeprägte Modulationen in der Einhüllenden im Zeitbereich auf.

Für jeden Sprachtest wurde ein sprachsimulierendes Rauschen erzeugt (*engl.* Speech-Adjusted Noise, SAN-Rauschen), das den spektralen Eigenschaften des jeweiligen Sprachmaterials entspricht und somit eine optimale Verdeckungswirkung aufweist [23]. Zur Generierung der SAN-Rauschen wurde das Sprachmaterial jedes Sprachtests in Analogie zu [24] 30fach überlagert. Die Rauschsignale stehen im Repositorium Zenodo zur Verfügung [25]. In Abbildung 2 sind die Differenzen der Leistungsdichtespektren der SAN-Rauschen zur jeweiligen Sprache dargestellt. Wie bei den STD-Rauschen wurden alle Messsignale auf den gleichen RMS kalibriert. Die Unterschiede zwischen SAN-Rauschen und den Sprachmaterialien liegen weit unterhalb von 1 dB.

Messaufbau

Alle Messungen fanden in einer gedämmten, doppelwandigen Hörkabine (IAC Acoustics) statt. Zur Durchführung der Sprachtests saßen die Probanden (18 bis 25 Jahre)

in einer Entfernung von 1,2 m frontal (0° -Richtung) vor einem Lautsprecher Genelec 8030 A (Audio Pro Heilbronn Elektroakustik, Heilbronn), der sowohl die Sprache als auch die Störgeräusche präsentierte. Die Ausgabe der Signale erfolgte über die Soundkarte Typ RME Fireface UC (Audio AG, Heimhausen). Die Störgeräusche wurden während jeder Testliste kontinuierlich bei einem Schalldruckpegel von 65 dB SPL dargeboten. Der Schalldruckpegel der Sprache variierte je nach SNR (s. Tabelle 1) für jede Testliste. Der Versuchsleiter gab die richtigen Antworten über einen Touchscreen (ProLite T1932MSC, Iiyama International Corporate, Hoofddorp, Niederlande), der mit dem PC (Latitude E6540, Dell GmbH, Frankfurt am Main) verbunden war, ein.

Tabelle 1: Verwendete Signal-Rausch-Abstände (SNR) aller Sprachtests im Standard-(STD)-Rauschen und im neu generierten sprachsimulierenden-(SAN)-Rauschen

Sprachtest mit zugehörigem Störgeräusch		SNR-Werte in dB				
		-16	-13	-10	-7	-4
OLSA	STD: OLnoise			x	x	x
	OLSA-SAN			x	x	x
FBE	STD: CCITT			x	x	x
	FBE-SAN			x	x	x
GÖSA	STD: GÖnoise			x	x	x
	GÖSA-SAN			x	x	x
WAKO	STD: GÖnoise	x	x	x	x	
	WAKO-SAN			x	x	x
HSM	STD: CCITT	x	x	x		
	HSM-SAN			x	x	x

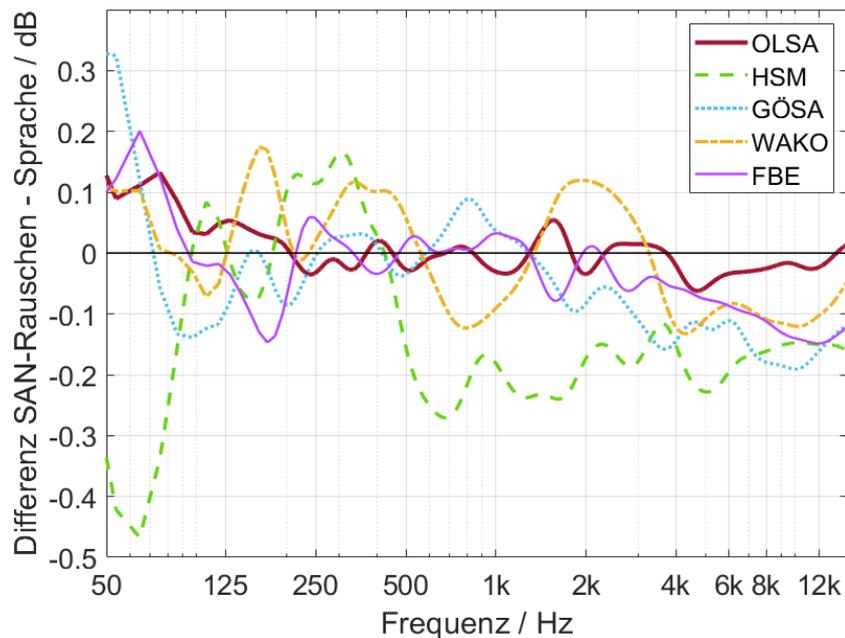


Abbildung 2: Frequenzabhängige Differenz der Leistungsdichtespektren zwischen den sprachsimulierenden Rauschen (SAN-Rauschen) und den jeweiligen Sprachmaterialien

Messdurchführung

Alle Sprachtests wurden mit beiden Störgeräuschen bei verschiedenen festen Signal-Rausch-Verhältnissen (Signal-to-Noise Ratio, SNR) durchgeführt. Um eine Diskriminationsfunktion anpassen und die Schwelle für 50%iges Verstehen (*engl.* Speech Recognition Threshold, SRT_{50}) mit dazugehöriger Steigung s_{50} bestimmen zu können, sollten die Messpunkte so gewählt werden, dass ein Sprachverstehen unter- und oberhalb von 50% erreicht wird. Aufgrund von Literaturwerten und Pilottests waren dafür bei den meisten Sprachtests dieselben drei SNR-Werte ausreichend (siehe Tabelle 1). Beim HSM und beim WAKO im STD-Rauschen mussten jedoch geringere SNR-Werte gewählt werden. Beim WAKO im STD-Rauschen wurde aufgrund der Pilottests vorsichtshalber noch ein vierter SNR ergänzt, um eine Anpassung der Diskriminationsfunktion für alle Probanden zuverlässig zu gewährleisten. Für den OLSA wurden vorab zwei Trainingslisten mit festem SNR (0 dB) durchgeführt, um den Trainingseffekt möglichst gering zu halten [4].

Um Reihenfolgeeffekte auszuschließen, erfolgte eine Randomisierung der Messkonditionen. Diese Randomisierung schloss die Reihenfolge der Sprachtests für jeden Probanden, das verwendete Störgeräusch und die SNR-Werte ein. Außerdem fand eine pseudorandomisierte Zuordnung der Testlistennummern zu den Probanden und Messkonditionen statt. Für keinen der verwendeten Sprachtests wurden vorab Testlisten ausgeschlossen, und alle Listen kamen im Mittel gleich häufig vor. Die Durchführung der Messungen erfolgte an zwei Terminen mit einer jeweiligen Dauer von maximal zwei Stunden inklusive Pausen.

Statistische Auswertung

Obwohl nur neun der erhobenen zehn Datensätze normalverteilt waren (nach Shapiro-Wilk-Test), wurde für alle statistischen Tests eine einfaktorische Varianzanalyse (*engl.* analysis of variance, ANOVA) und PostHoc der t -Test für abhängige Stichproben verwendet. Dieses Vorgehen wurde gewählt, um ein einheitliches Testverfahren für die gesamte statistische Auswertung anwenden zu können. Dies ist möglich, da sowohl die ANOVA als auch der t -Test relativ stabil gegenüber Abweichungen von der Normalverteilung sind. Das Signifikanzniveau lag bei 5%. Für Paarvergleiche wurde die Bonferroni-Korrektur angewandt, sodass das korrigierte Signifikanzniveau α' einem Wert von 0,5% entsprach.

SII-Berechnung

Neben dem Sprachverstehen wurde als objektives Maß der SII nach ANSI S3.5 [13] bestimmt. Diese Berechnung erfolgte für alle Sprachtests für das STD- und das SAN-Rauschen. Bei der SII-Berechnung wurde eine frequenzunabhängige Tonhörschwelle von 0 dB HL angenommen, da die an der Studie teilnehmenden Probanden alle normalhörend waren. Die tatsächliche Tonhörschwelle, gemittelt über alle linken und rechten Ohren, wich um maximal 2,7 dB bei der Messfrequenz 8 kHz von 0 dB HL ab. Außerdem wurde der vorgegebene Korrekturwert für binaurales Hören (1,7 dB) berücksichtigt [13]. Dadurch ergab sich eine frequenzunabhängige Schwelle von -1,7 dB HL, die dem binauralem Hören Normalhörender im Freifeld entsprach.

Nahezu alle frequenzabhängigen Gewichtungsfunktionen der ANSI S3.5 [13] weisen ein Maximum um 2 kHz auf und fallen sowohl unter- als auch oberhalb dieser Frequenz ab. Nur die Funktion „short passages of easy rea-

ding material“ zeigt einen abweichenden Verlauf mit einer höheren Gewichtung bei tieferen Frequenzen um 500 Hz und einer geringeren Gewichtung um 1 kHz. Um diejenige Gewichtungsfunktion zu finden, die die Sprachtestergebnisse am besten repräsentiert, wurden die Ergebnisse aller Gewichtungsfunktionen und einer Gleichgewichtung aller Terzbänder verglichen. Dabei zeigte sich, dass die Gleichgewichtung die Abhängigkeit des Sprachverstehens vom SII für die fünf Sprachtests und die verwendeten Störgeräusche am besten abbilden kann. Deshalb beziehen sich alle hier dargestellten Ergebnisse auf diese Gleichgewichtung.

Ergebnisse

Sprachverstehen

Die SNR wurden so gewählt, dass das Sprachverstehen möglichst unterhalb und oberhalb von 50% lag. Die Messergebnisse in Abbildung 3 (blaue Kreise) verdeutlichen, dass dieses Ziel für alle Sprachtests und beide Störgeräusche erreicht wurde. Für jede Messbedingung wurde an das erreichte prozentuale Sprachverstehen p in Abhängigkeit vom dargebotenen SNR (L_{SNR}) eine Diskriminationsfunktion nach Brand und Kollmeier [26] (s. Gleichung 2) angepasst (rote Kurve) und daraus der SNR für ein Sprachverstehen von 50% (SRT_{50}) sowie die Steigung in diesem Punkt (s_{50}) bestimmt.

Gleichung 2:

$$p(L_{SNR}, SRT_{50}, s_{50}) = \frac{100}{1 + e^{4 \cdot s_{50} \cdot (SRT_{50} - L_{SNR})}}$$

Schwellen und Steigung

In Abbildung 4 sind die Ergebnisse der SRT_{50} -Werte (oben) und die dazugehörigen Steigungen s_{50} (unten) bei individueller Anpassung der Diskriminationsfunktionen für jeden einzelnen Probanden dargestellt. Die linke Seite zeigt die Ergebnisse für die STD- und die rechte Seite die entsprechenden Werte für die SAN-Rauschen.

Die medianen SRT_{50} -Werte im STD-Rauschen reichen von $-12,9$ dB SNR (HSM) bis $-5,6$ dB SNR (GÖSA).

Die einfaktorielle ANOVA ergab signifikante Unterschiede ($F(2,75; 57,80)=369,17; p<0,001$), die sich auch im PostHoc-t-Test für alle Sprachtests bestätigten (siehe Tabelle 2, blau).

Tabelle 2: Statistische Auswertung (p-Werte) der Schwellen im Sprachverstehen SRT_{50} (blau) und der zugehörigen Steigungen s_{50} (rot) aller Sprachtests im Standard-(STD)-Rauschen

	OLSA	FBE	GÖSA	WAKO	HSM
OLSA	–	<0,001*	0,002*	<0,001*	<0,001*
FBE	<0,001*	–	<0,001*	<0,001*	<0,001*
GÖSA	<0,001*	<0,001*	–	<0,001*	<0,001*
WAKO	<0,001*	1,000	<0,001*	–	0,002*
HSM	<0,001*	0,461	<0,001*	0,012	–

Mit *-gekennzeichnete Werte sind signifikant unter Berücksichtigung der Bonferroni-Korrektur.

In Tabelle 3 sind die gemessenen SRT_{50} -Werte Literaturwerten gegenübergestellt. Dabei ergeben sich für FBE und WAKO große Differenzen. Wenn jedoch die unterschiedliche Kalibrierung der beiden Sprachtests berücksichtigt wird (korrigierte Literaturwerte), dann liegen die Differenzen bei maximal 0,8 dB (OLSA).

Tabelle 3: Gegenüberstellung der gemessenen SRT_{50} -Werte in dB SNR im STD-Rauschen mit Werten aus der Literatur; für die korrigierten Literaturwerte wurden Unterschiede in der Kalibrierung berücksichtigt.

Sprachtest	Literaturwert	Korrigierter Literaturwert	STD-Rauschen
OLSA [4]	-7,1	-7,1	-6,3
FBE [27]	-1,3	-7,8	-8,1
GÖSA [5]	-6,1	-6,1	-5,6
WAKO [12]	-11,3	-11,3	-11,6
HSM [6]	-7,2	-12,6	-12,9

Auch im SAN-Rauschen tritt die größte vorkommende Differenz zwischen den Medianen des HSM und GÖSA auf, hierfür beträgt der SRT_{50} im HSM $-8,9$ dB SNR und im GÖSA $-6,2$ dB SNR. Die einfaktorielle ANOVA konnte ebenfalls signifikante Unterschiede nachweisen ($F(1,95; 41,00)=34,13; p<0,001$). Im PostHoc-t-Test wurde jedoch deutlich, dass sich die signifikanten Unterschiede auf die beiden Sprachtests FBE und GÖSA im Vergleich zu den drei anderen Sprachtests (OLSA, HSM, WAKO) beschränken (siehe Tabelle 4, blau). Die Unterschiede zwischen FBE und GÖSA sowie zwischen OLSA, HSM und WAKO sind nicht signifikant.

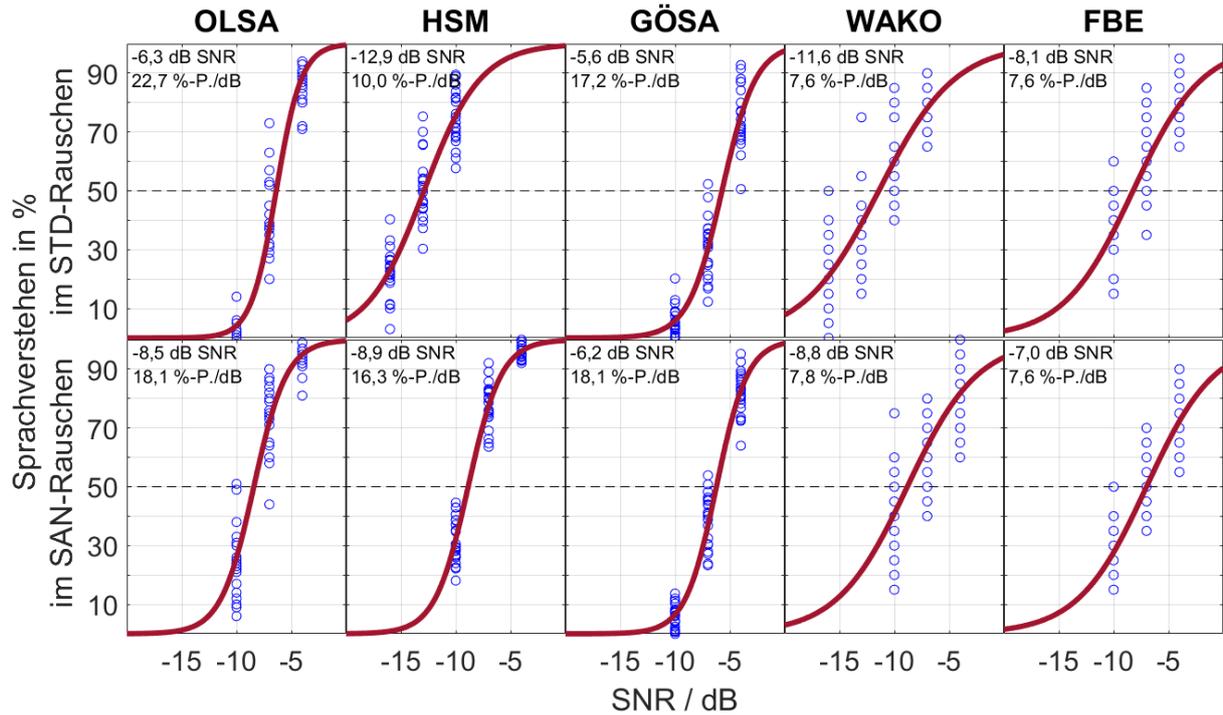


Abbildung 3: Sprachverstehen in % (blaue Kreise) für alle Probanden und alle Sprachtests im Standard-Rauschen (STD, oben) und im selbstgenerierten sprachsimulierenden Rauschen (SAN-Rauschen, unten); an die Messergebnisse wurden Diskriminationsfunktionen angepasst (rot). Die Zahlen in den Graphiken geben den SRT_{50} in dB SNR und die Steigung s_{50} in %-Punkte/dB an.

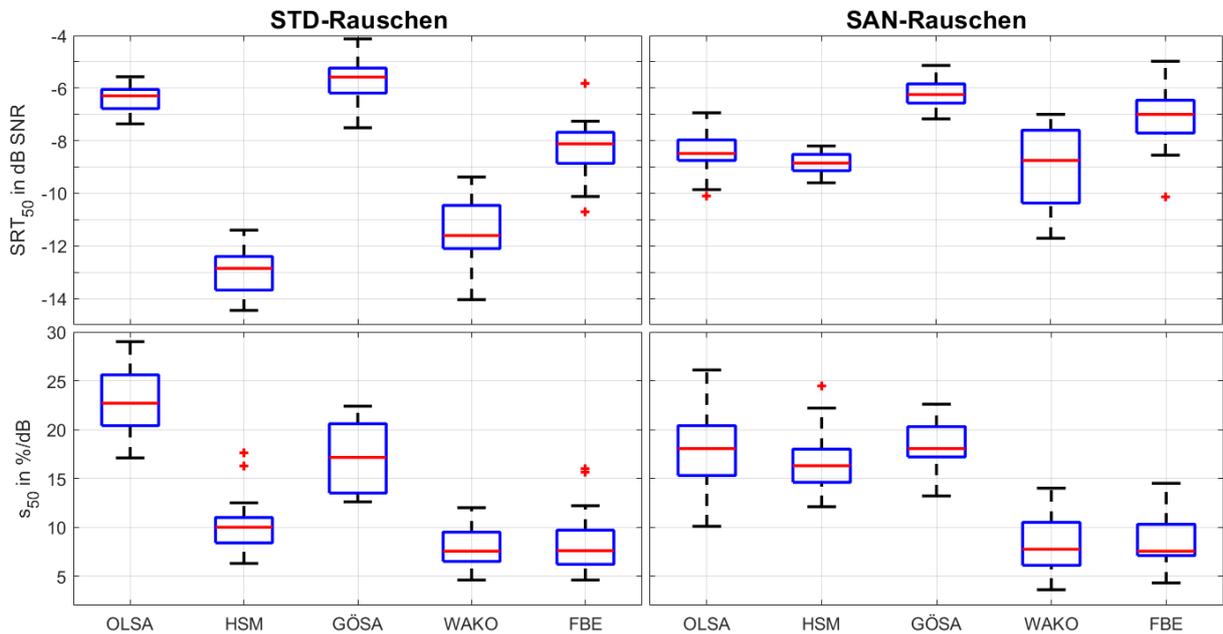


Abbildung 4: Individuelle Schwellen im Sprachverstehen (SRT_{50} , oben) und entsprechende Steigungen (s_{50} , unten) für alle Probanden und Sprachtests im Standardrauschen (STD-Rauschen, links) und im selbstgenerierten sprachsimulierenden Rauschen (SAN-Rauschen, rechts)

Tabelle 4: Statistische Auswertung (p-Werte) der Schwellen im Sprachverstehen SRT_{50} (blau) und der zugehörigen Steigungen s_{50} (rot) aller Sprachtests im sprachsimulierenden (SAN-)Rauschen

	OLSA	FBE	GÖSA	WAKO	HSM
OLSA	–	0,001*	<0,001*	0,949	0,065
FBE	<0,001*	–	0,017	<0,001*	<0,001*
GÖSA	1,000	<0,001*	–	<0,001*	<0,001*
WAKO	<0,001*	1,000	<0,001*	–	1,000
HSM	1,000	<0,001*	1,000	<0,001*	–

Mit *-gekennzeichnete Werte sind signifikant unter Berücksichtigung der Bonferroni-Korrektur.

Die Steigungen s_{50} weisen im STD-Rauschen einen Unterschied von bis zu 15,1%-Punkten/dB auf. Die Steigung des HSM befindet sich dabei in der Größenordnung der Worttests (FBE, WAKO). Die einfaktorielle ANOVA resultiert in signifikanten Unterschieden zwischen den Sprachtests ($F(4; 84,00)=128,91; p<0,001$). Im PostHoc-t-Test bestätigen sich die Unterschiede von OLSA und GÖSA zu den drei weiteren Sprachtests (HSM, FBE, WAKO) (siehe Tabelle 2, rot). Zwischen den drei letztgenannten Sprachtests konnten jedoch keine signifikanten Unterschiede nachgewiesen werden. Im SAN-Rauschen kommt es zu einer Erhöhung der Steigung im HSM um 6,3%-Punkte/dB und dadurch zu einer neuen Gruppierung, wobei sich die Satz- von den Worttests trennen. Die Steigungen der Satztests unterscheiden sich im SAN-Rauschen um maximal 1,8%-Punkte/dB und diejenigen der Worttests um 0,2%-Punkte/dB. Die einfaktorielle ANOVA ergibt einen signifikanten Unterschied ($F(4; 84,00)=62,17; p<0,001$), der im PostHoc-t-Test die Trennung in Satztests und Worttests bestätigt (siehe Tabelle 4, rot). Innerhalb der Wort- bzw. Satztests treten keine signifikanten Unterschiede in den Steigungen auf.

Beim Vergleich der SRT_{50} -Werte zwischen STD- und SAN-Rauschen ergab der t-Test für OLSA, HSM, GÖSA und WAKO jeweils einen signifikanten Unterschied (Tabelle 5). Der entsprechende Vergleich für den FBE zeigt keinen signifikanten Unterschied. Der t-Test zum Vergleich der s_{50} im STD- und SAN-Rauschen ergab nur für OLSA und HSM signifikante Unterschiede (siehe Tabelle 5).

Tabelle 5: p-Werte für den paarweisen Vergleich im t-Test der SRT_{50} und s_{50} -Werte für STD- und SAN-Rauschen der Sprachtests OLSA, FBE, GÖSA, WAKO und HSM

Vergleich STD-SAN	SRT_{50}	s_{50}
OLSA	<0,001*	<0,001*
FBE	0,057	1,000
GÖSA	0,010*	1,000
WAKO	0,010*	1,000
HSM	<0,001*	<0,001*

Mit *-gekennzeichnete Werte sind signifikant unter Berücksichtigung der Bonferroni-Korrektur.

SII-Berechnung

Abbildung 5 zeigt für alle Sprachtests im STD- und SAN-Rauschen das prozentuale Sprachverstehen in Abhängigkeit vom SII. Für den WAKO ergibt sich, unabhängig vom verwendeten Störgeräusch, ein beinahe linearer Zusammenhang zwischen dem Sprachverstehen und SII. Die Graphen des FBE, OLSA, GÖSA und HSM zeigen in Abhängigkeit vom verwendeten Störgeräusch leichte Abweichungen. Jedoch ist trotz der Verwendung verschiedener Rauschen ein gemeinsamer Anstieg des Sprachverstehens mit zunehmendem SII erkennbar.

Für 50%iges Sprachverstehen ergibt sich für alle Sprachtests ein SII zwischen 0,17 (HSM-STD) und 0,34 (FBE-SAN). Die größte Differenz zwischen den SII-Werten eines Sprachtests im STD- und SAN-Rauschen liegt bei 0,04 (OLSA), während die kleinste Differenz 0,01 (WAKO) beträgt.

Diskussion

Ziel dieser Studie war es, die Vergleichbarkeit verschiedener Sprachtests im Störgeräusch zu untersuchen. Dazu wurden neben den Messungen im STD-Rauschen auch Messungen im SAN-Rauschen durchgeführt, das durch 30fache Überlagerung des Sprachmaterials für jeden Sprachtest generiert wurde. Um eine bessere Vergleichbarkeit der Ergebnisse zu ermöglichen, erfolgte eine identische Kalibrierung aller Sprachmaterialien und Störgeräusche auf den RMS. Die Messungen wurden mit jungen normalhörenden Probanden durchgeführt, sodass eine Abhängigkeit der Ergebnisse vom individuellen Hörvermögen der Probanden ausgeschlossen werden kann. Zur Anpassung von Diskriminationsfunktionen wurden die zu messenden SNR-Werte für jeden Sprachtest im STD- und SAN-Rauschen so bestimmt, dass das Sprachverstehen sowohl unter- als auch oberhalb von 50% lag. Auf diese Weise konnten für jeden Sprachtest im STD- und SAN-Rauschen der SRT_{50} und die Steigung s_{50} ermittelt werden.

STD-Rauschen

Bei Verwendung der STD-Rauschen kommt es sowohl zwischen den SRT_{50} -Werten als auch zwischen den s_{50} der verschiedenen Sprachtests zu großen Unterschieden. Diese betragen für die SRT_{50} -Werte bis zu 7,3 dB SNR und für die s_{50} bis zu 15,1%-Punkte/dB. Unter der Voraussetzung, dass beim Vergleich der gemessenen SRT_{50} -Werte mit den Literaturwerten die verschiedenen Kalibrierungen berücksichtigt werden, beschränken sich die auftretenden Differenzen auf lediglich 0,8 dB. Dabei ist zu bedenken, dass die Präsentationsart des Störgeräuschs (kontinuierlich oder synchron) zu Unterschieden von ca. 1,4 dB führen kann [12]. Welche Präsentationsart zur Bestimmung der Literaturwerte für OLSA, HSM und GÖSA verwendet wurde, ist nicht dokumentiert. Nur für den Literaturwert des WAKO ist eindeutig beschrieben,

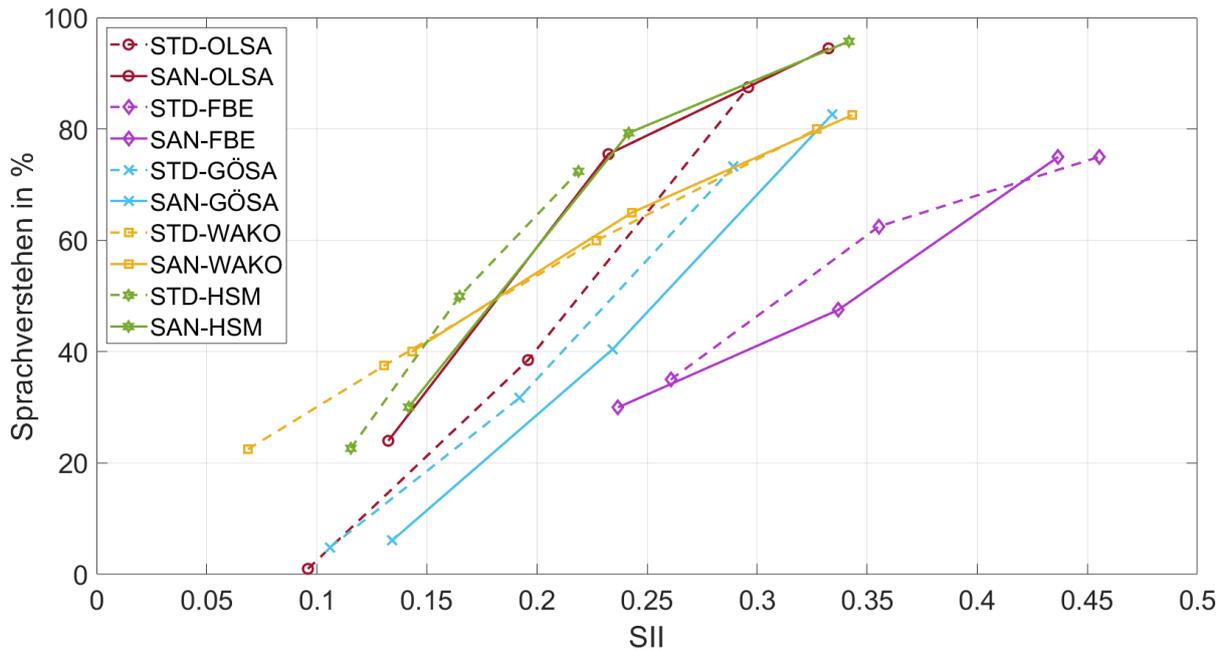


Abbildung 5: Medianes Sprachverstehen in Abhängigkeit vom SII bei Gleichgewichtung aller Terzbänder für alle Sprachtests im STD-Rauschen (o) und im SAN-Rauschen (x); die Symbole geben die Ergebnisse beim jeweiligen gemessenen SNR an.

dass dieser durch synchrone Präsentation des Störgeräuschs ermittelt wurde, bei der das Störgeräusch 0,5 s vor Beginn des Ankündigungssatzes einsetzte und 0,5 s danach endete [12].

Eine weitere Einflussgröße ist die Darbietungsart (monaural über Kopfhörer, diotisch über Kopfhörer, binaural im Freifeld). Durch den Lautstärkezuwachs, der beim Übergang von monauralem zu binauralem Hören entsteht, verschieben sich nach [27] die gemessenen SRT_{50} -Werte um ca. 2,5 dB zu niedrigeren Werten. Die Literaturdaten zum FBE und WAKO entstammen Freifeldmessungen mit Präsentation aus der 0° -Richtung, wie in dieser Studie. Die Literaturwerte für HSM und OLSA wurden mit vergleichbarer diotischer Kopfhörerdarbietung erhoben. Für den GÖSA nennen Kollmeier und Wesselkamp [5] die Kopfhörer-Präsentationsart „monaural“, während Wesselkamp [28] für die gleichen Daten „diotisch“ angibt. Da der Unterschied zwischen dem Literaturwert und dem Ergebnis dieser Untersuchung nur 0,5 dB beträgt, ist von einer diotischen Präsentationsart bei [5] auszugehen. Somit ist für keinen Sprachtest eine Korrektur für die Darbietungsart erforderlich.

Insgesamt konnten für die fünf Sprachtests bei Berücksichtigung der unterschiedlichen Kalibrierung die Literaturwerte mit nur geringen Abweichungen reproduziert werden. Zwischen den Sprachtests treten jedoch große Unterschiede auf. Als ein wesentlicher Einflussfaktor wurde die unterschiedliche Verdeckungswirkung der Störgeräusche vermutet, die sich aus den in Abbildung 1 dargestellten spektralen Unterschieden ergibt. Deshalb wurden für jeden Sprachtest neue sprachsimulierende Störgeräusche erzeugt.

SAN-Rauschen

Mit Hilfe der 30fachen Überlagerung der Sprachmaterialien konnte für jeden Sprachtest ein entsprechendes sprachsimulierendes Störgeräusch erzeugt werden. Die frequenzabhängigen Unterschiede zwischen SAN-Rauschen und dem jeweiligen Sprachmaterial betragen maximal 0,5 dB, so dass die SAN-Rauschen nach [19] auch zur Kalibrierung der Sprachmaterialien geeignet sind.

Insgesamt zeigte sich eine Angleichung der SRT_{50} -Werte bei der Verwendung der SAN- im Vergleich zu den STD-Rauschen. Die größte Verschiebung ergab sich für die beiden Sprachtests, die im STD-Rauschen die geringsten Schwellen aufwiesen: 4,0 dB für den HSM und 2,8 dB für den WAKO. Die beobachteten Veränderungen bestätigen die signifikante Abhängigkeit des Sprachverstehens vom verwendeten Störgeräusch sowohl für Einsilber [12] als auch für Satztests und höhere Schwellen bei besserer Verdeckung. Durch den Angleich der Rauschspektren an die Spektren der Sprache und durch dieselbe Kalibrierung lässt sich also die Vergleichbarkeit der Messergebnisse der verschiedenen Sprachtests verbessern.

Obwohl es sich bei dem zur Messung des OLSA verwendeten STD-Rauschen, dem OLnoise, bereits um ein sprachsimulierendes Störgeräusch handelt, das vergleichbar zum SAN-Rauschen generiert wurde, ergaben sich signifikante Unterschiede der SRT_{50} -Werte in Abhängigkeit vom Störgeräusch. Beim Vergleich der Spektren von Sprache und STD-Rauschen zeigt sich im Frequenzbereich von 0,5 bis 10 kHz ein bis zu 2,8 dB höherer Pegel des STD-Rauschens, der vermutlich zu einer höheren Verdeckung führt. Zusätzlich berichteten manche Probanden von einer höheren Schwierigkeit bei der Messung im STD-Rauschen, da eine weitere, nicht verständliche Stimme im Störgeräusch von der zu verstehenden Sprache ablen-

ken würde. Diese Beobachtung kann durch eine messtechnische Analyse unterstützt werden, die für das STD-Rauschen zwar einen sehr geringen, aber etwas höheren Modulationsgrad als für das SAN-Rauschen aufzeigt. Die Ursache für die Modulationsunterschiede zwischen STD- und SAN-Rauschen liegt vermutlich bei der geringfügig verschiedenen Generierung der Störgeräusche. Aus [2] geht hervor, dass zur Erzeugung des STD-Rauschens die einzelnen Wörter des OLSA aneinander gehängt und überlagert wurden. Dagegen wurden für das SAN-Rauschen ganze Sätze aneinander gereiht und zu Rauschen aufaddiert [24]. Dadurch kommt es bei der Rauscherzeugung zu verschiedenen Verhältnissen von Sprache und Pausen, wobei der Pausenanteil im STD-Rauschen höher ist. Durch zufällige konstruktive Überlagerungen kann es so zu einer Modulation des Rauschens kommen, die zusätzlich maskierend wirken kann.

Außerdem kommt es durch die Verwendung der SAN- im Vergleich zu den STD-Rauschen auch zu einer Verschiebung der Steigungen für OLSA und HSM. Die hohen Steigungen für den OLSA im STD-Rauschen resultieren möglicherweise für manche Probanden aus Bodeneffekten beim niedrigsten SNR. Die Differenz beim HSM kann dagegen mit der Reduzierung der spektralen Abweichungen zwischen Sprache und Störgeräusch bei der Verwendung des SAN-Rauschens erklärt werden. Bei Verwendung des STD-Rauschens führen die großen spektralen Unterschiede im tieffrequenten Bereich (bis ca. 400 Hz) zu einem Teilverstehen der Sätze auch bei niedrigem SNR. Für hohe prozentuale Verstehenswerte ist jedoch ein deutlich höherer SNR notwendig, da das standardmäßig zur Messung verwendete CCITT-Rauschen vor allem im Hauptsprachbereich (500 Hz bis 4 kHz) eine höhere Leistungsdichte besitzt als das Sprachmaterial. Dagegen ist im SAN-Rauschen auch für ein geringes Verstehen ein höherer SNR notwendig, während für gutes Verstehen ein ähnlicher SNR wie im STD-Rauschen ausreicht. Dadurch kommt es zu einer deutlich steileren Diskriminationsfunktion mit einem Steigungsunterschied zum STD-Rauschen von 6,3%-Punkten/dB.

Verbliebene Unterschiede zwischen Sprachtests

Insgesamt verbleiben bei Verwendung der SAN-Rauschen und gleicher Kalibrierung Unterschiede im SRT_{50} von weniger als 3 dB. Der größte Unterschied von 2,7 dB wurde zwischen HSM und GÖSA beobachtet, obwohl beide Sprachtests aus Alltagssätzen bestehen. Der HSM wurde im Vergleich zum GÖSA mit deutlicherer Artikulation und langsamerer Sprechrate aufgenommen (HSM: 222 Silben/min, GÖSA: 279 Silben/min) [29]. Eine messtechnische Analyse zeigt einen größeren Dynamikbereich mit längeren Sprechpausen und um 2 dB höhere Spitzenpegel für das HSM-Sprachmaterial bei gleichem RMS-Wert. Die Länge der Sprachpausen beeinflusst die Berechnung des RMS-Werts und führt zu einer höheren Verstärkung des HSM-Sprachmaterials im Vergleich zum GÖSA-

Sprachmaterial und SAN-Rauschen und somit zu besseren Schwellen.

Der OLSA wurde von dem gleichen Sprecher wie der WAKO und der GÖSA aufgesprochen. Gegenüber dem GÖSA weist der OLSA einen um 2,3 dB geringeren SRT_{50} auf. Es wird vermutet, dass die besseren Schwellen nicht auf die reduzierte Sprechrate des OLSA von 233 Silben/min, sondern auf die Matrixstruktur mit begrenztem Sprachmaterial zurückzuführen ist. Die Reduktion im SRT_{50} entspricht in etwa dem erwarteten Trainingseffekt nach zwei Testlisten von max. 2 dB [4].

Ein ähnlicher Unterschied von 1,8 dB ist zwischen WAKO und FBE zu beobachten, der vermutlich durch die geschlossene Durchführung des WAKO bedingt ist. Obwohl die Ergebnisse des WAKO um die Ratewahrscheinlichkeit von 20%, die durch die geschlossene Durchführung entsteht, korrigiert wurden, führt die Antwortauswahl zu besseren (niedrigeren) SRT_{50} -Werten. Dies kann damit begründet werden, dass die Probanden durch die Antwortauswahl wissen, welcher Teil des Wortes variabel ist. Damit reicht es aus, die Konzentration nur auf den Teil des Wortes zu lenken, der sich in den Antwortmöglichkeiten unterscheidet. Im Gegensatz dazu muss zum Verstehen eines Wortes im FBE die Aufmerksamkeit auf dem ganzen Wort liegen, um dieses korrekt wiederholen zu können. Ein weiterer Unterschied ist der Ankündigungssatz im WAKO, der die Aufmerksamkeit auf einen Zeitpunkt fokussieren könnte, obwohl Mallinger keinen signifikanten Unterschied bei Verwendung eines kurzen Ankündigungstons beobachtete [30].

Während sich die SRT_{50} -Werte aller Sprachtests unter gleichen Messbedingungen annähern, weisen die Steigungen s_{50} abhängig von der Art des Sprachmaterials deutliche signifikante Unterschiede auf (s. Tabelle 4). Die Steigungen der Worttests liegen bei 7,6%-Punkten/dB (WAKO) bzw. 7,8%-Punkten/dB (FBE). Dagegen erreichen die Steigungen der Satztests Werte zwischen 16,3 und 18,1%-Punkten/dB. Die Differenz der Steigungen zwischen Satz und Worttests liegt damit bei über 8,5%-Punkten/dB. Dieser Wert lässt sich vermutlich auf die größere Streuung im Wortverstehen bei den Einzelworttesten im Vergleich zum Verstehen in den Satztesten zurückführen [3]. Beim OLSA und GÖSA wurden die Diskriminationsfunktionen der Wörter innerhalb der Sätze aneinander angenähert, so dass geringere Unterschiede bestehen und gleichzeitig mehr Wörter innerhalb einer Testliste getestet werden. Daraus könnte gefolgert werden, dass die Diskriminationsfunktion des OLSA aufgrund des höheren j -Faktors gegenüber dem GÖSA eine höhere Steigung aufweisen müsste. Dies wurde jedoch im SAN-Rauschen genauso wenig beobachtet wie eine geringere s_{50} des HSM, für den keine Optimierung der Wortdiskriminationsfunktionen bekannt ist.

Modellierung mit SII

Die Modellierung des Sprachverstehens mittels SII lag für den WAKO mit beiden Störgeräuschen ähnlich wie bei Winkler und Holube [12] auf einer Kurve. Die SII-Werte

der anderen Sprachtests weisen jedoch Abweichungen abhängig vom verwendeten Störgeräusch auf. Dabei muss jedoch berücksichtigt werden, dass für alle Sprachtests bei der SII-Berechnung eine Gleichgewichtung der Frequenzbänder verwendet wurde. Andere ANSI-Gewichtungsfunktionen führten zu noch größeren Differenzen für manche der Sprachtests, so dass es notwendig erscheint, neue Gewichtungsfunktionen für die Sprachtests im Deutschen zu ermitteln. Insgesamt kann jedoch eine monotone Erhöhung des SII bei steigendem Sprachverstehen innerhalb der Sprachtests für beide Störgeräusche beobachtet werden, so dass die wesentlichen Unterschiede zwischen den Störgeräuschen aufgrund der unterschiedlichen Spektren durch den SII nachbildbar sind.

Fazit

Da sich die Messergebnisse der Sprachtests unter gleichen Messbedingungen aneinander annähern ist anzunehmen, dass die Wahl des Sprachmaterials eine untergeordnete Rolle spielt. Die verbleibenden Unterschiede im SRT_{50} von maximal 2,7 dB sind vermutlich durch die Eigenschaften der einzelnen Sprachtests (Trainingseffekt, offene/geschlossene Testdurchführung, ungleiche Dynamik/Spitzenpegel) bedingt. Ist eine hohe Genauigkeit bei einer adaptiven Pegelsteuerung zur Ermittlung des SRT_{50} erwünscht, bieten sich Satztests an, da diese eine um ca. 10%-Punkte/dB höhere Steigung als Worttests aufweisen. Insgesamt ergibt sich durch die Verwendung von Störgeräuschen mit gleicher Verdeckungswirkung, gleicher Kalibrierung und Berücksichtigung der Messbedingungen eine bessere Vergleichbarkeit der Ergebnisse.

Notes

Interessenkonflikte

Die Autorinnen erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Danksagung

Diese Untersuchung wurde gefördert aus dem Projekt VIBHear mit Mitteln des Europäischen Fonds für regionale Entwicklung (EFRE) und Mitteln des Landes Niedersachsen. Vielen Dank an alle Probanden für ihre Teilnahme an den Messungen. Die Korrektur der englischsprachigen Textanteile erfolgte durch <http://www.stels-ol.de/>.

Literatur

- Hahlbrock KH. Über Sprachaudiometrie und neue Wörtertests. *Archiv für Ohren-, Nasen- und Kehlkopfheilkunde*. 1953;162(5):394-431. DOI: 10.1007/bf02105664
- Wagener K, Kühnel V, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Z Audiol*. 1999;38(1):4-15.
- Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil II: Optimierung des Oldenburger Satztests. *Z Audiol*. 1999;38(2):44-56.
- Wagener K, Brand T, Kollmeier B. Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests. *Z Audiol*. 1999;38(3):86-95.
- Kollmeier B, Wesselkamp M. Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment. *J Acoust Soc Am*. 1997 Oct;102(4):2412-21. DOI: 10.1121/1.419624
- Schmidt M, Hochmair-Desoyer I, Schulz E, Moser L. Der HSM-Satztest. In: 23. Jahrestagung für Akustik, Kiel. Berlin: DEGA e.V.; 1997. p. 93-4.
- von Wallenberg EL, Kollmeier B. Sprachverständlichkeitsmessungen für die Audiologie mit einem Reimtest in deutscher Sprache: Erstellung und Evaluation von Testlisten. *Audiol Acoust*. 1989;28(2):50-65.
- Kollmeier B, Lenarz T, Winkler A, Zokoll MA, Sukowski H, Brand T, Wagener KC. Hörgeräteindikation und -überprüfung nach modernen Verfahren der Sprachaudiometrie im Deutschen [Indication for and verification of hearing aid benefit using modern methods of speech audiometry in German]. *HNO*. 2011 Oct;59(10):1012-21. DOI: 10.1007/s00106-011-2345-5
- Boothroyd A, Nittrouer S. Mathematical treatment of context effects in phoneme and word recognition. *J Acoust Soc Am*. 1988 Jul;84(1):101-14. DOI: 10.1121/1.396976
- ITU. ITU Recommendation G.227 – Conventional telephone signal. Geneva: ITU; 1988 (Zugriff: 2019 Aug 08). Available from: <https://www.itu.int/rec/T-REC-G.227-198811-1/en>
- Müller C. Perzeptive Analyse und Weiterentwicklung eines Reimtestverfahrens für die Sprachaudiometrie [Dissertation]. Göttingen: Georg-August-Universität; 1992.
- Winkler A, Holube I. Einfluss des Störgeräusches auf das Sprachverstehen von Einsilbern. *Z Audiol*. 2018;57(4):138-47. DOI: 10.4126/FRL01-006412917
- American National Standards Institute. ANSI S3.5 Methods for calculation of the Speech Intelligibility Index. 1997.
- Beutelmann R, Brand T. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am*. 2006 Jul;120(1):331-42. DOI: 10.1121/1.2202888
- Rhebergen KS, Lyzenga J, Dreschler WA, Festen JM. Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing. *J Acoust Soc Am*. 2010 Mar;127(3):1570-83. DOI: 10.1121/1.3291000
- Deutsches Institut für Normung. DIN EN ISO 8253-3. Akustik – Audiometrische Prüfverfahren – Teil 3: Sprachaudiometrie. Berlin: Beuth; 2012.
- Winkler A, Schlüter A, Gebauer T, Seifert J, Tuschen L, Radeloff A, Holube I. Einfluss von Sprechtempo und Störgeräusch auf das Sprachverstehen im Göttinger und im HSM-Satztest. *GMS Z Audiol (Audiol Acoust)*. 2021;3:Doc02. DOI: 10.3205/zaud000014
- Winkler A, Holube I. Der Freiburger Einsilbertest und die Norm DIN EN ISO 8253-3: Technische Analyse. *Z Audiol*. 2016;55(3):106-13.
- Deutsches Institut für Normung. DIN EN ISO 8253-1. Akustik – Audiometrische Prüfverfahren – Teil 1: Grundlegende Verfahren der Luft- und Knochenleitungs-Schwellenaudiometrie mit reinen Tönen (ISO 8253-1:2010-11). Berlin: Beuth Verlag; 2011.
- Technische Universität Dresden. 2021 (Zugriff: 2019 Dez 13). Zugriff unter: <https://versuch.file2.wcms.tu-dresden.de/w/index.php/Item-Ratekorrektur>

21. Deutsches Institut für Normung. DIN 45626-1. Tonträger mit Sprache für Gehörprüfung Teil 1: Tonträger mit Wörtern nach DIN 45621-1. Berlin: Beuth; 1995.
22. Brand T. Persönliche Mitteilung an die Autoren. 2019.
23. Festen JM, Plomp R. Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am.* 1990 Oct;88(4):1725-36. DOI: 10.1121/1.400247
24. Wagener K, Josvassen JL, Ardenkjaer R. Design, optimization and evaluation of a Danish sentence test in noise. *Int J Audiol.* 2003 Jan;42(1):10-7. DOI: 10.3109/14992020309056080
25. Zinner C, Winkler A, Holube I. Speech Adjusted Noises (SAN) for German speech recognition tests. 2021. DOI: 10.5281/zenodo.4609783
26. Brand T, Kollmeier B. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *J Acoust Soc Am.* 2002 Jun;111(6):2801-10. DOI: 10.1121/1.1479152
27. Brinkmann K, Diestel H. Untersuchung an Sprachaudiometern Teil III. Ergebnisse von Hörprüfungen. *Zeitschrift für Hörgeräte-Akustik.* 1970; 9: 114-26.
28. Wesselkamp M. Messung und Modellierung der Verständlichkeit von Sprache [Dissertation]. Göttingen: Georg-August-Universität;1994.
29. Müller-Deile J. Sprachverständlichkeitsuntersuchungen bei Kochleaimplantatpatienten [Speech intelligibility tests in cochlear implant patients]. *HNO.* 2009 Jun;57(6):580-92. DOI: 10.1007/s00106-009-1930-3
30. Mallinger E. Trainingseffekte und Listenäquivalenz des Freiburger Einsilbertests im Störschall [Dissertation]. Erlangen-Nürnberg: Friedrich-Alexander-Universität; 2011.
31. Winkler A, Holube I, Husstedt H. Der Freiburger Einsilbertest im Störschall [The Freiburg monosyllabic speech test in noise]. *HNO.* 2020 Jan;68(1):14-24. DOI: 10.1007/s00106-019-00763-6

Korrespondenzadresse:

Christina Zinner

Institut für Hörtechnik und Audiologie, Jade Hochschule, Ofener Straße 16/19, 26121 Oldenburg, Deutschland, Tel: +49 441 7708-3741

Christina.Zinner@uni-oldenburg.de

Bitte zitieren alsZinner C, Winkler A, Holube I. Vergleich von fünf Sprachtests im sprachsimulierenden Störgeräusch. *GMS Z Audiol (Audiol Acoust).* 2021;3:Doc04.

DOI: 10.3205/zaud000016, URN: urn:nbn:de:0183-zaud0000162

Artikel online frei zugänglich unter<https://www.egms.de/en/journals/zaud/2021-3/zaud000016.shtml>**Veröffentlicht:** 21.06.2021**Copyright**©2021 Zinner et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.