

EUROPÄISCHE UNION DER HÖRAKUSTIKER e. V.

# Förderpreis 2023

## Evaluation von Messverfahren für Sprachverständlichkeit und Höranstrengung in akustisch komplexen Szenen

Bachelorarbeit

Verfasserin: Sophie Jäker

Erstprüferin: Prof. Dr. rer. nat. Inga Holube

Zweitprüfer: Dr. rer. nat. Hendrik Kayser

Datum der Abgabe: 10. März 2023

**EUHA**

Europäische Union der  
Hörakustiker e.V.

Bachelorarbeit  
zur Erlangung des akademischen Grades  
Bachelor of Engineering  
Fachbereich Bauwesen Geoinformation Gesundheitstechnologie  
Abteilung Technik und Gesundheit für Menschen  
Jade Hochschule Wilhelmshaven/Oldenburg/Elsfleth  
Standort Oldenburg

Herausgeber: Europäische Union der Hörakustiker e. V.  
Saarstraße 52, 55122 Mainz, Deutschland  
Tel. +49 (0)6131 28 30-0  
Fax +49 (0)6131 28 30-30  
E-Mail: [info@euha.org](mailto:info@euha.org)  
Internet: [www.euha.org](http://www.euha.org)

Alle hier vorhandenen Daten, Texte und Grafiken sind urheberrechtlich geschützt. Eine Verwertung über den eigenen privaten Bereich hinaus ist grundsätzlich genehmigungspflichtig.

© EUHA 2023

# ZUSAMMENFASSUNGEN

## Deutsch

Die Sprachverständlichkeit wird gemessen, um eine Hörminderung zu diagnostizieren oder um den Nutzen von Hörgeräten zu bewerten. Typischerweise werden solche Messungen in einfachen akustischen Laborsituationen durchgeführt, d. h. Darbietung eines Zielsignals von vorne und maximal eine stationäre Störgeräuschquelle aus derselben Richtung (SON0) oder von einer Seite (SON90). Diese Labormessbedingungen spiegeln jedoch nicht die realen Hörumgebungen und Kommunikationssituationen von Menschen im Alltag wider. So werden beispielsweise Aspekte des räumlichen Hörens, wie binaurale Demaskierung und Hören mit dem besseren Ohr, im Vergleich zu realistischen Szenen nicht berücksichtigt (SON0) oder überspitzt dargestellt (SON90). Darüber hinaus wird bei einer Sprachaudiometrie nur die Spracherkennung berücksichtigt, die nur einen Aspekt der Sprachwahrnehmung abdeckt. Ein weiteres Maß, das Aufschluss über die Sprachwahrnehmung gibt, ist die Höranstrengung, die Aspekte der Kommunikation erfasst, auch wenn keine Abnahme der Spracherkennung gemessen wird. Als bekanntes Messverfahren für die Sprachverständlichkeit ist der Oldenburger Satztest (OLSA) und für die Messung der Höranstrengung das Adaptive Categorical Listening Effort Scaling (ACALES)-Verfahren bereits in Standardlaborsituationen validiert. Ziel dieser Arbeit war es, herauszufinden, ob die beiden Verfahren auch zur Anwendung in akustisch, komplexen Szenen geeignet sind. Dazu wurden Messungen zur Test-Retest-Reliabilität mit 15 Normalhörenden durchgeführt. Zudem werden interindividuelle Unterschiede der Messergebnisse mit Ergebnissen aus kognitiven Tests verglichen. Die komplexen Szenen unterscheiden sich durch die Anzahl und räumliche Anordnung der Störgeräuschquellen, sowie durch die Art des Störgeräuschs. Die Darbietung der Messsignale erfolgte mithilfe von virtueller Akustik. Die Ergebnisse werden vorgestellt und diskutiert. Insgesamt konnte festgestellt werden, dass das ACALES-Verfahren reproduzierbare Ergebnisse in allen gemessenen akustisch komplexen Hörsituationen liefern. Dies wurde mittels dem Intraklassen-Korrelationskoeffizienten bestimmt. Beim OLSA weisen Szenen mit räumlicher Trennung und/oder informationeller Maskierung keine guten Reproduzierbarkeiten im Gegensatz zu den anderen Szenen auf. Außerdem korreliert die selektive Aufmerksamkeit negativ mit der Höranstrengung. Je besser die selektive Aufmerksamkeit, desto höher ist die subjektiv bewertete Höranstrengung.

## English

Speech intelligibility is measured to diagnose hearing loss or to evaluate the benefit of hearing aids. Typically, such measurements are performed in simple acoustic laboratory situations, i.e., presentation of a target signal from the front and maximal one stationary noise source from the same direction (S0N0) or from one side (S0N90). However, these laboratory measurement conditions do not reflect the real listening environments and communication situations of people in everyday life. For example, aspects of spatial hearing, such as binaural desmasking and hearing with the better ear, are not considered (S0N0) or exaggerated (S0N90) compared to realistic scenes. Furthermore, speech audiometry only considers speech recognition, which covers only one aspect of speech perception. Another measure that provides information about speech perception is listening effort, which captures aspects of communication even if no decrease in speech recognition is measured. As a well-known measure of speech intelligibility, the Oldenburg Satz Test (OLSA) and for the measurement of listening effort, the Adaptive Categorical Listening Effort Scaling (ACALES) has already been validated in standard laboratory situations. The aim of this work was to find out whether the two procedures are also suitable for application in acoustically complex scenes. For this purpose, test-retest stability measurements were performed with 15 normal-hearing subjects. In addition, interindividual differences of the measurement results are compared with results from a cognitive test, Go/NoGo. The complex scenes differ in the number and spatial arrangement of the noise sources, as well as in the type of noise. Virtual acoustics were used to present the measurement signals. The results are presented and discussed. Altogether, it was found that the ACALES method provided reproducible results in all measured acoustically, complex listening situations. This was determined using the intraclass correlation coefficient. For the OLSA, scenes with spatial separation and/or informational masking do not show good reproducibilities in contrast to the other scenes. Selective attention correlates negatively with listening effort. The better the selective attention, the higher the subjectively rated listening effort.

**Keywords: Sprachverstehen, Höranstrengung, akustisch komplexe Szenen, OLSA, ACALES, Test-Retest-Reliabilität, kognitive Leistung, Go/NoGo, geteilte Aufmerksamkeit, Trail-Making-Test**

# INHALTSVERZEICHNIS

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretischer Hintergrund</b>	<b>3</b>
2.1	Sprachverstehen . . . . .	3
2.1.1	Freiburger Einsilbertest . . . . .	3
2.1.2	Göttinger Satztest . . . . .	4
2.1.3	Oldenburger Satztest . . . . .	4
2.2	Höranstrengung . . . . .	5
2.2.1	Physiologische Messungen . . . . .	5
2.2.2	Kognitive Messungen . . . . .	7
2.2.3	Fragebögen und Skalen . . . . .	7
2.2.4	ACALES . . . . .	7
2.3	Kognition und Sprachwahrnehmung . . . . .	8
2.4	Einflussfaktoren auf die Sprachwahrnehmung . . . . .	9
2.4.1	Spatial Release from Masking . . . . .	9
2.4.2	Dip Listening . . . . .	10
2.4.3	Nachhall . . . . .	10
2.4.4	Energetische und informationelle Maskierung . . . . .	11
<b>3</b>	<b>Methodik</b>	<b>12</b>
3.1	Versuchspersonen . . . . .	12
3.2	Stimuli . . . . .	12
3.3	Akustisch komplexe Szenen . . . . .	14
3.4	Ermittlung der Höranstrengung . . . . .	14
3.5	Ermittlung des Sprachverstehens . . . . .	16
3.6	Kognitive Tests . . . . .	16
3.7	Messapparatur . . . . .	18
3.8	Kalibrierung des Messaufbaus . . . . .	18
3.9	Messablauf . . . . .	19
3.10	Statische Analysen . . . . .	21
<b>4</b>	<b>Ergebnisse und Auswertung</b>	<b>22</b>
4.1	Test-Retest-Reliabilität . . . . .	22
4.1.1	OLSA . . . . .	22
4.1.2	ACALES . . . . .	26
4.1.3	Go/NoGo . . . . .	29

4.2	Korrelation geteilte Aufmerksamkeit und OLSA . . . . .	29
4.3	Korrelation geteilte Aufmerksamkeit und ACALES . . . . .	30
4.4	Korrelation TMT und OLSA . . . . .	31
4.5	Korrelation TMT und ACALES . . . . .	32
<b>5</b>	<b>Diskussion</b>	<b>33</b>
5.1	Test-Retest Reliabilität OLSA . . . . .	33
5.2	Test-Retest-Reliabilität ACALES . . . . .	35
5.3	Test-Retest Reliabilität Go/NoGo . . . . .	35
5.4	Korrelation Kognition und Sprachwahrnehmung . . . . .	36
<b>6</b>	<b>Fazit und Ausblick</b>	<b>38</b>
<b>A</b>	<b>Instruktionen</b>	<b>40</b>
<b>B</b>	<b>Testvorlage Trail Making Test</b>	<b>45</b>
<b>C</b>	<b>Ergebnisse OLSA</b>	<b>47</b>
<b>D</b>	<b>Ergebnisse ACALES</b>	<b>50</b>
<b>E</b>	<b>geteilte Aufmerksamkeit</b>	<b>51</b>
<b>F</b>	<b>Trail-Making-Test</b>	<b>53</b>
<b>G</b>	<b>Geräteliste</b>	<b>55</b>
	<b>Glossar</b>	<b>56</b>
	<b>Literatur</b>	<b>59</b>

# ABBILDUNGS- UND TABELLENVERZEICHNIS

## Abbildungsverzeichnis

3.1	Mittlere Hörschwelle der Versuchspersonen für die rechte (o) und linke (x) Ohrseite. Der grau hinterlegte Bereich gibt die Minimal- bzw. Maximalwerte an. . .	13
3.2	13-stufige Bewertungsskala für die subjektive Höranstrengungsmessung. Neben den acht ausgeschriebenen Stufen („müheles“ usw.) verfügt die Skala auch über sechs unbenannte Zwischenschritte (:). (Quelle: Software Oldenburger Messprogramm (OMA) F&E 2.3). . . . .	15
3.3	Anordnung der 16 Lautsprecher zur Messung der subjektiven Höranstrengung und des Sprachverstehens. Jeder Lautsprecher steht in einem Winkel von 22,5°-Abständen kreisförmig zu einander. Der Radius des Lautsprecherkreises beträgt 1,5 m. . . . .	17
3.4	Schematisch dargestellter Zusammenhang des Messaufbaus zur Wiedergabe der Signale mittels des Oldenburger Messprogramms (OMA), Matlab und TASCAR.	19
3.5	Schematische Darstellung des zeitlichen Ablaufs des ersten und zweiten Termins der Studie. Die beiden Termine unterscheiden sich hinsichtlich des Ablaufs der Vorbereitung (Tonaudiogramm und Anamnese) und der Durchführung der kognitiven Tests. . . . .	20
4.1	Streudiagramme der Signal-Rausch-Abstände (SRTs) der Test- und Retest-Messungen aller Versuchspersonen (n=15) für alle 10 Szenen. Die Abkürzungen der Szenenbeschreibung S0 steht für das Nutzsignal aus der Richtung 0°, N0/N90/N+90/N180 für das Störgeräusch welches entweder aus der Richtung 0°, 90°, ±90° oder 180° präsentiert wird. IFnoise entspricht dem <i>International Female Noise</i> , ICRA dem sprachfluktuierendem Geräusch mit einer weiblichen Gewichtung (ICRA4-250), IFFM dem <i>International Female Fluctuation Masker</i> , Caf. dem Cafeteria-Rauschen und kein Nachh. bedeutet kein Nachhhall. . . . .	23
4.2	Boxplots der Sprachverständlichkeitsschwellen (SRTs) für jede der 10 Szenen (Tabelle 3.1). Für jede Szene wurden sowohl die Ergebnisse aus dem ersten Termin (Test, schwarz) als auch für den zweiten Termin (Retest, grau) dargestellt. Der schwarze bzw. graue horizontale Strich kennzeichnet den Median. Das obere bzw. untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Die Antennen haben eine Länge von höchstens dem 1,5-fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet. . . . .	25

4.3	Streudiagramme der Signal-Rausch-Abstände (SNRs) bei ESCU 11 der Test- und Retest-Messungen aller Versuchspersonen (n=15) für alle 10 Szenen. Die Abkürzungen der Szenenbeschreibung S0 steht für das Nutzsignal aus der Richtung 0°, N0/N90/N+-90/N180 für das Störgeräusch welches entweder aus der Richtung 0°, 90°, ±90° oder 180° präsentiert wird. IFnoise entspricht dem <i>International Female Noise</i> , ICRA dem sprachfluktuierendem Geräusch mit einer weiblichen Gewichtung (ICRA4-250), IFFM dem <i>International Female Fluctuation Masker</i> , Caf. dem Cafeteria-Rauschen und kein Nachh. bedeutet kein Nachhhall. . . . .	27
4.4	Boxplots des Signal-Rausch-Abstandes (SNRs) für jede der 10 Szenen (Tabelle 3.1). Für jede Szene wurden sowohl die Ergebnisse aus dem ersten Termin (Test, schwarz) als auch für den zweiten Termin (Retest, grau) dargestellt. Der schwarze bzw. graue horizontale Strich kennzeichnet den Median. Das obere bzw. untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Die Antennen haben eine Länge von höchstens dem 1,5-fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet. . . . .	28
4.5	Streudiagramm für den Go/NoGo-Test. Aufgetragen sind die Mittelwerte der Reaktionszeiten (RZ) des ersten Termins (Test) in ms gegenüber den Mittelwerten der RZ des zweiten Termins in ms. . . . .	30
A.1	Instruktion Go/NoGo Test (Quelle: Zimmermann und Fimm, 2013) . . . . .	40
A.2	Instruktion geteilte Aufmerksamkeit (Quelle: Zimmermann und Fimm, 2013) . . . . .	40
A.3	Instruktion für die Messung der subjektiven Höranstrengung . . . . .	41
A.4	Instruktion für die Messung des Sprachverstehens . . . . .	42
A.5	Anleitung zur Durchführung des Trail Making Test A (TMT A) . . . . .	43
A.6	Anleitung zur Durchführung des Trail Making Test B (TMT B) . . . . .	44
B.1	Vorder- und Rückseite der Testvorlage für den Trail Making Test A (TMT A) . . . . .	45
B.2	Vorder- und Rückseite der Testvorlage für den Trail Making Test B (TMT B) . . . . .	46
C.1	Boxplots der Differenzen (Termin 1 - Termin 2) der Signal-Rausch-Abstände (SRTs) für jeder der 10 Szenen (Tabelle 3.1). Das obere und untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Der schwarze horizontale Strich innerhalb der Box kennzeichnet den Median. Die Antennen haben eine Länge von höchstens dem 1,5 fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet. . . . .	47
E.1	Streudiagramm der z-transformierten Reaktionszeit (RZ) des geteilten Aufmerksamkeittests in ms gegenüber dem Signal-Rausch-Abstand (SNR) des Messverfahrens zur subjektiven Höranstrengung (ACALES) bei ESCU 11 des zweiten Termins (Retest) in dB SNR. Zusätzlich wurden für die Szenen in denen eine Korrelation zwischen den Ergebnissen des geteilten Aufmerksamkeittests und ACALES auftreten, eine Regressiongerade eingefügt, gestrichelte Linie. Jedes der zehn Streudiagramme zeigt eine der zehn Szenen (vergleiche Tabelle 3.1). . . . .	51



E.2	Streudiagramme der z-transformierten Reaktionszeit (RZ) des geteilten Aufmerksamkeits-tests in ms gegenüber der Sprachverständlichkeitsschwelle (SRT) des Oldenburger Satztests (OLSA) des zweiten Termins (Retest) in dB SNR für jede der zehn Szenen (vergleiche Tabelle 3.1). . . . .	52
F.1	Aufgetragen sind die Composite Scores (Fuellgrabe et al., 2015) des Trail-Making-Tests (TMT) gegenüber den Sprachverständlichkeitsschwelle (SRT) des Oldenburger Satztestes (OLSA) des zweiten Termins (Retest) in dB SNR für jeder der zehn Szenen (vergleiche Tabelle 3.1). . . . .	53
F.2	Aufgetragen sind die Composite Scores (Fuellgrabe et al., 2015) des Trail-Making-Tests (TMT) gegenüber des Signal-Rausch-Abstandes (SNRs) des Messverfahren zur subjektiven Höranstrengung (ACALES) bei ESCU 11 des zweiten Termins (Retest) in dB SNR für jeder der zehn Szenen (vergleiche Tabelle 3.1). . . . .	54

## Tabellenverzeichnis

2.1	Wortmaterial des Oldenburger Satztests. Jede Wortkategorie, Name-Verb-Zahl-Adjektive-Objekt, beinhaltet zehn verschiedene Wörter (Wagener et al., 1999a, S.9). . . . .	5
3.1	Auflistung der verschiedenen Eigenschaften, aus denen sich die zehn unterschiedlichen Szenen zusammensetzen . . . . .	14
4.1	Ergebnisse des Paarvergleichs der Test-Retest-Ergebnisse für alle 10 Szenen des Oldenburger Satztests (OLSAs). Es ergibt sich kein signifikanter Unterschied bei Anwendung des nach Bonferroni korrigierten Signifikanzniveaus ( $\alpha = 0,005$ ). . . . .	24
4.2	Intraklassen-Korrelationskoeffizient (IKK) zwischen Termin 1 (Test) und Termin 2 (Retest) für alle 10 Szenen und die Bewertung nach Koo und Li (2016) für den Oldenburger Satztest (OLSA). . . . .	26
4.3	Intraklassen-Korrelationskoeffizient (IKK) zwischen Termin 1 (Test) und Termin 2 (Retest) für ACALES für alle 10 Szene bei ESCU 1, 3, 5, 7, 9, 11 und 13 der subjektiven Höranstrengungsmessung (ACALES) plus Mittelwert über alle ESCU-Kategorien. Die <b>fett</b> gedruckten Zahlen weisen auf die niedrigsten und die <u>unterstrichenen</u> Zahlen auf die höchsten IKKs innerhalb einer Szene hin. . . . .	29
4.4	Ergebnisse der wechselseitigen Beziehung des Signal-Rausch-Abstandes (engl.: Signal-to-noise ratio, SNR) und der Reaktionszeiten (RZ). In Spalte 2 sind die $p$ -Werte, in Spalte 3 die Korrelationskoeffizienten $r$ nach Pearson und in Spalte 4 das Bestimmtheitsmaß $R^2$ für jeder der 10 Szene aufgeführt. Eine Korrektur nach Benjamini-Hochberg wurde durchgeführt ( $q = 5\%$ , Benjamini und Hochberg, 1995) und die nach der Korrektur signifikanten $p$ -Werte <b>markiert</b> . . . . .	31
C.1	Test auf Normalverteilung mittels Shapiro-Wilk für den ersten Termin (Test) und den zweiten Termin (Retest) der Daten des Oldenburger Satztests (OLSA). . . . .	48
C.2	Paarvergleich der zehn unterschiedlichen Szenen (Oldenburger Satztest (OLSA)). Mit Bonferroni korrigiert. . . . .	49

D.1 Test auf Normalverteilung mittels Shapiro-Wilk für den ersten Termin (Test) und den zweiten Termin (Retest) der Daten der subjektiven Höranstrengungsmessung (ACALES). . . . .	50
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

Die Wahrnehmung und das Verstehen von Sprache ist eine der Grundfunktionen des menschlichen Lebens und ermöglicht Teilhabe am sozialen Leben (Pichora-Fuller et al., 2016). Die Fähigkeit, Sprache zu verarbeiten, verändert sich aufgrund von Alter und Hörverlust. Auch die Hörsituation, mit der Art und räumlicher Verteilung der Störgeräusche, hat Einfluss auf die Sprachwahrnehmung. Die Diagnose, Quantifizierung und Auswertung von Hörminderungen schließt routinemäßig Tests zum Sprachverstehen im Labor ein. Neben Sprachverständlichkeitstest geben auch Verfahren zur Messung der Höranstrengung Aufschluss über die Sprachwahrnehmung. Allerdings wird das Sprachverstehen und die Höranstrengung häufig in einfachen akustischen Szenarien (z.B. SON0) gemessen, welche nur in geringem Umfang reale Kommunikationssituationen widerspiegeln (Wagener, 2004; Meister, 2020). Aufgrund dessen ist der Mehrwert von Sprachverständlichkeits- und Höranstrengungsmessungen, welche einen besseren Vergleich zu realen Kommunikationssituationen geben, groß. So werden akustisch komplexere Laborsituationen gefordert, in denen die bereits evaluierten und standardisierten Sprach- und Höranstrengungsmessungen, wie z. B. der Oldenburger Satztest (OLSA, Wagener et al., 1999a,b,c) oder das Adaptive Categorial Listening Effort Scaling (ACALES, Krueger et al., 2017) Verfahren, angewendet werden können. Die Szenen sollen dabei durch unterschiedliche Störgeräuschkonfigurationen verschiedene akustische Aspekte aufweisen, wie z. B. räumliche Trennung von Ziel- und Störsprecher, spektrale oder zeitliche Unterschiede zwischen Ziel- und Störsprecher, und eine möglichst realistische Hörsituation aus dem Alltag im Labor abbilden. Dabei sollen systematische Effekte, welche die Sprachwahrnehmung beeinflussen, wie z. B. Nachhall, Dip-Listening, informationelle oder energetische Maskierung berücksichtigt werden, da dies in den standardmäßigen Messungen im Labor nicht der Fall ist. Die Generierung geeigneter Laborszenen spielt dabei eine große Rolle, denn diese akustischen Szenen sollen trotz ihrer Komplexität reproduzierbar sein. Ebenso sollen die Ergebnisse der Sprachverständlichkeits- und Höranstrengungsmessungen trotz der Komplexitäten der akustischen Szenen eine gute Test-Retest-Reliabilität vorweisen. Eine Möglichkeit akustisch, komplexe Szenen zu generieren besteht durch die Nutzung virtueller Akustik. In komplexen Hörsituationen wird die Sprachwahrnehmung nicht nur auf der peripheren Ebene beeinflusst, sondern auch zentrale auditorische Systeme sowie spezifische kognitive Prozesse auf höherer Ebene wie das Arbeitsgedächtnis (engl.: Working Memory, WM) haben einen Einfluss (Meister, 2020; Ronnberg et al., 2013; Dryden et al., 2017). Es wird davon ausgegangen, dass das Hören in komplexen Szenen auch mehr kognitive Leistung mit einbezieht als akustisch einfache Bedingungen (wie z. B. SON0).

Der OLSA und das ACALES-Verfahren stehen als bekannte und validierte Testverfahren zur Messung von Sprachverständlichkeit und Höranstrengung zur Verfügung. Diese Bachelorarbeit

beschäftigt sich mit der Fragestellung, ob die beiden Messverfahren auch zur Anwendung in akustisch, komplexen Szenen geeignet sind. Dazu werden Messungen zur Test-Retest-Reliabilität mit 15 Normalhörenden durchgeführt. Im Vorfeld dieser Bachelorarbeit wurden bereits zehn akustisch, komplexe Szenen konfiguriert. Zudem werden interindividuelle Unterschiede der Messergebnisse mit Ergebnissen aus einem kognitiven Test verglichen. Ebenso wird untersucht, ob eine Korrelation zwischen den Ergebnissen zweier kognitiver Tests und den Ergebnissen der Sprach- sowie Höranstrengungsmessung vorliegen. Die komplexen Szenen unterscheiden sich durch die Anzahl und räumlichen Anordnung der Störschallquellen, sowie durch die Art des Störgeräuschs. Die Ergebnisse werden mit deskriptiver Statistik dargestellt und mit Inferenz-Statistik verglichen.

In Kapitel 2 dieser Bachelorarbeit wird der theoretische Hintergrund zum Thema Sprachverstehen, Höranstrengung, Kognition und Sprachwahrnehmung sowie verschiedene Einflussfaktoren auf die Sprachwahrnehmung beschrieben. Die Messmethoden sowie Messablauf- und durchführung sind in Kapitel 3 näher beschrieben. Die Ergebnisse der Studie sind in Kapitel 4 aufgeführt. Dabei wird zunächst auf die Test-Retest-Reliabilität des OLSAs, des ACALES-Verfahrens und des kognitiven Tests Go/NoGo eingegangen. Anschließend werden die Zusammenhänge der geteilten Aufmerksamkeit und dem OLSA, der geteilte Aufmerksamkeit und ACALES, dem Trail-Making-Test und OLSA sowie dem Trail-Making-Test und ACALES untersucht. Im folgenden Kapitel 5 werden die Ergebnisse diskutiert und abschließend in Kapitel 6 ein Fazit der gesamten Studie gezogen und ein Ausblick gegeben.

Zunächst werden in diesem Kapitel die theoretischen Grundlagen zur Messung von Sprachverstehen und Höranstrengung erläutert. Dafür werden verschiedene Messverfahren dargestellt. Weiter werden die Aspekte Kognition und Sprachwahrnehmung näher beleuchtet und mögliche Einflussfaktoren auf die Sprachwahrnehmung aufgeführt.

## 2.1 Sprachverstehen

Die Sprachaudiometrie ist ein wesentlicher Bestandteil der Diagnostik von Hörstörungen und der Anpassung von Hörsystemen (Talbot und Larson, 1983; Sanchez Lopez et al., 2018), da sie die Fähigkeit einer Person zum akustischen Verstehen von Sprache misst. Es gibt verschiedene Verfahren in der Sprachaudiometrie, die im deutschsprachigen Raum Anwendung finden (Winkler und Holube, 2014). Neben dem Freiburger Einsilbertest (FBE, Hahlbrock, 1953) und dem Göttinger Satztest (GÖSA, Kollmeier und Wesselkamp, 1997) findet ebenfalls die Methode des Oldenburger Satztests (OLSA, Wagener et al., 1999a,b,c) Verwendung. Im Folgenden wird auf die Unterschiede in den zuvor benannten Tests, wie z. B. das Sprachmaterial, die Durchführung oder das verwendete Sprachsignal (Zinner et al., 2021), eingegangen und die Verfahren im einzelnen erklärt sowie die grundlegenden Eigenschaften beschrieben.

### 2.1.1 Freiburger Einsilbertest

Hahlbrock (1953) entwickelte den FBE, welcher bis heute überwiegend für die Diagnostik und bei der Hörgeräteanpassung verwendet wird (Winkler und Holube, 2014; Steffens, 2015). Die Anforderungen an Sprachtests wurde erstmalig im Jahr 1973 in der internationalen Norm DIN EN ISO 8253-3 beschrieben. Der FBE „ermittelt den Diskriminationsverlust für Sprache mittels einsilbiger Substantive“ (Winkler und Holube, 2014, S.1). Mit dem Diskriminationsverlust wird die Differenz zwischen 100 % und dem größtmöglichen Einsilberversuchen in Prozent beschrieben (Winkler und Holube, 2014). Als Testmaterial stehen 20 Gruppen mit je 20 Wörtern zur Verfügung. Die Auswahl der Wörter wurde so festgelegt, dass „nur bekannte und gebräuchliche einsilbige Substantive gewählt“ (Hahlbrock, 1953, S.414) wurden. Der FBE ist ein offener Test, d. h. die Versuchsperson wiederholt die Wörter verbal. Der Sprachtest kann sowohl in Ruhe als auch im Störgeräusch durchgeführt werden. Für die Messung im Störgeräusch wird das Rauschen gemäß

Comité Consultatif International Télégraphique et Téléphonique (sogenanntes CCITT-Rauschen, ITU (1988)) verwendet. Als Vorteil für den FBE kann die jahrzehntelange, weitverbreitete Anwendung und Interpretation des Tests genannt werden (Steffens, 2015). Durch verschiedene Studien wird sichtbar, dass nicht immer alle Normvorgaben zur Prüfung der Listenäquivalenz eingehalten wurden (Winkler und Holube, 2014).

### **2.1.2 Göttinger Satztest**

Kollmeier und Wesselkamp (1997) entwickelten einen Sprachtest, der im Gegensatz zu vorherigen Sprachtests die alltägliche Kommunikationssituation widerspiegelt, indem der Test als Testmaterial ganze Sätze präsentiert. Der GÖSA besteht aus 20 äquivalenten Testlisten mit jeweils zehn kurzen, sinnbehafteten Sätzen mit drei bis sieben Wörtern (Kollmeier und Wesselkamp, 1997). Mit dem GÖSA wird die Sprachverständlichkeitsschwelle (engl.: *Speech Recognition Threshold*, SRT) in Ruhe oder im Störgeräusch gemessen. Die SRT beschreibt den Signal-Rausch-Abstand (SNR), welcher zu einer Wortverständlichkeit von 50 % führt. Das bei dem GÖSA verwendete Störgeräusch ist laut Kollmeier und Wesselkamp (1997) ein kontinuierliches sprachsimulierendes Rauschen, welches das gleiche Langzeitspektrum wie das Sprachmaterials der Sätze aufweist. Zinner et al. (2021) ermittelten beim Vergleich der Leistungsdichtespektren vom GÖSA-Sprachmaterial und dem verwendeten Störgeräusch einen Unterschied von bis zu 16 dB bei ca. 11,5 kHz. Die Testlisten sind phonetisch ausgeglichen und stimmen in der Gesamtzahl der Wörter überein. Der Satztest kann mit einer adaptiven Pegelsteuerung oder einem festen SNR durchgeführt werden und jede Testliste kann nur einmal pro Versuchsperson durchgeführt werden. Nach Kollmeier et al. (2011) liegt der Vorteil von Alltagssätzen darin, dass keine Übung der Testlisten vor der eigentlichen Messung nötig ist. Der GÖSA kann nur als offener Test durchgeführt werden, d. h. die Versuchspersonen wiederholt frei die dargebotenen Wörter ohne aus mehreren vorgegebenen Wörtern auswählen zu können.

### **2.1.3 Oldenburger Satztest**

Als Sprachmaterial des OLSAs werden Sätze der Form Name, Verb, Zahlwort, Adjektiv, Objekt verwendet. Die Abfolge der Wörter ist eine zufällige Kombination aus einem Inventar von 50 Wörtern, für jede Wortart stehen zehn Wörter zur Verfügung (vergleiche Tabelle 2.1, Wagener et al., 1999a). Diese Matrix-Struktur und die daraus resultierenden 50 Wörter schränken den getesteten Wortschatz ein und führen zu Trainingseffekten (Wagener et al., 1999c). Um den Effekt zu berücksichtigen, werden daher zwei Trainingslisten zu Beginn der Messung durchgeführt (Wagener et al., 1999c). Der Sprachtest kann sowohl offen als auch geschlossen, d. h. der Proband wählt Wörter aus einer dargestellten Gesamtmatrix aus, sowie in Ruhe oder im Störgeräusch durchgeführt werden. Als Störgeräusch für den OLSA wird das OLnoise verwendet, welches durch 30-fache Überlagerung des zufällig zeitlich verschobenen gesamten Sprachmaterials generiert wurde. Das Langzeitspektrum des Rauschens entspricht dem des Testmaterials, wodurch die

**Tabelle 2.1:** Wortmaterial des Oldenburger Satztests. Jede Wortkategorie, Name-Verb-Zahl-Adjektive-Objekt, beinhaltet zehn verschiedene Wörter (Wagener et al., 1999a, S.9).

Name	Verb	Zahl	Adjektiv	Objekt
Peter	bekommt	drei	große	Blumen.
Kerstin	sieht	neun	kleine	Tassen.
Tanja	kauft	sieben	alte	Autos.
Ulrich	gibt	acht	nasse	Bilder.
Britta	schenkt	vier	schwere	Dosen.
Wolfgang	verleiht	fünf	grüne	Sessel.
Stefan	hat	zwei	teure	Messer.
Thomas	gewann	achtzehn	schöne	Schuhe.
Doris	nahm	zwölf	rote	Steine.
Nina	malt	elf	weiße	Ringe.

verdeckende Wirkung des Rauschens optimal ist (Wagener et al., 1999a). Ebenfalls entspricht es dem mittleren Langzeitspektrum von diversen Sprachen (engl.: *long-term average speech spectrum*, LTASS, nach Byrne et al., 1996). Das Design des Tests verhindert das Auswendiglernen der nicht sinnbehafteten Sätze, sodass der OLSA beliebig oft wiederholt werden kann (Wagener et al., 1999c).

## 2.2 Höranstrengung

Höranstrengung wird von Pichora-Fuller et al. (2016) als die bewusste Zuweisung mentaler Ressourcen zur Überwindung von Hindernissen bei der Verfolgung eines Zieles bei der Durchführung einer (Hör-)Aufgabe definiert. McGarrigle et al. (2014) beschreibt die Höranstrengung als die mentale Anstrengung, die erforderlich ist, um eine auditive Botschaft zu beachten und zu verstehen. Zur Bestimmung der Höranstrengung können objektive als auch subjektive Methoden herangezogen werden. Physiologische und kognitive Messungen zählen dabei zu den objektiven Messmethoden, während die subjektiven Verfahren Fragebögen und Skalen zur Erfassung der Anstrengung beinhalten (Pichora-Fuller et al., 2016).

### 2.2.1 Physiologische Messungen

Um Änderungen der Aktivität des autonomen und zentralen Nervensystems zu erfassen, werden physiologische Methoden angewendet (McGarrigle et al., 2014). Die Methoden beinhalten u.a. Messungen der Pupillenweite, der Herzaktivität und des Hautleitwerts. Des Weiteren werden Messungen der Hirnaktivität mithilfe der Elektroenzephalografie (EEG) durchgeführt (Pichora-Fuller et al., 2016; McGarrigle et al., 2014).

Die Aufmerksamkeit und Belastung bei Durchführung einer Aufgabe ist durch die Weitung der Pupille erkennbar (Pichora-Fuller et al., 2016). Somit ist die Pupillenweite eine geeignete Indikation für Höranstrengung, da sie die kognitive Anstrengung bei der Lösung einer (Hör-)Aufgabe widerspiegelt (Kahnemann und Beatty, 1966, zitiert nach Fiedler et al., 2021). Die Messung der Sprachverständlichkeit über einen breiten Bereich von SNRs ergab, dass die Beziehung der Sprachverständlichkeit (Bereich von korrekter Antworten von 0 % bis 100 %) und der maximalen Pupillenerweiterung einer umgekehrten U-förmigen Kurve folgt. Die größte Pupillenerweiterung tritt auf, wenn die Verständlichkeit bei einer korrekten Leistung von etwa 50 % liegt. Dieses Muster wurde sowohl bei normalhörenden (NH), als auch bei schwerhörigen (SH) Personen beobachtet (Ohlenforst et al., 2017; McMahon et al., 2016). Zekveld et al. (2014) zeigten, dass eine geringere Sprachverständlichkeit zu einer erhöhten Pupillengröße führt. Zudem führen verschiedene Maskertypen ebenfalls zu einer Veränderung der Pupillengröße (Koelewijn et al., 2014, zitiert nach Fiedler et al., 2021). Wang et al. (2017) beobachteten, dass höhere Müdigkeit und schlechtere Sprachverständlichkeitsschwellen mit einer geringeren Pupillenweite verbunden waren.

Der Hautleitwert, welche die elektrische Leitfähigkeit der Haut beschreibt, kann zur Bestimmung von Höranstrengung genutzt werden (Pichora-Fuller et al., 2016), ist jedoch nur bedingt aussagekräftig, da der Hautleitwert nicht sensitiv für die Veränderungen des SNR ist (Mackersie und Calderon-Moultrie, 2016; Seeman und Sims, 2015). Des Weiteren ist die Bestimmung der elektrischen Leitfähigkeit bei älteren Menschen aufgrund von Hautveränderungen schwierig (Holube et al., 2016).

Mit Hilfe des EEGs werden elektrische Potentiale direkt an der Kopfhaut abgegriffen. Das EEG kann in verschiedene Frequenzbänder zerlegt werden. Wellen des Frequenzbandes von 8 - 13 Hz werden als Alpha-Wellen bezeichnet. Eine Zunahme der Amplitude kann als steigende Höranstrengung gedeutet werden (Obleser et al., 2012). Zudem kann anhand der unterschiedlichen Amplituden der Alpha-Welle der rechten und linken Hirnhälfte bestimmt werden, zu welcher Seite eine Person die Aufmerksamkeit richtet. Bei Experimenten ohne räumliche Trennung von relevanten und ablenkenden auditiven Ereignissen stellten Obleser et al. (2012) und Seifi Ala et al. (2020) eine höhere Alpha-Leistung bei ungünstigeren Bedingungen und verminderter Sprachverständlichkeit fest, da verschlechterte Sprache zu einer höheren Belastung des Arbeitsgedächtnisses führt. Im Gegensatz dazu wurde von Miles et al. (2017) und Hjortkjaer et al. (2018), zitiert nach Fiedler et al. (2021), festgestellt, dass eine höhere Arbeitsgedächtnisbelastung und ungünstige Bedingungen auch zu einer Abnahme der Alpha-Leistung führen. Des Weiteren besteht ein Zusammenhang zwischen dem Rhythmus der Sprache und der Aktivität der Alpha-Wellen. Je besser die Aktivität der Wellen mit dem Sprachrhythmus übereinstimmt, desto mehr Worte werden korrekt verstanden (Woestmann et al., 2016). Neben den Alpha-Welle werden auch ereigniskorrelierte Potentiale im EEG abgebildet. Als ereigniskorrelierte Potentiale werden Potentiale beschrieben, welche beim Einwirken eines Reizes (z. B. akustisch) auf eine Person entstehen. Das auditorisch evozierte Potential P3, welches nach 250 bis 450 ms in Form eines positiven Potentials erkennbar ist, kann ebenfalls zur Bestimmung von Höranstrengung herangezogen werden. Ein Anstieg des Potentials ist ein Indiz für eine erhöhte Höranstrengung (Bertol und Bodmer, 2016).



## 2.2.2 Kognitive Messungen

Kognitive Messungen zur Bestimmung der Höranstrengung beinhalten sowohl die Durchführung von Single- als auch Dual-Task-Paradigmen. Dabei basiert die Methode des Dual-Task-Paradigmas auf Kahnemans Theorie zur Verteilung mentaler Kapazitäten. Die Verwendung von Dual-Task-Ansätzen, bei denen mindestens eine Aufgabe das Zuhören von Sprache ist, lässt Rückschlüsse auf die kognitive Belastung durch die Höraufgabe zu. Es wird angenommen, dass eine abnehmende Leistung in der zweiten Aufgabe mit einer zunehmenden Höranstrengung zusammenhängt (Übersicht von Gagné et al., 2017).

## 2.2.3 Fragebögen und Skalen

Mittels des „Speech, Spatial and Qualities of Hearing Scale“ (SSQ), ein Fragebogen, welcher Fragen zum Sprachverstehen, zum räumlichen Hören und zu weiten Bereichen der Hörqualität beinhaltet, wird auch die Höranstrengung der befragten Personen erfasst. Dabei befassen sich drei der 49 Fragen mit dem Thema der Höranstrengung. Die Befragten haben die Möglichkeit, die wahrgenommene Höranstrengung mit einer beliebigen Zahl zwischen 0 und 10 zu bewerten. Eine höhere Zahl bedeutet eine geringere Höranstrengung (Gatehouse und Noble, 2004). Der „National Aeronautics and Space Administration Task Load Index“ (NASA-TLX), entwickelt von Hart und Staveland (1988), verfügt über Fragen zur physischen, mentalen und zeitlichen Anforderung der Probanden an eine Aufgabe. Zudem werden Leistungen, Anstrengung und Frustration abgefragt. Die Antwortmöglichkeiten erstrecken sich auf einer Skala von „sehr wenig“ bis „sehr viel“ (Hart und Staveland, 1988). Eine 13-stufige Skala mit Werten von 0 („keine Anstrengung“) bis 6 („sehr anstrengend“) inklusive unbenannter Zwischenschritte wurde von Luts et al. (2010) entwickelt, um die subjektive Höranstrengung zu bestimmen.

## 2.2.4 ACALES

Krueger et al. (2017) entwickelte eine adaptive kategoriale Skalierung zur Erfassung der subjektiven Höranstrengung (engl.: *Adaptive Categorical Listening Effort Scaling*, ACALES), welche das erste adaptive Verfahren zur Messung der subjektiven Höranstrengung ist. Im Gegensatz zum statistischen Verfahren, werden die SNRs von ACALES für jede Präsentation adaptiv auf der Basis früherer Bewertungen angepasst, um die gesamte Bandbreite möglicher Kategorien individuell abzudecken. Die Antwortskala entspricht der Skala von Luts et al. (2010) und umfasst sieben benannte Kategorien von „mühelos“ bis „extrem anstrengend“, sowie sechs unbenannte Unterkategorien zwischen den benannten Kategorien. Ebenfalls wird die Kategorie „nur Störgeräusch“ mit in die Skala aufgenommen. Jeder Kategorie wird ein Zahlenwert der kategorischen Einheit des Aufwandes (engl.: *Effort Scale Categorical Units*, ESCU) zugeordnet. Die Werte reichen von 1 ESCU für die Kategorie „mühelos“ bis 14 ESCU für die Kategorie „nur Störgeräusch“. Als Sprachmaterial werden der Versuchsperson Sätze des OLSAs (Wagener et al., 1999b)

im Störgeräusch dargeboten. ACALES misst die individuell wahrgenommene Höranstrengung für verschiedene SNR-Bereiche und berechnet adaptiv die nächste SNR-Präsentation auf der Grundlage der Teilnehmerantwort.

### 2.3 Kognition und Sprachwahrnehmung

Kognition spielt bei der Wahrnehmung von Sprache eine Rolle (Roberts und Allen, 2016, zitiert nach Dryden et al., 2017). Laut Zekveld et al. (2014) ist es schwierig, Gesprächen in einer lauten Umgebung zu folgen und die erforderliche (Hör-)Anstrengung nimmt mit der Hörminderung zu. Auch Meister (2020) beobachtete, dass die kognitive Leistung und das Sprachverstehen in akustisch schwierigen Situationen verknüpft sind. So beeinflussen neben einem möglichen Hörverlust (engl.: *hearing loss*, HL) und dem Alter auch kognitive Leistungen, wie die Aufmerksamkeitsfähigkeit und das WM, das Sprachverstehen, da Sprache auf kortikaler Ebene verarbeitet und verstanden wird (Meister, 2020; Ronnberg et al., 2013). Jede Variation, z. B. Art der Reaktion (offener oder geschlossener Test), das Sprachmaterial (Wörter oder Sätze) oder der SNR kann sich auf verschiedene Art und Weise auf das Ausmaß auswirken, indem kognitive Ressourcen für die Wahrnehmung der Sprachbotschaft erforderlich sind (Dryden et al., 2017). Kognition im Zusammenhang mit Sprachwahrnehmung kann durch viele verschiedene Messgrößen bewertet werden und in verschiedene kognitive Teilbereiche wie z. B. Aufmerksamkeit, exekutive Prozesse, WM, Intelligenz und Verarbeitungsgeschwindigkeit unterteilt werden (Dryden et al., 2017). Die Verarbeitungsgeschwindigkeit ist die Geschwindigkeit, mit der Informationen verarbeitet werden, um eine Aufgabe auszuführen. Salthouse (2000) vermutet, dass sie eine entscheidende Rolle bei der Erklärung altersbedingter Veränderungen der Kognition spielt (zitiert nach Dryden et al., 2017). Die Verarbeitungsgeschwindigkeit wird mit der Sprachwahrnehmung in Verbindung gebracht, da die sequenzielle Natur des Sprachsignals eine schnelle und wiederholte Aktivierung anderer kognitiver Prozesse erfordert, wie z. B. des WMs und des sprachlichen Wissens (Wingfield, 1996, zitiert nach Dryden et al., 2017). Das WM ist unter anderem wichtig für Sprache im Lärm, da ein/e Zuhörer\*in bei längeren Sprachsignalen eine Sprachspur im Gedächtnis behalten muss, um sie mit zuvor gehörten oder abgerufenen Informationen abzugleichen (Goldinger, 1996, zitiert nach Dryden et al., 2017). WM ist ein Prozess mit begrenzter Kapazität, mit dem Personen gleichzeitig Informationen speichern, verarbeiten und manipulieren, die für die Erfüllung komplexer Aufgaben erforderlich sind (Daneman und Carpenter, 1980, zitiert nach Dryden et al., 2017; Kahnemann, 1973, zitiert nach Rudner et al., 2012). Das Modell des leichten Sprachverstehens geht davon aus, dass WM eine Rolle bei der Wiederherstellung von verschlechterten Sprachsignalen und bei der Hemmung von Maskierungssignalen spielt (Ronnberg et al., 2013). Der Zusammenhang zwischen WM und dem Sprachverstehen ist möglicherweise nicht so allgegenwärtig, wie manchmal angenommen wird, sondern kann je nach Alter oder Hörstatus der Person erheblich variieren (Dryden et al., 2017). Rudner et al. (2012) fanden heraus, dass Personen mit einer größeren WM-Kapazität das Hören in Lärm als weniger anstrengend empfanden als Personen mit geringerer WM-Kapazität. Unabhängig von der WM-Kapazität nahm die subjektive Höranstrengung ab, wenn sich die SNRs verbessern.

Alerting (Alarmierung) ist die Fähigkeit, die Aufmerksamkeit auf ein Signal mit hoher Priorität vorzubereiten und aufrechtzuerhalten (Posner und Petersen, 1990, zitiert nach Dryden et al., 2017). Alerting könnte für die Wahrnehmung von Sprache in Lärm wichtig sein, weil es den Hörern ermöglicht, sich in einer Umgebung mit anderen Geräuschquellen auf das Sprachziel zu konzentrieren (Binder et al., 1994; Heald und Nusbaum, 2014, zitiert nach Dryden et al., 2017). Möglicherweise spielt sie eine besonders wichtige Rolle für komplexere Zielsignale (z. B. ganze Sätze), weil sie eine längere Aufmerksamkeit erfordern. Fuellgrabe und Rosen (2016) fanden keinen klaren Unterschied bei jungen Erwachsenen im Vergleich zu älteren Personen im Bezug auf Kognition und Sprachwahrnehmung in akustisch schwierigen Situationen.

## 2.4 Einflussfaktoren auf die Sprachwahrnehmung

Das menschliche Hörsystem hat die Aufgabe, aus komplexen Signalen sinnvolle Informationen zu extrahieren, insbesondere dann, wenn ein Nutzsignal in einer störrauschbehafteten Umgebung verstanden werden soll (Bronkhorst, 2015). Das Selektieren von Störfaktoren fällt NH leichter als SH. Hörsituationen, welche mit mehreren Sprecher\*innen verbunden sind und Schwierigkeiten beim Sprachverstehen hervorrufen, werden oft mit dem „Cocktailparty-Effekt“ beschrieben (Bronkhorst, 2015; J. F. Culling und Stone, 2018). Jedoch nicht nur das Extrahieren von Sprache aus störenden Sprachsignalen oder Umgebungsgeräuschen erweist sich als schwierig, sondern auch Hörsituationen mit anderen Einflussfaktoren, wie z. B. Nachhall.

### 2.4.1 Spatial Release from Masking

Sprache ist bei Hintergrundgeräuschen leichter zu verstehen, wenn die Nutzschaquelle und die Störschaquelle(n) räumlich getrennt sind, als wenn sie sich am selben Ort befinden. Der Unterschied im SRT zwischen diesen beiden Fällen ist als „räumliche Demaskierung“ (engl.: *spatial release from masking*, SRM) bekannt (J. F. Culling und Stone, 2018). SRM entsteht durch viele verschiedene Komponenten wie z. B. das Hören mit dem besseren Ohr (engl.: *better-ear-listening*) oder die binaurale Demaskierung, sowie räumliche Auflösung von Informationsmaskierung und die räumliche Trennung von konkurrierenden Geräuschquellen (J. F. Culling und Stone, 2018). Die Mechanismen des Besser-Ohr-Hörens und der binauralen Demaskierung werden im Folgenden näher beschrieben.

**Better-ear-listening:** Better-ear-listening liegt vor, wenn die räumliche Trennung von Ziel- und Störgeräusch zu unterschiedlichen SNRs an jedem Ohr führt. Dieser Unterschied ergibt sich aus dem Vorhandensein des Kopfes und dessen Wirkung auf die Schallpegel und liegt sowohl im Pegel- als auch im Zeitbereich vor. Die interaurale Pegeldifferenz (engl.: *interaural level*

*differences*, ILD) und die interaurale Zeitdifferenz (engl.: *interaurale time differences*, ITD) geben Hinweise zur Lokalisation einer Schallquelle, welche jedoch nicht notwendig sind, damit eine Demaskierung stattfinden kann (Edmonds und Culling, 2005, zitiert nach J. F. Culling und Stone, 2018). Stattdessen scheinen Lokalisation und Demaskierung unabhängig voneinander durch ganz unterschiedliche Mechanismen zu funktionieren (J. F. Culling und Stone, 2018).

**Binaurale Demaskierung:** Binaurale Demaskierung tritt auf, wenn sich die interauralen Phasen sowie die Pegel des Ziels und des Maskierers voneinander unterscheiden. Die interauralen Phasen entstehen durch die ITDs, die auftreten, wenn es einen Unterschied in der Weglänge vom Ort der Schallquelle zu jedem Ohr gibt. Wenn sich zwei Schallquellen am selben Ort befinden, sind ihre interauralen Phasenunterschiede identisch. Wenn sie jedoch räumlich voneinander getrennt sind, sind die Unterschiede in der Weglänge zu jedem Ohr für jeden Ort der Schallquelle unterschiedlich. Unterschiede hinsichtlich des Pegels entstehen durch die ILDs, da der Kopf den Pegel auf der kontralateralen Seite dämpft (J. F. Culling und Stone, 2018). Die binaurale Demaskierung erfordert eindeutig eine gewisse Verarbeitung durch das Gehirn.

## 2.4.2 Dip Listening

Unter *dip listening* wird die Fähigkeit verstanden, kurze „akustische Einblicke“ in Sprache und andere Geräusche zu erhalten, wenn die schwankenden Hintergrundgeräusche kurzzeitig im Pegel abnehmen (J. F. Culling und Stone, 2018). Die Ausnutzung von Einbrüchen in natürlichen Geräuschkchwankungen trägt dazu bei, dass Personen das „Cocktail-Party-Problem“ des Sprachverständnisses in sozialen Umgebungen mit vielen Sprechern überwinden können. Laut Cueille et al. (2022) sind NH in der Lage, die „Lücken“ in modulierten Geräuschen zu nutzen, um die Zielsprache besser zu verstehen. Dieser Vorteil wird oft als „dip listening advantage“ oder „fluctuation masker benefit“ bezeichnet (J. F. Culling und Stone, 2018; Cueille et al., 2022). Hall et al. (2012) belegten, dass NH einen signifikant höheren Nutzen des Dip-Listenings haben im Vergleich zu SH, dies beobachteten auch Bernstein und Grant (2009).

## 2.4.3 Nachhall

Die Sprachverständlichkeit in einer lauten Umgebung hängt von mehreren Faktoren ab (Kidd und Colburn, 2018), die durch Nachhall beeinflusst werden können. Bei Nachhall können zwei monaurale Effekte beschrieben werden. Zum einen die zeitliche Verschiebung der Zielsprache, die die Sprache weniger verständlich macht, und zum anderen die zeitliche Verschmierung des Lärms, die die Möglichkeit des Zuhörers in den Masker-Dips verringert. Es wurde nachgewiesen, dass diese Phänomene bei NH das Sprachverstehen beeinträchtigen (Cueille et al., 2022). Helfer und Wilber (1990) beobachteten, dass bei Zunahme des Nachhalls ein Rückgang der Sprachverständlichkeit folgte und dies sowohl bei Nachhall in der Zielsprache als auch im Störgeräusch.

Der Vorteil des binauralen Hörens wird durch Nachhall verringert (Lavandier et al., 2012 und Beutelmann et al. 2006, zitiert nach Cueille et al., 2022).

#### **2.4.4 Energetische und informationelle Maskierung**

In störgrauschbehafteter Umgebung treten zwei Arten von Maskierung auf, energetische Maskierung (EM) und informationelle Maskierung (IM). Nach Amiri und Jarollahi (2020) entsteht EM durch die Überlagerung der Energie des Zielsignals mit dem konkurrierenden Störsignal in Zeit und Frequenz, sodass eine spektrale Überlagerung beider Schallsignale auf der Basilarmembran auftritt. Dies verursacht eine Einschränkung der Frequenzselektion auf peripherer Ebene. Dagegen entsteht IM in der zentralen Ebene des auditorischen Systems. IM resultiert aus einer Ungewissheit bzw. Zufälligkeit des Störsignals und möglicherweise in einer Ähnlichkeit zum Zielsignal, sodass die Trennung beider Signale erschwert wird und nur mit größerer Anstrengung möglich ist. IM kann durch die Interferenz beider Signale bei der kognitiven Verarbeitung häufig zusätzlich zu der EM zur Verschlechterung des Sprachverstehens führen. Für Normalhörende stellt dies im Alltag in der Regel kein Problem dar. Für Kinder, Ältere und Personen mit einer Hörminderung ist der Einfluss des IMs jedoch wesentlich schwieriger und behindert die Betroffenen an der bewussten oder unterbewussten Analyse von Hörsituationen (Amiri und Jarollahi, 2020). Der Übergang von EM und IM ist fließend und kann nicht genau bestimmt werden. Nach Brungart et al. (2001) ist bei der Bestimmung von IM wichtig, dass die verwendeten Sätze keinen inhaltlichen Zusammenhang haben, während Dai et al. (2017), zitiert nach Amiri und Jarollahi (2020), beobachtet haben, dass eine größere inhaltliche Bedeutung der konkurrierenden Sprache zu mehr IM führen kann. In mehreren Studien wurde bestätigt, dass die Variabilität zwischen verschiedenen Personen für IM wesentlich größer ist als für EM (Leek et al., 1991; Neff et al., 1995, zitiert nach Amiri und Jarollahi, 2020).

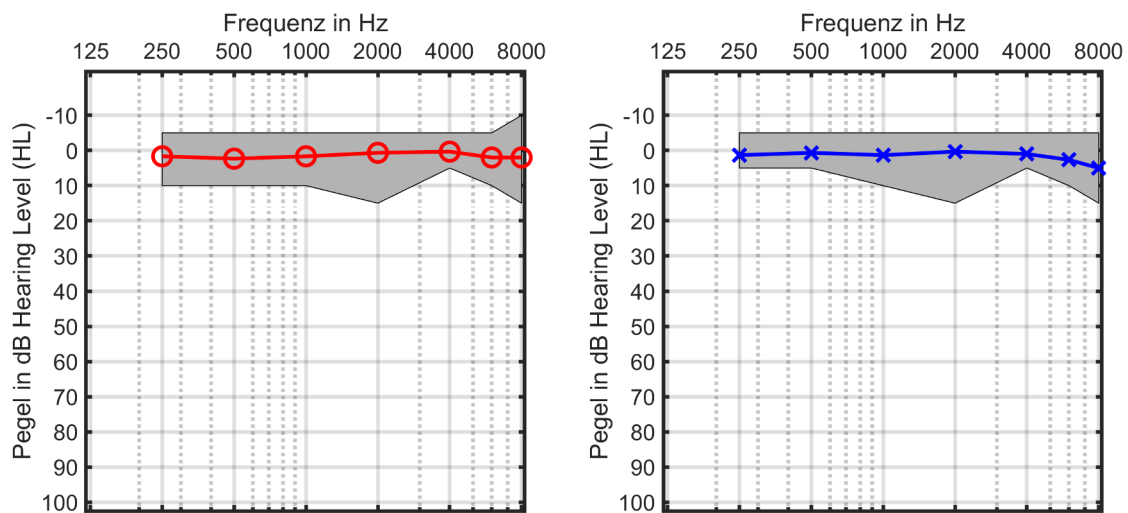
Dieses Kapitel beschreibt die im Rahmen der Studie verwendeten Messmethoden sowie den Messablauf und die Durchführung. Es werden die verwendeten Stimuli und deren Darbietung aufgeführt, welche für die Sprachverständlichkeitsmessung und zur subjektiven Höranstrengungsmessung verwendet wurden. Zudem werden die kognitiven Tests beschrieben, welche ebenfalls Teil der Studie waren.

### 3.1 Versuchspersonen

Insgesamt 15 teilnehmende Personen, acht davon weiblich und sieben männlich, im Alter zwischen 20 und 29 Jahren (Durchschnittsalter: 24,1 Jahre) wurden für zwei Sitzungen, mit einem Abstand von mind. fünf Tagen zwischen dem ersten und dem zweiten Termin, in das Hörzentrum Oldenburg eingeladen. Alle Personen hatten ein normales Gehör, definiert als *Pure Tone Average* ( $PTA_4$ ), was dem Mittelwert des Hörverlustes im Tonaudiogramm bei den Frequenzen 0,5, 1, 2 und 4 kHz entspricht, und einen Pegel kleiner gleich 20 dB HL aufweisen muss (vergleiche Abbildung 3.1). Der mittlere  $PTA_4$  betrug 0,8 dB HL für das rechte Ohr (Standardabweichung (SD) = 3,4 dB) und 1,3 dB HL für das linke Ohr (SD = 3,7 dB). Die Hälfte der Versuchspersonen gab an, bereits Erfahrung mit dem OLSA zu haben. Lediglich eine Person hatte bereits Erfahrung mit dem ACALES-Verfahren. Alle Versuchspersonen erhielten eine Aufwandsentschädigung von 12 € pro Stunde.

### 3.2 Stimuli

Der OLSA (siehe Kapitel 2.1.3) wurde mit einer weiblichen Sprecherin (Wagener et al., 2014) als Nutzsignal sowohl bei der Messung des Sprachverstehens als auch bei der Höranstrengungsmessung verwendet. Bei der Sprachverständlichkeitsmessung erfolgte nur eine Darbietung der Sätze, bei der Höranstrengungsmessung wurden hingegen zwei Sätze hintereinander präsentiert. Die Höranstrengungsbewertung und der Sprachverständlichkeitstest wurden mit vier verschiedenen Hintergrundgeräuschen durchgeführt. Als Maskierer wurden das stationäre International Female Noise (IFnoise, Holube et al., 2009), das fluktuierende ICRA4-250-Rauschen (Wagener et al., 2006), der fluktuierende International Female Fluctuation Masker (IFFM, Holube, 2011)



**Abbildung 3.1:** Mittlere Hörschwelle der Versuchspersonen für die rechte (o) und linke (x) Ohrseite. Der grau hinterlegte Bereich gibt die Minimal- bzw. Maximalwerte an.

sowie ein Cafeteria-Rauschen (Grimm und Hohmann, 2019) genutzt. Das IFnoise wurde von Holube et al. (2009) entwickelt und ist ein stationäres Rauschen, welches das Langzeitspektrum von internationalen Sprecherinnen aufweist (Bryne et al. 1994, zitiert nach Holube et al., 2009). Aus der Arbeitsgruppe des International Collegium of Rehabilitative Audiology (ICRA) ging das dreibandige, sprachfluktuierende Rauschen mit einem weiblich gewichteten, idealisierten Sprachspektrum, das ICRA4-Signal, hervor, welches die Modulation einer weiblichen Sprecherin repräsentiert (Dreschler et al., 2001). Die Hüllkurve des ICRA4-Signals wurde von einem einzelnen Sprecher aufgenommen, aber die Feinstruktur ist rauschähnlich. Für ICRA4-250 wurde das ICRA4-Signal durch Verkürzung der maximalen Pausendauer auf 250 ms modifiziert (Wagener et al., 2006). Das IFFM-Rauschen entspricht dem International Speech Test Signal (ISTS, Holube et al., 2010), jedoch mit einer verkürzten Pausendauer von 250 ms. Das ISTS wurde aus Aufnahmen natürlicher Sprache generiert. Die Sprache von sechs weiblichen Sprecherinnen mit unterschiedlicher Muttersprache (amerikanisches Englisch, Arabisch, Chinesisch, Deutsch, Französisch und Spanisch) wurde beim Lesen von „Nordwind und Sonne“ aufgezeichnet. Anhand der Aufzeichnungen wurde ein Signal generiert, das Ausschnitte aus jeder Sprache enthält und deshalb die gleichen Eigenschaften wie Sprache beinhaltet, jedoch nicht verständlich für den/die Zuhörer\*in ist. Die Reihenfolge der Sprachsegmente variiert, um die Unterscheidbarkeit des ISTS zu erhöhen und dadurch einen unverständlichen Maskierer zu generieren. Das gemittelte Langzeitspektrum entspricht dabei dem weiblicher Sprache (Holube, 2011). Das Cafeteria-Rauschen ist ein räumlich diffuses, fluktuierendes Störgeräusch, welches während der Mittagspause in der Cafeteria des Wissenschaftscampus der Universität Oldenburg, Deutschland, aufgenommen wurde. Es liegt als räumliche Aufnahme (Ambisonics) vor und die Datei wurde bearbeitet, um verständliche Unterhaltungen zu entfernen (Grimm und Hohmann, 2019).

**Tabelle 3.1:** Auflistung der verschiedenen Eigenschaften, aus denen sich die zehn unterschiedlichen Szenen zusammensetzen

Szene	räumliche Kondition	Maskierer	Nachhall	Start SNR OLSA-Messung
1	S0N0	IFnoise	nein	5 dB
2	S0N90	IFnoise	nein	0 dB
3	S0N90	IFnoise	ja	0 dB
4	S0N±90	IFnoise	ja	5 dB
5	S0N±90	ICRA4-250	ja	3 dB
6	S0N±90	IFFM	ja	0 dB
7	S0Ndiff	Cafeteria	ja	5 dB
8	S0Ndiff & S0N90	Cafeteria & IFFM	ja	5 dB
9	S0Ndiff & S0N±90	Cafeteria & IFFM	ja	5 dB
10	S0Ndiff & S0N180	Cafeteria & IFFM	ja	5 dB

### 3.3 Akustisch komplexe Szenen

Mit Hilfe von virtuellen akustischen Szenen können realistische Hörsituationen im Labor simuliert werden. Eine Möglichkeit zur Realisierung von komplexen, akustischen Szenen liefert die *Toolbox for acoustic scene creation and rendering* (TASCAR). Mit Hilfe dieser Software können Szenen in Echtzeit aufbereitet und über ein beliebiges Wiedergabegerät abgespielt werden. Für die Simulation eines Schallfeldes werden einige physikalische Phänomene wie Reflexion, Luftabsorption oder Beugung berücksichtigt.

Die Darbietung des Sprachmaterials des OLSA als Nutzsignal (engl.: *Signal*, S) erfolgt aus der Richtung 0° (S0). Als Störgeräusch (engl.: *Noise*, N) werden die vier Geräusche, IFnoise, ICRA4-250, IFFM und Cafeteria verwendet, welche näher im Kapitel 3.2 beschrieben sind. Je nach Szene werden Störgeräusche aus ± 90°, 0° oder 180° dargeboten. Außerdem wird in vier Szenen das Cafeteria-Rauschen abgespielt. Die Eigenschaften der akustisch komplexen Szenen sind der Tabelle 3.1 zu entnehmen. Die Realisierung des Halls in den Szenen wurde ebenfalls durch TASCAR umgesetzt. Zur Umsetzung wurde das Feedback Delay Network verwendet, welches eine Nachhallzeit von ca. 1250 ms simulierte. Die Zeit der Nachhallzeit wurde dabei so gewählt, dass diese mit der Nachhallzeit der Cafeteria Wechloy (1250 ms) vergleichbar war (Kayser et al., 2009), in deren Räumlichkeiten auch das Cafeteria-Rauschen aufgenommen wurde.

### 3.4 Ermittlung der Höranstrengung

Die Ermittlung der subjektiven Höranstrengung erfolgte mithilfe von ACALES (Krueger et al., 2017). Der Start-SNR wurde auf den Wert 0 dB festgesetzt. Anhand eines adaptiven Verfahrens



Wie anstrengend ist es, dem Sprecher zu folgen?	
nur Störgeräusch	
extrem anstrengend	
:	
sehr anstrengend	
:	
deutlich anstrengend	
:	
mittelgradig anstrengend	
:	
wenig anstrengend	
:	
sehr wenig anstrengend	
:	
müheless	

**Abbildung 3.2:** 13-stufige Bewertungsskala für die subjektive Höranstrengungsmessung. Neben den acht ausgeschriebenen Stufen („müheless“ usw.) verfügt die Skala auch über sechs unbenannte Zwischenschritte (:). (Quelle: Software Oldenburger Messprogramm (OMA) F&E 2.3).

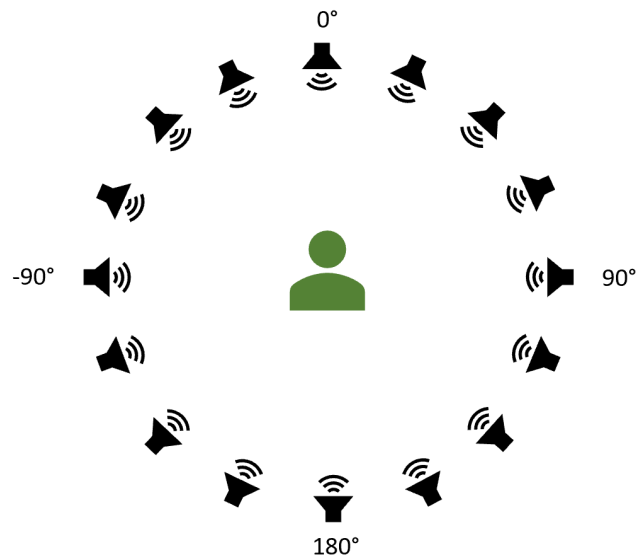
wurde der SNR je nach Antwort der Versuchsperson verändert. Nach jeder Präsentation von zwei OLSA-Sätzen bewerte die Versuchsperson die wahrgenommene Höranstrengung auf der Skala (siehe Abbildung 3.2). Der Pegel des Maskierers blieb konstant, während der Pegel der Sprecherin adaptiv variierte. Der Messablauf von ACALES lässt sich in drei Phasen unterteilen. In der ersten Phase wurden die SNRs für die Bewertungen „müheless“ und „extrem anstrengend“ durch adaptiv wechselnde SNRs in zwei verschachtelten Suchprozessen ermittelt. Ausgehend von einem anfänglichen SNR von 0 dB (anfänglicher Sprach-/Lärmpegel: 65 dB SPL) wurde der SNR um 3 dB erhöht, bis „müheless“ oder „extrem anstrengend“ oder „nur Störgeräusch“ gewählt wurden. In der zweiten Phase wurden sieben SNRs in zufälliger Reihenfolge präsentiert. Die SNRs wurden durch eine lineare Interpolation zwischen „müheless“ und „extrem anstrengend“ geschätzt, und die SNRs entsprechen den Kategorien „müheless“, „sehr wenig anstrengend“, „wenig anstrengend“, „mittelgradig anstrengend“, „deutlich anstrengend“, „sehr anstrengend“ und „extrem anstrengend“ präsentiert. In der dritten Phase wurden die Grenzen sowie die SNRs durch lineare Regression erneut geschätzt und die SNRs, die mit den geschätzten Ziel-Höranstrengungswerten korrelierten, wurden zufällig präsentiert. Die dritte Phase wurde zweimal durchgeführt. Um zu leise oder zu laute Pegel für das Sprachsignal zu vermeiden, wurden die Grenzen der Präsentations-SNRs auf -30 und 25 dB festgelegt. Vor der ersten Messung wurde eine Trainingsphase in der Szene 1 (vergleiche Tabelle 3.1) durchgeführt, die die Auswertung der SNRs für die Bewertung „müheless“ und „extrem anstrengend“ beinhaltet (Messphase 1) und der Demonstration des Verfahrens diente. Die Instruktion von ACALES wurde den Versuchspersonen nur in schriftlicher Form (siehe im Anhang Abbildung A.3) vorgelegt, um einen Einfluss unterschiedlicher Versuchsleiter\*innen (z. B. Wortwahl, Sprachstil, Art der Erklärung) auf die Messergebnisse zu vermeiden (Krueger et al., 2017).

### 3.5 Ermittlung des Sprachverstehens

Für die Ermittlung des Sprachverstehens wurde der OLSA (Wagener et al., 1999a,b,c) verwendet. Die Aufgabe der Versuchspersonen bestand darin, die Sätze oder Teile davon verbal (offene Testform) zu wiederholen (Wortbewertung, d. h. jedes der fünf Wörter hat dieselbe Gewichtung und es ist nicht ausschlaggebend, dass der komplette Satz verstanden werden muss). Während der Messung blieb der Masker-Pegel konstant bei 65 dB SPL, der Sprachpegel wurde adaptiv angepasst, um die individuelle SRT zu bestimmen. Da die Sätze bei verschiedenen SNRs dargeboten wurden, lag der Gesamtpegel des Maskierers und des Satzes je nach gewähltem SNR bei 65 dB SPL oder höher. Je nach Szene wurde der Start-SNR unterschiedlich gewählt, siehe Tabelle 3.1. Um die Genauigkeit zu erhöhen, wurden zwei Trainingslisten in Szene 1 durchgeführt, wie von Wagener et al. (1999b) empfohlen. Eine Testliste umfasste 20 Sätze. Die Einweisung in das Verfahren wurde den Versuchspersonen nur in schriftlicher Form vorgelegt (siehe im Anhang Abbildung A.4), um einen Einfluss auf die Messergebnisse zu vermeiden.

### 3.6 Kognitive Tests

Neben ACALES und OLSA wurden ebenfalls drei kognitive Tests durchgeführt: der geteilte Aufmerksamkeits-Test (Zimmermann und Fimm, 2013), der Go/NoGo (Zimmermann und Fimm, 2013) sowie der Trail Making Test (TMT, Reitan, 1992, zitiert nach Nuesse et al., 2018). Bei dem geteilten Aufmerksamkeits-Test, ein crossmodales Verfahren aus der Testbatterie für Aufmerksamkeitsprüfungen (TAP), muss parallel eine visuelle und eine auditive Aufgabe von den Versuchspersonen bearbeitet werden. Die Aufgaben waren in zwei synchron dargebotene Stimuluskanäle (visuell und auditiv) aufgeteilt. Die Versuchsperson musste auf beide Zielreize mit einem Tastendruck reagieren. Die visuelle Aufgabe wurde auf einem Bildschirm präsentiert und beinhaltete die Beobachtung einer quadratischen Anordnung mit einem 4x4-Punktemuster. Alle 2 s erschienen sechs bis acht Kreuze, die unterschiedliche Muster bildeten. Sobald sich aus vier zusammenhängenden Positionen ein Quadrat aus den Kreuzen bildete, waren die Versuchspersonen aufgefordert, die Antworttaste zu drücken. Bei der auditiven Aufgabe wurden zwei Sinustöne (450 und 1070 Hz) abwechselnd im Abstand von 1 s für 433 ms präsentiert. Die Aufgabe der Versuchsperson bestand darin, die Antworttaste zu drücken, wenn eine Frequenz zweimal hintereinander dargeboten wurde. Als Ergebnisvariablen standen die auditiven und visuellen Reaktionszeiten sowie die Anzahl der Fehler zur Verfügung. Die Einweisung erfolgte durch eine bereits in der TAP vorhandene Instruktion auf dem Bildschirm des Laptops (siehe im Anhang Abbildung A.2), auf dem die TAP installiert war. Vor der eigentlichen Messung wurde ein abgekürzter Probedurchlauf durchgeführt. Für die auditive Aufgabe wurden vor Beginn der Durchführung die beiden Töne den Versuchspersonen akustisch dargeboten und die Lautstärke auf „gut hörbar“ eingestellt. Der zweite Test stammt ebenfalls aus der TAP und beurteilte die Fähigkeit, unter Zeitdruck eine angemessene Reaktion auszuführen und gleichzeitig einen inadäquaten Verhaltensimpuls zu kontrollieren. Beim Go/NoGo kommt es darauf an, eine durch externe Reize getriggerte Reaktion zugunsten einer intern kontrollierten Verhaltensweise zu unterdrücken. Bei



**Abbildung 3.3:** Anordnung der 16 Lautsprecher zur Messung der subjektiven Höranstrengung und des Sprachverstehens. Jeder Lautsprecher steht in einem Winkel von  $22,5^\circ$ -Abständen kreisförmig zu einander. Der Radius des Lautsprecherkreises beträgt 1,5 m.

dem Go/NoGo-Paradigma ist der Aufmerksamkeitsfokus auf das vorhersehbare Erscheinen von Reizen gerichtet, die dann eine selektive Reaktion erfordern, d. h. auf die entweder zu reagieren oder nicht zu reagieren ist. Auf dem Bildschirm wurden in wechselnder Abfolge ein stehendes („+“) und ein liegendes Kreuz („×“) dargeboten. Die Versuchsperson musste bei Erscheinen des liegenden Kreuzes so schnell wie möglich die Antworttaste drücken, beim stehenden Kreuz sollte die Person nicht reagieren. Insgesamt wurden 40 Reize, davon 20 Zielreize, dargeboten. Die Darbietung der Reize war 200 ms kurz, um eine schnelle Reaktion zu provozieren. Zuvor wurde ein verkürzter Probedurchlauf gestartet. Die Einweisung erfolgte ebenfalls durch die TAP auf dem Bildschirm des Laptops mit der installierten TAP (vergleiche im Anhang Abbildung A.1). Als Ergebnisvariablen gab die TAP den Mittelwert und den Median der Reaktionszeit, sowie die Standardabweichung an. Zur Untersuchung der exekutiven Funktionen der Versuchspersonen wurde der von Reitan (1992) eingeführte TMT verwendet (zitiert nach Nuesse et al., 2018). Der Stift- und Papiertest für Erwachsene bestand aus zwei Teilen (A und B) mit 25 Kreisen, die Zahlen oder Buchstaben enthielten. Ziel war es, die Kreise so schnell wie möglich zu verbinden (vergleiche im Anhang Abbildung B.1 und B.2). Im Testteil A (TMT-A) sollten die Zahlen von 1 bis 25 in aufsteigender Reihenfolge verbunden werden. Im zweiten Teil (TMT-B) sollten die Zahlen von eins bis 15 und die Buchstaben von A bis L abwechselnd in aufsteigender Reihenfolge (1-A-2-B-3-C usw.) verbunden werden (Reitan, 1992, zitiert nach Nuesse et al., 2018). Diese Aufgabe erforderte in erster Linie kognitive Flexibilität und exekutive Funktionen. Es lassen sich aber auch Rückschlüsse auf das Arbeitsgedächtnis bei Vergleich der beiden Tests (A,B) ableiten. Die Ergebnisvariable war die Zeit, die für jeden Testabschnitt benötigt wurde. Die Einweisung erfolgte durch die Versuchsleiterin nach einer Vorlage, siehe im Anhang Abbildung A.5 und A.6.

### 3.7 Messapparatur

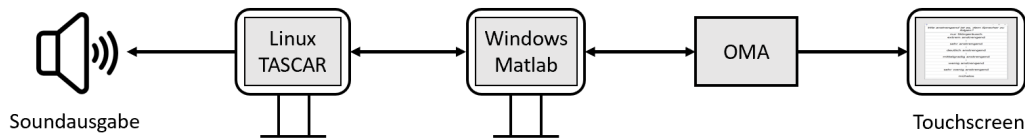
Für die Messung eines Tonaudiogramms der jeweiligen Versuchsperson wurde das Audiometer Aurical II von Otometrics verwendet. Die Stimuli wurden dabei über den Kopfhörer HDA200 von Sennheiser dargeboten. Die Reintonaudiometrie fand in einer Audiometrikabine des Hörzentrums Oldenburg statt. Der Go/NoGo-Test sowie der Test für die geteilte Aufmerksamkeit wurden mittels der Version 2.3.1 der TAP auf einem Windows-Rechner mit dem Betriebssystem Windows 7 durchgeführt. Für die Soundausgabe wurde eine Focusrite Soundkarte und ein Fostex 6301B Lautsprecher verwendet. Zur Durchführung stand ein schallisolierter Raum ohne Fenster zur Verfügung, damit die Ablenkungsparameter für die kognitiven Tests möglichst klein gehalten werden konnten. Der TMT wurde ebenfalls in dem Raum durchgeführt, hierzu war kein Rechner erforderlich, da der Test ein Papier-und-Bleistifttest ist.

Für die Durchführung des ACALES-Messverfahrens wurde das Oldenburger Messprogramm Forschung & Entwicklung 2.3 (OMA F&E) der Hörzentrum Oldenburg gGmbH verwendet. Die Versuchsperson nahm in der Mitte eines Lautsprecher-Arrays, bestehend aus 16 Lautsprechern, platz. In Abbildung 3.3 kann der Messaufbau nachvollzogen werden. Die Lautsprecher (Genelec des Typs 8030B) waren mit  $22,5^\circ$ -Abständen zueinander kreisförmig mit einem Radius von 1,5 m angeordnet. Die akustische Ausgabe der Signale erfolgte mittels TASCAR (siehe Kapitel 3.3). Die ACALES-Messung wurde mittels der OMA auf einem Windows-Rechner gestartet. Mit der Software *Matlab* wurde eine Schnittstelle zwischen dem Windows-Rechner und einem Linux-Rechner hergestellt. Auf dem Linux-Rechner wurden die Signale mittels TASCAR generiert und im Anschluss mittels einer Soundkarte (ADI-8 PRO von RME) ausgegeben. Eine schematische Darstellung des Messaufbaus ist in Abbildung 3.4 gezeigt. Die Eingabe der subjektiven Höranstrengung erfolgte nach der Stimuluswiedergabe durch die Versuchsperson auf einem vor ihm platzierten Touchscreen mit der bekannten Oberfläche der Skala zur Bewertung der Höranstrengung (siehe Abbildung 3.2).

Die Sprachverstehens-Messung erfolgte ähnlich, jedoch wurde hier anstelle des Touchscreens ein Ansteckmikrofon verwendet, da die Versuchsperson, wie bereits in Kapitel 2.1 beschrieben, den OLSA offen durchführte. Das Ansteckmikrofon wurde mittels Vorverstärker an einen Kopfhörer (Sennheiser HD580) übermittelt, welcher von der Versuchsleiterin während der OLSA-Messung getragen wurde. Die Eingabe der richtig verstandenen Wörter erfolgte durch die Versuchsleiterin außerhalb vom Labor am Rechner.

### 3.8 Kalibrierung des Messaufbaus

Der Messaufbau wurde mittels TASCAR und einem Messmikrofon (Freifeldmikrofon Falcon Typ 4134 von Brüel & Kjær) in der Mitte des Lautsprecher-Arrays kalibriert. Dazu wurden die

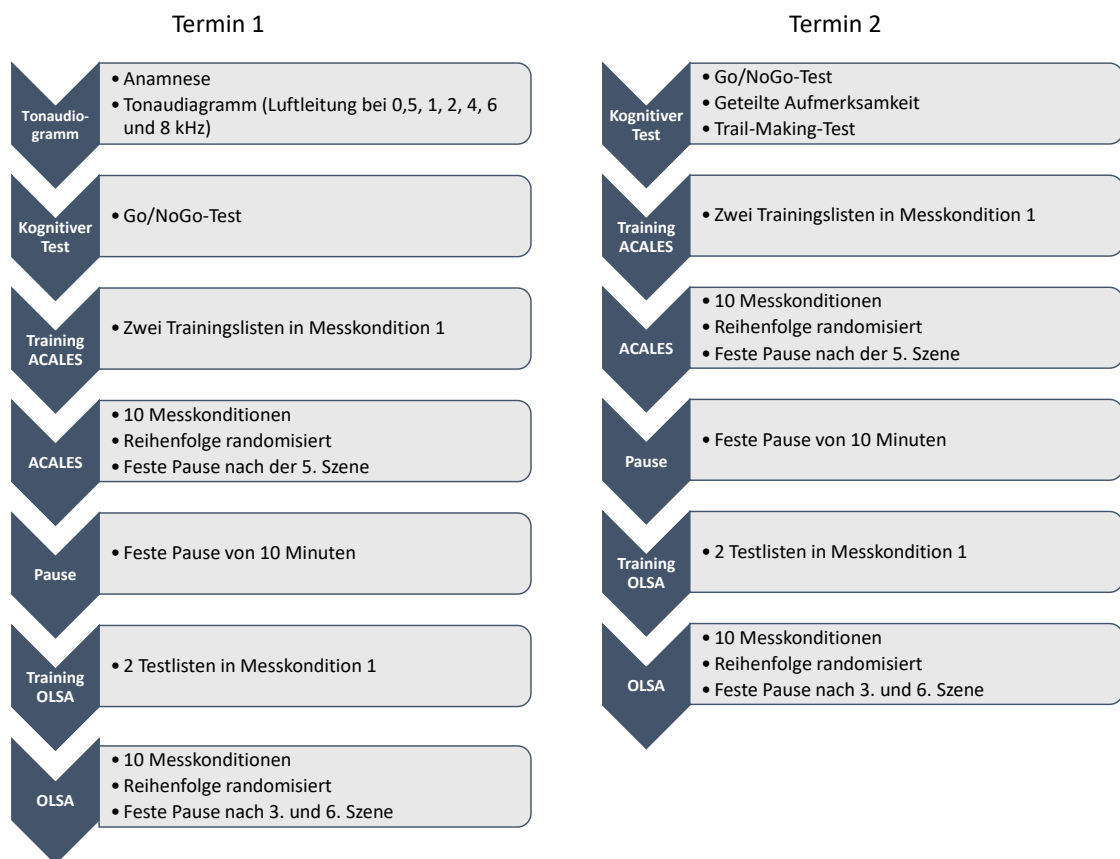


**Abbildung 3.4:** Schematisch dargestellter Zusammenhang des Messaufbaus zur Wiedergabe der Signale mittels des Oldenburger Messprogramms (OMA), Matlab und TASCAR.

Lautsprecher entzerrt und ein Abgleich der Pegel-Simulation aus TASCAR und den tatsächlichen am Mikrophon ankommenden Pegel durchgeführt und ggf. angepasst. Die Kalibrierung des Sprachmaterials erfolgte ebenfalls mit TASCAR, welches auf einen Pegel von 65 dB SPL in 1,5 m Entfernung zu dem Lautsprecher auf 0° kalibriert wurde. Die Kalibrierung wurde mit dem Rauschen mit gleichem Langzeitspektrum des Sprachmaterials des weiblichen OLSA durchgeführt. Auch die Kalibrierung der Störgeräusche wurde mittels TASCAR durchgeführt. Der Pegel jedes einzelnen Störgeräusches bei Darbietung über einen Lautsprecher (0°, ±90° oder 180°) wurde in 1,5 m Entfernung zum Lautsprecher auf 65 dB SPL kalibriert. Für fluktuierende Maskierer wurden ähnliche stationäre Signale mit gleichem Spektrum verwendet. Beim diffusen Schallfeld wurde die Kalibrierung mit dem Cafeteria-Rauschen durchgeführt. Jede der Störquellen wurde einzeln kalibriert, so ergeben sich in den Szene mit mehreren Störquellen (vergleiche Tabelle 3.1) höhere Pegel am Platz der Versuchsperson.

### 3.9 Messablauf

Die Versuchspersonen wurden jeweils zu zwei Terminen à zwei Stunden in das Hörzentrum eingeladen. Am ersten Termin stand die Überprüfung der Hörschwelle mittels einer Reintonaudiometrie an, um zu überprüfen, ob die Versuchsperson die Kriterien für eine Normalhörigkeit erfüllt. Es wurde die Luftleitungshörschwelle über Kopfhörer für das rechte und das linke Ohr bei den Messfrequenzen 250, 500, 1000, 1500, 2000, 4000, 6000 und 8000 Hz gemessen. Vorab wurde die Versuchsperson über die Studie und den Ablauf informiert, eine kurze Anamnese, um mögliche Ohrerkrankungen oder Tinnitus auszuschließen, sowie eine Otoskopie durchgeführt. Nach der Audiometrie erfolgte die Durchführung der kognitiven Tests. Am ersten Termin wurde lediglich der Go/NoGo Test aus der TAP durchgeführt. Am zweiten Termin mussten alle drei kognitiven Tests, Go/NoGo, geteilte Aufmerksamkeit und der TMT, als Aufgabe absolviert werden (vergleiche Abbildung 3.5). Im Anschluss daran erfolgte die Messung der subjektiven Höranstrengung mittels ACALES. Die Reihenfolge der Konditionen wurde an beiden Terminen randomisiert. Vor Beginn der Messung wurde den Versuchspersonen eine kurze Audiosequenz der zehn verschiedenen Szenen vorgespielt. Anschließend erfolgte eine Trainingsliste in Szene 1 (vergleiche Tabelle 3.1). Nach Abschluss des Trainings wurde die Messung gestartet. Nach



**Abbildung 3.5:** Schematische Darstellung des zeitlichen Ablaufs des ersten und zweiten Termins der Studie. Die beiden Termine unterscheiden sich hinsichtlich des Ablaufs der Vorbereitung (Tonaudiogramm und Anamnese) und der Durchführung der kognitiven Tests.

jeweils jeder dritten Messkondition wurde eine Pause von bis zu 5 min durch die Versuchsleiterin eingeleitet. Nach Abschluss der subjektiven Höranstrengungsmessung wurde die Versuchsperson gebeten, den Raum zu verlassen, und eine längere Pause von 10 min wurde eingelegt. Anschließend wurde die Messung des Sprachverstehens mittels OLSA in dem selben Raum durchgeführt. Die Versuchsperson erhielt zusätzlich ein Ansteckmikrofon, da der Sprachtest offen durchgeführt wurde und die Versuchsleiterin, welche sich außerhalb des Labors befand, die Antworten an den Computer übergab. Auch beim OLSA wurde eine Trainingsliste in Szene 1 (vergleiche Tabelle 3.1) durchgeführt und Pausen nach jeder dritten Konditionen eingelegt. Die Messung der Höranstrengung und des Sprachverstehens erfolgten in einer Freifeldkabine des Hörzentrums.

### 3.10 Statische Analysen

Zur Erstellung der deskriptiven Statistik wurde die Mathworks Software Matlab (Version 2022a) verwendet. Um die Mediane und Interquartile der SRTs beim ersten (Test) und zweiten (Retest) Termin grafisch darzustellen, wurden Boxplots generiert. Die Box symbolisiert dabei das untere sowie obere Quartil und wird dabei durch die Interquartilsdistanz  $F_s$  angegeben. Der durchgezogene Strich in der Mitte der Box zeigt die Lage des Medians. Die Länge der Whisker (Antennen) wird durch  $1,5 \cdot F_s$  bestimmt und auf den Wert des oberen Quartils addiert und dem Wert des unteren Quartils subtrahiert. Die Enden liegen allerdings bei dem kleinsten und größten tatsächlichen Wert. Ausreißer werden mittels eines  $+$  dargestellt. Um die Ergebnisse aus Test- und Retest-Messungen auf Korrelation zu überprüfen, wurden Scatterplots erstellt und der Zusammenhang der beiden Wertepaare anhand des Pearson'schen Korrelationskoeffizienten und den Intraklassen-Korrelationskoeffizienten (IKK) überprüft.

Die statistische Auswertung erfolgte mittels der IBM SPSS-Software (Version 25). Für alle durchgeführten Tests wurde ein Signifikanzniveau von  $\alpha = 0,05$  gewählt. Bei Post-Hoc Tests wurde das Signifikanzniveau nach Bonferroni korrigiert. Die Korrektur der Korrelationstests der kognitiven Daten wurden nach Benjamini-Hochberg (Benjamini und Hochberg, 1995) korrigiert. Alle erhobenen Daten sind nach dem Shapiro-Wilk-Test normalverteilt. Ein Datensatz von einer Versuchsperson bei einer Szene war unvollständig und wurde aus den Berechnungen ausgeschlossen.

In diesem Kapitel werden die Ergebnisse der erhobenen Daten vorgestellt. Zunächst erfolgt eine Analyse der Reproduzierbarkeit der Messergebnisse und anschließend eine statistische Untersuchung der Ergebnisse für alle Messkonditionen. Ebenfalls wurden Zusammenhänge zwischen den kognitiven Daten und den Sprachverständlichkeits- sowie Höranstrengungsmessungen untersucht.

#### 4.1 Test-Retest-Reliabilität

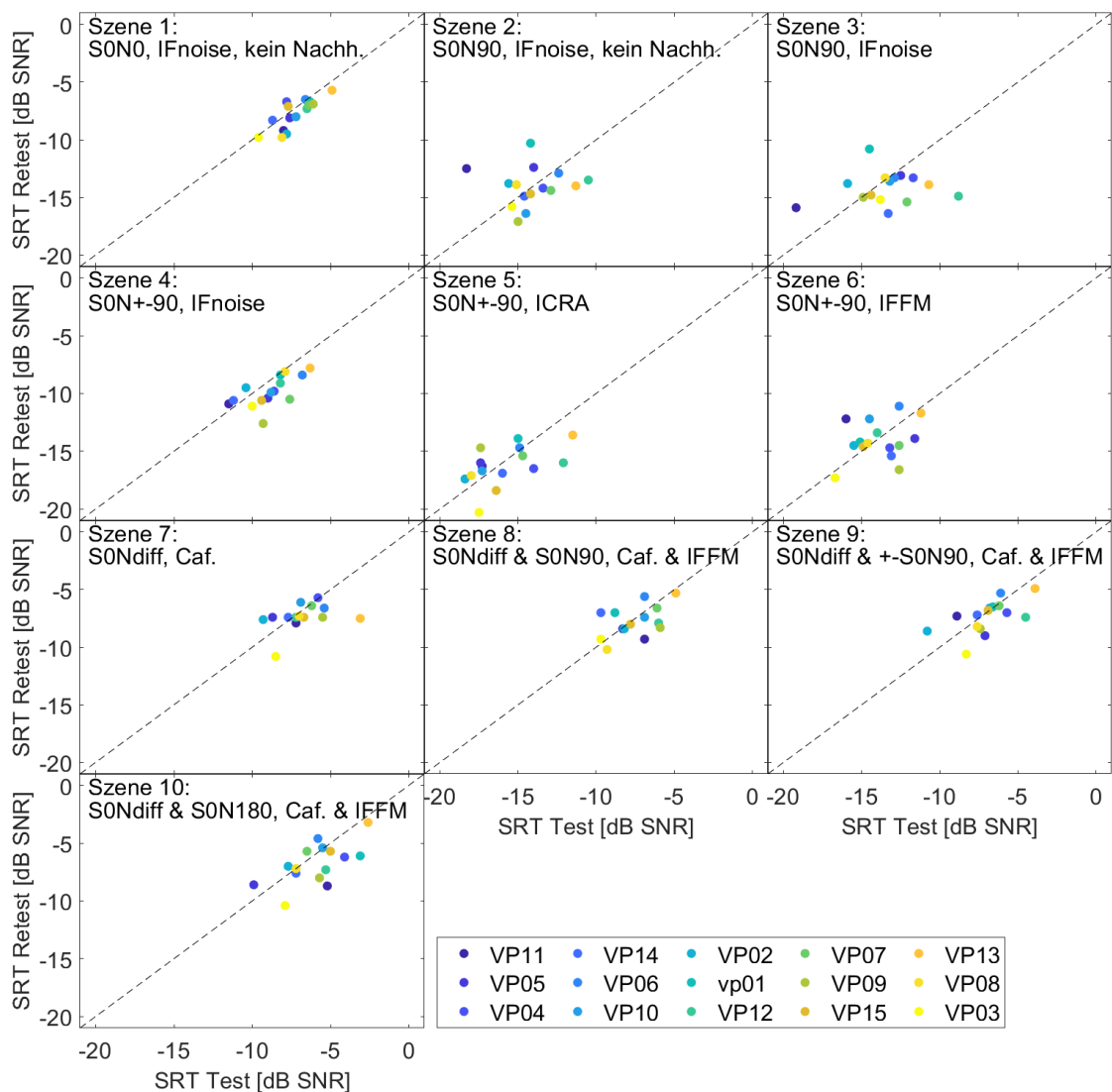
Die Reproduzierbarkeit der gemessenen SRTs und SNRs zwischen Test- und Retest-Messungen wird im Folgenden für die OLSA-Messung und anschließend für die ACALES-Messungen, sowie für den Go/NoGo-Test untersucht. Für jedes Messverfahren werden zunächst die Streudiagramme betrachtet. Des Weiteren werden für den OLSA Boxplots betrachtet, anhand derer die Ergebnisse von den Test und den Retest-Messungen verglichen werden können. Neben der deskriptiven Statistik wird auch die Inferenzstatistik näher beleuchtet.

##### 4.1.1 OLSA

Ein Vergleich der SRTs des ersten Termins (Test) gegenüber der SRTs des zweiten Termins (Retest) für alle 15 Versuchspersonen und 10 Szenen ist anhand von Streudiagrammen in Abbildung 4.1 dargestellt. Ein Streudiagramm bezieht sich dabei jeweils auf eine der 10 Szenen (Tabelle 3.1). Auf der *x-Achse* sind die SRT-Werte des ersten Termins in dB SNR aufgetragen und auf der *y-Achse* die SRT-Werte des zweiten Termins ebenfalls in dB SNR. Es zeigen sich grafisch Unterschiede in der Streuung der Daten über die 10 Szenen. Szene 1 weist eine geringe Streuung und eine Verteilung nahe der Winkelhalbierenden im Vergleich zu Szene 2 auf, welche eine größere Streuung um die Winkelhalbierende aufweist. Ebenso ergeben sich unterschiedliche SRT-Wertebereiche der Messergebnisse. So siedelt sich die Punktwolke der Szene 1 im Vergleich zu Szene 8 in einem höheren SRT-Bereich an. Einen breiteren SRT-Wertebereich (-10 dB SNR bis -2 dB SNR) weist Szene 10 im Vergleich zu Szene 1 (-10 dB SNR bis -5 dB SNR) auf.

Die Daten der OLSA-Messungen sind sowohl für den ersten als auch für den zweiten Termin für alle 10 Messkonditionen gemäß dem Shapiro-Wilk-Test (überwiegend) normalverteilt ( $p > 0,05$ , vergleiche im Anhang Tabelle C.1). Durch die Normalverteilung der Daten kann eine





**Abbildung 4.1:** Streudiagramme der Signal-Rausch-Abstände (SRTs) der Test- und Retest-Messungen aller Versuchspersonen ( $n=15$ ) für alle 10 Szenen. Die Abkürzungen der Szenenbeschreibung S0 steht für das Nutzsignal aus der Richtung  $0^\circ$ , N0/N90/N+-90/N180 für das Störgeräusch welches entweder aus der Richtung  $0^\circ$ ,  $90^\circ$ ,  $\pm 90^\circ$  oder  $180^\circ$  präsentiert wird. IFnoise entspricht dem *International Female Noise*, ICRA dem sprachfluktuierendem Geräusch mit einer weiblichen Gewichtung (ICRA4-250), IFFM dem *International Female Fluctuation Masker*, Caf. dem Cafeteria-Rauschen und kein Nachh. bedeutet kein Nachhall.

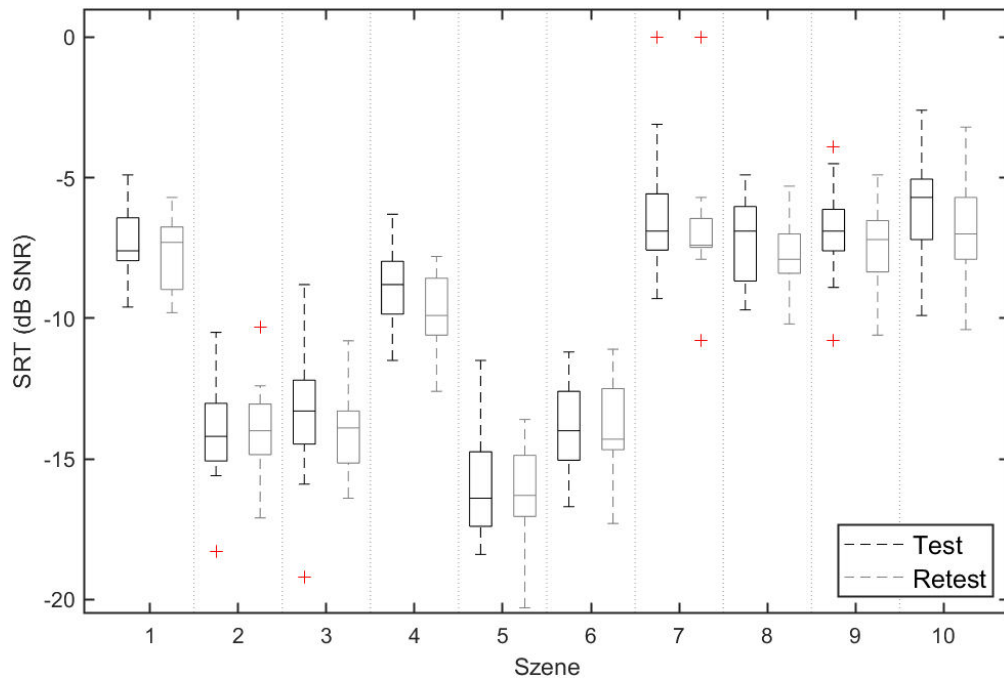
zweifaktorielle Varianzanalyse (ANOVA, engl.: *Analysis of Variance*) mit Messwiederholung (Test, Retest) für die OLSA-Daten mit den Innersubjektfaktoren Messkondition (Szene 1 bis Szene 10, Tabelle 3.1) durchgeführt werden. Die ANOVA bestätigte einen signifikanten Effekt der Messwiederholung zwischen Test- und Retest-Ergebnissen ohne spezifisch die Konditionen zu betrachten [ $F(1,13) = 5,304$ ;  $p = 0,038$ ]. Ein signifikanter Effekt bestätigte sich bei den Messkonditionen [ $F(9,117) = 304,729$ ;  $p < 0,01$ ]. Es liegt keine Interaktion zwischen Messwiederholung und Messkondition vor.

Als Post-hoc Test wurden t-Tests für verbundene Stichproben für die 10 unterschiedlichen Messkonditionen durchgeführt. Das Signifikanzniveau wurde dazu nach Bonferroni korrigiert ( $\alpha = 0,005$ , siehe im Anhang Tabelle C.2). Der Paarvergleich wird in dieser Arbeit nicht näher beleuchtet. Weitere t-Tests für verbundene Stichproben für die zwei Messwiederholungen (Test, Retest) in jeder der 10 Messkonditionen wurden durchgeführt. Das Signifikanzniveau wurde dazu nach Bonferroni korrigiert ( $\alpha = 0,005$ ). In keiner Szene unterscheiden sich die Ergebnisse von Termin 1 zu Termin 2 signifikant (siehe Tabelle 4.1). Die ist auch in Abbildung 4.2 sichtbar, welche für jede Szene zwei Boxplots darstellt, jeweils für den ersten (schwarz) und den zweiten Termin (grau). Beim Vergleich der Mediane zwischen Test- und Retest-Ergebnissen fällt kein großer Unterschied innerhalb der Szenen auf. Die Differenzen der SRTs von Termin 1 zu Termin 2 sind im Anhang in Abbildung C.1 grafisch dargestellt. Die Differenzen bewegen sich in einem Bereich von 1 dB bis hin zu 3,5 dB.

**Tabelle 4.1:** Ergebnisse des Paarvergleichs der Test-Retest-Ergebnisse für alle 10 Szenen des Oldenburger Satztests (OLSAs). Es ergibt sich kein signifikanter Unterschied bei Anwendung des nach Bonferroni korrigierten Signifikanzniveaus ( $\alpha = 0,005$ ).

Test/Retest Szene	$p$ -Wert
1	0,032
2	0,950
3	0,279
4	0,007
5	0,427
6	0,763
7	0,212
8	0,381
9	0,303
10	0,053

Um genauere Aussagen über die Test-Retest-Reliabilität treffen zu können, wurde der IKK (zweifach, gemischt und absolute Übereinstimmung) für die 10 Messkonditionen (Tabelle 3.1) berechnet (siehe Tabelle 4.2) und nach Koo und Li (2016) bewertet. Dabei werden IKK-Werte  $< 0,5$  mit „schlecht“,  $0,5 - 0,75$  mit „durchschnittlich“,  $0,75 - 0,9$  mit „gut“ und Werte  $> 0,9$  mit „sehr gut“ bewertet. Mit  $0,123$  ist der mittlere IKK der Szene 2 am niedrigsten und wird nach Koo und Li (2016) mit „schlecht“ bewertet. Gefolgt mit  $0,204$  und  $0,445$  sind die mittleren IKKs der Szene



**Abbildung 4.2:** Boxplots der Sprachverständlichkeitsschwellen (SRTs) für jede der 10 Szenen (Tabelle 3.1). Für jede Szene wurden sowohl die Ergebnisse aus dem ersten Termin (Test, schwarz) als auch für den zweiten Termin (Retest, grau) dargestellt. Der schwarze bzw. graue horizontale Strich kennzeichnet den Median. Das obere bzw. untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Die Antennen haben eine Länge von höchstens dem 1,5-fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet.

3 und 6 ebenfalls niedrig und werden ebenfalls mit „schlecht“ bewertet. Szene 4, 5, 8 und 10 weisen „durchschnittliche“ Werte nach Koo und Li (2016) zwischen 0,639 und 0,738 auf. Szene 1, 7 und 8 weisen IKK-Werte zwischen 0,752 und 0,857 auf und werden in „gut“ einkategorisiert.

**Tabelle 4.2:** Intraklassen-Korrelationskoeffizient (IKK) zwischen Termin 1 (Test) und Termin 2 (Retest) für alle 10 Szenen und die Bewertung nach Koo und Li (2016) für den Oldenburger Satztest (OLSA).

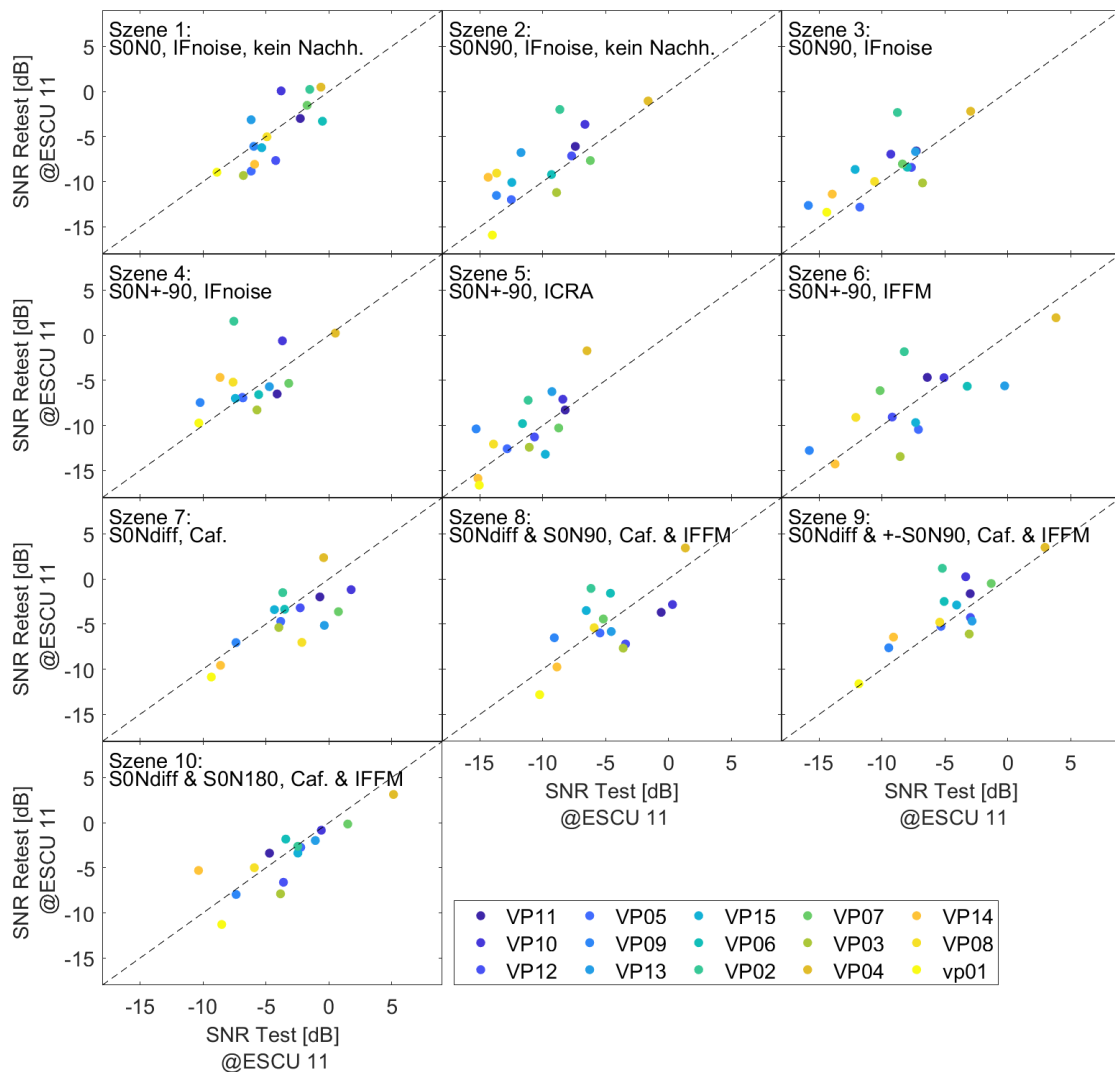
Szene	IKK	Bewertung
1	0,857	gut
2	0,123	schlecht
3	0,204	schlecht
4	0,691	durchschnittlich
5	0,688	durchschnittlich
6	0,445	schlecht
7	0,825	gut
8	0,639	durchschnittlich
9	0,752	gut
10	0,738	durchschnittlich

#### 4.1.2 ACALES

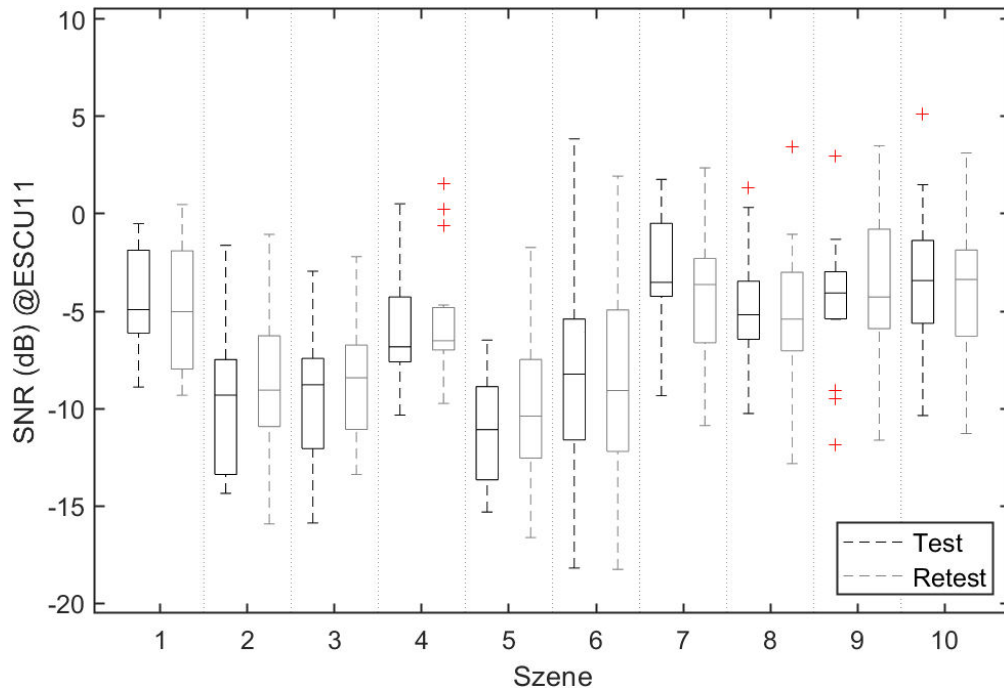
Ein Vergleich der SNR-Werte des ersten Termins (Test), aufgetragen auf der *x-Achse* in dB bei ESCU 11, gegenüber den SNR-Werten des zweiten Termins (Retest), aufgetragen auf der *y-Achse* ebenfalls in dB bei ESCU 11, für alle 15 Versuchspersonen und 10 Szenen sind mittels Streudiagrammen in Abbildung 4.3 dargestellt. Die Kategorie ESCU 11 wurde gewählt, da hier der SRT des OLSAs mit den SNRs des ACALES-Verfahren korreliert und die Daten gut vergleichbar sind (Krueger et al., 2017). Unterschiede in der Anordnung der Punktwolken um die Winkelhalbierende sind zu erkennen. So erstrecken sich Messergebnisse in einigen Szenen über den gesamten Wertebereich wie z. B. in Szene 6 oder 10. Kleinere Wertebereiche, wie in Szene 1 oder 5, treten ebenfalls auf. Die Lage der Punktwolken unterscheiden sich auch. So siedeln sich die Messergebnisse in Szene 1 im höheren SNR-Bereich an als die Messergebnisse in Szene 5, hier liegt eine Verschiebung in den negativen SNR-Bereich vor.

Die Daten der ACALES-Messung für den ersten und zweiten Termin für alle 10 Szenen bei ESCU 11 waren gemäß dem Shapiro-Wilk-Test (überwiegend) normalverteilt ( $p > 0,05$ , vergleiche im Anhang Tabelle D.1). Durch die Normalverteilung der Daten kann eine zweifaktorielle ANOVA mit Messwiederholung (Test, Retest) mit den Innersubjektfaktoren Messkondition (Szene 1 bis Szene 10, Tabelle 3.1) bei ESCU 11 durchgeführt werden. Die ANOVA bestätigt keinen signifikanten Unterschied in den Messwiederholungen [ $F(1,14) = 2,972$ ;  $p = 0,107$ ]. Ein signifikanter Unterschied zwischen den Messkonditionen liegt vor [ $F(9,126) = 34,090$ ;  $p \leq 0,001$ ].

Abbildung 4.4 zeigt für jede Szene zwei Boxplots, jeweils für den ersten (schwarz) und zweiten (grau) Termin. Beim Vergleich der Mediane zwischen Test- und Retest-Ergebnissen fällt kein



**Abbildung 4.3:** Streudiagramme der Signal-Rausch-Abstände (SNRs) bei ESCU 11 der Test- und Retest-Messungen aller Versuchspersonen (n=15) für alle 10 Szenen. Die Abkürzungen der Szenenbeschreibung S0 steht für das Nutzsignal aus der Richtung 0°, N0/N90/N+90/N180 für das Störgeräusch welches entweder aus der Richtung 0°, 90°, ±90° oder 180° präsentiert wird. IFnoise entspricht dem *International Female Noise*, ICRA dem sprachfluktuierendem Geräusch mit einer weiblichen Gewichtung (ICRA4-250), IFFM dem *International Female Fluctuation Masker*, Caf. dem Cafeteria-Rauschen und kein Nachh. bedeutet kein Nachhall.



**Abbildung 4.4:** Boxplots des Signal-Rausch-Abstandes (SNRs) für jede der 10 Szenen (Tabelle 3.1). Für jede Szene wurden sowohl die Ergebnisse aus dem ersten Termin (Test, schwarz) als auch für den zweiten Termin (Retest, grau) dargestellt. Der schwarze bzw. graue horizontale Strich kennzeichnet den Median. Das obere bzw. untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Die Antennen haben eine Länge von höchstens dem 1,5-fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet.

großer Unterschied auf.

Um genauere Aussagen über die Test-Retest-Reliabilität treffen zu können, wurde der IKK (zweifach, gemischt und absolute Übereinstimmung) für die 10 Szenen (Tabelle 3.1) für ESCU 1, 3, 5, 7, 9, 11 und 13 berechnet (siehe Tabelle 4.3). Im Mittel liegt der IKK bei allen Szenen zwischen 0,815 und 0,888 und wird nach Koo und Li (2016) mit „gut“ bewertet. Mit 0,815 ist der mittlere IKK der Szene 4 am geringsten, währenddessen der mittlere IKK der Szene 6 mit 0,888 am höchsten ist. Szene 1 weist über alle ESCUs IKKs zwischen 0,820 und 0,872 auf. Ähnliche Werte ergeben sich in Szene 3 mit IKKs zwischen 0,760 und 0,837. In Szene 3 liegen die IKKs im Bereich von 0,811 und 0,930. Szene 4 weist die größten Schwankungen innerhalb der Szene über die ESCUs auf. Bei ESCU 1, 3, 5, 7 und 9 liegen die IKKs zwischen 0,864 und 0,931, Bei ESCU 11 liegt der Wert bei 0,659 und bei ESCU 13 bei 0,291. Einen ähnlichen, jedoch nicht so starken Abfall der IKKs findet sich auch in Szene 5. Hier liegen die IKKs zwischen 0,595 und 0,919. Der niedrigste Wert tritt wieder bei ESCU 13 auf. Szene 6 weist eine gleichmäßige Verteilung der IKKs zwischen 0,849 und 0,905 auf. Szene 7 weist Werte zwischen 0,834 und 0,931 auf. Bei ESCU 13 ist der IKK am niedrigsten mit 0,649. Auch in Szene 8 ist der IKK mit 0,776 bei ESCU 13 an niedrigsten. Die anderen Werte liegen zwischen 0,881 und 0,920. Szene

10 weist IKKs zwischen 0,860 und 0,905 auf. Bei allen Szenen, außer Szene 4, liegen die IKKs im Bereich „gut“ bis „sehr gut“, Bewertung nach (Koo und Li, 2016). In der Tabelle sind die **Minimal** und Maximalwerte pro Szene markiert.

**Tabelle 4.3:** Intraklassen-Korrelationskoeffizient (IKK) zwischen Termin 1 (Test) und Termin 2 (Retest) für ACALES für alle 10 Szene bei ESCU 1, 3, 5, 7, 9, 11 und 13 der subjektiven Höranstrengungsmessung (ACALES) plus Mittelwert über alle ESCU-Kategorien. Die **fett** gedruckten Zahlen weisen auf die niedrigsten und die unterstrichenen Zahlen auf die höchsten IKKs innerhalb einer Szene hin.

Szene	IKK ESCU 1	IKK ESCU3	IKK ESCU 5	IKK ESCU 7	IKK ESCU 9	IKK ESCU 11	IKK ESCU 13	IKK Mittel
1	0,868	<u>0,872</u>	0,848	<b>0,820</b>	0,831	0,861	0,829	0,866
2	<u>0,837</u>	0,835	0,816	0,799	0,810	0,818	<b>0,760</b>	0,817
3	<u>0,930</u>	0,922	0,888	0,845	0,841	0,853	<b>0,811</b>	0,871
4	0,914	0,903	<u>0,931</u>	0,912	0,864	0,659	<b>0,291</b>	0,815
5	0,886	0,905	<u>0,918</u>	<u>0,919</u>	0,906	0,831	<b>0,595</b>	0,856
6	0,892	0,905	0,902	0,898	<u>0,907</u>	0,902	<b>0,849</b>	0,888
7	0,834	0,907	0,937	0,931	0,918	0,838	<b>0,649</b>	0,862
8	0,848	0,884	0,907	<u>0,913</u>	0,903	0,820	<b>0,645</b>	0,852
9	0,902	<u>0,920</u>	0,914	<u>0,902</u>	0,906	0,881	<b>0,776</b>	0,883
10	0,866	<u>0,896</u>	0,887	0,868	0,885	<u>0,905</u>	<b>0,860</b>	0,881

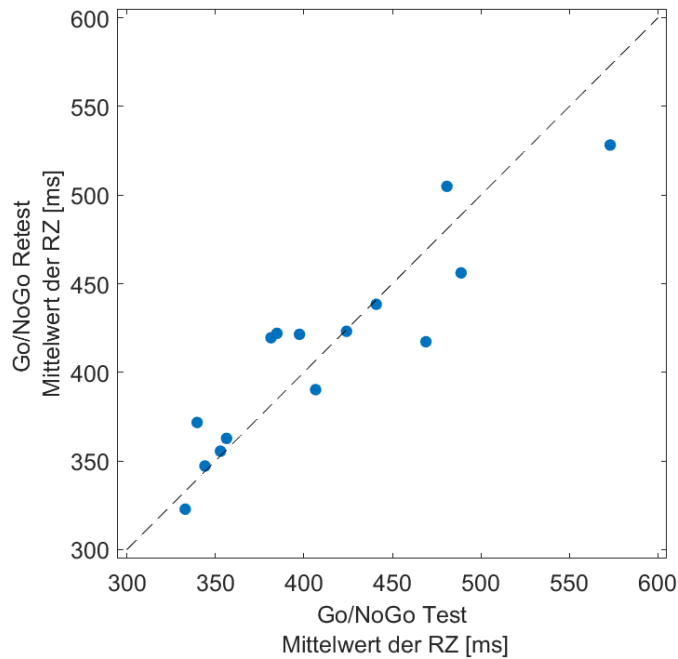
### 4.1.3 Go/NoGo

Die Mittelwerte der Reaktionszeiten (RZ) des Go/NoGo-Test des ersten Termins gegenüber dem zweiten Termin für alle 15 Versuchspersonen sind in Abbildung 4.5 dargestellt. Auf der *x-Achse* sind die Mittelwerte der RZ in ms des ersten Termins (Test) und auf der *y-Achse* sind die Mittelwerte der RZ des zweiten Termins (Retest) in ms aufgetragen. Überwiegend siedeln sich die Messpunkte auf bzw. um die Winkelhalbierende an. Es gibt einige wenige kleine Ausreißer, sowohl unterhalb als auch oberhalb der Winkelhalbierenden. Der IKK bestätigt die hohe Test-Retest-Reliabilität zwischen den RZ des ersten und zweiten Termins mit 0,95, welcher nach Koo und Li (2016) als „sehr gut“ bewertet werden kann.

## 4.2 Korrelation geteilte Aufmerksamkeit und OLSA

Als Ergebnisvariable des geteilten Aufmerksamkeitstestes wurden die Mittelwerte der RZ, der visuelle sowie der auditiven Aufgabe, mittels der Formel

$$z = \frac{x - \mu}{\sigma} \quad (4.1)$$



**Abbildung 4.5:** Streudiagramm für den Go/NoGo-Test. Aufgetragen sind die Mittelwerte der Reaktionszeiten (RZ) des ersten Termins (Test) in ms gegenüber den Mittelwerten der RZ des zweiten Termins in ms.

z-transformiert, nachdem die Ergebnisse auf Normalverteilung geprüft wurden. Dabei ist  $x$  der zu standardisierende Wert,  $\mu$  der Mittelwert und  $\sigma$  die Standardabweichung. Die z Standardisierung ist ein statistisches Verfahren, mit dem die Werte unterschiedlicher Maßeinheiten und Stichproben in eine gemeinsame Einheit überführt werden können, um diese vergleichbar zu machen. Anschließend wurde die Korrelation zwischen den SRTs des OLSA des zweiten Termins und den z-transformierten RZ des geteilten Aufmerksamkeitstests berechnet. Es liegt in keiner der 10 Szenen eine Korrelation der beiden Variablen vor. Die Streudiagramme für jede Szene sind im Anhang in Abbildung E.1 dargestellt. Die Ergebnisse des Tests der geteilten Aufmerksamkeit werden in der Regel anhand der Genauigkeit (Fehler) interpretiert und ausgewertet. Da die Versuchspersonen keine Fehler gemacht haben, wurden in dieser Studie die RZ ausgewertet.

### 4.3 Korrelation geteilte Aufmerksamkeit und ACALES

Die  $p$ -Werte sowie der Korrelationskoeffizient  $r$  nach Pearson der SNRs von ACALES aus dem zweiten Termin wurden mit den bereits mit der Formel 4.1 berechneten z-Werte der RZ des geteilten Aufmerksamkeitstests, der visuellen und auditiven Aufgabe, berechnet. Die Ergebnisse sind in Tabelle 4.4 aufgeführt. Es wurden anstatt der Fehler die RZ der Versuchspersonen untersucht, da nur wenige bis gar keine Fehler von den Versuchspersonen gemacht wurden. Die korrigierten  $p$ -Werte ( $q$ -Werte) wurden mit Hilfe der Benjamini-Hochberg-Korrektur berechnet,



Falscherkennungsrate (engl.: false discovery rate, FDR) von 5 % (Benjamini und Hochberg, 1995), und in der Tabelle 4.4 **hervorgehoben**. Nach der Korrektur korrelieren die RZ und SNRs in Szene 9 und 10 stark negativ miteinander. In Szene 1, 3, 6, 7 und 8 korrelieren die RZ und die SNRs negativ miteinander. Je länger die RZ, desto schlechter ist die geteilte Aufmerksamkeit und desto geringer ist die subjektive Höranstrengung. Die Streudiagramme für jede der 10 Szenen sind im Anhang in Abbildung E.1 dargestellt.

Neben dem Korrelationskoeffizienten  $r$  wurde zusätzlich das Bestimmtheitsmaß  $R^2$  berechnet.  $R^2$  gibt an, wie gut die durch ein Regressionsmodell vorhergesagten Werte mit den tatsächlichen Beobachtungen übereinstimmen und kann einen Wert zwischen 0 und 1 annehmen. Je kleiner  $R^2$  ist, desto geringer ist der lineare Zusammenhang. Der kleinste lineare Zusammenhang liegt in Szene 5 mit  $R^2 = 0,14$  und der größte in Szene 9 mit 0,49 vor.

**Tabelle 4.4:** Ergebnisse der wechselseitigen Beziehung des Signal-Rausch-Abstandes (engl.: Signal-to-noise ratio, SNR) und der Reaktionszeiten (RZ). In Spalte 2 sind die  $p$ -Werte, in Spalte 3 die Korrelationskoeffizienten  $r$  nach Pearson und in Spalte 4 das Bestimmtheitsmaß  $R^2$  für jeder der 10 Szene aufgeführt. Eine Korrektur nach Benjamini-Hochberg wurde durchgeführt ( $q = 5\%$ , Benjamini und Hochberg, 1995) und die nach der Korrektur signifikanten  $p$ -Werte **markiert**.

Szene	$p$ -Wert	Korrelationskoeffizient $r$ nach Pearson	Bestimmtheitsmaß $R^2$
<b>1</b>	<b>0,041</b>	<b>-0,533</b>	<b>0,29</b>
2	0,070	-0,480	0,23
<b>3</b>	<b>0,023</b>	<b>-0,583</b>	<b>0,34</b>
4	0,063	-0,491	0,24
5	0,116	-0,377	0,14
<b>6</b>	<b>0,026</b>	<b>-0,572</b>	<b>0,33</b>
<b>7</b>	<b>0,015</b>	<b>-0,572</b>	<b>0,37</b>
<b>8</b>	<b>0,007</b>	<b>-0,614</b>	<b>0,38</b>
<b>9</b>	<b>0,004</b>	<b>-0,699</b>	<b>0,49</b>
<b>10</b>	<b>0,008</b>	<b>-0,654</b>	<b>0,44</b>

#### 4.4 Korrelation TMT und OLSA

Um die Ergebnisse aus dem TMT-A und -B mit den Ergebnissen des OLSAs zu vergleichen, wurde der Composite Score (Fuellgrabe et al., 2015) aus den Zeiten, die für jeden Testabschnitt (A oder B) benötigt wurden, mittels der Formel

$$\text{Composite Score} = \frac{B - A}{A} \quad (4.2)$$

berechnet und für die weitere Auswertung verwendet. Nachfolgend wurde die Korrelation zwischen den Composite Scores und den SRTs des zweiten Termins der OLSA-Messung berechnet. Es ergab sich in keiner der 10 Szenen ein Zusammenhang der beiden Messergebnisse. Die Streudiagramme für jede der 10 Szenen sind im Anhang in Abbildung F.1 dargestellt.

## **4.5 Korrelation TMT und ACALES**

Auch für die Überprüfung des Zusammenhanges der TMT-Daten und der ACALES-Daten wurde der Composite Score (wie bereits in Kapitel 4.4 näher beschrieben) verwendet. Es ergab keine Signifikanz zwischen den Composite Scores und den SNRs der ACALES-Messung bei ESCU 11. Die Streudiagramme für jede der 10 Szenen sind im Anhang in Abbildung F.2 dargestellt.

### 5.1 Test-Retest Reliabilität OLSA

Beim Vergleich der Test- und Retest-Ergebnisse des OLSAs konnte kein signifikanter Unterschied in einer der 10 Szenen (vergleiche Tabelle 3.1) festgestellt werden. Die IKKs zwischen Test- und Retest-Ergebnissen weisen deutliche Unterschiede in den 10 Szenen auf (vergleiche Tabelle 4.2). So weist Szene 1 mit 0,857 den höchsten IKK auf. Das Ergebnis lässt sich mit dem von Wagener und Brand (2005) vergleichen und bestätigt dieses, welche für die SON0-Kondition ebenfalls eine hohe Test-Retest-Reliabilität feststellten. Jedoch wurde hier die Test-Retest-Differenz betrachtet und nicht der IKK. Szene 2 und 3 zeigen die niedrigsten IKKs, 0,123 und 0,204, und werden nach Koo und Li (2016) als „schlecht“ bewertet. Im direkten Vergleich von Szene 1 zu Szene 2, 0,857 und 0,123, ist der Unterschied am größten. Der einzige Unterschied dieser beiden Szenen liegt in der räumliche Trennung von Ziel- und Störquelle. Dies scheint der ausschlaggebende Effekt zu sein, der zu den unterschiedlichen Messergebnisse führt. Mögliche Einflüsse könnten bereits leichte Kopfbewegungen nach rechts oder links, vorne oder hinten sein, die evtl. zu einem noch größeren Unterschied der unterschiedlichen SNRs an jedem Ohr führen und das Better-ear-listening und/oder die Binaurale Demaskierung begünstigen und so zu unterschiedlichen Ergebnissen an beiden Terminen führen. Grange und J. Culling (2016) fanden heraus, dass einige jüngere normalhörende Versuchspersonen ihren Kopf spontan bewegten, um die räumliche Befreiung aus der Maskierung zu lösen. Da die Kopfbewegungen in dieser Studie nicht mit aufgezeichnet wurden, wäre dies in einer spätere Studie von Interesse.

Szene 2 zu Szene 3 unterscheidet lediglich der Effekt des Nachhalls. Dieser scheint keinen sonderlichen Effekt auf die Reproduzierbarkeit und die Sprachverständlichkeit zu haben. Von Szene 3 (SON90) zu Szene 4 (SON±90) kommt eine zusätzliche Störquelle bei - 90° hinzu. Dadurch hat der Einfluss möglicher Kopfbewegungen geringere Ausmaße und es lassen sich höhere Reproduzierbarkeiten erzielen. Der IKK der Szene 4 liegt im Vergleich zu Szene 3 mit 0,691 deutlich höher und wird nach Koo und Li (2016) als „durchschnittlich“ bewertet.

In Szene 5 kommt das ICRA4-250 Signal als Störgeräusch zum Einsatz. Der IKK wird nach Koo und Li (2016) ebenfalls als „durchschnittlich“ eingestuft. Wagener und Brand (2005) stellten in ihrer Studie die größte Test-Retest-Differenz (2 dB) zwischen den SRT-Werten der normalhörenden Versuchspersonen bei dem stark schwankenden Geräusch ICRA5 fest. Das ICRA5 ist vergleichbar mit dem ICRA4-250, welches in dieser Studie verwendet wurde. Lediglich die Pausendauer wurde

auf 250 ms verkürzt und das Spektrum entspricht dem einer weiblichen Sprecherin im Vergleich zu einem männlichen Sprecher beim ICRA5. Wagener und Brand (2005) nannten als einen möglichen Grund für diese große SRT-Variabilität die Tatsache, dass die Verständlichkeit stark variieren würde, wenn die Wörter synchron zu einem Maximum oder Minimum der Geräuschamplitude präsentiert würden. Im Vergleich zu Szene 5 fällt der IKK von Szene 6 von 0,688 auf 0,445 ab und wird nach Koo und Li (2016) mit „schlecht“ bewertet. Hinsichtlich der Szenenkonfiguration ändert sich hier das Störgeräusch von ICRA4-250 auf IFFM. Ein Effekt, der zusätzlich zu den Effekten aus Szene 5 hinzukommt, ist IM, welches nicht im ICRA4-250 vorhanden ist. Anhand der Ergebnisse lässt sich sagen, dass IM einen Einfluss auf die Reproduzierbarkeit hat. Eine weitere mögliche Erklärung dieses Ergebnisses wäre, dass der IFFM den Versuchspersonen nicht so gut vertraut war. So schilderten Versuchspersonen von Nuesse et al. (2018), dass es schwierig war, den IFFM zu unterdrücken, da das Signal ungewohnt war. Versuchspersonen aus dieser Studie äußerten ähnliche Aussagen über den stark fluktuierenden Maskierer.

In Szene 7 wird als Störgeräusch ein Cafeteria-Rauschen mit einem diffusen Schallfeld verwendet. Der IKK ist mit 0,825 der zweithöchste Wert (nach Szene 1). Nach Koo und Li, 2016 wird der Wert als „gut“ bewertet. Die Auswirkungen aus dem diffusen Schallfeld scheinen keinen negativen Einfluss auf die Reproduzierbarkeit zu haben. Bei Szene 8 kommt ein zusätzliches Störgeräusch (IFFM, wie in Szene 6) von 90° hinzu. Der IKK verringert sich merklich. Auch hier könnte sich die Verringerung der Reproduzierbarkeit durch IM erklären lassen. Ähnliche Ergebnisse lassen sich in Szene 9 und 10 beobachten. Hier ist der IKK jedoch etwas höher als in Szene 8 jedoch geringer als in Szene 7.

Die geringen Korrelationen, welche in einigen Szenen ersichtlich sind, könnten damit erklärt werden, dass die Versuchspersonen sehr ähnliche SRTs über beide Termine erzielen und die Differenzen zwischen Test und Retest in der gleichen Größenordnung wie die Unterschiede zwischen den jungen normalhörenden Versuchspersonen liegen. Die Ergebnisse könnten mit schwerhörigen Versuchspersonen eine bessere Test-Retest-Reliabilität vorweisen.

Wagener und Brand (2005) stellten einen Trainingseffekt zwischen Test- und Retest-Messungen fest, trotz der Durchführung zweier Trainingslisten, welche in deren Berechnung mit einbezogen wurde. Die Herausrechnung des Trainingseffektes war in dieser Studie durch die Randomisierung nicht möglich. Weitere Recherche konnte keine vergleichbare Studie mit Test-Retest-Messungen und Berechnung des IKKs im Hinblick auf den OLSA finden.

Das Messverfahren OLSA ist nicht für alle 10 Szenen geeignet. In Szene 1, 7 und 9 lassen sich gute reproduzierbare Ergebnisse erzielen. Auch die SRT-Differenzen aus Termin 1 zu Termin 2 weisen eine kleine Differenz von ca. 1,5 dB auf. In Szene 4, 5, 8 und 10 ist die Reproduzierbarkeit durchschnittlich. Die Differenzen der SRTs befinden sich zwischen 1,5 dB und 2,5 dB. Der größte Einfluss hat dabei das Störgeräusch IFFM mit Effekten des IM. Szene 2, 3 und 6 eignen sich nur bedingt und weisen schlecht reproduzierbare Ergebnisse auf. Bei Betrachtung der SRT-Differenzen zeigen sich auch in Szene 2, 3 und 6 die höchsten Werte mit  $\geq 3$  dB. Dies lässt darauf schließen, dass das Messverfahren nicht für alle Szenen geeignet ist. In Szenen ohne IM ist das Verfahren sehr genau.

## 5.2 Test-Retest-Reliabilität ACALES

Zwischen dem ersten und zweiten Messtermin konnte unabhängig von den 10 Szenen (vergleiche Tabelle 3.1) kein signifikanter Unterschied festgestellt werden. Der über alle ESCUs gemittelte IKK jeder einzelnen Szene lag zwischen 0,815 und 0,888, was auf eine gute Zuverlässigkeit hinweist (Koo und Li, 2016). Krueger et al. (2017) erzielten einen gemittelten IKK von 0,9, welcher somit vergleichbar mit denen in dieser Studie erhaltenen Werte ist. Krueger et al. (2022) ermittelten einen IKK von 0,89 für ihre Kondition SON0 bei Verwendung des OLnoise als Störgeräusch, welche mit der Szene 1 dieser Studie vergleichbar ist, die einen IKK von 0,866 aufweist. Der IKK bei SON90 bei Krueger et al. (2022) blieb ähnlich hoch mit 0,88, während in dieser Studie der IKK in Szene 2 etwas abfiel auf 0,817. Bei Betrachtung der einzelnen IKKs bei den unterschiedlichen ESCU-Werten der Szene 2 fällt auf, dass der niedrigste IKK bei ESCU 13 auftritt mit 0,760. Alle anderen IKKs waren höher. Da ESCU 13 eine Randgruppe der Rankingkategorien ist, ist es nicht auszuschließen, dass es sich hier um Deckeneffekte handelt. Bis auf Szene 1 weisen alle anderen neun Szenen den niedrigsten IKK bei ESCU 13 auf (siehe Tabelle 4.3), was Deckeneffekte bestärken könnte. Sobald die Versuchspersonen in der ersten Sitzung eine interne Definition von Höranstrengung erstellt hatten, konnten sie ihre Bewertungen zu einem späteren Zeitpunkt recht gut reproduzieren, so auch bei Krueger et al. (2022) und Krueger et al. (2017). Die interindividuellen Unterschiede in den SNRs sind beim ACALES-Verfahren wesentlich größer als die interindividuellen SRTs beim OLSA. Und damit kann eher ein hoher Korrelationskoeffizient beim ACALES-Verfahren als beim OLSA erreicht werden. Das Messverfahren ACALES scheint somit für alle 10 Szenen gut reproduzierbare Ergebnisse zu erzielen.

## 5.3 Test-Retest Reliabilität Go/NoGo

Ein sehr hoher IKK von 0,95 lässt auf sehr gut reproduzierbare Ergebnisse des kognitiven Tests Go/NoGo schließen. Da der Test sowohl am ersten als auch am zweiten Termin durchgeführt wurde, kann eine Aussage über die Tagesperformance der Versuchspersonen getroffen werden. Da es minimale Unterschiede zwischen den Ergebnissen beider Termine gibt, kann davon ausgegangen werden, dass die Versuchspersonen an beiden Terminen die gleiche kognitive Leistung erbringen konnten. Ausreißer und große Streuung gab es keine. Es wird vermutet, dass diese Ergebnisse anders ausfallen würden, wenn der Test mit älteren Menschen mit einer Hörminderung durchgeführt werden würde, da bei der Probandengruppe die kognitiven Leistungen eingeschränkt sein könnten. Da es sich hier um junge normalhörende Versuchspersonen handelt, wird davon ausgegangen, dass die kognitive Leistung (sehr) hoch ist. In der Regel wird beim Go/NoGo-Test die Anzahl der Fehler ausgewertet. Es ist zu erkennen, dass diese Personengruppe wenige bis keine Fehler macht, weshalb die Auswertung der Fehler in dieser Studie nicht möglich war. Daher stellt sich die Frage, ob der Go/NoGo-Test für die Probandengruppe geeignet war. Ebenso sollte die Motivation der einzelnen Versuchspersonen im Auge behalten werden. Eine hohe Motivation könnte zu einer besseren Leistung führen. Ein Großteil der Versuchspersonen waren Bekannte

der Versuchsleiterin, wodurch die Versuchspersonen wahrscheinlich motivierter waren.

Wie in Kapitel 4.1.1 beschrieben, konnten beim OLSA für einige Szenen Unterschiede in den individuellen SRTs zwischen Test- und Retest-Messung festgestellt werden. Laut den Ergebnissen aus dem Go/NoGo-Test können diese Ergebnisse nicht durch Unterschiede in kognitiver Tagesform erklärt werden. Für ACALES wurde eine hohe Test-Retest-Reliabilität festgestellt, ähnlich wie beim Go/NoGo, was das Ergebnis dieses kognitiven Tests bestärkt.

## 5.4 Korrelation Kognition und Sprachwahrnehmung

In keiner der 10 Szenen (Tabelle 3.1) liegt zwischen den SRTs des OLSAs und den RZ des Tests zur geteilten Aufmerksamkeit eine Korrelation vor. Wie bereits zuvor erwähnt, wurde in dieser Studie nicht die Streuung sondern die RZ betrachtet. Dies kann darauf hindeuten, dass dieser kognitive Test für die jungen normalhörenden Versuchspersonen nicht aussagekräftig genug ist. Es wird davon ausgegangen, dass ältere (schwerhörige) Versuchspersonen andere Ergebnisse erzielen würden, hinsichtlich der Fehlerrate der Daten, wie es auch Nuesse et al. (2018) beobachteten. Nuesse et al. (2018) fanden heraus, dass ältere normalhörende Versuchspersonen mit besseren Ergebnissen in Tests zur geteilten Aufmerksamkeit unter Messbedingungen mit einem räumlich getrennten Maskierer schlechter Sprache erkennen. Dies konnte in dieser Studie mit jungen normalhörenden Versuchspersonen nicht festgestellt werden. Die RZ aus dem geteilten Aufmerksamkeitstest korrelieren negativ mit den SNRs der ACALES-Messung, bei sieben der zehn Szenen. Insbesondere bei Szenen in denen IM auftritt. Dies lässt darauf schließen, dass die geteilte Aufmerksamkeit zwar keinen Zusammenhang mit dem Sprachverstehen bei jungen normalhörenden Versuchspersonen aufweist, jedoch die Fähigkeit zur geteilten Aufmerksamkeit Auswirkung auf die subjektive Höranstrengung hat. Versuchspersonen mit einer besseren geteilten Aufmerksamkeit, kürzere RZ, weisen eine höhere subjektive Höranstrengung auf als Versuchspersonen mit einer schlechten geteilten Aufmerksamkeit, längere RZ, bei gleichem SNR. Dies lässt vermuten, dass Personen, die eine schlechtere geteilte Aufmerksamkeit haben, in Szenen mit IM weniger Konzentration aufbringen müssen, um dem Zielsprecher zu folgen, als Personen mit einer besseren geteilten Aufmerksamkeit, da erstere weniger durch Störsprecher abgelenkt werden. Frühere Forschungsarbeiten haben keine signifikante Korrelation zwischen verschiedenen Messmethoden der Höranstrengung gezeigt, was darauf hindeutet, dass jede Messmethode unterschiedliche Aspekte der kognitiven Verarbeitung bewertet (Alhanbali et al., 2019).

Zwischen den Ergebnissen des TMT und des OLSAs sowie ACALES konnte keine signifikante Korrelation festgestellt werden. Der TMT misst die exekutiven Funktionen. Eine Studie von Ellis et al. (2016) hat gezeigt, dass ein Zusammenhang zwischen exekutiven Funktionen und dem Sprachverstehen besteht (zitiert nach Nuesse et al., 2018). Ein solcher Zusammenhang konnte in dieser Studie nicht festgestellt werden. Beim TMT werden die Zeiten notiert, die die Versuchsperson benötigt, um die Aufgabe zu erledigen. Fehler werden lediglich durch die benötigte Zeit notiert, die die Versuchsperson länger brauchen. Die Versuchspersonen haben (fast) keine

Fehler gemacht und den Test womöglich schneller absolviert, als ältere Menschen, weshalb der Test evtl. zu einfach für diese Gruppe war. Die Streuung der Daten war sehr gering. Es kann nicht ausgeschlossen werden, dass es sich hier um Deckeneffekte handelt.

Im Rahmen dieser Studie wurden der OLSA, ein Messverfahren zur Sprachverständlichkeit, und das ACALES-Verfahren, ein Messverfahren zur Ermittlung der subjektiven Höranstrengung, in zehn akustisch, komplexen Hörsituationen gemessen. Die zehn Messkonditionen unterschieden sich dabei hinsichtlich der räumlichen Trennung von Ziel- und Störsprecher, den verwendeten Maskierern und der Anzahl der Störquellen. Insgesamt 15 junge, normalhörende Versuchspersonen führten sowohl den OLSA als auch das ACALES-Verfahren an zwei aufeinander folgenden Terminen durch. Anhand der Ergebnisse wurde die Test-Retest-Reliabilität der beiden Verfahren in akustisch komplexen Szenen untersucht. Zusätzlich zu der Sprachverständlichkeit- und der Höranstrengungsmessung wurden drei verschiedene kognitive Tests, welche unterschiedliche Aspekte der Kognition erfassen, durchgeführt. Der Go/NoGo-Test misst die kognitive Tagesform, der geteilte Aufmerksamkeitstest gibt Rückschlüsse über die geteilte Aufmerksamkeit und der TMT untersucht die exekutiven Funktionen. Zusammenfassend lässt sich folgern:

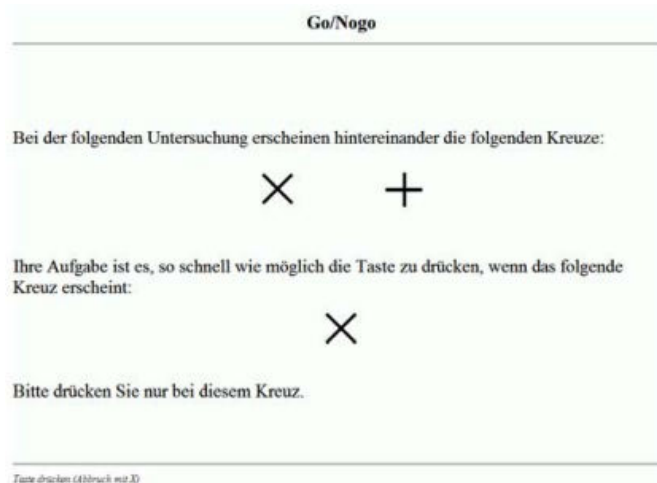
- Der OLSA wies unterschiedlich gute Test-Retest-Reliabilitäten in den 10 verschiedenen Szenen auf. Die Test-Retest-Reliabilität sinkt mit steigendem IM im Störsignal und der räumlichen Trennung von Ziel- und Störsprechern. Die SRT-Differenzen aus Termin 1 zu Termin 2 liegen in einem Wertebereich von 1,5 dB bis  $\geq 3$  dB. Szenen mit einer Differenz  $>2$  dB sind nur bedingt geeignet. Der OLSA ist lediglich für die Szenen 1, 4, 5, 7, 8, 9, 10 gut geeignet.
- ACALES wies über alle 10 Szene eine gute Test-Retest-Reliabilität auf. Das Verfahren ist geeignet, um subjektive Höranstrengung in akustisch, komplexen Szenen valide zu messen.
- Es wurden keine Korrelationen zwischen dem OLSA und den kognitiven Tests festgestellt. Dies kann auf das Alters der Versuchspersonen und die Auswahl der kognitiven Tests zurückgeführt werden, da die ausgewählten Tests evtl. nicht aussagekräftig bzw. sensitiv genug sind, um Unterschied bei jungen normalhörenden Versuchspersonen festzustellen. Es ist jedoch auch möglich, dass bei jungen normalhörenden Versuchspersonen die kognitiven Leistungen im Bezug auf das Sprachverstehen keinen Einfluss haben.
- Je besser die geteilte Aufmerksamkeit der jungen, normalhörenden Versuchspersonen war, desto höher war die Höranstrengung in komplexen Hörsituationen.

In zukünftigen Studien könnte statt einer rein akustischen Darbietung auch eine audiovisuelle

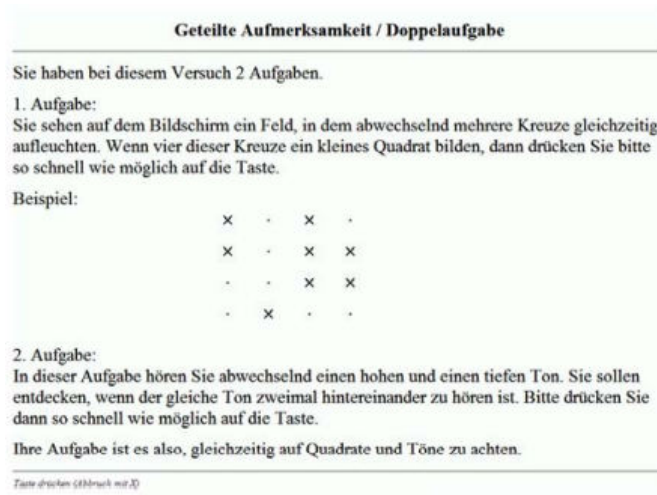


Darbietung des Zielsprechers übertragen werden. Dadurch könnte das Sprachverstehen in komplex akustischen Hörsituationen für die Versuchspersonen erleichtert werden. Die Unterstützung des Mundbildes spielt auch in den realen Hörsituationen eine große Rolle. Ebenso könnten mittels eines Headtrackers die Kopfbewegungen der Versuchspersonen aufgezeichnet werden und so Rückschlüsse über die Bewegungen in verschiedenen Szenen, insbesondere in Szenen, in denen Ziel- und Störsprecher räumlich getrennt sind, geben. Die Auswahl der kognitiven Tests sollte im Vorfeld einer weiteren Studie besser gewählt werden und die Eigenschaften der Tests hinsichtlich der Ergebnisse bei jungen Versuchspersonen im Vorfeld näher betrachtet werden.

# INSTRUKTIONEN



**Abbildung A.1:** Instruktion Go/NoGo Test (Quelle: Zimmermann und Fimm, 2013)



**Abbildung A.2:** Instruktion geteilte Aufmerksamkeit (Quelle: Zimmermann und Fimm, 2013)

## Instruktion Höranstrengungsmessung

Mit diesem Test wird die **Anstrengung** in verschiedenen Hörsituationen untersucht. Wir möchten von Ihnen wissen, wieviel Anstrengung sie aufwenden, bzw. in welchem Ausmaß Sie sich dabei **konzentrieren** müssen. Bitte hören Sie sich dazu je **2** Sätze gut an und bewerten Sie anschließend

### „wie anstrengend es für Sie war dem Sprecher zu folgen“.

Sie brauchen die Sätze nicht nachzusprechen. Drücken Sie zur Bewertung der Höranstrengung die entsprechende Taste der unten gezeigten Skala, die von „mühe-los“ bis „extrem anstrengend“ reicht. Versuchen Sie immer dem Sprecher zu folgen und fokussieren Sie bitte den Lautsprecher direkt vor Ihnen.

Sie können auch die Zwischenschritte (:) auswählen. Falls Sie keine Sprache wahrnehmen, drücken Sie bitte „nur Störgeräusch“.

Die Tasten zur Bewertung sind aktiv, wenn die Schrift von hellgrau auf schwarz übergeht, dies geschieht nachdem der 2. Satz abgespielt wurde.

Wie anstrengend ist es, dem Sprecher zu folgen?	
nur Störgeräusch	
extrem anstrengend	
:	
sehr anstrengend	
:	
deutlich anstrengend	
:	
mittelgradig anstrengend	
:	
wenig anstrengend	
:	
sehr wenig anstrengend	
:	
mühe-los	

Falls Sie keine Sprache wahrnehmen

Ausmaß der Konzentrationsleistung / Grad der Höranstrengung, um den Sprecher zu folgen

Eine Hörsituation kann z.B. als „mühe-los“ angesehen werden, wenn Sie dem Sprecher ohne Mühe über einen längeren Zeitraum folgen können

Bitte bewerten Sie nicht, wie gut Sie die Sätze **verstehen** können, sondern wie **anstrengend** es für Sie ist dem Sprecher zu folgen.

Es gibt keine „richtigen“ oder „falschen“ Antworten. Die Antworten bilden Ihre individuelle Wahrnehmung ab.

Copyright © 2022 Hörzentrum Oldenburg gGmbH

**Abbildung A.3:** Instruktion für die Messung der subjektiven Höranstrengung

## Instruktion Sprachverstehen

Bei diesem Test wird das **Sprachverstehen** in geräuschvollen Umgebungen ermittelt. Ihnen werden 20 Sätze dargeboten. Die Sätze werden von einer weiblichen Stimme gesprochen und von vorne dargeboten.

Jeder Satz besteht aus 5 Wörtern der Struktur:

	Name	Verb	Zahl	Adjektiv	Objekt
z.B.	Ulrich	schenkt	sieben	schwere	Sessel

Die Sätze ergeben nicht unbedingt einen Sinn. Zusätzlich zu der Sprache wird ein Rauschen abgespielt.

Bitte hören Sie sich die Sätze gut an und wiederholen Sie nach der Darbietung eines Satzes **jedes** einzelne Wort, das Sie verstanden haben. Sie dürfen gerne raten aber fokussieren Sie bitten den Lautsprecher vor Ihnen mit der weiblichen Sprecherin.

Während der Messung wird die Sprache in der Lautstärke verändert. Wenn Sie wenig Wörter verstehen, wird der Sprecher lauter. Wenn Sie viele Wörter verstehen, wird der Sprecher leiser. Lassen Sie sich dadurch nicht entmutigen. Für die Aussagekraft der Messung ist es wichtig, unter schwierigen Bedingungen zu messen.

Vielen Dank für Ihre Mitarbeit!

Copyright © 2022 Hörzentrum Oldenburg gGmbH

**Abbildung A.4:** Instruktion für die Messung des Sprachverstehens

## Trail Making Test A

Original-Instruktion von Reitan

Zahlen (Test A) oder Zahlen und Buchstaben (Test B) sollen in aufsteigender Reihenfolge so schnell wie möglich verbunden werden. Die benötigte Zeit wird gestoppt.

### Übungsblatt A

**„Auf diesem Blatt gibt es verschiedene Zahlen (zeigen). Beginnen Sie bei der Zahl 1 (zeigen) und zeichnen Sie einen Strich von 1 nach 2 (zeigen), von 2 nach 3 (zeigen), von 3 nach 4 (zeigen), in der richtigen Reihenfolge bis Sie am Ende (zeigen) sind. Zeichnen Sie die Linien so schnell wie Sie können. Fertig- START!“**

Wenn die Versuchsperson (VP) das Beispiel richtig auf eine Art und Weise durchführt, dass sie damit zeigt, dass sie verstanden hat, was der Versuchsleiter (VL) sagt. Dann sagen Sie:

**„Gut. Lassen Sie uns das nächste probieren.“**

Drehen Sie das Beispielblatt um und geben Sie Aufgabenblatt A (TMT A).

Falls die VP im Übungsbeispiel einen Fehler macht, wird sie sofort darauf aufmerksam gemacht!

1. **Sie haben mit dem falschen Kreis begonnen. Hier ist der Start (zeigen Zahl 1).**
2. **Sie haben einen Kreis ausgelassen (zeigen Sie den Kreis, den die VP ausgelassen hat). Sie sollen von Zahl 1 (zeigen) zu Zahl 2 (zeigen), von 2 zu 3 (zeigen) und so weiter, bis sie den Kreis sehen, der mit End (zeigen) markiert ist.**

Wenn die VP Beispiel A nicht selbständig lösen kann. Nehmen Sie die Hand und führen den umgedrehten Stift (ohne dass dieser zeichnet) durch das Beispiel. Dann sagen Sie: **„Jetzt probieren Sie!“**

**Geben sie den Stift der VP richtig herum zurück und sagen:**

**„Denke Sie daran, Sie beginnen bei der Zahl 1 (zeigen) und zeichnen einen Strich von 1 zu 2 (zeigen) 2 zu 3 (zeigen), 3 zu 4 (zeigen) und so weiter in der Reihenfolge, bis Sie den Kreis erreichen, der mit „End“ (zeigen) gekennzeichnet ist. Springen Sie nicht herum, sondern gehen von einer Zahl zu der nächsten Zahl in der richtigen Reihenfolge. Erinnern Sie sich bitte: Sie sollen so schnell wie sie könne arbeiten. Fertig -Start.“**

Wenn Übungsblatt A verstanden wurde, zur Testung übergehen. Wenn die VP weiterhin Schwierigkeiten hat, wiederholen Sie die Vorgehensweise, bis die Aufgabe erfolgreich durchgeführt wurde – oder es wird deutlich, dass die VP die Aufgabe nicht durchführen kann.

Drücken Sie die Stoppuhr, sobald Sie das Startzeichen zum Anfangen („Fertig -Los!“) gegeben haben. Der VL muss die VP genau beobachten, um gleich eingreifen zu können, wenn ein Fehler beobachtet wird.

Wenn die VP einen Fehler macht, sofort darauf aufmerksam machen und Fehler korrigieren lassen, d.h. zum letzten richtigen Kreis zurückkehren und von dort aus den Test weiterführen lassen. Zeit weiterlaufen lassen.

Bewertung: Zeit in Sekunden; Fehler werden vor allem indirekt, durch die zusätzliche aufgewendete Zeit bewertet. Die Striche der VP sollen die Kreise mindestens berühren.

### Abbildung A.5: Anleitung zur Durchführung des Trail Making Test A (TMT A)

## Trail Making Test B

Original-Instruktion von Reitan

### Übungsblatt b

**„Auf diesem Blatt sind Zahlen und Buchstaben. Beginnen Sie bei der Zahl 1 (zeigen) und zeichnen Sie einen Strich von 1 nach A (zeigen), von A nach 2 (zeigen), von 2 nach B (zeigen), von B nach 3 (zeigen), von B nach 3 (zeigen), von 3 nach C (zeigen) und so weiter in der richtigen Reihenfolge bis Sie am Ende (zeigen) sind.**

**Beachten Sie: erst haben Sie eine Zahl (auf 1 zeigen), dann einen Buchstaben (auf A zeigen), dann eine Zahl (auf 2 zeigen) und dann einen Buchstaben (auf B zeigen), und so weiter. Zeichnen Sie die Linien so schnell wie sie können. Fertig -Los.“**

Wenn die Versuchsperson (VP) das Beispiel richtig auf eine Art und Weise durchführt, dass sie damit zeigt, dass sie verstanden hat, was der Versuchsleiter (VL) sagt. Dann sagen Sie:

**„Gut. Lassen Sie uns das nächste probieren.“**

Drehen Sie das Beispielblatt um und geben Sie Aufgabenblatt A (TMT A).

Falls die VP im Übungsbeispiel einen Fehler macht, wird sie sofort darauf aufmerksam gemacht!

1. **Sie haben mit dem falschen Kreis begonnen. Hier ist der Start (zeigen Zahl 1).**
2. **Sie haben einen Kreis ausgelassen (zeigen Sie den Kreis, den die VP ausgelassen hat). Sie sollen von Zahl 1 (zeigen) zum Buchstaben A (zeigen), von A zu 2 (zeigen) und so weiter, bis sie den Kreis sehen, der mit End (zeigen) markiert ist.**

Wenn die VP Beispiel A nicht selbständig lösen kann. Nehmen Sie die Hand und führen den umgedrehten Stift (ohne dass dieser zeichnet) durch das Beispiel. Dann sagen Sie: **„Jetzt probieren Sie!“**

Geben sie den Stift der VP richtig herum zurück und sagen:

**„Denke Sie daran, Sie beginnen bei der Zahl 1 (zeigen) und zeichnen einen Strich von 1 zu A (zeigen), A zu 2 (zeigen), 2 zu B (zeigen) und so weiter in der Reihenfolge, bis Sie den Kreis erreichen, der mit „End“ (zeigen) gekennzeichnet ist. Erinnern Sie sich bitte: Sie sollen so schnell wie sie könne arbeiten. Fertig -Start.“**

Wenn Übungsblatt B verstanden wurde, zur Testung übergehen. Wenn die VP weiterhin Schwierigkeiten hat, wiederholen Sie die Vorgehensweise, bis die Aufgabe erfolgreich durchgeführt wurde – oder es wird deutlich, dass die VP die Aufgabe nicht durchführen kann.

Drücken Sie die Stoppuhr, sobald Sie das Startzeichen zum Anfangen („Fertig -Los!“) gegeben haben. Der VL muss die VP genau beobachten, um gleich eingreifen zu können, wenn ein Fehler beobachtet wird.

Wenn die VP einen Fehler macht, sofort darauf aufmerksam machen und Fehler korrigieren lassen, d.h. zum letzten richtigen Kreis zurückkehren und von dort aus den Test weiterführen lassen. Zeit weiterlaufen lassen.

Bewertung: Zeit in Sekunden; Fehler werden vor allem indirekt, durch die zusätzliche aufgewendete Zeit bewertet. Die Striche der VP sollen die Kreise mindestens berühren.

### Abbildung A.6: Anleitung zur Durchführung des Trail Making Test B (TMT B)

TESTVORLAGE TRAIL MAKING TEST

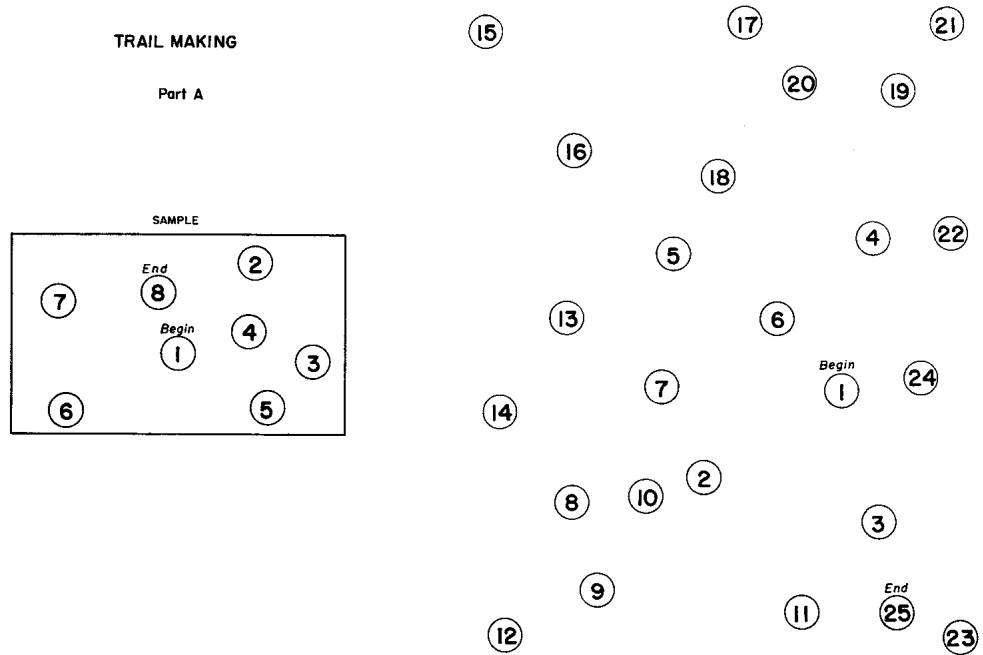
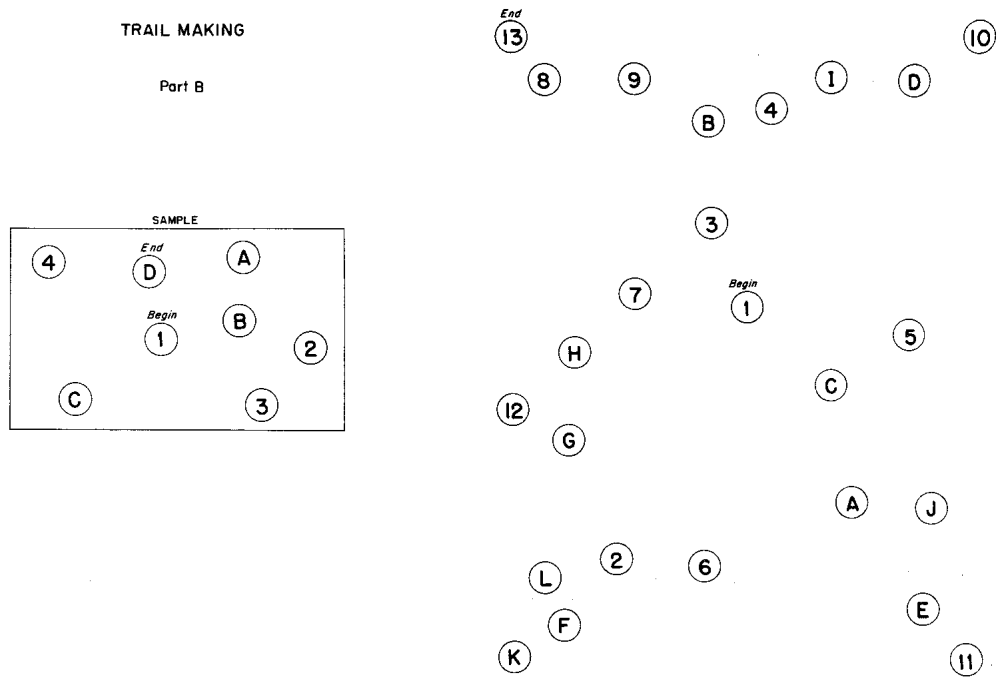
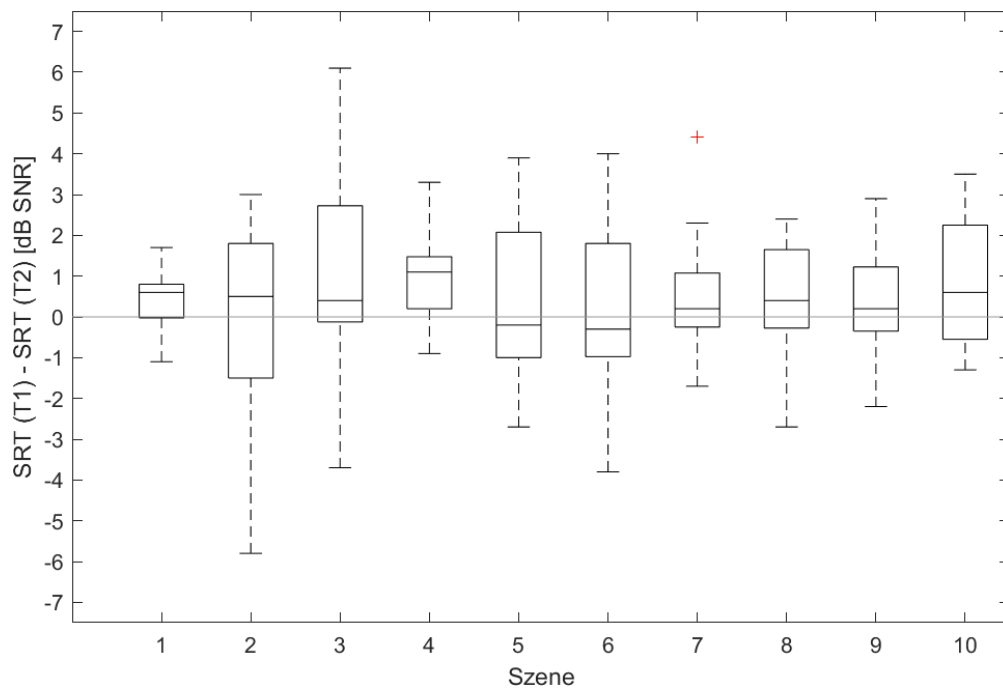


Abbildung B.1: Vorder- und Rückseite der Testvorlage für den Trail Making Test A (TMT A)



**Abbildung B.2:** Vorder- und Rückseite der Testvorlage für den Trail Making Test B (TMT B)





**Abbildung C.1:** Boxplots der Differenzen (Termin 1 - Termin 2) der Signal-Rausch-Abstände (SRTs) für jeder der 10 Szenen (Tabelle 3.1). Das obere und untere Ende des Boxplots gibt das obere bzw. untere Quartil an. Der schwarze horizontale Strich innerhalb der Box kennzeichnet den Median. Die Antennen haben eine Länge von höchstens dem 1,5 fachen des Interquartilsabstands. Ausreißer sind mit einem roten Kreuz gekennzeichnet.

**Tabelle C.1:** Test auf Normalverteilung mittels Shapiro-Wilk für den ersten Termin (Test) und den zweiten Termin (Retest) der Daten des Oldenburger Satztests (OLSA).

Test		Retest	
Szene	Irrtumswahrscheinlichkeit $p$	Szene	Irrtumswahrscheinlichkeit $p$
1	0,923	1	0,336
2	0,798	2	0,611
3	0,628	3	0,179
4	0,986	4	0,757
5	0,051	5	0,752
6	0,850	6	0,711
7	0,697	7	0,003
8	0,304	8	0,798
9	0,919	9	0,924
10	0,948	10	0,999

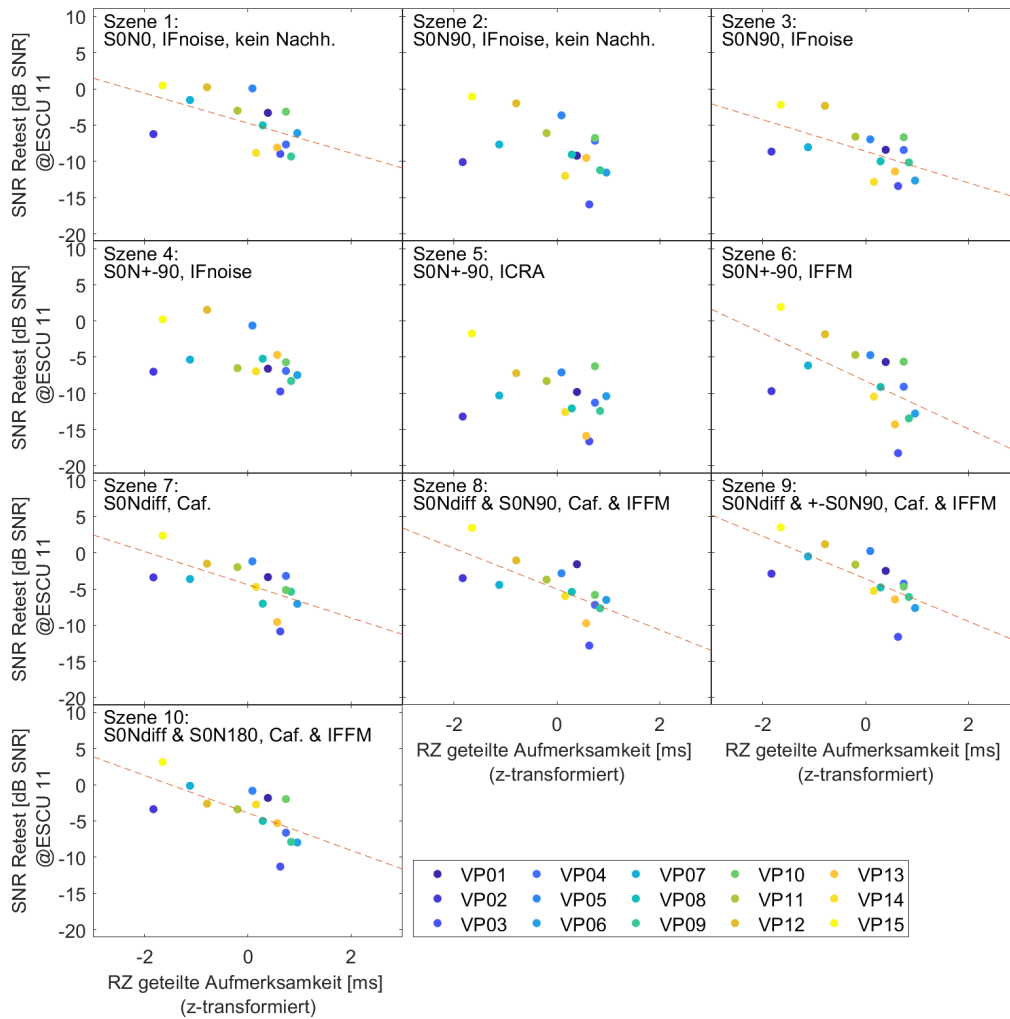
**Tabelle C.2:** Paarvergleich der zehn unterschiedlichen Szenen (Oldenburger Satztest (OLSA)). Mit Bonferroni korrigiert.

Szene	Irrtumswahrscheinlichkeit $p$	Szene	Irrtumswahrscheinlichkeit $p$		
1	2	0,000*	6	1	0,000*
	3	0,000*		2	1,000
	4	0,003*		3	1,000
	5	0,000*		4	0,000*
	6	0,000*		5	0,000*
	7	0,455		7	0,000*
	8	1,000		8	0,000*
	9	1,000		9	0,000*
	10	0,070		10	0,000*
	2	1		0,000*	7
3		1,000	2	0,000*	
4		0,000*	3	0,000*	
5		0,001*	4	0,000*	
6		1,000	5	0,000*	
7		0,000*	6	0,000*	
8		0,000*	8	1,000	
9		0,000*	9	1,000	
10		0,000*	10	1,000	
3		1	0,000*	8	
	2	1,000	2		0,000*
	4	0,000*	3		0,000*
	5	0,004*	4		0,005
	6	1,000	5		0,000*
	7	0,000	6		0,000*
	8	0,000*	7		1,000
	9	0,000*	9		1,000
	10	0,000*	10		0,045
	4	1	0,003*		9
2		0,000*	2	0,000*	
3		0,000*	3	0,000*	
5		0,000*	4	0,000*	
6		0,000*	5	0,000*	
7		0,000*	6	0,000*	
8		0,005*	7	1,000	
9		0,000*	8	1,000	
10		0,000*	10	0,448	
5		1	0,000*	10	
	2	0,001*	2		0,000*
	3	0,004*	3		0,000*
	4	0,000*	4		0,000*
	6	0,000*	5		0,000*
	7	0,000*	6		0,000*
	8	0,000*	7		1,000
	9	0,000*	8		0,045
	10	0,000*	9		0,448

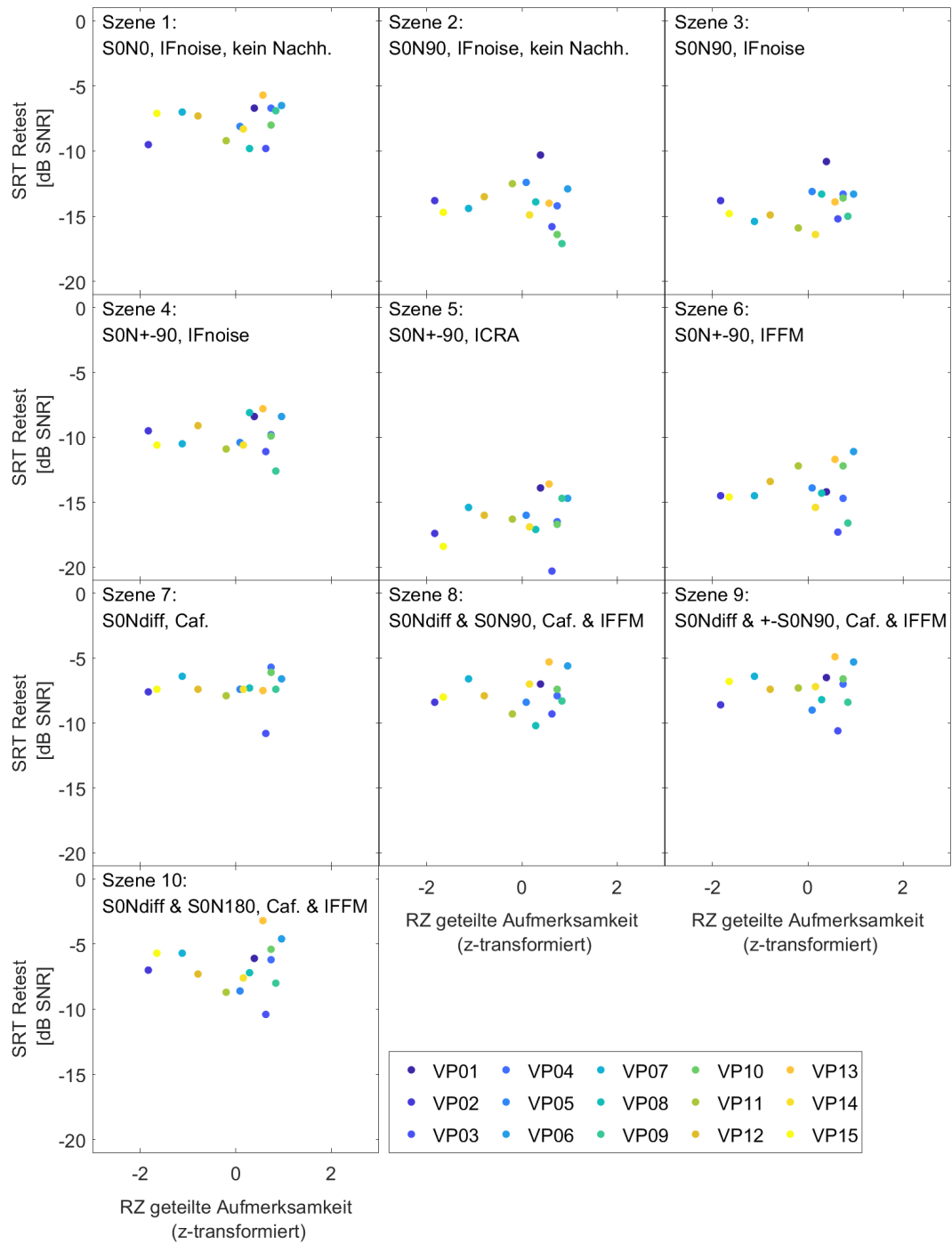
**Tabelle D.1:** Test auf Normalverteilung mittels Shapiro-Wilk für den ersten Termin (Test) und den zweiten Termin (Retest) der Daten der subjektiven Höranstrengungsmessung (ACALES).

Test		Retest	
Szene	Irrtumswahrscheinlichkeit $p$	Szene	Irrtumswahrscheinlichkeit $p$
1	0,401	1	0,163
2	0,193	2	0,908
3	0,673	3	0,325
4	0,652	4	0,029
5	0,455	5	0,876
6	0,986	6	0,990
7	0,438	7	0,913
8	0,635	8	0,975
9	0,312	9	0,998
10	0,958	10	0,971

GETEILTE AUFMERKSAMKEIT

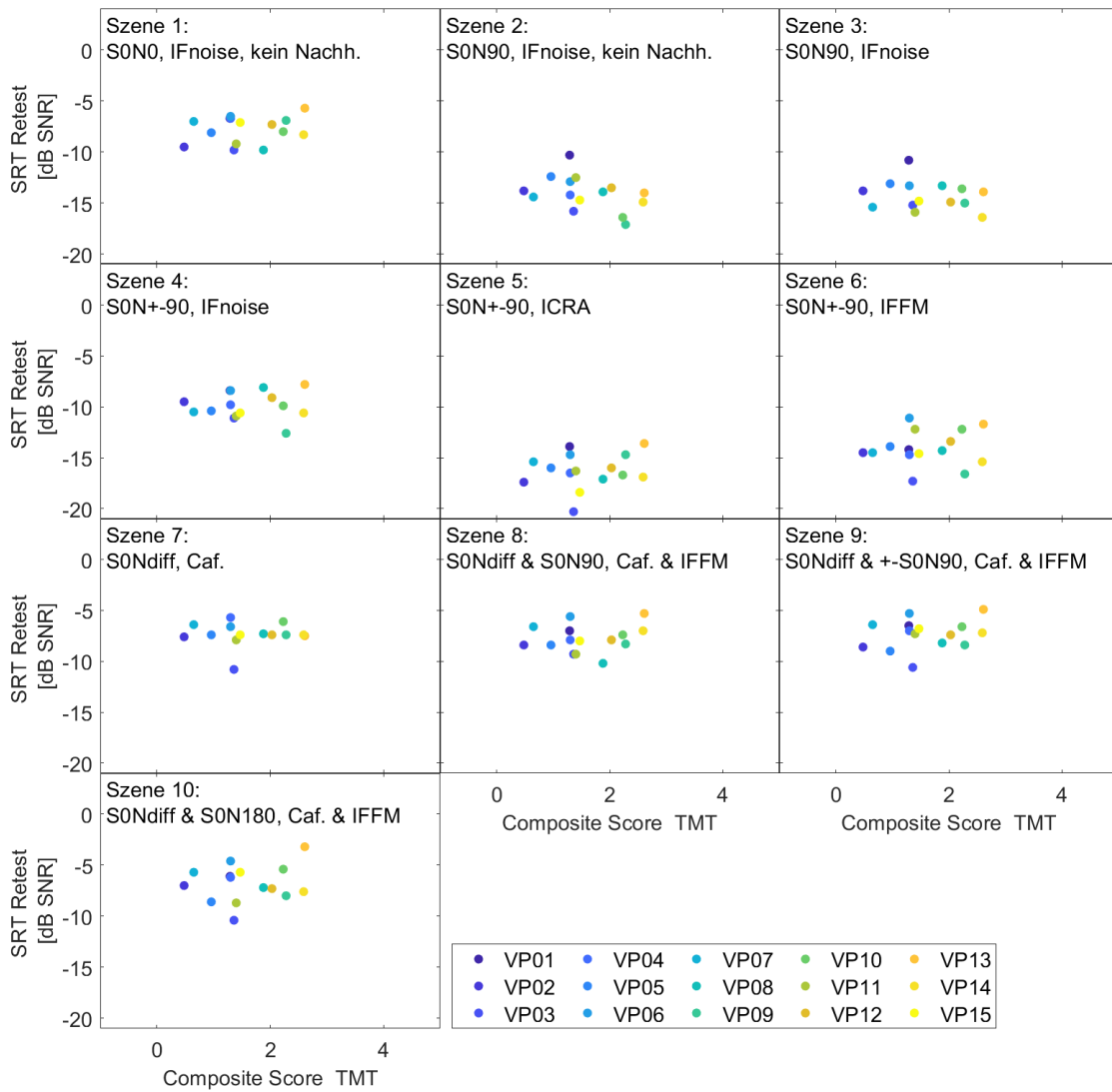


**Abbildung E.1:** Streudiagramm der z-transformierten Reaktionszeit (RZ) des geteilten Aufmerksamkeits-tests in ms gegenüber dem Signal-Rausch-Abstand (SNR) des Messverfahrens zur subjektiven Höranstrengung (ACALES) bei ESCU 11 des zweiten Termins (Retest) in dB SNR. Zusätzlich wurden für die Szenen in denen eine Korrelation zwischen den Ergebnissen des geteilten Aufmerksamkeits-tests und ACALES auftreten, eine Regressionsgerade eingefügt, gestrichelte Linie. Jedes der zehn Streudiagramme zeigt eine der zehn Szenen (vergleiche Tabelle 3.1).

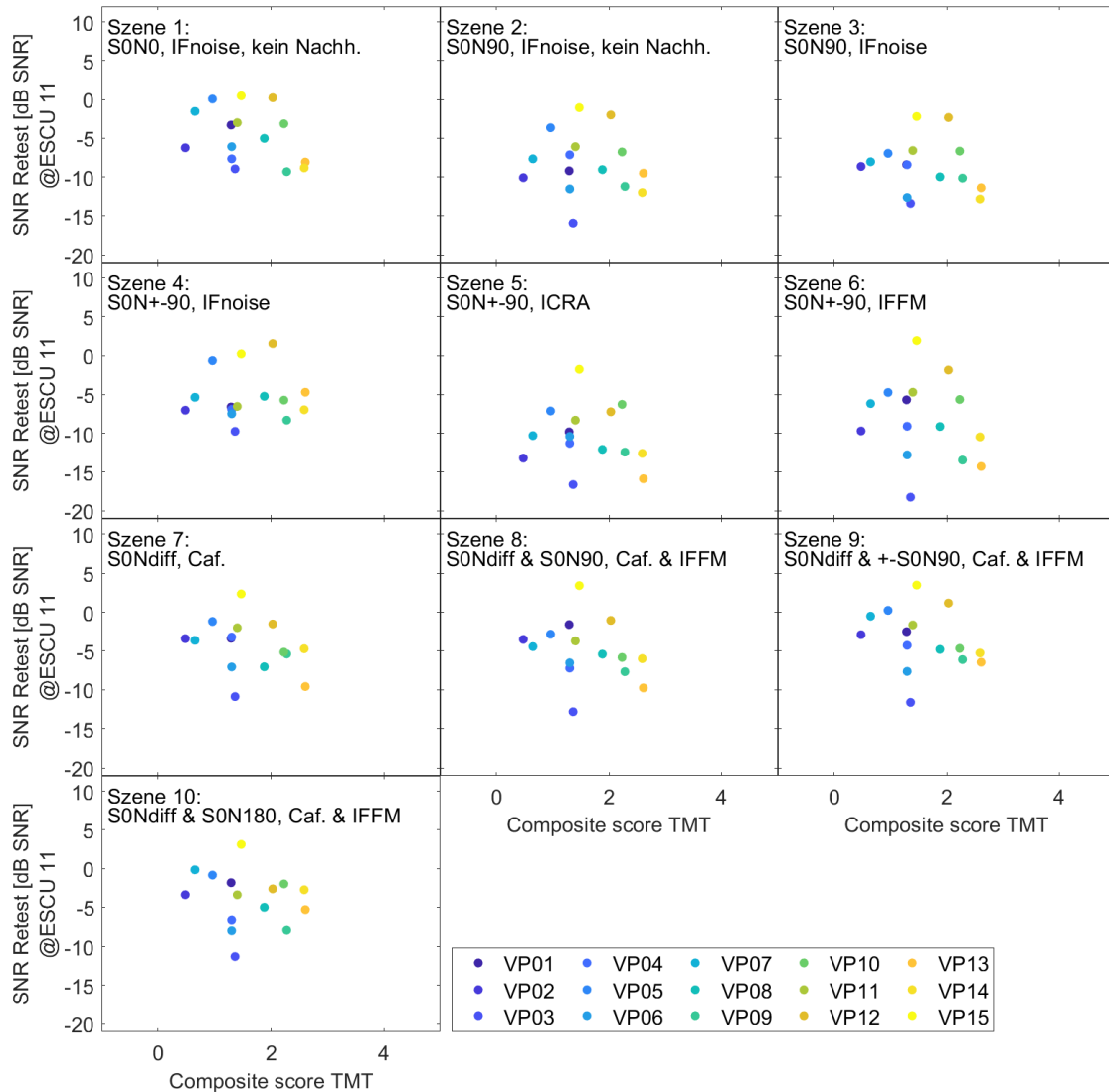


**Abbildung E.2:** Streudiagramme der z-transformierten Reaktionszeit (RZ) des geteilten Aufmerksamkeitstests in ms gegenüber der Sprachverständlichkeitsschwelle (SRT) des Oldenburger Satztests (OLSA) des zweiten Termins (Retest) in dB SNR für jede der zehn Szenen (vergleiche Tabelle 3.1).

TRAIL-MAKING-TEST



**Abbildung F.1:** Aufgetragen sind die Composite Scores (Fuellgrabe et al., 2015) des Trail-Making-Tests (TMT) gegenüber den Sprachverständlichkeitsschwelle (SRT) des Oldenburger Satztestes (OLSA) des zweiten Termins (Retest) in dB SNR für jeder der zehn Szenen (vergleiche Tabelle 3.1).



**Abbildung F.2:** Aufgetragen sind die Composite Scores (Fuellgrabe et al., 2015) des Trail-Making-Tests (TMT) gegenüber des Signal-Rausch-Abstandes (SNRs) des Messverfahren zur subjektiven Höranstrengung (ACALES) bei ESCU 11 des zweiten Termins (Retest) in dB SNR für jeder der zehn Szenen (vergleiche Tabelle 3.1).



## GERÄTELISTE

- Oskop: HEINE mini 3000, Heine, Gilching, Deutschland
- Audiometer: Aurical II, GN Otometrics, Taastrup, Dänemark
- Kopfhörer: HDA 200, Sennheiser, Wedemarkt, Deutschland
- D/A-Wandler: ADI-8 pro, RME Audio Interface, Haimhausen, Deutschland
- Software: Oldenburger Messaparatur (OMA) F&E 2.3
- Lautsprecher: 8030B, Genelec, Iisalmi, Finnland
- Software Windows-Rechner: Matlab, Version 2022a, Mathworks
- Software Linux-Rechner: TASCAR 0.277.1.2
- Kopfhörer: HD580, Sennheiser, Wedemarkt, Deutschland
- Ansteckmikrofon: MCE 100, EKULIT, Ostfildern, Deutschland

## GLOSSAR

**ACALES** engl. *Adaptive Categorical Listening Effort Scaling*, adaptive kategoriale Skalierung zur Erfassung der subjektiven Höranstrengung

**ESCU** engl. *Effort Scale Categorical Units*, kategorische Einheiten der Aufwandsskala

**ANOVA** engl. *Analysis of Variance*, Varianzanalyse

**EEG** Elektroenzephalografie

**EM** engl. *Energetic Masking*, energetische Maskierung

**FBE** Freiburger Einsilbertest

**GÖSA** Göttinger Satztest

**HL** engl. *hearing loss*, Hörverlust

**IFnoise** engl. *International Female Noise*, stationäres Rauschen, welches das Langzeitspektrum internationaler Sprecherinnen aufweist

**IKK** Intraklassen-Korrelationskoeffizient

**ILD** engl. *Interaural Level Differences*, interaurale Pegeldifferenzen

**IM** engl. *Informational Masking*, informationelle Maskierung

**ISTS** engl. *International Speech Test Signal*, internationale Sprachtestsignal

**ITD** engl. *Interaural Time Differences*, interaurale Zeitdifferenzen

**LTASS** engl. *Long-Term Average Speech Spectrum*, Langzeitspektrum der Sprache

**NH** Normalhörende Person

**OLnoise** Rauschen, welches aus Sätzen des Oldenburger Satztests generiert wurde

**OLSA** Oldenburger Satztest

**OMA** Oldenburger Messprogramm des Hörzentrum Oldenburg gGmbH

**PTA<sub>4</sub>** engl. *Pure Tone Average*, Reintonmittelwert, Mittelwert über 0,5, 1, 2 und 4 kHz sort

**RZ** Reaktionszeit(en)

**ICRA4-250** sprachfluktuierendes Geräusch mit einem weiblich gewichteten, idealisierten Sprachspektrum einer weiblichen Sprecherin

**IFFM** engl. *International Female Fluctuation Masker*,

**SD** Standardabweichung

**SH** Schwerhörige Person

**SNR** engl. *Signal-to-Noise ratio*, Signal-Rauschabstand. Das Verhältnis von Signal- zu Rauschleistung.

**SRM** engl. *Spatial Release from Masking*, räumliche Befreiung von der Maskierung

**SRT** engl. *Speech Recognition Threshold*, Sprachverständlichkeitsschwelle (d.h. dem SNR, bei dem 50 % der Testwörter korrekt wiedergegeben werden können

**NASA-TLX** „Natioanl Aeronautics and Space Administration Task Load Index “

**SSQ** „Speech, Spatial and Qualities of Hearing Scale“, Fragebogen, welcher Fragen zum Sprachverstehen, zum räumlichen Hören und zu weiten Bereichen der Hörqualität beinhaltet

**TAP** Testbatterie zur Aufmerksamkeitsprüfung

**TASCAR** engl. *Toolbox for acoustic scene creation and rendering*, Toolbox für die Erstellung und das Aufbereiten von akustischen Szenen

**TMT** Trail Making Test

**WM** engl. *Working Memory*, Arbeitsgedächtnis

## LITERATUR

- Alhanbali, S., Dawes, P., Millman, R. E. und Munro, Kevin J. (2019). „Measures of Listening Effort Are Multidimensional“. In: *Ear and Hearing* 40(5), S. 1084–1097.
- Amiri, M. und Jarollahi, F. (2020). „An overview on informational masking“. In: *Auditory and Vestibular Research* 29, S. 128–139.
- Benjamini, Yoav und Hochberg, Yosef (1995). „Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing“. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, S. 289–300.
- Bernstein, J. G. W. und Grant, K. W. (2009). „Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners“. In: *The Journal of the Acoustical Society of America* 125.5, S. 3358–3372.
- Bertol, S. und Bodmer, D. (2016). „Effects of age and task difficulty on ERP responses to novel sounds presented during a speech-perception-in-noise test“. In: *Clinical Neurophysiology* 127.1, S. 360–368.
- Bronkhorst, A. W. (2015). „The cocktail-party problem revisited: early processing and selection of multi-talker speech“. In: *Attention, Perception, & Psychophysics* 77, S. 1465–1487.
- Brungart, D., Simpson, B., Ericson, M. und Scott, K. (2001). „Informational and energetic masking effects in the perception of multiple simultaneous talkers.“ In: *The Journal of the Acoustical Society of America* 110 5 Pt 1, S. 2527–2538.
- Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hayerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Notby, M., Abdelnasser, N., El-Kholy, W., Oyer, H., Powell, R., Stephens, d., Meredith, R. und Ludvigsen, C. (1996). „An international comparison of long-term average speech spectra“. In: *Journal of The Acoustical Society of America* 96, S. 2108–2120.
- Cueille, R., Lavandier, M. und Grimault, N. (2022). „Effects of reverberation on speech intelligibility in noise for hearing-impaired listeners“. In: *Royal Society Open Science* 9, S. 1–20.
- Culling, J. F. und Stone, M. A. (2018). „Energetic Masking and Masking Release“. In: *The Auditory System at the Cocktail Party*. Springer Handbook of Auditory Research. Middlebrooks, J. C. u. a. Kap. 3, S. 41–73.
- Dreschler, W., Herschuure, H., Ludvigsen, C. und Westermann, S. (2001). „ICRA Noises: Artificial Noise Signals With Speech-Like Spectral and Temporal Properties for Hearing Aid Assessment“. In: *Audiology: official oragan of the International Society of Audiology* 40, S. 148–157.
- Dryden, A., Allen, H.-A., Henshaw, H. und Heinrich, A. (2017). „The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis“. In: *Trends in Hearing* 21, S. 1–20.

- Fiedler, L., Seifi Ala T. Graversen, C., Alickovic, E., Lunner, T. und Wendt, D. (2021). „Hearing Aid Noise Reduction Lowers the Sustained Listening Effort During Continuous Speech in Noise—A Combined Pupillometry and EEG Study“. In: *Ear and Hearing* 42(6), S. 1590–1601.
- Fuellgrabe, C., Moore, B. C. J. und Stone, M. A. (2015). „Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition“. In: *Frontiers in Aging Neuroscience* 6, S. 1–25.
- Fuellgrabe, C. und Rosen, S. (2016). „On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds“. In: *Frontiers in Psychology* 07, S. 1–8.
- Gagné, J. P., Besser, J. und Lemke, U. (2017). „Behavioral Assessment of Listening Effort Using a Dual-Task Paradigm: A Review“. In: *Trends in Hearing* 21, S. 1–25.
- Gatehouse, S. und Noble, W. (2004). „The Speech, Spatial and Qualities of Hearing Scale (SSQ)“. In: *International Journal of Audiology* 43.2, S. 85–99.
- Grange, J. und Culling, J. (2016). „The benefit of head orientation to speech intelligibility in noise“. In: *The Journal of the Acoustical Society of America* 139, S. 703–712.
- Grimm, G. und Hohmann, V. (2019). *First Order Ambisonics field recordings for use in virtual acoustic environments in the context of audiology*. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 B1.
- Hahlbrock, K. H. (1953). „Über Sprachaudiometrie und neue Wörterteste“. In: *Archiv für Ohren-, Nasen- und Kehlkopfkunde* 162(5), S. 394–431.
- Hall, J., Buss, E., Grose, J. und Roush, P. (2012). „Effects of Age and Hearing Impairment on the Ability to Benefit From Temporal and Spectral Modulation“. In: *Ear and hearing* 33, S. 340–348.
- Hart, S. G. und Staveland, L. E. (1988). „Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research“. In: *Ear and Hearing* 52, S. 139–183.
- Helfer, K. S. und Wilber, L. A. (1990). „Hearing Loss, Aging, and Speech Perception in Reverberation and Noise“. In: *Journal of Speech, Language, and Hearing Research* 33.1, S. 149–155.
- Holube, I. (2011). „Speech intelligibility in fluctuating maskers“. In: *International Symposium on Auditory and Audiological Research (ISAAR), Nyborg, Dänemark*.
- Holube, I., Blab, S., Fürsen, K., Gürtler, S., Meisenbacher, K., Nguyen, D. und Taesler, S. (2009). „Einfluss des Maskierers und der Testmethode auf die Sprachverständlichkeitsschwelle von jüngeren und älteren Normalhörenden“. In: *Zeitschrift für Audiologie* 48, S. 120–127.
- Holube, I., Fredelake, S., Vlaming, M. und Kollmeier, B. (2010). „Development and analysis of an International Speech Test Signal (ISTS)“. In: *International Journal of Audiology* 49.12, S. 891–903.
- Holube, I., Haeder, K., Imbery, C. und Weber, R. (2016). „Subjective Listening Effort and Electrodermal Activity in Listening Situations with Reverberation and Noise“. In: *Trends in hearing* 20, S. 1–15.
- ITU (1988). „Conventional telephone signal“. In: *ITU Recommendation G.227*. (aufgerufen am 13.12.2022). URL: <https://www.itu.int/rec/T-REC-G.227-198811-I/en>.
- Kayser, H., Ewert, S., Anemüller, Jörn, Rohdenburg, Thomas, Hohmann, V. und Kollmeier, B. (2009). „Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural

- Room Impulse Responses“. In: *EURASIP Journal on Advances in Signal Processing* 2009, S. 1–10.
- Kidd, G. und Colburn, H. (2018). „Informational masking in speech recognition“. In: *The Auditory System at the Cocktail Party*. Springer Handbook of Auditory Research. Middlebrooks, J. C. u. a. Kap. 4, S. 77–109. ISBN: 9783319847115.
- Kollmeier, B., Lenarz, T., Winkler, A., Zokoll, M. A., Sukowski, H., Brand, T. und Wagener, K. C. (2011). „Hörgeräteindikation und -überprüfung nach modernen Verfahren der Sprachaudiometrie im Deutschen“. eng. In: *HNO* 59.10, S. 1012–1021.
- Kollmeier, B. und Wesselkamp, M. (1997). „Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment“. In: *The Journal of the Acoustical Society of America* 102.4, S. 2412–2421.
- Koo, T. K. und Li, M. Y. (2016). „A guideline of selecting and reporting intraclass correlation coefficients for reliability research“. In: *Journal of Chiropractic Medicine* 15(2), S. 155–163.
- Krueger, M., Schulte, M. und Brand, T. (2022). „Assessing and Modeling Spatial Release From Listening Effort in Listeners With Normal Hearing: Reference Ranges and Effects of Noise Direction and Age“. In: *Trends in Hearing* 26, S. 1–19.
- Krueger, M., Schulte, M., Brand, T. und Holube, I. (2017). „Development of an adaptive scaling method for subjective listening effort“. eng. In: *The Journal of the Acoustical Society of America* 141.6, S. 4680–4693.
- Luts, H., Eneman, K., Wouters, J., Schulte, M., Vormann, M., Buechler, M., Dillier, N., Houben, R., Dreschler, W. A., Froehlich, M., Puder, H., Grimm, G., Hohmann, V., Leijon, A., Lombard, A., Mauler, D. und Spriet, A. (2010). „Multicenter evaluation of signal enhancement algorithms for hearing aids“. Englisch. In: *Journal of the Acoustical Society of America* 127.3, S. 1491–1505.
- Mackersie, C. L. und Calderon-Moultrie, N. (2016). „Autonomic Nervous System Reactivity During Speech Repetition Tasks: Heart Rate Variability and Skin Conductance“. In: *Ear and Hearing* 37, 118S–125S.
- McGarrigle, R., Munro, K. J., Dawes, P., Stewart A., J., Moore, D. R., Barry, J. G. und Amitay, S. (2014). „Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group ‘white paper’“. In: *International Journal of Audiology* 53.7, S. 433–445.
- McMahon, Ca. M., Boisvert, I., Lissa, P. de, Granger, L., Ibrahim, R., Lo, C. Y., Miles, K. und Graham, P. L. (2016). „Monitoring Alpha Oscillations and Pupil Dilation across a Performance-Intensity Function“. eng. In: *Frontiers in psychology* 7, S. 745–745.
- Meister, H. (2020). „Sprachverstehen und kognitive Leistungen in akustisch schwierigen Situationen“. ger. In: *HNO* 68.3, S. 171–176.
- Nuesse, T., Steenken, R., Neher, T. und Holube, I. (2018). „Exploring the Link Between Cognitive Abilities and Speech Recognition in the Elderly Under Different Listening Conditions“. In: *Frontiers in Psychology* 9, S. 1–17.
- Obleser, J., Woestmann, M., Hellbernd, N., Wilsch, A. und Maess, B. (2012). „Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network“. In: *Journal of Neuroscience* 32.36, S. 12376–12383.
- Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., Versfeld, N. J. und Kramer, Sophia E. (2017). „Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation“. In: *Hearing Research* 351, S. 68–79.

- Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L. und Wingfield, A. (2016). „Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)“. In: *Ear and hearing* 37.1, 5S–27S.
- Ronnberg, J., Lunner, T., Zekveld, A. A., Sorqvist, P., Danielsson, H., Lyxell, B., Dahlstrom, O., Signoret, C., Stenfelt, S., Pichora-Fuller, M. K. und Rudner, M. (2013). „The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances“. Undefined/Unknown. In: *Frontiers in Systems Neuroscience* 7, S. 1–17.
- Rudner, M., Lunner, T., Behrens, T., Thorén, E. und Rönnberg, J. (Sep. 2012). „Working Memory Capacity May Influence Perceived Effort during Aided Speech Recognition in Noise“. In: *Journal of the American Academy of Audiology* 23, S. 577–89.
- Sanchez Lopez, R., Bianchi, F., Fereczkowski, M., Santurette, S. und Dau, T. (2018). „Data-Driven Approach for Auditory Profiling and Characterization of Individual Hearing Loss“. In: *Trends in Hearing* 22, S. 1–12.
- Seeman, S. und Sims, R. (2015). „Comparison of Psychophysiological and Dual-Task Measures of Listening Effort“. In: *Journal of Speech, Language, and Hearing Research* 58.6, S. 1781–1792.
- Seifi Ala, T., Graversen, C., Wendt, D., Alickovic, E., Whitmer, W. und Lunner, T. (2020). „An exploratory Study of EEG Alpha Oscillation and Pupil Dilation in Hearing-Aid Users During Effortful listening to Continuous Speech“. In: *PLOS ONE* 15, S. 1–15.
- Steffens, T. (2015). „Der Freiburger Sprachtest: ein Klassiker mit Stärken und Schwächen“. In: *HNO Nachrichten* 45, S. 22–25.
- Talbott, R. E. und Larson, V. D. (1983). „Research needs in speech Audiometry“. In: *Semin Hear* 4(3), S. 299–308.
- Wagener, K. C. (2004). *Factors Influencing Sentence Intelligibility in Noise*. B.S. ISBN: 9783814208978.
- Wagener, K. C. und Brand, T. (2005). „Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters“. eng. In: *International journal of audiology* 44.3, S. 144–156.
- Wagener, K. C., Brand, T. und Kollmeier, B. (1999a). „Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests“. In: *Zeitschrift für Audiologie* 2, S. 44–56.
- Wagener, K. C., Brand, T. und Kollmeier, B. (1999b). „Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests“. In: *Zeitschrift für Audiologie* 3, S. 86–95.
- Wagener, K. C., Brand, T. und Kollmeier, B. (2006). „The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners“. In: *International Journal of Audiology* 45.1, S. 26–33.
- Wagener, K. C., Hochmuth, S., Ahrlich, M. und Kollmeier, B. (2014). „Der weibliche Oldenburger Satztest“. In: *gehalten auf der 17. Jahrestagung der DGA, Oldenburg*.
- Wagener, K. C., Kühnel, V. und Kollmeier, B. (1999c). „Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests“. In: *Zeitschrift für Audiologie* 1, S. 4–15.



- Wang, Y., Naylor, G., Kramer, S. E., Zekveld, A. A., Wendt, D., Ohlenforst, B. und Lunner, T. (2017). „Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During a Speech Perception in Noise Task“. eng. In: *Ear and hearing* 39.3, S. 573–582.
- Winkler, A. und Holube, I. (2014). „Was wissen wir über den Freiburger Sprachtest?“ In: *Zeitschrift für Audiologie (Audiological Acoustics)* 53, S. 146–154.
- Woestmann, M., Herrmann, B., Maess, B. und Obleser, J. (2016). „Spatiotemporal dynamics of auditory attention synchronize with speech“. In: *Proceedings of the National Academy of Sciences* 113.14, S. 3873–3878.
- Zekveld, A. A., Rudner, M., Kramer, S. E., Lyzenga, J. und Rönnerberg, J. (2014). „Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech“. In: *Frontiers in Neuroscience* 8, S. 1–11.
- Zimmermann, P. und Fimm, B. (2013). „Testbatterie zur Aufmerksamkeitsprüfung Version 2.3 Teil 1“. In: *Herzogenrath: Psyttest*. 2.Version.
- Zinner, C., Winkler, A. und Holube, I. (2021). „Vergleich von fünf Sprachtests im sprachsimulierenden Störgeräusch“. In: *GMS Zeitschrift für Audiologie - Audiological Acoustics*. URL: <https://www.egms.de/static/en/journals/zaud/2021-3/zaud000016.shtml>.

## DANKSAGUNG

An dieser Stelle möchte ich allen danken, die mich bei Verfassen dieser Bachelorarbeit unterstützt haben. Insbesondere bei Prof. Dr. rer. nat. Inga Holube und Dr. rer. nat. Hendrik Kayser. Ihr habt es mir ermöglicht, meine Bachelorarbeit im Hörzentrum Oldenburg zu schreiben. Außerdem möchte ich mich bei euch sowie Theresa Jansen, Melanie Krüger und Laura Hartog für alle Ideen und Ratschläge, die ihr mir gegeben habt, bedanken. Ihr hattet immer ein offenes Ohr für mich. Außerdem möchte ich mich bei allen weiteren Mitarbeitern am Hörzentrum und meinen Probanden bedanken, die mich bei der Durchführung meiner Studie und Erstellung meiner Arbeit unterstützt haben. Ich habe mich zu jedem Zeitpunkt sehr gut aufgehoben gefühlt und habe mich bei Fragen immer sehr gerne an euch gewandt. Außerdem möchte ich mich bei meinen Freunden bedanken. Danke für eure Geduld und die aufmunternden Worte. Mein letzter Dank gilt meinen Eltern, die mich in jeder Lebenslage unterstützen und mir alles ermöglicht haben. Danke!

# EIGENSTÄNDIGKEITSERKLÄRUNG

**Erklärung gemäß §1 (5) Allgemeiner Teil (Teil A) der Prüfungsordnung für die Bachelor-Studiengänge (BPO) an der Jade Hochschule Wilhelmshaven/Oldenburg/Elsfleth in der Fassung vom 21.10.2014 (Verköndungsblatt Nr. 56/2014 der Jade Hochschule)**

Diese Bachelorarbeit ist eine Einzelarbeit. Ich erkläre hiermit, diese Bachelorarbeit selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

.....  
Ort, Datum, Unterschrift (Sophie Jäker Matr.-Nr. 6032270)

# LIZENZ

© Sophie Jäker

Diese Arbeit ist lizenziert unter einer  
**Creative Commons Namensnennung 4.0 International** Lizenz.

Eine Kopie der Lizenz sollte dieser Arbeit beiliegen. Falls nicht, ist sie hier abrufbar:  
<https://creativecommons.org/licenses/by/4.0/>