

Aus der Klinik und Poliklinik für Hals-, Nasen- und Ohren-
krankheiten, plastische und ästhetische Operationen
der Universität Würzburg

Direktor: Prof. Dr. med. Dr. h. c. Rudolf Hagen

**Kurz- und langfristige
Lern- und Habituationseffekte
bei der Anwendung des Oldenburger Satztests
an jungen, normalhörenden Probanden**

Dissertation
zur Erlangung der Doktorwürde der
Medizinischen Fakultät
der
Julius-Maximilians-Universität Würzburg
vorgelegt von

Johannes Heyn

aus Bamberg

Würzburg, Dezember 2019

Referent: Prof. Dr. med. Dr. h. c. Rudolf Hagen
Koreferent: Priv.-Doz. Dr. med. Regina Taurines
Dekan: Prof. Dr. med. Matthias Frosch

Tag der mündlichen Prüfung: 27. April 2020

Der Promovend ist Arzt

Für meine Familie!

Inhaltsverzeichnis

1	Einleitung	1
1.1	Das auditive System	2
1.1.1	Anatomische Grundlagen	2
1.1.2	Physiologie des Hörens	4
1.1.3	Pathophysiologie des Hörens	4
1.2	Hördiagnostik	5
1.2.1	Klinische Orientierungstests	5
1.2.2	Reintonaudiometrie	6
1.2.3	Sprachaudiometrie	6
1.3	Oldenburger Satztest	12
1.3.1	Entwicklung und Eigenschaften	12
1.3.2	Kurz- und langfristige Lerneffekte	14
1.3.3	Bisherige, relevante Studien zu Lerneffekten	15
1.3.4	Kritische Bewertung	20
1.4	Fragestellungen der Dissertation	21
1.5	Wissenschaftspolitischer Hintergrund der Arbeit	23
2	Versuchsplanung und Methodik	25
2.1	Probanden	25
2.2	Zeitlicher Ablauf der Tests	26
2.3	Testmethoden	27
2.3.1	Reintonaudiogramm	27
2.3.2	Durchführung des Oldenburger Satztests in dieser Arbeit	28
2.4	Testumfeld und technische Ausstattung	29
2.5	Versuchsdesign	30
2.5.1	Versuchsaufbau	30
2.5.2	Versuchsablauf	33
2.5.3	Testleitung, Steuerung und Dokumentation	35

2.6	Statistische Auswertung	36
3	Ergebnisse	39
3.1	Tonaudiogramme der Probanden	39
3.2	Bestimmung der selektierten Hauptgruppe der Probanden	41
3.3	Einflussfaktoren auf das Sprachverständnis	43
3.4	Darstellung der Absolutwerte der Schalldruckpegel des Sprachsignals .	56
3.5	Korrelation der Testsitzungen und -durchläufe hinsichtlich des SRT-Werts	57
3.6	Gesamtdauer der Testdurchführung und Antwortzeit	59
4	Diskussion	69
4.1	Reintonaudiometrie	69
4.2	Kurz- und langfristige Lerneffekte beim Oldenburger Satztest	70
4.2.1	Diskussion um den Begriff „Lerneffekt“	70
4.2.2	Untersuchte Einflussfaktoren auf den Lerneffekt	72
4.2.3	Nicht untersuchte Einflussfaktoren auf den Lerneffekt	76
4.3	Kurz- und langfristige Lerneffekte bei Sprachverständnistests	78
4.4	Lösungsvorschläge zur Eradikation des Lerneffekts	84
4.5	Interpretation der Berechnungsvorschrift des SRT-Werts	87
4.6	Untersuchung der Antwortzeit und Testzeit	88
4.7	Überlegungen zum Oldenburger Satztest	91
4.8	Stellenwert des Oldenburger Satztests	93
4.9	Abschluss und Ausblick	94
5	Zusammenfassung	95
6	Literaturverzeichnis	97
A	Rohdaten	107
A.1	SRT-Werte aller Probanden aller Testsitzungen aller Durchläufe	107
A.2	3D-Darstellung der Mittelwerte der SRT-Berechnung	129
A.3	SRT-Verlaufswerte der nicht selektierten Probanden	155
B	Vollständige Darstellung der Varianzanalysen	159

Danksagung

Lebenslauf

1 Einleitung

Kommunikation – ermöglicht durch unser Gehör als wichtigem Sinnesorgan – bildet eine wesentliche Säule der sozialen Integration. Das Herausfiltern von Stimmen aus einer Umgebung voll von Störgeräuschen und Umgebungslärm jeglicher Art ist dabei eine zentrale, entscheidende Anforderung an das menschliche Gehör. Im Alltag stellt diese Aufgabe je nach Geräuschkulisse, Alter, Konzentrationsvermögen und anderen individuellen Voraussetzungen eine mehr oder weniger große Herausforderung dar. Gerade ältere Menschen klagen dabei über Schwierigkeiten; seltener sind auch junge Menschen von relevanten Hörstörungen betroffen. Für eine bleibende soziale Teilhabe sorgte die Erfindung des elektrischen Hörgeräts Anfang des 20. Jahrhunderts (König, 1979). In Anbetracht der demographischen Entwicklung mit einer Umverteilung der Alterspyramide in Richtung Urnen- bzw. Zwiebelform kann zukünftig von einer größeren Patientenklientel ausgegangen werden, bei der die Indikation für die Versorgung mit einem Hörgerät bestehen wird. Genauso wichtig wie das Hören selbst, ist die Aufnahme und Verarbeitung des Gehörten (Hörverständnis und Sprachperzeption), welche die Analyse und Anwendung des Gehörten bzw. das Erfassen des Inhalts garantiert. So wird der Erwerb neuer Informationen (bewusstes und unbewusstes Lernen) zum großen evolutionstechnischen Vorteil.

Bestimmte Hörtests zielen exakt auf die Testung der speziellen Filterfunktion des Gehörs ab, indem zusätzlich zu einer aufgesprochenen Stimme ein Störgeräusch präsentiert wird. Auf diese Weise kann in der Sprachaudiometrie das Sprachverständnis quantifiziert werden.

In dieser Dissertation werden die kurz- und langfristigen Lern- und Habituationseffekte eines solchen sprachaudiometrischen Tests, des Oldenburger Satztests (abgekürzt: OLSA) untersucht. Dieser Test wurde im Jahre 1999 von einer Arbeitsgruppe am Department für Medizinische Physik und Akustik der Universität Oldenburg entwickelt (Wagner et al., 1999a,b,c).

Aus Gründen der Zugehörigkeit und besseren Übersicht wird der Oldenburger Satztest in Kapitel 1.3 auf Seite 12 näher beschrieben und sein Funktionsprinzip explizit erläutert.

1.1 Das auditive System

Das Ohr ist das empfindlichste Sinnesorgan des Menschen (Schmidt und Lang (2007, S. 344)).

Schallwellen als periodische Druckschwankungen in der Luft sind der adäquate Reiz für das Gehör (Schünke et al. (2009, S. 132)).

Der gerade noch wahrnehmbare Schalldruck p_0 ist der Bezugsschalldruck und liegt bei 2×10^{-5} Pascal und entspricht in etwa der Hörschwelle (Huppelsberg und Walter (2013, S. 337 ff.)).

Der Hörbereich liegt beim Menschen im Bereich zwischen 16 Hz und 16 bis 20 kHz (Pschyrembel, 2007).

1.1.1 Anatomische Grundlagen

Das Ohr setzt sich aus einem peripheren und einem zentralen Anteil zusammen (Lenarz und Boenninghaus (2007, S. 6-27)):

1. Den peripheren Teil bilden das äußere Ohr, das Mittelohr, das Innenohr sowie der Nervus vestibulocochlearis.
2. Zum zentralen Teil werden die Hörbahn, die Vestibularbahnen und die zentralen Anteile des Nervus facialis gezählt.

Peripherer Anteil des Ohrs: Das äußere Ohr besteht aus Ohrmuschel, äußerem Gehörgang und Trommelfell. Die Paukenhöhle, die Ohrtrumpete (mit einer Verbindung zum Nasenrachenraum) und die Kette der Gehörknöchelchen (Malleus, Incus und Stapes) bilden das Mittelohr. Im Innenohr ist das cochleäre Labyrinth mit dem membranösen Schnecken gang und der knöchernen Hülle lokalisiert.

Der etwa 35 mm lange Schneckenkanal ist 2,5-fach um seine eigene Achse (den Modiolus) gewunden. Im Querschnitt durch die Cochlea können drei flüssigkeitsgefüllte Räumen (die Skalen) voneinander abgegrenzt werden: Die Scala tympani und die Scala vestibuli enthalten Perilymphe und sind an der Schneckenspitze (dem Helicotrema) miteinander verbunden. Die Scala media (der Ductus cochlearis) enthält Endolymphe und endet blind. Die Flüssigkeiten unterscheiden sich hinsichtlich der Kalium⁺- und Natrium⁺-Ionenkonzentrationen. Die Basilarmembran und die Reissner-Membran begrenzen den Ductus cochlearis zur Scala tympani bzw. zur Scala vestibuli.

Der Begriff der Tonotopie besagt, dass auf der zur Schneckenspitze immer breiter werdenden Basilarmembran unterschiedliche Tonhöhen (Frequenzen) an verschiede-

nen Orten repräsentiert sind: tiefe Töne werden an den breiten Teilen der Basilarmembran (d. h. zur Schneckenspitze hin) erfasst, wohingegen hohe Töne an den schmalen Teilen (d. h. zur Schneckenbasis hin) detektiert werden. Einer der Gründe für die Tonotopie ist, dass die Basilarmembran basisnah starr ist und im weiteren Verlauf zur Spitze an Elastizität gewinnt. Auf der Basilarmembran liegt das Corti-Organ, als Abgrenzung des Ductus cochlearis zur Scala tympani. Das Corti-Organ ist das eigentliche Hörorgan und besteht neben funktionsunterstützenden Zellen aus den inneren und den äußeren Haarzellen. Auf den Haarzellen sitzen jeweils 50 bis 100 Stereozilien, über die sich die Tektorialmembran erstreckt. Die inneren, überwiegend afferent innervierten Haarzellen sind Rezeptoren und wandeln die Schwingung der Basilarmembran in elektrische Impulse um und leiten diese über das Ganglion cochleare und den Nervus cochlearis zum Hirnstamm weiter. Die äußeren, überwiegend efferent innervierten Haarzellen verstärken die durch Schallwellen ausgelöste Wanderwelle, indem sie ihre Länge verändern (Gekle (2010, S. 657-667), Benninghoff und Drenckhahn (2008, S. 538-551), Schünke et al. (2009, S. 126-141)).

Zentraler Anteil des Hörsystems: Die Hörbahn, die Vestibularbahnen und die zentralen Anteile des Nervus facialis sind als Teil des zentralen Nervensystems an der Reizverarbeitung und der bewussten Wahrnehmung akustischer Signale beteiligt. Das afferente Nervensystem besteht aus vier Neuronen, die die inneren Haarzellen über das Ganglion spirale cochleae (1), die Nuclei cochleares, die Oliva superior (2), den Nucleus accessorius, den seitlichen Schleifenkern, den Colliculus inferior (3) und das Corpus geniculatum mediale (Thalamus, 4) mit der primären Hörrinde (in den Heschl-Windungen des Temporallappens) verbindet. Von dort werden sekundäre Hörregionen angesteuert, welche das Analysieren von Klängen, das Vergleichen von ähnlichen akustischen Reizen und das Lauschen ermöglichen.

In den drei Nuclei cochleares (anteroventral, posteroventral, dorsal) findet die tonotopie Zuordnung der verschiedenen Frequenzen statt. Hier kommt es außerdem zur lateralen Hemmung und infolgedessen zur Kontrastierung und Rauschunterdrückung. Durch den Vergleich der Schallintensität und der Laufzeitlatenz des rechten und linken Ohres in der oberen Olive und dem Nucleus accessorius, wo die Neurone im Verlauf der Hörbahn erstmalig nach kontralateral kreuzen, wird das Richtungshören möglich. Die überwiegend nach kontralateral gekreuzten Neurone im Colliculus inferior koordinieren Reflexe und stellen die Verbindung zwischen Auge und Ohr her.

Das efferente Nervensystem modifiziert über die äußeren Haarzellen akustische Reize (Lenarz und Boeninghaus (2007, S. 6-27), Silbernagl (2012)).

1.1.2 Physiologie des Hörens

Zusammenfassend kann der Hörvorgang aus physiologischer Sicht in Reizaufnahme (Luftleitung oder Knochenleitung), Reiztransport (mechanische Übertragung und Leitung in Flüssigkeit), Reizverteilung (Transduktion), Reiztransformation (elektrische Übertragung) und Reizwahrnehmung im Gehirn unterteilt werden (Silbernagl (2012, S. 384-391)).

Schallwellen werden im Medium Luft (Luftleitung) vom äußeren Ohr aufgefangen und durch den Gehörgang zum Trommelfell geleitet. Das äußere Ohr schützt dabei nicht nur das Trommelfell vor mechanischen Reizen, sondern bewirkt durch den Schallübergang vom freien Raum zum engen Gehörgang auch eine Erhöhung des Schalldruckpegels um etwa 20 dB SPL. Im Mittelohr wird der Schall mechanisch durch die Gehörknöchelchenkette bis zur Fenestra ovalis fortgeleitet und auf die Perilymphe der Schnecke im Innenohr (wässriges Medium) übertragen. Der Schallwellenwiderstand ist in Luft geringer als in Flüssigkeit. Daher müssen die Schallwellen entsprechend verstärkt werden (Impedanzwandlung). Dies geschieht zum einen durch den Größenunterschied von Trommelfell und Fenestra ovalis und zum anderen durch einen Hebeleffekt der Gehörknöchelchen. Ist der Schalldruckpegel zu stark (v. a. im niederfrequenten Bereich) so sind der M. tensor tympani und der M. stapedius in der Lage, diesen durch Kontraktion zu senken und somit das Ohr vor zu starken Reizen zu schützen. Im Innenohr wird die Basilarmembran innerhalb der Schnecke in Schwingung versetzt; es entsteht eine Wanderwelle. Die Haarzellen des Corti-Organ werden gereizt, indem sich die Tektorialmembran und die Basilarmembran gegeneinander verschieben und so die Rezeptoren der Haarzellen (Stereozilien) aktivieren. Es kommt zur Ausschüttung von Glutamat an den basalen Teilen der Zelle. Daraufhin entstehen elektrische Impulse an den afferenten Neuronen, die über den N. cochlearis zum Gehirn weitergeleitet werden. Die Schallwelle wird über die Scala tympani und das runde Fenster wieder aus dem Innenohr heraus geleitet (Schmidt und Lang (2007), Benninghoff und Drenckhahn (2008), Schünke et al. (2009, S. 126-141), Zervos-Kopp (2013, S. 388-396), Paulsen und Waschke (2011), Silbernagl (2012)).

1.1.3 Pathophysiologie des Hörens

Schwerhörigkeit bezeichnet eine Herabsetzung des Hörvermögens.

Unterschieden werden dabei v. a. zwei Formen:

1. Bei der Schallleitungsstörung (Mittelohrschwerhörigkeit) lokalisiert sich der Scha-

den im Mittelohr (z. B. Otitis media) oder im Gehörgang (z. B. Cerumen).

2. Bei der Schallempfindungsstörung (Innenohrschwerhörigkeit) liegt die Pathologie entweder im Innenohr (z. B. Morbus Menière), im Hörnerv (z. B. Vestibularis-schwannom) oder im zentralen Nervensystem.

Diagnostiziert werden diese krankhaften Veränderungen durch Hörprüfungen (1.2). Die Hauptrolle spielt hier die Audiometrie.

Aus der Art der Pathologie ergeben sich folgende Therapieoptionen: Bei Schalleitungsstörungen werden Hörgeräte oder operative Verfahren (z. B. Stapesplastik, Tympanoplastik) eingesetzt. Patienten mit einem irreparablen Schaden der Haarzellen und konsekutiver Taubheit werden mit Cochleaimplantaten versorgt (Pschyrembel (2007), Benninghoff und Drenckhahn (2008), Silbernagl (2012)).

1.2 Hördiagnostik

Es stehen umfangreiche diagnostische Mittel zur Detektion von Pathologien des Ohres und des Hörorgans zur Verfügung. Sie unterscheiden sich hinsichtlich der Komplexität der Untersuchung, der dafür benötigten Ausrüstung und den Grundkenntnissen über deren Auswertung. Die *klassischen Hörprüfungen* bestehen im Wesentlichen aus der *Reintonaudiometrie* und den *Sprachverständnistests*.

1.2.1 Klinische Orientierungstests

Im klinischen Alltag werden zwei verschiedene Stimmgabeltests zur Differenzierung zwischen Schalleitungsstörung und Schallempfindungsstörung angewandt: Bei dem Versuch nach Weber wird die angeschlagene Stimmgabel mittig auf den Kopf aufgesetzt. Während Hörgesunde den Ton mittig wahrnehmen, lateralisieren Patienten mit einer Schalleitungsstörung (Pathologie im Mittelohr) auf das kranke Ohr und Patienten mit einer Schallempfindungsstörung (Pathologie im Innenohr) auf das gesunde Ohr. Die Tatsache, dass Luftleitung besser ist als Knochenleitung, nutzt der Versuch nach Rinne aus: Nachdem die angeschlagene, auf dem Mastoid aufgesetzte Stimmgabel über die Knochenleitung nicht mehr registriert wird, wird diese umgedreht vor das äußere Ohr gehalten. Hörgesunde nehmen den Schall wieder wahr (= Rinne positiv). Patienten mit Schalleitungsstörung hören den Schall nicht mehr (= Rinne negativ). Patienten mit Schallempfindungsstörung erhalten im Rinne-Test ein positives Resultat. (Die Luftleitung ist in diesem Fall immer noch besser als die Knochenleitung.)

1 Einleitung

Bei der Hörweitenprüfung wird das Verstehen von leiser Flüster- und Umgangssprache getestet.

Eingesetzt werden diese audiometrischen Tests zur schnellen, diagnostischen Orientierung und zur groben Analyse der Lokalisation der Hörstörung (Lenarz und Boenninghaus (2007), Schmidt und Lang (2007), Reiß (2009, S. 139-297), Silbernagl (2012)).

1.2.2 Reintonaudiometrie

Die (Rein-) Tonaudiometrie stellt das am häufigsten verwendete und wichtigste Verfahren der Audiometrie dar. Sie dient der Bestimmung der Hörschwelle. Die ermittelten Werte werden in Form eines (Reinton-) Audiogramms graphisch dargestellt (Schmidt und Lang (2007), Reiß (2009, S. 139-297), Silbernagl (2012)).

1.2.3 Sprachaudiometrie

Bei sprachaudiometrischen Verfahren wird das Sprachverständnis (die Sprachverständlichkeit) getestet. Die Messungen geben einen relativ realen, alltagsnahen Überblick über die tatsächliche Hörleistung und eine graduelle Einordnung des Hörverstehens in Konversationen. Es werden sämtliche Komponenten des komplexen, peripheren Hörprozesses inklusive der zentralen Verarbeitung des Gehörten getestet: Hörbarkeit, Verständlichkeit und Verstehen.

Dem Patienten werden aufgesprochene Wörter oder ganze Sätze in einer bestimmten Lautstärke unter veränderbaren Bedingungen präsentiert. Geprüft wird der Grad des Verstehens, welcher Rückschlüsse über die Fähigkeit zur Kommunikation bzw. des Zuhörens in alltäglichen Situationen zulässt. Auf diese Weise kann das genaue Ausmaß einer individuellen Hörstörung quantifiziert werden (Lenarz und Boenninghaus (2007), Reiß (2009, S. 139-297), Lehnhardt und Laszig (2009), Steffens (2017a)). Voraussetzungen sind die Kenntnis der Testsprache und eine intellektuelle Mindestleistungsfähigkeit (Mrowinski, 2006).

Die praktische Durchführung der Sprachaudiometrie gestaltet sich wie folgt: Über einen Kopfhörer oder im freien Schallfeld werden dem Patienten (mit oder ohne Hörgerät) Wörter und/oder Sätze präsentiert, welche dieser nachsprechen soll. Die Darbietungslautstärke wird dabei ohne Veränderung der Klangfarbe modifiziert (Mrowinski, 2006).

Mit diesen Grundlagen ergeben sich folgende Indikationen zur Anwendung sprachaudiometrischer Verfahren (Mrowinski (2006), Reiß (2009, S. 139-297), Lehnhardt und Laszig (2009)):

- Indikationsstellung zu Hörhilfen
- Anpassung, Kontrolle, Verlaufsbeurteilung und Reevaluation von Hörhilfen
- Erstellung von Gutachten
- Präoperative Diagnostik vor Ohroperationen
- Überprüfung unklarer Reintonaudiogramme (z. B. bei Simulation oder Aggravation)

Steffens (2017a) betont in diesem Zusammenhang die Relevanz einer konkreten Fragestellung bei der Verfahrenswahl.

Obwohl sprachaudiometrische Verfahren das Hörverstehen verglichen mit der Reintonaudiometrie besser abbilden, existieren auch kritische Stimmen: Der deutlich höhere Zeitaufwand und die größeren Ansprüche in der Durchführung sowie die Tatsache, dass die Sprachaudiometrie bei Patienten mit eingeschränkten kognitiven Fähigkeiten nur limitiert einsetzbar sind, werden als Argumente angeführt (Mrowinski (2006), Probst (2013)). Kritik an sprachaudiometrischen Verfahren zugunsten der Tonaudiogramms übt Probst (2013), da „die Sprachkompetenz grundsätzlich nichts mit dem Hörverlust zu tun“ habe.

Sprachverständnis bei Umgebungslärm

Das Prinzip des Oldenburger Satztests kann aus folgenden Grundlagen der Akustik und Aufmerksamkeitspsychologie abgeleitet werden (Schweizer (2006, S. 17-18)):

Das *cocktail party problem* bzw. der *Cocktail-Party-Effekt* oder *selektives Hören* (nach Edward Colin Cherry (1953)) beschreibt die Fähigkeit, sich während eines Gespräches in lauter Umgebung (z. B. auf einer Cocktailparty mit einer signifikanten Geräuschkulisse aus Stimmen und Hintergrundmusik) auf die Stimme und das Gesprochene des Gegenübers konzentrieren zu können. Die Sprache als ein relevantes, akustisches Signal bzw. als eine erwünschte Schallinformation wird herausgefiltert, während die Umgebungsgeräusche, bestehend aus Störgeräuschen, anderen Stimmen und Schallquellen mehr oder weniger ausgeblendet werden können (Cherry (1953, S. 975-979), Kuyper (1972), Arons (1992)).

Diese Leistung ist ein Korrelat des binauralen Hörens, wodurch ein besseres Hörverständnis und ein besseres Richtungs- und Bewegungshören sowie die Suppression des akustischen Kopfschattens mit dem Individuum als Zentrum im Raum gelingt (Kuyper (1972), Akeroyd (2006), Klockgether et al. (2011), Kießling (2002, S. 23), Schweizer (2006, S. 17-18), Claßen und Schnitzler (2012, S. 263)).

Die binaurale Analyse beruht dabei auf einer Analyse und Verarbeitung der Laufzeit- und Schallpegeldifferenzen zwischen beiden Ohren (Blauert (1972)). Durch das binaurale System können Laufzeitdifferenzen von 30 μ s zwischen den beiden Ohren wahrgenommen und verarbeitet werden. Das entspricht einer 3° neben der Mittellinie liegenden Schallquelle (Behrends (2010, S. 692)).

Lewald und Getzmann (2015) identifizierten das elektrophysiologische Korrelat des selektiven Hörens: Das Cocktail-Party-Hören basiert (im Vergleich zu dem Ortungshören einer einzigen Schallquelle) auf einer komplexen Abfolge von sukzessiven, elektrischen Erregungen innerhalb eines spezifischen kortikalen Netzwerkes, welches am räumlichen Hören in komplexen Situationen beteiligt ist.

Zusammenfassend beruft sich Akeroyd (2006) auf Green (1976): „Bestimmte Fakten und Ergebnisse können erklärt werden, allerdings ist es nicht möglich, einzelne Mechanismen oder Theorien des Phänomens des binauralen Hörens gänzlich zu durchleuchten.“

Sprachaudiometrie im Störgeräusch

Sprachaudiometrische Verfahren, bei denen das akustisch dargebotene Material aus Sprachsignal und Störgeräusch besteht, arbeiten möglichst realitätsnah, was eine genauere Beurteilung beispielsweise in Gutachten ermöglicht und skizzieren ein noch exakteres Bild des tatsächlichen Hörverstehens in alltagsrelevanten Kommunikationssituationen (Schorn (2004), Müller-Deile (2009), Kollmeier et al. (2011), Batsoulis und Lesinski-Schiedat (2017), Rahne (2017), Steffens (2017a)). Die Messung wird „reliabler, valider und objektiver“ (Steffens, 2015). In diesem Sinne vertreten einige Arbeitsgruppen die Meinung, dass ältere Verfahren schrittweise abgelöst werden können (Studie von Kollmeier et al. (2011) am Freiburger Einsilber Test, Göttinger Satztest, Oldenburger Satztest und dem Einsilber-Reimtest nach Sotschek/von Wallenberg/Kollmeier).

Grundsätzlich existieren verschieden definierte Störgeräusche. Das Rauschen besteht aus vielen, zeitgleich präsentierten Tönen wechselnder Intensität. Man unterscheidet unter anderem Breitbandrauschen, Schmalbandrauschen oder Sprachüberlagerungsrauschen (Mrowinski (2006), Lehnhardt und Laszig (2009)). Beim Oldenburger Satztest wird das Störgeräusch aus Wortgewirr gebildet, welches bei der 30-fachen Überlagerung des gesamten Sprachmaterials entsteht (Schorn, 2004). Einzelne Wörter werden zufällig überlagert, wodurch das gleiche Langzeitspektrum entsteht (HörTech, 2011).

Tabelle 1.1: Deutsche Sprachverständnistests in Satzform

Satztest	Charakteristika
Innsbrucker Satztest (Wallenberg, 1987)	– geschulte Sprecherin (und damit gut geeignet für schlecht verstehende Cochlea-Implantat-Patienten)
Göttinger Satztest (Wesselkamp u. Kollmeier, 1994)	– 10 kurze, sinnhafte Sätze – steile Diskriminationsfunktion wg. hoher Listenäquivalenz – ungeschulter Sprecher
Hochmair-Schulz-Moser-Test (HSM-Test) (Hochmair et al., Schmidt et al., 1997)	– 20 kurze, sinnhafte Sätze – CCITT-Rauschen als Störgeräusch – geschulter Sprecher
Oldenburger Satztest (Wagener et al., 1999)	– 10 syntaktisch gleiche, fraglich sinnhafte Sätze – ungeschulter Sprecher

Übersicht über sprachaudiometrische Verfahren in Deutschland

Geordnet nach der Art des Sprachmaterials liefert die folgende Aufzählung in aufsteigender Komplexität eine Übersicht über die wichtigsten in deutscher Sprache konzipierten sprachaudiometrischen Untersuchungen: Logatomtests, Einsilber-Tests (Freiburger Einsilbertest), Zwei- und Mehrsilber- sowie Zahlen-Tests (Freiburger Zahlen-Test), Satztests (Göttinger Satztest, Hochmair-Schulz-Moser-Test, Innsbrucker Satztest, Oldenburger Satztest) und Fließtext-Tests (Kollmeier (1992), Mrowinski (2006, S. 51-63), Lehnhardt und Laszig (2009, S. 147-164), Steffens (2017a)).

Tabelle 1.1 bietet eine Übersicht über die wichtigsten, deutschen Sprachverständnistests in Satzform (Kollmeier (1992), Mrowinski (2006, S. 51-63), Lehnhardt und Laszig (2009, S. 147-164)). Diese Tests werden meist in Ruhe präsentiert, können aber auch im Störschall durchgeführt werden. Der Göttinger Satztest und der Oldenburger Satztest unterscheiden sich im Wesentlichen durch die Art des Satzbaus und die Vorhersagbarkeit: Während der Oldenburger Satztest aus syntaktisch korrekten, fraglich sinnhaften, semantisch nicht vorhersagbaren Sätzen besteht, sind die Alltagssätze des Göttinger Satztests und des HSM-Satztests in sich schlüssig, sinnhaft und vorhersagbar (Lehnhardt und Laszig (2009, S. 149f.)). Der Innsbrucker Satztest ist aufgrund der langsamen Sprechweise und der extrem deutlichen Artikulation für die Anfangsphase der Anpassung von Cochlea-Implantaten für schlecht hörende Cochlea-Implantat-Patienten konzipiert. Zu den Charakteristika des Oldenburger Satztests zählen eine Sprechgeschwindigkeit von 233 Silben pro Minute. Mit dieser Anzahl liegt der Oldenburger Satztest zwischen dem HSM-Test und dem Göttinger Satztest (Steffens, 2017a). Die Sätze

des Göttinger Satztests sind wegen ihrer schnellen Sprechgeschwindigkeit und der im Vergleich undeutlichen Artikulation eher im Bereich der Umgangssprache anzusiedeln und daher für Cochlea-Implantat-Patienten schwer zu verstehen. Der Schwierigkeitsgrad des HSM-Satztests liegt für Cochlea-Implantat-Patienten zwischen dem Innsbrucker und dem Göttinger Satztest. Die Qualität des Sprachsignals ist beim HSM-Satztest durch einen geschulten Sprecher am Höchsten anzusiedeln. Der Göttinger Satztest und der HSM-Satztest zeigen bei wiederholter Durchführung aufgrund der semantisch sinnhaften Sätze und der relativ geringen Anzahl an Testlisten, die die Wiederholbarkeit im Rahmen ausführlicher Testungen einschränken, einen Erinnerungseffekt (gesprochen wird von einem *Lerneffekt*) (Mrowinski (2006, S. 60)). Verglichen mit den oben genannten Satztests ist der Oldenburger Satztest für ausländische und schwerhörige Patienten leichter zu verstehen (Mrowinski (2006, S. 60)).

Freiburger Test

Der Freiburger Sprachtest (Freiburger Sprachverständnistest; FST), entwickelt von Karl-Heinz Hahlbrock (1953), stellt das Standardverfahren der Sprachaudiometrie in Deutschland dar. Er ist das in Deutschland am häufigsten eingesetzte sprachaudiometrische Verfahren (Kießling et al. (2008), Löhler et al. (2013), Steffens (2015), Hoth (2016), Baljić et al. (2016)). Sein hoher Bekanntheitsgrad, eine breite Verfügbarkeit und eine gute Interpretierbarkeit sind Gründe dafür, dass der Freiburger Sprachtest trotz seiner geringen Messgenauigkeit den diagnostischen Standard und die rechtliche Grundlage bei Hörverlust darstellt (Feldmann (2004), Steffens (2015)).

Die Industrienorm DIN 45621 legt den Freiburger Sprachtest zur Indikationsstellung, Kontrolle und Begutachtung von Hörhilfen fest (Müller-Deile (2009), Löhler et al. (2013)).

Der Freiburger Test ist ein offener Test und setzt sich aus 10 Gruppen zu je 10 zweistelligen Zahlen und 20 Gruppen zu je 20 einsilbigen Wörtern zusammen. Die folgenden Beispielreihen zeigen den Aufbau und das Testprinzip: Zahlenreihe, Mehrsilber (10): 98 22 54 19 86 71 35 47 80 63 sowie Wörterreihe, Einsilber (20): Ring Spott Farm Hang Geist Zahl Hund Bach Floh Lärm Durst Teig Prinz Aas Schreck Nuß Wolf Braut Kern Stich. In der klinischen Anwendung wird mit dem Zahlentest die Sprachverständlichkeitsschwelle (Schalldruckpegel in dB SNR, bei dem 50 % der Zahlen verstanden werden) angegeben, während mit dem Einsilbertest oft die Sprachverständlichkeit (in %) bei einem festen Schallpegel (meist 65 dB SPL) angegeben wird.

Zusätzlich zu den oben genannten Vorteilen des Freiburger Tests werden der Standard nach DIN 45621 und somit eine genormte Reproduzierbarkeit und die jahrelange

Erfahrung im Einsatz mit einer konsekutiv guten Studienlage sowie die Anwendbarkeit für Erwachsene und Schulkinder genannt (Reiß (2009, S. 168-170), Löhler et al. (2013)).

Kießling (2010) kritisiert den Freiburger Test, da keine realistische Gesprächssituation nachgestellt wird. Weitere Nachteile des Freiburger Tests sind die unnatürliche Artikulation, die unausgewogene Äquivalenz der Testlisten, die phonetische Unausgewogenheit des Testmaterials, die fehlende Variabilität der Wortfolge und die fehlende Möglichkeit zur Untersuchung der Phonemverwechslungen (Kießling et al. (2008), Lehnhardt und Laszig (2009, S. 149), Reiß (2009, S. 168-170), Hoth (2016), Baljić et al. (2016)). Löhler et al. (2013) führen als weitere Kritikpunkte die regional unterschiedliche Bekanntheit der Wörter, das Bedenken der Probanden bestimmte Wörter wiederzugeben (z. B. Sau) und die fehlende Möglichkeit der Anwendung im Störgeräusch an. Dadurch kam es zu der Empfehlung, die Verwendung bestimmter Listen (Liste 3, 5, 14, 15) zu vermeiden. Aufgrund eines Wortrepertoires aus dem 19. Jahrhundert (45 % der Einsilber werden in der Alltagssprache nicht mehr gebraucht) und einer ungleichen Verteilung dieser Wörter zwischen den Testlisten wird das Testergebnis abhängig vom individuellen Wortschatz der Testperson und sollte reevaluiert werden (Steffens, 2016).

Schmidt und Baljić (2016) betonen, dass bei der praktischen Anwendung des Freiburger Einsilber Tests aufgrund des prozeduralen Lerneffekts vor Beginn der eigentlichen Messung zwei Trainingslisten zu Übungszwecken getestet werden sollten. Dadurch erreiche man mit hoher Wahrscheinlichkeit eine Reliabilitätsverbesserung der Messung. In einer Studie mit dem Freiburger Einsilbertest an 40 normalhörenden Probanden zeigte er bei den ersten beiden Testläufen einen relevanten Trainingseffekt. (Auch Ahlich (1985) berichtete von einem Lerneffekt der Testwörter bei mehrfacher Durchführung des Freiburger Sprachverständnistest an einem Probanden.)

Bei Satztests, welche mit alltäglich gebrauchten Sätzen arbeiten, tritt ein besonders großer Erinnerungseffekt auf. Dieser kann unter Umständen über viele Wochen und Monate bestehen bleiben (Yund und Woods, 2010). Gemäß Müller-Deile (2009) sollten traditionelle Satztests frühestens nach einem Jahr erneut angewendet werden.

Um den Einfluss des Erinnerungseffekts zu minimieren, ist die Wahl von Matrixtests indiziert (Brand und Wagener, 2017).

Zusammenfassend müssen aufgrund seiner zahlreichen Nachteile die Grenzen der Anwendung des Freiburger Tests bekannt sein. Trotzdem bleibt seine alltägliche Relevanz und Wichtigkeit bestehen (Hoth (2016), Baljić et al. (2016)).

1.3 Oldenburger Satztest

Der Oldenburger Satztest mit seinen positiven und negativen Seiten soll in dieser Dissertation mit besonderem Augenmerk auf den kurz- und langfristigen Habituations- und Lerneffekten beleuchtet werden.

1.3.1 Entwicklung und Eigenschaften

Bei dem Oldenburger Satztest (kurz: OLSA) handelt es sich um ein von [Wagener et al.](#) entwickeltes, sprachaudiometrisches Verfahren aus dem Jahr 1999, bei dem das Sprachverständnis getestet wird. Die Testidee basiert auf dem Konzept des schwedischen Hörtests nach [Hagerman](#) aus dem Jahr 1982. [Schorn \(2006\)](#) beschreibt den Oldenburger Satztest als „Sprachverständnistest der Zukunft“.

Der Oldenburger Satztest wurde mittels des sogenannten Matrixprinzips konzipiert. Jeder Satz setzt sich aus fünf Wörtern zusammen und ist nach demselben Muster aufgebaut: Ein Name bildet das Subjekt, gefolgt von einem Verb als Prädikat, abgeschlossen von einem dreiteiligen Objekt aus Zahlwort, Adjektiv und Substantiv. Ein theoretisch möglicher Satz könnte z. B. „Peter hat drei grüne Autos“ lauten. Für jede der fünf Satzpositionen stehen zehn verschiedene Wörter zur Verfügung, die prinzipiell beliebig miteinander kombiniert werden können. Daraus ergibt sich eine theoretische Anzahl von insgesamt 100.000 Sätzen. Tatsächlich existieren nur 100 verschiedene Sätze. Da die Sätze zwar syntaktisch gleich aufgebaut, aber semantisch nicht vorhersehbar sind, ist gewährleistet, dass die Probanden sich keine ganzen Testlisten merken können, geschweige denn deren Wiederholung überhaupt erkennen. Daher können die Listen auch wiederholt gemessen werden. Da sich die Sätze durch eine sehr geringe Vorhersagbarkeit auszeichnen, werden in einer bestimmten Zeit mehrere voneinander unabhängige Variablen untersucht, weshalb der Oldenburger Satztest besonders für adaptive Messungen geeignet ist. Dargeboten werden die Sätze parallel zu einem Störuschen definierter Lautstärke und Tonhöhe ([Wagener et al., 1999a,b,c](#)). Das Störuschen wurde durch ein willkürliches Überlagern der aufgesprochenen Wörter produziert. Da das Rauschen das gleiche Langzeitspektrum wie das Sprachsignal besitzt, folgt eine steile Diskriminationsfunktion und eine hohe Messgenauigkeit.

Der Oldenburger Satztest wird ursprünglich von einer männlichen Stimme präsentiert; zusätzlich ist der Test inzwischen auch von einer weiblichen Sprecherin gesprochen erhältlich ([Steffens, 2017a](#)). Eine im Vergleich relativ undeutliche Artikulation entsteht durch den ungeschulten Sprecher Jochem Sotschek ([Steffens, 2017a](#)). Das Sprach-

signal wurde in mittlerer Sprechgeschwindigkeit aufgenommen. Unter Berücksichtigung des Koartikulationseffekts wurden 100 Sätze so aufgesprochen, dass zehn verschiedene Versionen (jeweils in Abhängigkeit von dem darauffolgenden Wort) entstanden sind. Bei der Generierung des Satzes bzw. bei der Satzkonstruktion konnte durch dieses Verfahren stets auf das koartikulationstechnisch passende Wort zurückgegriffen werden. Die Phonemverteilung der Listen entspricht der mittleren Phonemverteilung der deutschen Sprache.

Mit dem Ziel positive Eigenschaften bestehender Tests zusammenzuführen, wurden folgende Ansprüche formuliert und umgesetzt (Wagener et al., 1999c):

1. **Durchführbarkeit im Störgeräusch:** Da schwerhörige Patienten die größten Kommunikations- und Sprachverständnisprobleme in geräuschvoller Umgebung zu bewältigen haben, simuliert der Test die reale Situation möglichst detailgetreu, indem zusätzlich zum Sprachsignal ein Störgeräusch präsentiert wird.
2. **Steile Diskriminationskurve:** Eine steile Steigung der Diskriminationsfunktion ist als Maß für die Effizienz eines Tests mit einer hohen Messleistung (hohe Genauigkeit in bestimmter Zeit) gleichzusetzen. Eine hohe Äquivalenz des Testmaterials ist als Voraussetzung gegeben.
3. **Hohe Anzahl und Wiederholbarkeit der Testlisten:** Für den Gebrauch des Tests zu Forschungszwecken oder für die Anpassung von Hörgeräten sind oft wiederholte Messungen nötig. Dafür spielt die Verfügbarkeit ausreichend vieler perzeptiv äquivalenter Listen eine zentrale Rolle. Idealerweise sind unendlich viele Wiederholungen durchführbar.
4. **Mittlere Sprechgeschwindigkeit des Sprachsignals:** Da schnell Gesprochenes v. a. Cochlea-Implantat-Patienten mit großen Herausforderungen bzw. Schwierigkeiten konfrontiert, soll eine mittlere Sprechgeschwindigkeit die Anwendung des Tests bei diesem Kollektiv garantieren.
5. **Verwendung gebräuchlicher Umgangssprache:** Um eine natürliche Gesprächssituation zu imitieren, wird gewöhnliche Umgangssprache verwendet. Überartikulation würde die Resultate verzerren.

Aufgrund des Testdesigns, konzipiert nach dem Satztest nach Hagerman (1982) zeichnet sich der Oldenburger Satztest durch diese Eigenschaften aus:

1. Die Verwendung desselben Sprachmaterials garantiert die Homogenität der Testlisten.

1 Einleitung

2. Die Tatsache, dass ein Auswendiglernen der Testlisten praktisch nicht möglich ist, bestätigt die Wiederholbarkeit des Tests.
3. Die Generierung des Rauschens aus dem Sprachmaterial sorgt für eine spektral gleichmäßige Maskierung des Testsignals durch das Störgeräusch.

Bei der Generierung und Bearbeitung der Sätze des Oldenburger Satztests wurden (mit dem Ziel einen möglichst natürlichen Klang und einen fließend gesprochenen Sprachfluss zu produzieren) die Koartikulationseffekte (Übergänge zwischen den einzelnen Lauteinheiten) mit einbezogen (Wagener et al., 1999c).

Die Steigung der Gesamt-Diskriminationsfunktion (Maß für die Effizienz) des Oldenburger Satztests beträgt 17,2 % / dB (Wagener et al., 1999a). Im Vergleich dazu beträgt die Steigung der Gesamt-Diskriminationsfunktion beim Göttinger Satztest 19,2 % / dB und im Satztest nach Hagerman 25 % / dB. Der mittlere L_{50} -Wert liegt bei der extrem gut trainierten Probandengruppe bei -8,4 dB SNR (Wagener et al., 1999a).

1.3.2 Kurz- und langfristige Lerneffekte

Der Lerneffekt, der bei der mehrmaligen Durchführung des Oldenburger Satztests beobachtet werden kann, war schon in der Vergangenheit und ist noch gegenwärtig Gegenstand vieler Studien, Experimente, Publikationen und Dissertationen. Die Autoren Wagener et al. (1999b), Brand et al. (2004), Wagener und Brand (2005), Wagener et al. (2006), Müller-Deile (2009), Grimm (2014) kamen dabei seit Entwicklung des Oldenburger Satztests 1999 (19 Jahre) zu größtenteils deckungsgleichen Resultaten.

Im Folgenden wird zunächst der Lerneffekt definiert und genauer unterteilt, anschließend werden die einzelnen Artikel und Studien in chronologischer Reihenfolge vorgestellt und zusammengefasst.

Definition des Lerneffekts

Unter dem Lerneffekt (auch: Habituationseffekt, Trainingseffekt) eines Tests versteht man eine „Verbesserung der Verständlichkeit des Sprachmaterials des Tests“ (nach Schlüter et al. (2012)), welche bei wiederholter Durchführung des Tests gemessen werden kann.

In deutschsprachigen Artikeln wird von einem *Lerneffekt* und in Publikationen in englischer Sprache von einem *training effect* gesprochen.

In Anlehnung an Schlüter et al. (2012) und Yund und Woods (2010) kann der Lerneffekt in verschiedene zeitliche und inhaltliche Komponenten unterteilt werden:

1. Kurzzeit-Lerneffekt: kurzfristiger Lerneffekt, Lerneffekt innerhalb einer Sitzung; intrasessionaler Lerneffekt
2. Langzeit-Lerneffekt: langfristiger Lerneffekt, Lerneffekt über mehrere Sitzungen hinweg; intersessionaler Lerneffekt
3. *content learning* („Wortrepertoire-Erlernen“): Lerneffekt, welcher durch das Erlernen der vorkommenden Wörter bzw. Sätze des Tests entsteht
4. *procedural learning* („Testprozedur-Erlernen“): Lerneffekt, der durch das Kennenlernen der Hörsituation, der Eigenschaften und Eigenheiten des Störgeräusches und der Stimme des Sprechers, durch Artikulation, durch syntaktische und semantische Gesetzmäßigkeiten im Satzmaterial und durch jegliche anderen Aspekte der Testmethoden, deren Kenntnis die Testung vereinfachen könnte, entsteht

1.3.3 Bisherige, relevante Studien zu Lerneffekten beim Oldenburger Satztest

Entwicklung und Evaluation eines Satztests für die deutsche Sprache, Teil I - III: Design, Optimierung und Evaluation des Oldenburger Satztests (Wagner et al., 1999a,b,c): In der dritten Originalpublikation zum Oldenburger Satztest wird folgendes Versuchsdesign gewählt: Die zwölf Testlisten zu je zehn Sätzen wurden zu sechs Doppellisten zu je 20 Sätzen zusammengefasst und an 20 hörgesunden Probanden getestet. Die Resultate zeigen einen signifikanten Gewöhnungs- und Trainingseffekt von 1 bis maximal 2 dB, der v. a. während des Abspielens der ersten beiden Doppellisten (40 Sätze) stattfindet. Diese Effekte können durch das Präsentieren von ein bis zwei Übungslisten auf weniger als 1 dB begrenzt werden. Der L_{50} -Wert (Signal-Rausch-Verhältnis, bei dem 50% verstanden wird) betrug in der Evaluationsarbeit -8,4 dB SNR, in der Ursprungsarbeit (Wagner et al., Teil I) nur -7,1 dB SNR. Das Ergebnis lässt sich (zumindest teilweise) durch das hörtechnisch stärker trainierte Probandenkollektiv, durch das quasi-geschlossene Format und eventuell auch durch unterschiedliche Störgeräuschpegel erklären.

Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen (Brand et al., 2004): Der Artikel beschäftigt sich mit der Variation des Testdesigns und dessen Folgen. In der *geschlossenen* Version eines Tests kennt die Testperson alle Antwortalternativen. Im Gegensatz dazu wird ein Test als *offen* bezeichnet, wenn nicht alle Antwortmöglichkeiten bekannt sind. In *quasi geschlossenen* Tests (wie z. B. dem Freiburger Zahlentest und dem Oldenburger Satztest) wird den Probanden ohne deren Wissen ein begrenztes Antwortmaterial vorgespielt. Der anfangs offene

wird mit der Zeit de-facto zum geschlossenen Test. Die Vorteile geschlossener Tests bestehen in der Möglichkeit der Durchführung am Computer (ohne Versuchsleiter) und in der Möglichkeit der Auswertung von Phonemverwechslungsmatrizen. In dieser Studie wurden 20 normalhörende Personen (16- bis 35-jährig) dem Oldenburger Satztest in fünf Messungen (viermal geschlossen, dann einmal offen) zu je 20 Sätzen pro Messung unterzogen und der L_{50} -Wert ermittelt. Im Laufe der vier geschlossenen Messungen sinkt der L_{50} -Wert um 1 dB SNR. Der L_{50} -Wert betrug nach der 4. Messung bei [Wagener et al.](#) (offen) -7,1 dB SNR. Bei [Brand et al.](#) (geschlossen) lag dieser Wert bei nur -6,3 dB SNR. Begründet liegt diese Diskrepanz vermutlich in dem ausgiebigeren Training bei [Wagener et al.](#). Insgesamt zeigte sich also ein ähnlich stark ausgeprägter Trainingseffekt.

Der Zeitverbrauch wird als Maß für die Schwierigkeit des Tests hinzugezogen. Bei geschlossenen Durchführungen lag der Zeitverbrauch für eine Liste (20 Sätze) bei 4,9 Minuten, während Probanden bei der Testung im offenen Format nur 3,1 Minuten benötigten. Zusammenfassend macht es keinen Unterschied, ob der Oldenburger Satztest als offene oder geschlossene Version präsentiert wird. Der gewünschte positive Einfluss der geschlossenen Version auf den Lerneffekt zeigte sich nicht, was bedeutet, dass auch im geschlossenen Format ein Training nötig ist. Zum Göttinger Satztest wird geraten, wenn ein Training aufgrund von wenig benötigten Messungen nicht lohnenswert erscheint. Im geschlossenen (computergestützten) Format wurde ein deutlich größerer Zeitverbrauch verbucht als in der offenen Version (mit Testleiter). Da diese Resultate die Reaktionszeit und das Handling des Tests junger Personen widerspiegelt, bleibt die Frage, wie ältere Personen mit den Testmodalitäten umgehen können.

Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters ([Wagener und Brand, 2005](#)): Bei Test-Retest-Messungen zu je 30 Sätzen an zwei verschiedenen Tagen („performed on two different days“) wurden zehn normalhörende Probanden (fünf männliche und fünf weibliche Probanden im Alter zwischen 22 und 40 Jahren, bei einem Durchschnittsalter von 26,5 Jahren) monaural getestet. Ebenso wurden zehn hörgeschädigte Probanden (sieben männliche und drei weibliche Probanden im Alter zwischen 59 und 79 Jahren, bei einem Durchschnittsalter von 70 Jahren) untersucht. Test-Retest-Differenzen zwischen den SRT-Werten waren bei 84 % der Messungen bei den normalhörenden Probanden und bei 61 % der hörgeschädigten Patienten erkennbar. Bei den normalhörenden Probanden zeigte sich dabei ein Trainingseffekt von 0,67 dB. Dieser Effekt ist insofern verwunderlich, als die beiden Tests an verschiede-

nen Tagen durchgeführt wurden und als zwei Listen (60 Sätze) vor der ersten Messung zu Übungszwecken durchgeführt wurden. Bei hörgeschädigten Probanden zeigte sich hingegen ein vernachlässigbarer Trainingseffekt von 0,2 dB. Dieser Wert ist geringer als die Messgenauigkeit. Dennoch gab es auch hörgeschädigte Probanden mit großen Differenzen zwischen dem eigentlichen Test und der Wiederholungsmessung.

Dieser Trainingseffekt ist vergleichbar mit dem, der bei dem schwedischen Satztest nach [Hagerman](#) auftritt: Dort zeigte sich ein Lerneffekt von 0,1 dB pro Liste (10 Sätze) bei normalhörenden Probanden und ein Lerneffekt von 0,3 dB pro Liste (10 Sätze) bei hörgeschädigten Probanden. (Hagerman and Kinnefors, 1995 und Hagerman, 2002). [Wagner und Brand](#) schreiben den Lerneffekt der Struktur des Tests zu: Durch das quasi-geschlossene/semi-offene Konstrukt des Tests ist den Versuchspersonen das Wortrepertoire mehr oder weniger bekannt. Aufgrund einer erhöhten Ratewahrscheinlichkeit im dem Fall, wenn nur ein Teil eines Wortes verstanden wurde profitieren Probanden in einem quasi-geschlossenen Testdesign, im Vergleich zu einem offenen Testkonzept, von der reduzierten Wortmenge ([Bronkhorst et al., 2002](#)). [Puglisi et al. \(2014\)](#) begründen die besseren Ergebnisse bei der geschlossenen Durchführung ihres italienischen Satztests damit, dass der Proband durch die visuelle Präsentation der Antwortmöglichkeiten zuvor akustisch Aufgenommenes besser benennen kann.

Daher könnten normalhörende Probanden mehr Übungssätze und somit ein längeres Training benötigen. Es muss noch geprüft werden, ob durch die Präsentation von 80 Übungssätzen vor der eigentlichen Datenerhebung (anstelle von hier 60) der verbleibende Lerneffekt beseitigt werden kann. Weiterhin muss in einer Langzeitstudie untersucht werden, ob dieses erweiterte Übungsprogramm nur einmal durchgeführt werden muss. Das wäre der Fall, wenn die Probanden dadurch „für ihr ganzes Leben vor-trainiert“ wären und deren Wissensstand später nur durch ein paar Sätze wieder aufgefrischt werden könnte. Auch wenn dieser Trainingseffekt eindeutig den Gebrauch des vorliegenden Sprachtests in der Klinik und zu Forschungszwecken einschränkt, ist kein alternativer Test verfügbar. Satztests mit Sätzen mit hoher Vorhersagbarkeit (wie beispielsweise der Göttinger Satztest oder der Hearing In Noise Test) können von den Probanden für gewöhnlich nur einmal durchgeführt werden, da die bedeutungsvollen Sätze leicht erinnerlich sind und die Wörter im Kontext erschlossen werden können. Das hätte Resultate mit falsch niedrigen SRT-Werten zur Folge. Eine wiederholte Messung mit derselben Liste ist solange nicht möglich, bis ausreichend Zeit vergangen ist (z. B. ein halbes Jahr oder sogar länger). Probanden, welche mit den bedeutungsvollen Sätzen vertraut sind, werden immer niedrigere SRT-Werte zeigen als naive Versuchspersonen mit einem vergleichbaren Gehör. Diese Unterschiede sind nicht in demselben

Ausmaß ersichtlich, wenn der Satzbau festgesetzt ist, aber semantisch nicht vorhersagbare Sätze verwendet werden. Es sollte in der Zukunft untersucht werden, ob der individuelle Trainingsstatus durch eine Analyse der durchgeführten Messungen abgeschätzt werden kann, um anschließend eine Korrekturvorschrift, durch die das Lernen ausgeglichen werden kann, zu berechnen. Wenn also Sprachverständlichkeitstests mit nicht vorhersagbaren Sätzen, aber begrenztem Wortmaterial verwendet werden, ist es notwendig, ein geeignetes Training vor der Datenerhebung durchzuführen. Es sollte in der Zukunft untersucht werden, ob es möglich ist, das Lernen durch eine adäquate Analyse der Messungen abzuschätzen und auszugleichen.

The role of silent intervals for intelligibility in fluctuating noise in hearing - impaired listeners (Wagener et al., 2006): Die meisten Test-Retest-Messungen zeigten einen Trainigseffekt von 1 dB.

Sprachverständlichkeitsuntersuchungen bei Cochlea-Implantat-Patienten (Müller-Deile, 2009): In dieser Studie über den Einfluss und die Auswirkungen der Signalverarbeitung auf die Sprachverständlichkeit wurden in einer Test-Retest-Situation Differenzen der L_{50} -Werte ermittelt. Die Untersuchungsreihenfolge der Hörsituationen wurde in der zehnköpfigen Gruppe CIC 1 gemäß HS1, HS2, HS1, HS2, HS3, HS4, HS3, HS4, HS5, HS6, HS5, HS6 und in der 36 Probanden starken Gruppe CIC 2 mit gemäß randomisierter Folge von HS1 und HS2 bzw. HS3 bis HS6 gewählt. Das Training der Patienten der Gruppe CIC 1 war zu Beginn der Messungen noch nicht beendet. Es zeigte sich ein leichter, nicht erwarteter, aber signifikanter Lerneffekt. Die starke Verbesserung der Patienten der Gruppe CIC 1 unterstreicht die Wichtigkeit eines ausreichenden Trainings, bevor die eigentliche Daten erhoben werden können. Test-Retest-Differenzen wie in Gruppe CIC 2 konnten in Gruppe CIC 1 erst am Ende der Messungen erzielt werden. Müller-Deile (2009) zieht Schlussfolgerungen und ein Fazit für die Praxis: Der Oldenburger Satztest ist zur Evaluation des Kommunikationsvermögens von Cochlea-Implantat-Patienten der am besten geeignete Test. Bei Cochlea-Implantat-Patienten ergeben sich Diskriminationsfunktionen mit flacheren Steigungen als bei Normalhörenden. Dennoch sind diese im Vergleich zu ähnlichen Tests steil genug, sodass der Oldenburger Satztest auch bei Cochlea-Implantat-Patienten geeignet ist. Es wird die Empfehlung ausgesprochen, das Störgeräusch bei 65 dB SPL konstant zu halten. Als Nachteil des Oldenburger Satztests wird der Lerneffekt gesehen, weshalb vorgeschlagen wird, vor der Testung im Störgeräusch mindestens 140 Sätze in Ruhe zu präsentieren. Weiterhin sollte vor jeder neuen Testsitzung eine Liste in Ruhe gemessen werden, um den Pro-

banden bzw. den Patienten wieder mit der Situation und dem Prinzip des Tests vertraut zu machen.

Lerneffekte des Oldenburger Satztests: Die Einflüsse von Schwerhörigkeit und Zeitkompression auf Kurz- und Langzeitlerneffekte (Holube und Schlüter, 2012):

Der Lerneffekt wird in dieser Arbeit in die Komponenten Kurzzeit-Lerneffekt und Langzeit-Lerneffekt unterteilt. Holube und Schlüter (2012) zeigten in einem Experiment an 24 normalhörenden Probanden einen Kurzzeit-Lerneffekt von -1,8 dB (während der Sitzung 1) und einen Langzeit-Lerneffekt von -1,15 dB. Die Testkonditionen waren dabei wie folgt gewählt: fünf Termine im Abstand von zwei Tagen (\pm ein Tag), Testrahmen insgesamt: neun bis 21 Tage, Testung von je sechs 30er-Listen.

In einem zweiten Experiment an neun schwerhörigen Probanden betrug der Kurzzeit-Lerneffekt -1,6 dB (während der Sitzung 1) und der Langzeit-Lerneffekt -0,90 dB. Die Testkonditionen waren dabei dieselben wie im ersten Experiment. Die Verbesserung des SRT-Werts geschieht bei schwerhörigen Patienten langsamer als bei Normalhörenden.

In einem dritten Experiment an zwölf Normalhörenden, diesmal mit verändertem Sprechtempo betrug der Kurzzeit-Lerneffekt -9,2 dB (während der Sitzung 1) und der Langzeit-Lerneffekt -2,95 dB. Die Testkonditionen waren dabei wie oben gewählt, mit dem Unterschied, dass hier sieben 30er-Listen in einer schnellen Version getestet wurden. Nur der intersessionelle Lerneffekt zeigte sich bei Normalhörenden signifikant größer als bei Hörgeschädigten.

Auswirkungen eines Hörtrainings auf das Sprachverstehen im Störgeräusch: (Nagel, 2011):

Nagel untersuchte den Einfluss des Hörtrainings "audiclick" der Firma Meditech auf die SRT-Resultate bei der sechsmaligen Durchführung des Oldenburger Satztests. Getestet wurden je eine 30er Liste im Abstand von je einer Woche an 30 hörgeschädigten Probanden, von denen zwölf ein digitales Hörgerät besaßen (insgesamt sechs Sitzungen innerhalb von sechs Wochen). Im Rahmen dieser Studie zeigte die zehnköpfige Hörtraining-Gruppe einen Trainingserfolg von -1,5 dB, die zehnköpfige Placebo-Gruppe -1,0 dB und die zehnköpfige Kontroll-Gruppe -0,5 dB. Weiterhin war der Trainingserfolg bei Männern größer als bei Frauen, bei jüngeren größer als bei älteren und bei längerer Trainingsdauer zunehmend.

Trainingseffekte bei normaler und schneller Sprache (Schlüter et al., 2012): Bei der Durchführung des Oldenburger Satztests treten inter- und intrasessionale Trainings-

effekte auf. Die Studie an 36 jungen, Normalhörenden und neun älteren Hörgeschädigten, welche an fünf Terminen (im Abstand von zwei Tagen (\pm ein Tag) jeweils sechs Listen des Oldenburger Satztests durchführten zeigte bei den Normalhörenden einen Trainingseffekt von 1,4 dB (bzw. 3,0 dB bei schnellerer Präsentation des Tests) und bei den Hörgeschädigten von 0,9 dB. Der intrasessionale Trainingseffekt konnte durch Präsentation von zwei Testlisten so verringert werden, dass nur noch Abweichungen von maximal 0,5 dB auftraten.

Kernaussagen der Studienlage

Beim Oldenburger Satztest existiert ein Lerneffekt, der die Leistung der Patienten um 2 bis 4 dB verbessern kann (Wagener et al. (1999b), Brand et al. (2004), Wagener und Brand (2005), Wagener et al. (2006), Müller-Deile (2009), Nagel (2011), Holube und Schlüter (2012), Schlüter et al. (2012), Grimm (2014)). Dabei beträgt der Trainingseffekt des Oldenburger Satztests in der Version *adaptive Veränderung des Sprachsignalpegels bei konstantem Störgeräuschpegel* bei naiven Versuchspersonen innerhalb der ersten 30 getesteten Sätze bis zu 2 dB. In der Variante *adaptive Veränderung des Störgeräuschpegels bei fixem Sprachsignalpegel* ist ein geringer ausgeprägter Trainingseffekt zu beobachten. Dieser wird im Laufe der Testungen kontinuierlich kleiner (Wagener et al., 1999a,b,c). Da aber im Rahmen der Cochlea-Implantat-Anpassung die Messungen sehr häufig wiederholt werden (müssen), spielt dieser Lerneffekt eine untergeordnete Rolle (Lehnhardt und Laszig (2009, S. 256)).

1.3.4 Kritische Bewertung

Trotz vieler Vorteile birgt der Oldenburger Satztest sowohl die allgemein bei Satztests auftretenden als auch konkret testspezifische Nachteile.

Der von vielen Autoren beschriebene Lerneffekt und der auch teilweise damit verbundene, relativ hohe zeitliche Aufwand (aufgrund von Übungsmessungen) für eine valide Datenerhebung stellen die zentralen Kritikpunkte am Oldenburger Satztest dar (Schorn (2004), Müller-Deile (2009), Kießling (2010), Löhler et al. (2014), Hey et al. (2014)).

Im Gegensatz zum optionalen Training bei dem Freiburger Test ist beim Oldenburger Satztest vor jeder Messung ein Training zwangsläufig erforderlich (Schmidt und Baljić, 2016).

Da das Wortrepertoire ein- und zweisilbige Wörter enthält, besteht eine Asymmetrie der pro Satz präsentierten Silben. Es werden pro Satz zwischen 8 und 10 Silben vorgespielt. Schraud (2014) zeigte, dass bestimmte Wörterkombinationen innerhalb des

Tests häufiger vorkommen als andere.

Das planmäßige Gelingen eines Satztests ist maßgeblich von einem ausreichenden auditiven Arbeitsgedächtnis abhängig. Für Patienten mit einer geringen Merkfähigkeit (vorwiegend ältere Menschen) kann es aufgrund einer Beeinträchtigung des deklarativen Verbalgedächtnisses schwierig sein, sich an alle abgespielten (und auch verstandenen) Wörter zu erinnern. Dadurch können falsch negative Resultate entstehen (Meister et al. (2011), Löhler et al. (2014), Meister (2017)).

Ferner werden Satztests in der Praxis allgemein als „belastend für Patienten und Untersucher“ eingestuft (Löhler et al., 2014).

1.4 Fragestellungen der Dissertation

Viele Fragen, die den kurzfristigen Lerneffekt betreffen werden durch die zitierten Studien größtenteils gut geklärt und ausführlich erläutert. Zusammenfassend wird ein kurzfristiger Lerneffekt von allen Autoren beschrieben.

Lerneffekte spielen demnach beim Oldenburger Satztest eine große Rolle. Ihr Vorliegen wird nicht in Frage gestellt, wohingegen ihre Dimension in Bezug auf Zeit und Ausmaß noch nicht abschließend untersucht wurde. Das Ziel dieser Dissertation stellt das strukturierte Erfassen des kurz- und langfristigen Lerneffekts bei der mehrfachen Durchführung des Oldenburger Satztests dar.

Der Aspekt des langfristigen Lerneffekts wurde in bisherigen Arbeiten noch nicht abgedeckt.

Frühere Lösungsansätze: Der Lerneffekt müsste durch eine Vortestung mit 20 bis 140 Sätzen (je nach Quelle in Ruhe oder im Störgeräusch) vor der eigentlichen Datenerhebung beseitigt werden.

Die Frage, ob das „Hörtraining“ vor jeder neuen Sitzung wiederholt werden muss, blieb bisher offen. Diese für die alltägliche Praxis relevante Lücke der Forschung, wo bezüglich des langfristigen Lerneffekts noch Verbesserungsbedarf im Sinne einer etwaigen Anpassung der individuellen Resultate bestehen könnte, muss sowohl für Normalhörende, als auch für Hörgeschädigte auch unter kompensationsstechnischen Aspekten genauer unter die Lupe genommen werden.

In Anlehnung an die Arbeit von Wagener und Brand (2005) werden einige Fragen und Theorien zum Lerneffekt sowie Überlegungen zu dessen Elimination aufgeworfen: Lernen die Probanden den Test besser kennen oder lernen sie tatsächlich etwas dazu? Welche Einflussfaktoren können identifiziert werden? Interagieren Einflussfaktoren

1 Einleitung

miteinander?

Weiterhin wurde noch nicht untersucht, ob und in welchem Ausmaß der kurzfristige und langfristige Lerneffekt zusammenhängen. Die zeitlichen Abstände in bis dato durchgeführten Anwendungen des Oldenburger Satztests beschränken sich auf wenige Tage. Was geschieht, wenn die Abstände verlängert werden? Weiter konnte bisher nicht abschließend festgesetzt werden, ob und inwieweit eine Lernsättigung ermittelt werden kann.

Konsekutiv wird die hinsichtlich Folgeuntersuchungen klinisch relevante Frage nach Testdurchläufen bzw. Probedurchgängen vor der eigentlichen Datenerhebung aufgegriffen. Zusätzlich werden die Unterschiede zwischen einer offenen versus einer geschlossenen Durchführung aufgegriffen, indem der Einfluss einer Vorlagenkarte detektiert wird. Dabei wird die nicht beantwortete Frage, hinsichtlich unterschiedlicher Instruktionen des Anwendens der Vorlagenkarte (Lesen Sie die Karte spaltenweise bzw. zeilenweise.) aufgegriffen. Aus Voruntersuchungen wird die Frage nach dem Einfluss des Geschlechtes überprüft. Genauer beleuchtet wird im Folgenden auch die Frage des Faktors Zeit auf die Ergebnisse. Werden die Probanden schneller? Als Zusatzfrage kann zumindest aufgeworfen werden, ob durch die Analyse des Lerneffekts ein Faktor oder ein Verfahren ermittelt werden kann, durch welchen/welches der Lerneffekt weggerechnet bzw. kompensiert werden kann? Kann der Lerneffekt durch eine Korrekturvorschrift ausgeglichen werden?

Benötigen normalhörende Probanden mehr Übungssätze/ein längeres Training? Kann in einer Langzeitstudie untersucht werden, ob dieses erweiterte Übungsprogramm nur einmal oder mehrfach (beispielsweise vor jeder Sitzung) durchgeführt werden muss? Sind Probanden durch ein Übungsprogramm ein ganzes Leben lang „vor-trainiert“ und können ihren „Wissensstand“ in späteren Messungen nur durch ein paar Sätze wieder auffrischen? Kann der individuelle Trainingsstatus durch eine Analyse der durchgeführten Messungen eingestuft werden? Gelingt eine Elimination des Lerneffekts durch die Präsentation einer bestimmten Anzahl an Übungssätzen vor der Datenerhebung?

Im weiteren Kontext stellte sich außerdem die Frage, ob der Oldenburger Satztest als modernes sprachaudiometrisches Verfahren zum wiederholten Gebrauch geeignet ist und die Vorgaben für eine DIN-Norm erfüllt.

Durch das Studiendesign sollen die hier neu aufgeworfenen Fragen beantwortet werden.

Abgeleitet aus den Voruntersuchungen ergeben sich mehrere Hypothesen:

1. Es existiert ein Kurzzeit-Lerneffekt.

2. Es existiert ein Langzeit-Lerneffekt.
3. Der Lerneffekt wird über die Durchgänge und Sitzungen hinweg kleiner.
4. Der Lerneffekt kann durch die Vorlage einer Karte mit den benutzten Wörtern beeinflusst werden.
5. Der Lerneffekt ist unabhängig von der Instruktion der Leserichtung dieser Karte.
6. Der Lerneffekt ist unabhängig vom Geschlecht.
7. Der Zeitverbrauch der Probanden mit Karte ist größer.
8. Der Lerneffekt kann quantifiziert werden und mit Hilfe eines Faktors neutralisiert werden.

In Kapitel 4 wird auf diese Hypothesen eingegangen sowie deren Richtigkeit bewertet und diskutiert.

1.5 Wissenschaftspolitischer Hintergrund der Arbeit

Konträre Positionen werden von [Löhler et al. \(2013\)](#) und [Kollmeier et al. \(2014\)](#) in einem unlösbaren, wissenschaftlichen Dissens vertreten.

In ihrer Arbeit „Evaluation des Freiburger Einsilbertests im Störschall“ testen [Löhler et al.](#), ob der Freiburger Einsilbertest, der derzeit bei sprachaudiometrischen Untersuchungen in Deutschland Standard ist, auch im Störschall zur Anpassung von Hörgeräten eingesetzt werden kann. Als Fazit wird die Eignung des Freiburger Einsilbertests bei der Beurteilung der Funktion von Hörgeräten sowohl in Ruhe, als auch im Störschall bestätigt. Ferner seien keine Unterschiede zwischen den einzelnen Wortlisten zu erkennen.

Im Leserbrief „Zur Diskussion um den Freiburger Einsilbertest“ von [Kollmeier et al.](#) und der anschließenden Erwiderung von [Löhler et al.](#) wird auf den vorherigen Artikel Stellung bezogen. [Kollmeier et al.](#) kritisieren die Studie von [Löhler et al.](#) insofern, als der Freiburger Einsilbertest durch sein Abweichen von der ISO Norm 8253-3 nicht mehr zeitgemäß sei und nicht für eine Anwendung im Störschall geeignet sei. Weiter würden die beiden Kernaussagen nicht durch die Resultate belegt werden. [Kollmeier et al.](#) befürchten, dass [Zitat:] [...], „neuere Verfahren, die zur Erfassung des Sprachverstehens im Störgeräusch besser geeignet sind, keinen Eingang in die audiometrische Praxis, insbesondere auch beim niedergelassenen HNO-Arzt finden werden.“

In der dem Leserbrief angehängten Erwiderung stellten [Löhler et al.](#) klar, dass nicht in den Dimensionen „richtig“ und „falsch“ gedacht werden darf, sondern, dass die Anwendungsgebiete der unterschiedlichen sprachaudiometrischen Untersuchungen beachtet

werden müssen. Er weist sämtliche Kritikpunkte zurück und verweist auf die einzelnen Abschnitte der DIN EN ISO 8253-3. Der Zweifel an der wissenschaftlichen Arbeit wird als „perfide“ und „unhaltbar“ zurückgewiesen. [Löhler et al.](#) weiter [Zitat]: „Es handelt sich bei der umfänglichen, aber unseriösen Polemik von Kollmeier et al. um den untauglichen Versuch, die im deutschsprachigen und auch im internationalen Maßstab üblichen und im großen Stil eingesetzten Worttests in Störschall [...] zu diskreditieren und den flächendeckenden Einsatz von (z. B. in Oldenburg entwickelten und von dort aus mit eigener Vertriebsorganisation vertriebenen) Satztests nach Gutsherrenrecht einzufordern.“ „Wissenschaft lebt von der Diskussion [...], nicht von pseudopäpstlichen Bullen mit Alleinvertretungsanspruch in Teilgebieten, auch wenn man durch Stimmenmehrheit in einzelnen DIN-Ausschüssen versuchen kann, hierfür Normen zu kreieren.“

Zusammenfassend stellt sich auf wissenschaftspolitischer Ebene die Frage, ob zur Erfassung des Sprachverständnisses (in Ruhe und im Störschall) der Freiburger Test (1953) oder der Oldenburger Satztest (1999) besser geeignet sind und welcher Test in der praktischen Anwendung sowohl in Kliniken, als auch bei niedergelassenen Hals-Nasen-Ohren-Ärzten zum diagnostischen Goldstandard bestimmt werden sollte.

Nach der 2017 aktualisierten Richtlinie des gemeinsamen Bundesausschusses über die Verordnung von Hilfsmitteln in der vertragsärztlichen Versorgung wird erwähnt, dass der Oldenburger Satztest für die Indikationsstellung bei der beidohrigen Hörgeräteversorgung eingesetzt werden *kann* ([Gemeinsamer Bundesausschuss, 2017](#)).

Für die Indikation von Hörhilfen und Hörimplantaten stellen Satztests somit gemäß der aktuellen Hilfsmittelrichtlinien keine zwingende Voraussetzung dar. Als Anwendungsbedingung wird die Ausschaltung des Trainingseffektes genannt ([Hoppe und Hast, 2017](#)).

Bei der Erstellung von Gutachten werden sprachaudiometrische Verfahren zur Ermittlung des Hörverlustes verwendet. Genutzt werden der Freiburger Einsilbertest und der Freiburger Zahlentest. Nur in Grenzfällen wird zusätzlich das Tonaudiogramm mitberücksichtigt. Wenig beachtet werden sprachaudiometrische Verfahren im Störgeräusch. Diese werden bisher in der Praxis zur Überprüfung von Hörsystemen eingesetzt ([Batsoulis und Lesinski-Schiedat, 2017](#)). [Steffens \(2015\)](#) wünscht sich ein breiteres Anwendungsgebiet.

In diesen Rahmen fügt sich diese Dissertation durch weitere, den Lerneffekt des Oldenburger Satztests betreffende Untersuchungen ein.

2 Versuchsplanung und Methodik

Um der zentralen Frage nach dem kurz- und langfristigen Lern- und Habituationseffekt systematisch nachzugehen, wurde eine Vorgehensweise entworfen, die potentielle Störfaktoren möglichst ausschließen sollte. So wurde bspw. eine homogene, normalhörende Probandengruppe vorgesehen, auf Geschlechtergleichverteilung geachtet und die zu untersuchenden Einflussfaktoren systematisch permutiert.

Ethikvotum: Auf die Beantragung eines Ethikvotums wurde aus folgenden Gründen verzichtet:

Die Richtwerte für die zulässige Schallbelastung (vorgegeben durch die Richtlinie 2003/10/EG) beziehen sich auf einen Dauerschallpegel von 80 dB bzw. einen Spitzenschalldruckpegel von 135 dB. Diese jeweiligen Pegel beliefen sich in der vorliegenden Arbeit auf 65 dB (Störrauschen während des Satzes für wenige Sekunden) bzw. bei wenigen Probanden bei einzelnen Sätzen auf 70 dB (einzelner Stimulus). Es wurden für die Messungen nur als Medizinprodukt deklarierte Geräte verwendet. Aufgrund der Wartung und Kalibrierung der Apparatur sind die oben genannten Schalldruckpegel garantiert; höhere, das Gehör potentiell schädigende Pegel können somit sicher ausgeschlossen werden. Eine Schädigung des Gehörs der Probanden ist im folgenden Versuchsaufbau nicht möglich. Weiterhin handelt es sich bei der Durchführung des Oldenburger Satztests um keine neue Methode der Sprachaudiometrie, sondern um ein im klinischen Alltag etabliertes, häufig und regelmäßig eingesetztes Verfahren. Die Einverständniserklärung und Probandeninstruktion (inklusive Erklärungen zu oben Beschriebenem) wurden von jedem Probanden vor Durchführung der Tests unterschrieben.

2.1 Probanden

Als Probanden wurden 41 Frauen und Männer, die zum Zeitpunkt der ersten Testsitzung zwischen 20 und 29 Jahre alt waren, rekrutiert. Für die Beantwortung der Fragestellung war eine Kerngruppengröße von 32 Probanden erforderlich, aufgrund der geplanten drei Testsitzungen (s. Abschnitt 2.2, S. 26) musste aber damit gerechnet werden, dass

einige der Probanden ausfallen könnten. Daher wurde initial diese größere Zahl rekrutiert. Drei der Probanden fielen aufgrund verschiedener Umstände (siehe unten) schon früh aus, so dass der erweiterte Kreis aller Probanden von 41 auf 38 schmolz. Daher beziehen sich alle nachfolgenden Angaben und Berechnungen auf die Probandenzahl von 38.

Im Durchschnitt waren die 19 Frauen bei der ersten Testsitzung 24,10 Jahre (8797 Tage) und die 19 Männer 24,92 Jahre (9098 Tage) alt. Das mittlere Alter aller 38 Probanden betrug 24,51 Jahre (8947 Tage).

Das Probandenkollektiv bestand aus Studenten, größtenteils der Humanmedizin, die nach Erläuterung des Testkonzepts motiviert und interessiert, unentgeltlich an der Studie teilnehmen wollten.

Neben dem Alter zwischen 20 und 29 Jahren stellten Normalhörigkeit, Deutsch als Muttersprache und die Partizipation an allen drei Testsitzungen weitere Teilnahme Kriterien dar. Außerdem durfte der Oldenburger Satztest den Probanden nicht bekannt sein, d. h. vor der ersten Testsitzung noch nicht von diesen absolviert worden sein. Weiterhin galt die anamnestisch erfragte Abwesenheit (offensichtlicher) pathologischer Veränderungen des Ohres als Einschlusskriterium.

Zur Quantifizierung des Hörvermögens bzw. zur Bestätigung der Normalhörigkeit wurde bei jedem Probanden vor jeder Testsitzung ein Tonaudiogramm durchgeführt (Details unter [2.3.1](#)).

Bei nur drei Ausnahmen erfüllten 38 Studienteilnehmer diese Bedingungen. Die Drop-Outs kamen folgendermaßen zustande: Ein rekrutierter Proband musste aufgrund eines nicht tolerierbaren Hörverlusts im Tonaudiogramm schon vor den eigentlichen Tests ausgeschlossen werden. Aus Eigeninteresse und zum Vergleich mit den deutschen Muttersprachlern, wurde ein mexikanischer Proband, welcher Spanisch als Muttersprache spricht, ebenfalls getestet. Eine Probandin absolvierte aus privaten Gründen lediglich die erste der zur Auswertung der Arbeit nötigen drei Testsitzungen.

2.2 Zeitlicher Ablauf der Tests

Zur Untersuchung der *kurzfristigen* Lerneffekte wurden in einer Testsitzung fünf Testdurchläufe des Oldenburger Satztests durchgeführt. Zur Untersuchung der *langfristigen* Lerneffekte wurden drei solche Testsitzungen in einem Zeitabstand von ca. 50 Tagen angestrebt. Die Probanden führten somit 15 Testdurchläufe durch.

2.3 Testmethoden

Im Rahmen dieser Dissertation kamen zwei audiometrische Testverfahren zum Einsatz: das Tonaudiogramm und der Oldenburger Satztest. Der Unterschied besteht darin, dass in der *Tonaudiometrie* zwischen „Ton gehört“ und „Ton nicht gehört“ differenziert wird, wohingegen bei der *Sprachaudiometrie* (in diesem Fall bei dem Oldenburger Satztest) die Menge des Herausgehörten quantifiziert wird.

2.3.1 Reintonaudiogramm

Das Reintonaudiogramm (auch: Audiogramm oder Hörkurve) stellt die subjektive Hörschwelle in Abhängigkeit von verschiedenen Frequenzen dar.

In einem Standard-Tonaudiogramm werden die folgenden 12 Frequenzen (als Sinusschwingungen, in Kilohertz) für jedes Ohr einzeln untersucht: 0,125, 0,25, 0,5, 0,75, 1, 1,5, 2, 3, 4, 6, 8 und 12,5.

Der Versuchsleiter präsentiert dem Probanden über einen Kopfhörer diese Frequenzen als unterschiedlich tiefe bzw. hohe Töne in zunehmender Lautstärke. Angegeben wird die Hörschwelle in der Einheit Dezibel (dB HL), einem logarithmischen Verhältnismaß zweier Schallpegel. Als frequenzspezifische Bezugsschallpegel dienen hierbei genormte Referenzwerte, die die mittleren Hörschwellen junger Normalhörender repräsentieren.

Erhöht wird die Lautstärke der Töne in 5-dB-HL-Schritten. Sobald der durch den sukzessive gesteigerten Prüfschalldruck lauter gewordene Ton vom Probanden vernommen wird, drückt dieser auf einen Knopf. Daraus ergibt sich das Tonaudiogramm: In einem kartesischen Koordinatensystem (Abszisse: Frequenzen in Hertz; Ordinate: Schalldruckpegel in dB HL in Abhängigkeit von der Bezugsgröße) können die Werte dargestellt werden. Eine Verbindungslinie der individuellen Ergebnispunkte ergibt die Hörschwellenkurve des jeweiligen Probanden. Die Nulllinie entspricht dabei der Hörschwelle normalhörender Jugendlicher. Abweichungen nach unten (höherer Schalldruck zum Wahrnehmen des Tones nötig) repräsentieren (ab bestimmten Werten) eine Pathologie des Ohres im Sinne eines Hörverlustes.

Das Tonschwellenaudiogramm kann sowohl für die Luftleitung (mit Kopfhörern), als auch für die Knochenleitung (mit einem Knochenleitungshörer, welcher hinter dem Ohr platziert wird) durchgeführt werden (Mrowinski (2006), Lenarz und Boenninghaus (2007), Friesecke (2014)).

In dieser Studie erfolgte die Durchführung des Tonaudiogramms mithilfe eines Kopf-

Tabelle 2.1: Maximal tolerierter Hörverlust in der Tonaudiometrie

Frequenz des abgespielten Tons (in kHz)	0,125	0,25	0,5	0,75	1	1,5	2	3	4	6	8	12,5
maximal tolerierter Hörverlust (in dB HL)	30	20	20	20	20	20	20	20	20	25	30	30

hörers, getrennt für jedes Ohr. Sobald der Ton gehört wurde, sollte auf einen Knopf gedrückt werden. Der so ermittelte Schallpegel wurde für jede Frequenz solange wiederholt getestet, bis mindestens zwei Messungen denselben Wert ergaben.

Probandeninstruktion bei der Durchführung eines Tonschwellenaudiogramms:

Sie hören jetzt unterschiedlich hohe und tiefe Töne, welche über den Kopfhörer für jedes Ohr einzeln dargeboten werden. Die Lautstärke der Töne wird langsam erhöht. Sobald die Töne laut genug sind, sodass Sie sie hören können, drücken Sie sofort auf den Druckknopf. Zuerst wird das rechte Ohr getestet, anschließend das linke.

Die Bedingung der Normalhörigkeit erfüllte der Studienteilnehmer, wenn er im Tonaudiogramm eine im Vorfeld definierte Schwelle nicht unterschritt. Die Definition „normalhörend“ (modifiziert in der DIN-Norm [ISO 7029:2000 \(2000\)](#)) hatte sich in den Dissertationen von [Domes \(2009\)](#) und [Qualen \(2010\)](#) bewährt und wurde in einer leicht modifizierten Form übernommen: Im Hauptsprachbereich mit Frequenzen zwischen 250 Hz und 4 kHz wurde ein Hörverlust von maximal 20 dB HL toleriert. In den Nebenfrequenzen 125 Hz, 8 kHz und 12,5 kHz durften die Probanden maximal einen Hörverlust von 30 dB HL aufweisen. Bei 6 kHz betrug der maximal geduldete Hörverlust 25 dB HL. Der maximal tolerierte Hörverlust ist in [2.1](#) zusammengefasst. Alle 38 rekrutierten Studienteilnehmer (ohne Drop-outs) erfüllten diese Voraussetzungen und wurden somit in die Studie eingeschlossen.

2.3.2 Durchführung des Oldenburger Satztests in dieser Arbeit

Der Oldenburger Satztest wurde in dieser Studie in seiner computerisierten Form durchgeführt: Abgespielt wurden 40 Testlisten zu je 30 Sätzen.

Um die Ergebnisse mit anderen Arbeiten vergleichen zu können und etwaiger Kritik vorzubeugen, wurde wie im ursprünglichen Testdesign vorgesehen, das Störgeräusch konstant auf demselben Lautstärkepegel (65 dB SPL) gehalten, während der Pegel der gesprochenen Sprache adaptiv verändert wurde (Ursprungsconfigurationen). Aus der Pegeldifferenz ergibt sich der Signal-Rausch-Abstand (signal-noise-ratio, SNR). Der Schalldruckpegel (in dB SPL) des Störgeräusches wird dabei vom Schalldruckpegel (in

dB SPL) des Sprachsignals subtrahiert.

Gemessen wurde die Sprachverständnisschwelle (SVS) = speech reception threshold (SRT) als derjenige SNR-Wert, bei dem 50 % der dargebotenen Stimuli richtig verstanden werden.

Diese Sprachverständnisschwelle wurde berechnet als Mittelwert aus den letzten (abgespielten) 19 SNR-Werten (Sätze 12-30) und dem vorausberechneten SNR-Wert des nicht mehr abgespielten (und nicht existierenden) 31. Satzes. Diese Berechnungsvorschrift wurde bewusst aus der originalen OLSA-Anleitung übernommen.

Detailliertere Angaben zum Oldenburger Satztest sind dem Kapitel 1.3 zu entnehmen.

2.4 Testumfeld und technische Ausstattung

Sämtliche Messungen fanden in der Testkabine C, einem Audiometriemessraum für klinische Untersuchungen und Forschungen in der Klinik für Hals-, Nasen- und Ohrenkrankheiten der Universitätsklinik Würzburg statt. In demselben Raum fand auch die Datenerhebung von Grimm (2014) statt.

Verwendet wurde ein Intel @Core™ Computer mit 3.09 GHz und 2,89 GB Ram (Betriebssystem Windows XP), von dem aus das Durchführungsprotokoll der Reintonaudiometrie und des Oldenburger Satztests gesteuert wurden. Im Rahmen der Tonaudiometrie mit dem Audiometer WESTRA CAD-03 wurden die Kopfhörer Beyerdynamic DT 48 CAD-03, deren rechte Seite mit einem roten Punkt und deren linke Seite mit einem blauen Punkt gekennzeichnet ist, zur Präsentation der Töne und das Programm WESTRA zu deren Aufzeichnung verwendet. Zur Wiedergabe und Aufzeichnung der Daten des Oldenburger Satztests diente die LOST-Software sowie die Matlab-Software. Die LOST-Software speichert die Rohdaten der Tests auf allen Ebenen in Dateien. Bei der Auswertung wurden diese Daten benutzt.

Der verwendete Lautsprecher WESTRA Audiometriebox LAB-1001 erzeugt durch sein geringes Eigenresonanzverhalten und dadurch geringen Klirrfaktor einen relativ verzerrungsfreien Klang über alle Frequenzen hinweg (Papke, 2013).

Tabelle 2.2: Versuchsaufbau: Gruppeneinteilung

Gruppennummer (Zufällige Randomisierung des Probanden anhand des Zeitpunkts seiner Zusage für die Teilnahme bei der Studie)	Testmodus (Vorlage der Karte mit allen vorkommenden Wörtern)	Leserichtung (Verwendungsmodus der Karte)	Anzahl der getesteten Männer (vorab geplante Anzahl bzw. selektierte Hauptgruppe hinterlegt)	Anzahl der getesteten Frauen (vorab geplante Anzahl bzw. selektierte Hauptgruppe hinterlegt)
1 o. 4 (nur methodisch getrennt, gleicher Testmodus)	ohne Karte (offener Testmodus)	–	9 8	8 8
2	mit Karte (geschlossener Testmodus)	spaltenweise	5 4	6 4
3	mit Karte (geschlossener Testmodus)	zeilenweise	5 4	5 4
			19 16	19 16
			38 = Hauptgruppe 32 = selektierte Hauptgruppe (41 = Gesamtgruppe) (alle rekrutierten Probanden)	

2.5 Versuchsdesign

2.5.1 Versuchsaufbau

Der Versuchsleiter teilte die 19 männlichen und 19 weiblichen Probanden in die verschiedenen Studiengruppen ein. Dies geschah anhand des Zeitpunkts deren Zusage, an der Studie teilnehmen zu wollen. Eine Zufallsrandomisierung war damit garantiert.

Tabelle 2.2 verdeutlicht die Gruppeneinteilung.

Den Probanden der Gruppe 1 und 4 wurde die Karte während aller Messungen vorenthalten. Der Oldenburger Satztest blieb im Prinzip während aller Testungen offen. De facto ist aber davon auszugehen, dass der Test durch das Kennenlernen des Wortmaterials im Laufe der 15 Testlisten zu einem *geschlossen* bzw. *semi-offenen* Test wurde. Die Trennung in Gruppe 1 und 4 ist nur aus methodischen Zwecken vorgenommen worden, der Testmodus ist derselbe.

Den Teilnehmern der Gruppe 2 und 3 wurde vor der Durchführung des Oldenburger Satztests die Karte mit dem Wortrepertoire (2.3) gezeigt. Darauf standen alle vorkommenden Wörter in Form einer Art Tabelle bzw. Matrix, welche entweder spaltenweise (vertikal) oder zeilenweise (horizontal) gelesen, aufgenommen und interpretiert werden konnte. So wurde in diesen Gruppen der Oldenburger Satztest direkt von einem *offenen* in einen *geschlossenen* Test überführt. Unterteilt wurde diese Gruppe durch

Tabelle 2.3: Oldenburger Satztest: Wort-/Satzrepertoire

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

die Art der Anweisung, wie mit der Karte umgegangen werden sollte: Eine Untergruppe wurde angewiesen, die fünf Spalten verschiedener Worttypen der Reihe nach, d. h. spaltenweise zu lesen (Gruppe 2). Die andere Untergruppe sollte die Wörter als zehn Fünf-Wort-Sätze, d. h. zeilenweise lesen (Gruppe 3).

Das Studiendesign sah vor, bei der Durchführung des Oldenburger Satztests von jedem Probanden die folgenden Parameter (abhängige Variablen) zu erheben:

- Sprachverständlichkeit,
- Sprachsignalschallpegel (in dB SPL),
- Störgeräuschkpegel (in dB SPL),
- Signal-Rausch-Abstand (signal - noise - ratio, SNR) (in dB),
- Sprachverständnisschwelle (speech reception threshold, SRT) (in dB SNR),
- Anzahl der korrekt verstandenen Wörter (0 bis 5),
- Listennummer des Oldenburger Satztests (1 bis 40),
- Satznummer innerhalb der jeweiligen Liste (1 bis 30),
- Antwortlatenz (Zeit bis zum ersten Klick) (in Sekunden) und
- Testzeit, die im Verlauf aller Testungen durch das Programm Matlab protokolliert wurde

Diese sollten in Abhängigkeit von diesen potentiellen Einflussfaktoren (unabhängige Variablen) untersucht werden:

- Geschlecht (σ oder φ),
- Testsitzung (1 bis 3),
- Testdurchlauf (1 bis 5),

- Vorlage der Karte (mit oder ohne) und dabei die Leserichtung (spaltenweise oder zeilenweise)

Der Zeitplan des Versuchsdesigns sah vor, sämtliche Probanden zu drei verschiedenen Zeitpunkten im Abstand von acht Wochen (\pm eine Woche) bzw. nach rund 50 und 100 Tagen zu testen.

Pro Testsitzung wurden jedem Probanden fünf Listen à 30 Sätze des Oldenburger Satztests vorgespielt.

Permutation der Testlisten: Ausgewählt wurden die Listen anhand der Probandennummer. Das Prinzip sah vor, mit der der Probandennummer entsprechenden Liste zu beginnen und von dort ab die darauffolgenden 14 Listen zu nutzen. Nach Liste 40 wurde wieder von vorne mit Liste 1 begonnen. Beispielsweise erhielt Proband Nummer 13 in der ersten Testsitzung die Listen 13 bis 17, in der zweiten 18 bis 22 und in der dritten 23 bis 27.

Zustandekommen der Stichprobengröße

Das Zustandekommen der Stichprobengröße $n = 32$ erklärt sich im Hinblick auf empirische, praktische und statistische Hintergründe wie folgt: Ursprünglich wurde eine erforderliche Stichprobengröße $n = 24$ gewählt. Aufgrund des langen Testzeitraums und der damit verbundenen Annahme, es könnten Probanden unerwartet wegfallen wurden jedoch von vornherein mehr Individuen ausgewählt (41) bzw. getestet (38). Letztlich schied in dieser Gruppe kein relevanter Anteil aus, so dass die größere Stichprobengröße (32) in die Auswertung einfließen konnte. Unter Berücksichtigung der gewünschten Reliabilität flossen das Vorhandensein der zeitlichen, örtlichen und individuellen Ressourcen, der Informationsmehrgewinn, die Heterogenität der Probanden (männlich und weiblich) und das Versuchsdesign, inklusive der Gruppeneinteilung in die Entscheidung, ursprünglich eine Stichprobengröße $n = 24$ zu wählen, ein. Die Vergleichbarkeit mit und die Erfahrung aus ähnlichen Studien sowie die praktische und realistische Umsetzbarkeit und die schlichtweg größere Aussage- und Überzeugungskraft bei erhöhter Stichprobengröße spielten zusätzlich bei der Wahl der endgültigen Stichprobengröße $n = 32$ eine Rolle. Es bleibt dennoch zu vermerken, dass durch einen größeren Stichprobenumfang ($n > 32$) eine bessere Verlässlichkeit hervorgeht.

Bestimmung der Hauptgruppe der Probanden

Aus den insgesamt 41 rekrutierten Probanden bzw. aus der potentiell für die Hauptgruppe in Frage kommenden 38 Probanden wurde schließlich die Hauptgruppe mit 32 Probanden selektiert. Dies geschah anhand eines zusätzlichen Kriteriums, welches in Kapitel 3 ab Seite 39 beschrieben wird.

2.5.2 Versuchsablauf

Auswahl: Bei der Vorauswahl der Testpersonen wurde bereits eine kurze Anamnese vorgenommen, in der die Einschlusskriterien abgefragt wurden. Nach Rekrutierung des Probanden wurden die Testprozedur ausführlich mündlich erläutert und dem Probanden zusätzlich eine schriftliche Einverständniserklärung vorgelegt. In dieser wurden die Methoden vorgestellt und betont, dass durch Drosselung der Lautstärke nie Schalldruckpegel über 90 dB SPL erreicht werden, sodass eine Hörschädigung durch die Tests garantiert ausgeschlossen werden konnte. Die Probanden unterschrieben die Einverständniserklärung vor der ersten Testsitzung. Nachfolgend wurden die Einschlusskriterien nochmals abgefragt und ergänzend bestätigt (siehe unten).

Testung: Zu Beginn jeder Testsitzung in der Kabine C wurde der Proband angewiesen, auf dem Stuhl in der Mitte der Testkabine Platz zu nehmen. Zur Überprüfung der richtigen, genormten Sitzposition wurde darauf geachtet, dass sich der Kopf der Testperson direkt unter einer Schraubenmutter befand, die mit einem Faden an der Decke befestigt worden war und die für die Testungen richtige Kopfposition markierte. Es wurde während der Testsitzung darauf geachtet, dass sich der Kopf des jeweiligen Probanden stets unterhalb der Schraubenmutter befand. Vor dieser standardisierten Position stand in einem Abstand von 1 m ein leicht gekippter Lautsprecher aufgebaut, dessen Vorderseite zum Probanden zeigt (Grimm (2014, S. 13, Abb. 2.1)).

Einverständniserklärung und Probandeninstruktion

Im Anschluss erhielt die Versuchsperson Instruktionen über den Versuchsablauf: Die Einverständniserklärung und Probandeninstruktion wurden schriftlich ausgehändigt und lauteten wie folgt:

Der Zweck des Versuches ist es, den Lerneffekt bei wiederholter Durchführung des Oldenburger Satztests zu quantifizieren. Hierfür sind drei Testsitzungen im Abstand von

2 Versuchsplanung und Methodik

8 Wochen (+/- 1 Woche) nötig. Zum Gelingen der Arbeit müssen alle drei Testsitzungen durchgeführt werden; andernfalls wären die bis dahin erhobenen Daten und Ergebnisse unbrauchbar. Bei allen Messungen befinden sich die verwendeten Schallpegel stets unter 90 dB SPL, sodass eine Hörschädigung ausgeschlossen ist. Vor der Ausführung des Sprachtests wird ein Hörtest durchgeführt, um festzustellen, ob Sie als Proband „normalhörend“ und somit für die Tests geeignet sind. Der Oldenburger Satztest stellt fest, wie gut Sie Sprache in geräuschvoller Umgebung verstehen können. Hierzu werden Ihnen Sätze von einer männlichen Stimme über einen Lautsprecher vorgesprochen. Jeder Satz besteht aus fünf Wörtern folgender Struktur: Name (als Subjekt), Verb (Prädikat in unterschiedlichen Tempora), Zahlwort, Adjektiv und Substantiv (als Objekt). Die Sätze sind nicht unbedingt sinnvoll. (Ein Beispielsatz wäre: „Otto siebt fünfzehn feuchte Lappen.“) Zusätzlich zu der Sprache wird ein in der Lautstärke gleich bleibendes Rauschen dargeboten. Bitte wiederholen Sie nach der Darbietung den Satz oder jedes Wort, welches Sie verstanden haben. Wenn Sie unsicher sind, dürfen Sie gerne auch raten. Ich werde die Wörter festhalten, die korrekt wiedergegeben wurden. Während der Tests wird die Sprache in der Lautstärke verändert. Sie kann dadurch teilweise sehr leise erscheinen. Lassen Sie sich dadurch nicht entmutigen. Für die Aussagekraft der Messung ist es wichtig, diese Messung unter schwierigen Bedingungen durchzuführen. Der Oldenburger Satztest wird pro Testsitzung jeweils fünf Mal, teilweise unter verschiedenen Bedingungen, durchgeführt. Beispielspielweise werden Sie einmal naiv getestet, dann mit Kenntnis der Wörter, die vorkommen können. Versuchen Sie bitte, während des gesamten Tests möglichst konzentriert zu sein. Pausen zwischen den einzelnen Tests sind jederzeit möglich. Sie können den Test jederzeit abbrechen. Die bis dahin erhobenen Daten werden dann allerdings unbrauchbar und können nicht mehr ausgewertet werden. Die Daten werden zu rein wissenschaftlichen Zwecken erhoben; sie werden verschlüsselt und nur in anonymisierter Form ausgewertet. Wir verpflichten uns zur Einhaltung der datenschutzrechtlichen Bestimmungen. Schon im Voraus vielen Dank für Ihre Bemühungen, Ihre Teilnahme und Ihr Durchhaltevermögen! Ich habe den oben stehenden Text gelesen und zur Kenntnis genommen und erkläre meine Einwilligung zur Teilnahme an den Versuchen.

[Datum, Teilnehmer]

[Datum, Versuchsleiter]

Als Vorlage hierfür dienten ähnliche Erläuterungen, welche aus der thematisch verwandten Dissertationsarbeit von Grimm (2014) passend umformuliert wurden .

Mündliche Ergänzungen zu der schriftlichen Vorlage:

Summarisch wurde jede/r Proband/in über den Ablauf und das Prinzip des Oldenburger Satztests unterrichtet: All die Worte, die er/sie meinte, trotz des Störgeräusches verstanden zu haben, sollten laut nachgesprochen werden. Falls etwas akustisch nicht verstanden wurde, durfte entweder geraten werden oder unbeantwortet zur nächsten Frage fortgeschritten werden. Es wurde betont, dass die Sinnhaftigkeit der Kurzsätze nicht im Vordergrund steht und auch vermeintlich Sinnloses wiederholt werden soll. Wert gelegt wurde weiterhin auf eine gute Konzentration.

Die anschließende Darbietung der Sätze des Oldenburger Satztests (fünf Listen zu je 30 Sätzen) erfolgte unter Berücksichtigung der oben beschriebenen Prinzipien.

2.5.3 Testleitung, Steuerung und Dokumentation

Sämtliches Satzmaterial sowie die Frequenzen der Tonaudiometrie spielte der Testleiter computergesteuert von einem Audiometrie-pult aus ab. Jeder laut wiederholte vollständige oder unvollständige Satz bzw. jedes einzelne wiedergegebene Wort wurde durch den Versuchsleiter mit einer durch das Programm präsentierten Vorlage überprüft. Richtig Verstandenes wurde einmal und Falsches zweimal angeklickt. Dadurch wurde jedes einzelne Wort im Programm entweder grün oder rot gefärbt und vermerkt. Ähnlich klingende Wörter (beispielsweise Jutta statt Britta, zwei statt drei, Petra anstelle von Peter) mussten trotz der Ähnlichkeit als „nicht richtig“ gewertet werden. Indem der Testleiter neben der Dokumentation der Antworten ebenfalls wahllose, nichts dokumentierende Klicks durchführte, konnte der Proband nicht aufgrund der Anzahl oder der Regelmäßigkeit der Klicks auf die Richtigkeit seines Ergebnisses schließen.

In Abhängigkeit von der Anzahl der korrekt verstandenen Wörter ($x/5$) wird die Lautstärke verändert. Wird keines, eines oder zwei der fünf Wörter richtig verstanden, erhöht sich Schalldruckpegel und der nächste Satz wird konsekutiv lauter präsentiert, während im Falle der korrekten Wiedergabe von drei, vier oder fünf Wörtern der Schalldruckpegel verringert wird und ein leiserer Satz generiert wird. Liegt das Sprachverständnis dementsprechend bei 0, 20 oder 40 %, wird das Sprachsignal lauter; liegt es bei 60, 80 oder 100 % wird es hingegen leiser.

Die Differenz zwischen dem gegenwärtig und dem nachfolgend präsentierten Schalldruckpegel wird in der dritten Originalpublikation zum Oldenburger Satztest beschrieben und nach folgendem Prinzip berechnet (Wagner et al. (1999b), Brand und Kollmeier (2002)):

$$\Delta L = - \frac{f_i(SV-T)}{m} \quad (\text{mit } f_i = 1,5 \times \sqrt{2}^{(i)})$$

Dabei gelten folgende Definitionen: ΔL = Veränderung des SNR-Werts, f = „Geschwindigkeit“ der Steuerung, i = Anzahl der Wendepunkte, SV = Satzverständlichkeit im vorhergehenden Satz, T = Ziel-Richtig-Antwort-Wahrscheinlichkeit, m = Steigung.

Der Pegel des Störrauschens wurde, wie im ursprünglichen Oldenburger Satztest-Design nicht verändert und lag konstant bei 65 dB SPL. Das Störrauschen startete 0,5 Sekunden vor der Präsentation des Satzes. Außerdem wurde am Ende des Satzes das Rauschen innerhalb von 0,5 Sekunden wieder ausgeblendet.

Jeder Proband nahm wie im Versuchsdesign vorgesehen an drei Testsitzungen teil. Abgesehen von etwas ausführlicheren Erläuterungen zu Beginn des ersten Termins gestaltete sich die Prozedur (Dauer, Ablauf, Durchführung) während aller drei Sitzungen gleich.

Alle Testungen fanden innerhalb von sechs Monaten im Zeitraum zwischen März und August 2013 statt (exakter Zeitraum: 14. März bis 21. August 2013). Die ersten Testsitzungen erfolgten zwischen März und Mai 2013 (exakter Zeitraum: 14. März bis 15. Mai 2013).

Die meisten Testungen wurden unter der Woche (an Arbeitstagen) nachmittags nach 16 Uhr durchgeführt, da die Kabine C vorher im klinischen Alltag die meiste Zeit belegt war. Einige Testungen erfolgten aus terminlichen Gründen seitens der Probanden an Wochenenden. Die Basiseinstellungen wurden durch ihre Zeitposition nicht beeinflusst.

2.6 Statistische Auswertung

Von jedem Probanden wurden die oben beschriebenen Werte erfasst und gespeichert. Die Rohdaten wurden auf verschiedenen Ebenen extrahiert, analysiert und zusammengeführt. Die erste Ebene enthält die Daten eines jeden Probanden (38), die zweite Ebenen befasst sich mit dem Testdurchlauf ($38 \times 15 = 570$). Die Satzebene ($38 \times 15 \times 30 = 17100$) und die Wortebene ($38 \times 15 \times 30 \times 5 = 85500$) stellen das dritte und vierte Level dar.

Die Tonaudiogramme wurden mittels Microsoft Excel in Tabellen umgeformt. Die ermittelten Resultate der Sprachaudiometrie wurden auf den verschiedenen Ebenen mit-

Tabelle 2.4: Vorgehen bei der Datenauswertung

Angewandeter Test	Testfunktion	Signifikanzniveau
1. Shapiro-Wilk-Test	Testung der Normalverteilung	5%
2. Levene-Test	Testung der Varianzhomogenität	5%
3. Varianzanalyse (ANOVA)	Testung der Varianz (Besteht ein Unterschied zwischen den Faktoren?)	5%
4. Tukey HSD Test	Post-hoc-Analyse (Paarweiser Vergleich der einzelnen Faktoren)	

tels der LOST-Software dem Statistikprogramm R Project ([R Core Team \(2018\)](#)) durchgeführt und ausgewertet. R als freie Programmiersprache dient dabei dem Programm R Project zur Berechnung statistischer Analysen und zur Erstellung von Grafiken. So wurden zu Beginn der statistischen Auswertung mehrfaktorielle univariante Varianzanalysen (englisch: analysis of variance, ANOVA) errechnet.

Die Nullhypothese in dieser Arbeit lautete: Die Faktoren Geschlecht, Testsitzung, Testdurchlauf und die Vorlage einer Karte beeinflussen die Gesamtergebnisse des Oldenburger Satztests.

Als Voraussetzung für die Durchführung der Varianzanalyse müssen die folgenden Bedingungen gegeben sein ([Bortz und Schuster, 2010](#)):

1. Die abhängige Variable muss *metrisch* sein, d. h. quantitativ in Form von Zahlen darstellbar (= gegeben durch das Versuchsdesign)
2. Die unabhängige Variable muss *kategorial* sein, d. h. mindestens zwei Faktorstufen besitzen (= gegeben durch das Versuchsdesign)
3. Die Varianzhomogenität (Homoskedasität) muss gegeben sein (Prüfung durch Levene-Test)
4. Die Stichproben müssen normalverteilt sein (Prüfung durch Shapiro-Wilk-Test)

Bei allen Tests (Levene-Test, Shapiro-Wilk-Test, Varianzanalyse) wurde mit einem Signifikanzniveau von 5 % gearbeitet.

In der Post-hoc-Analyse kam der Tukey-HSD-Test zum Einsatz. Mit diesem Signifikanztest wurden die Einflussfaktoren paarweise miteinander verglichen und berechnet, welche Werte sich signifikant voneinander unterscheiden ([Bortz und Schuster, 2010](#)).

3 Ergebnisse

Von 38 Probanden stehen die Resultate aus fünf Testdurchläufen für jede der drei Sitzungen vollständig zur Verfügung. Damit existieren aus insgesamt 114 Testsitzungen 570 SRT-Werte aus den Messdaten. Auf Satzebene können 17100 Sätze und auf Wortebene 85500 Wörter analysiert werden. In der aus 32 Probanden bestehenden selektierten Hauptgruppe liegen 480 SRT-Werte bzw. 14400 Sätze bzw. 72000 Wörter zur Datenauswertung vor.

Die individuellen 38 SNR-Verlaufskurven können im Anhang eingesehen werden.

3.1 Tonaudiogramme der Probanden

In Kapitel 2 ab Seite 25 wurde beschrieben, dass bei jedem Probanden vor jeder Testsitzung ein Tonaudiogramm erstellt wurde. Dessen Ergebnis stellte ein Selektionskriterium für die Teilnahme an der Studie dar. Probanden, deren Hörverlust die im Voraus definierte, zulässige Grenze überschritt, wurden aus der Studie ausgeschlossen.

Die individuellen Resultate der in die Arbeit aufgenommenen Testpersonen (selektierte Hauptgruppe) sind für jede Testsitzung einzeln in Form eines üblichen Tonaudiogramms in den Abbildungen 3.1, 3.2 und 3.3 auf den Seiten 40 und 41 dargestellt. Aufgeteilt wurde nach Sitzung Nummer 1, 2 und 3 sowie dem rechten und linken Ohr. Die individuellen Hörkurven der 38 Probanden sind als hellgraue Linien eingezeichnet. Die schwarzen Linien fassen die errechneten Mittelwerte der Hörkurven der Probanden zusammen. (Sie sind in kleinem Maße frequenzabhängig und betragen etwa 5 dB HL.) Die rote Linie repräsentiert den maximal geduldeten Hörverlust (vgl. 2.1). Die hellgrauen Linien variieren dabei in allen drei Abbildungen je nach Frequenz zwischen -15 und 30 dB HL und bleiben damit immer oberhalb oder auf der roten Linie. Damit ist der Nachweis erbracht, dass alle Probanden das Kriterium der Normalhörigkeit erfüllen. (Werte unterhalb der roten Linie führten zum Ausschluss aus der Studie.)

Unterschiede der Hörleistung in Abhängigkeit von der Testsitzung bestehen nicht.

Eine Varianzanalyse, die den Zusammenhang zwischen den Resultaten der Tonaudiometrie und den im Oldenburger Satztest erzielten SRT-Werten untersucht, erwies sich als nicht signifikant.

3 Ergebnisse

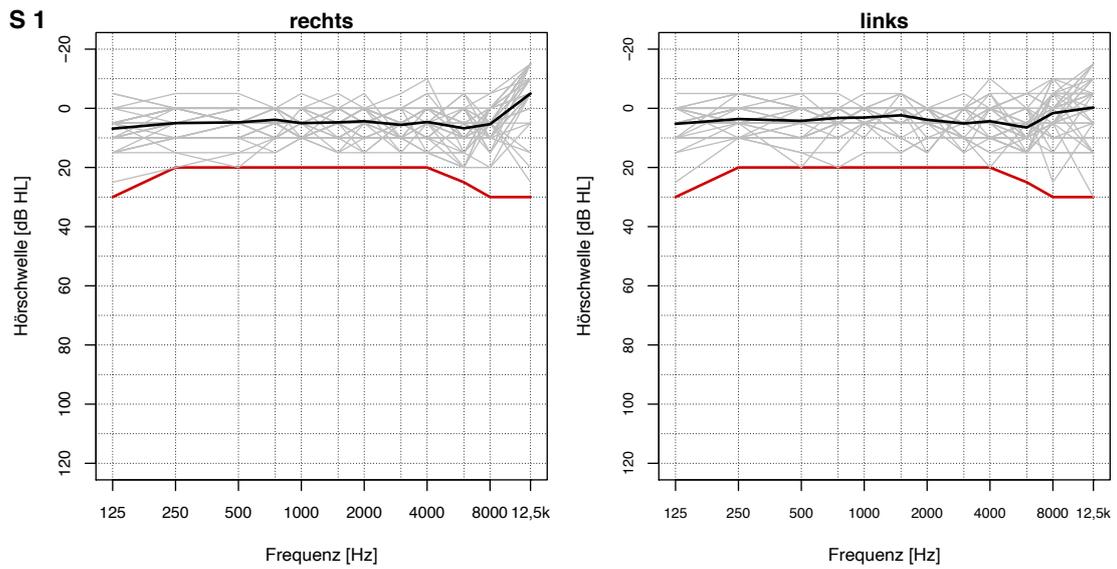


Abbildung 3.1: Audiogramme der (selektierten) Probanden der ersten Sitzung

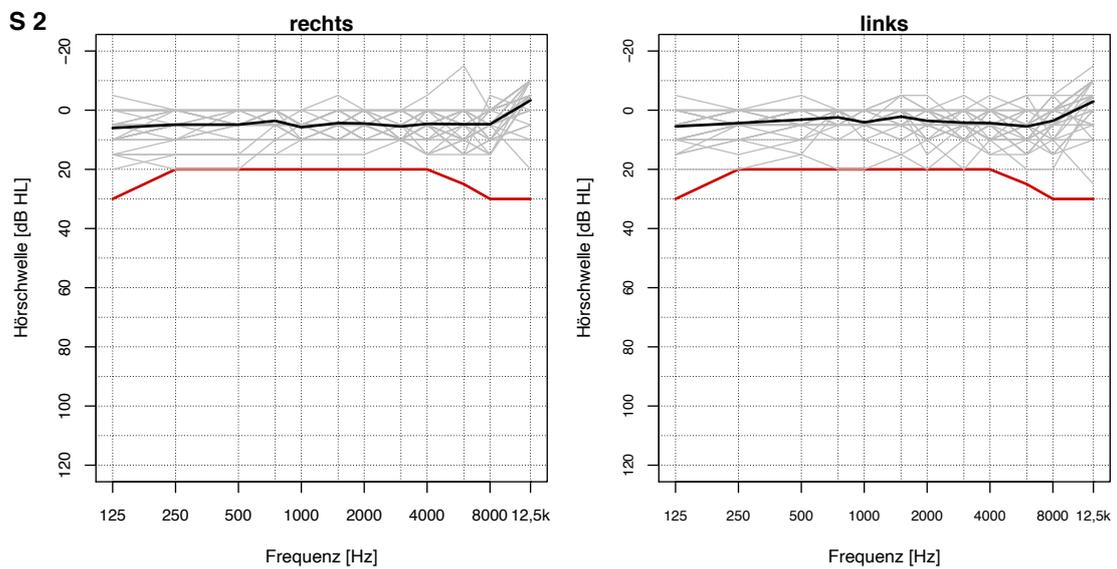


Abbildung 3.2: Audiogramme der (selektierten) Probanden der zweiten Sitzung

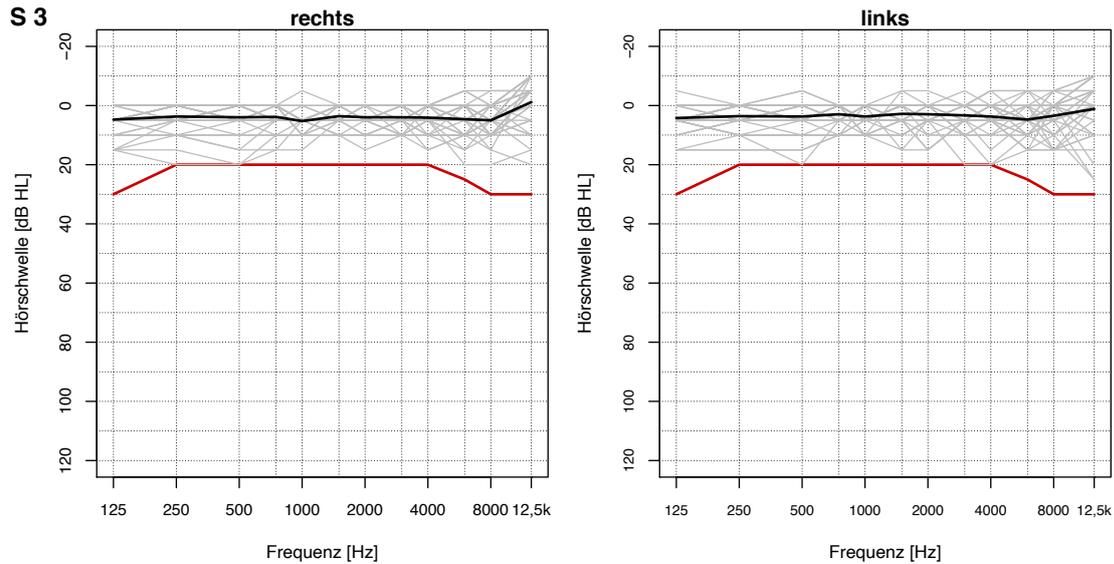


Abbildung 3.3: Audiogramme der (selektierten) Probanden der dritten Sitzung

3.2 Bestimmung der selektierten Hauptgruppe der Probanden

Insgesamt wurden 41 Probanden rekrutiert. 3 davon wurden aus der Studie ausgeschlossen: Ein Proband (ohne Nummer) verfehlte die tolerierte Hörschwelle. Ein Proband (Nr. 31) absolvierte lediglich die erste der drei Testsitzungen. Ein Proband (Nr. 34) spricht Spanisch als Muttersprache und wurde nur aus Eigeninteresse getestet. Alle anderen 38 Testpersonen kamen als potentiell geeignete Kandidaten für die selektierte Hauptgruppe in Frage.

Aus Gründen der praktischen Anwendbarkeit wurden aus dieser Gruppe der 38 Probanden diejenigen 32 ausgewählt, bei welchen der Abstand zwischen den Testsitzungen Nummer 1 und 2 bzw. Nummer 2 und 3 idealerweise (möglichst genau) 50 Tage betrug. Das Prinzip der Selektion und die daraus resultierende, selektierte Hauptgruppe von 32 Probanden ist in Abbildung 3.4 auf Seite 42 dargestellt. In der Auswertung berücksichtigte Probanden sind als schwarze Punkte markiert und stellen die selektierte Hauptgruppe dar. Probanden, die aufgrund des Kriteriums des zeitlichen Abstands herausfallen sind als weiße Punkte abgebildet. In dieser Abbildung (3.4) ist Proband Nr. 31 wegen der Unvollständigkeit der Testsitzungen nicht dargestellt und Proband Nr. 34 wegen seiner Muttersprache nicht berücksichtigt worden.

3 Ergebnisse

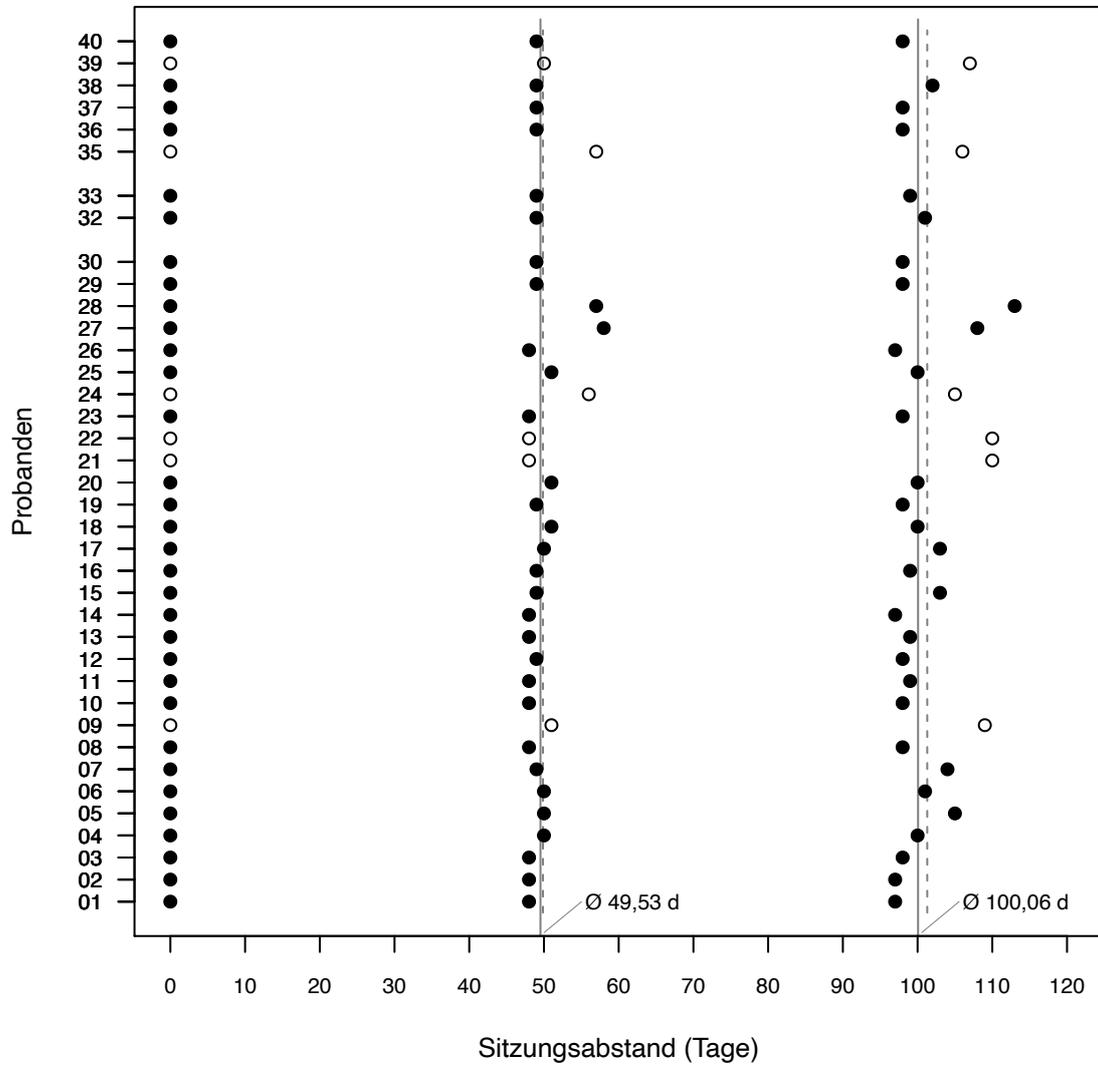


Abbildung 3.4: Liste der selektierten Probanden, eingeteilt aufgrund der Zeitpunkte der Testsitzungen

3.3 Einflussfaktoren auf das Sprachverständnis

Das Sprachverständnis ist mit seinen potentiellen Einflussfaktoren zentrales Element dieser Dissertation. Die Kalkulation des Sprachverständnisses erfolgt beim Oldenburger Satztest mittels der Sprachverständlichkeitsschwelle, der *speech reception threshold* (SRT). Der SRT-Wert berechnet sich dabei wie in 2.5.3 beschrieben.

Darstellung der SRT-Verlaufsdaten

Die individuellen SRT-Verlaufskurven der Hauptgruppe können im Anhang A (Rohdaten) ab Seite 107 eingesehen werden.

Die Abbildungen 3.5 und 3.6 zeigen die SRT-Verlaufsdaten der einzelnen Testsitzungen und Testdurchläufe derjenigen Probanden, welche der selektierten Hauptgruppe zugeteilt wurden. Abbildung 3.5 zeigt die SRT-Werte der männlichen Probanden, Abbildung 3.6 die der weiblichen Probanden.

Jeweils in der linken Spalte sind die acht ausgewählten Probanden vertreten, denen die Vorlagenkarte (Karte mit allen vorkommenden Wörtern) zu keinem Zeitpunkt gezeigt wurde. In der mittleren Spalte sind die vier Probanden vertreten, denen die Vorlagenkarte zu Beginn des Tests mit der Anweisung diese zeilenweise (also in Form von zehn Fünf-Wort-Sätzen) zu lesen gezeigt wurde. In der rechten Spalte sind diejenigen vier Probanden vertreten, denen die Vorlagenkarte zu Beginn des Tests mit der Anweisung diese spaltenweise (also fünf Wortgruppen-Spalten) zu lesen gezeigt wurde.

In jeder Kachel des Diagramms sind die Resultate eines Probanden aufgezeichnet. Auf der Abszisse sind die Testdurchläufe innerhalb einer Sitzung (1-5) aufgetragen. Die Ordinate zeigt den SRT-Wert. Die erste Testsitzung ist durch hellgraue Punkte und Verbindungslinien hervorgehoben, die zweite durch dunkelgraue und die dritte durch schwarze Punkte.

Diese beiden Abbildungen (3.5 und 3.6) ergeben den Ausgangspunkt für die weitere Auswertung. In den drei folgenden Abbildungen werden die Resultate weiter zusammengefasst, wobei die Unterteilung in männlich und weiblich außer Acht gelassen wird.

In Abbildung 3.7 sind die SRT-Verlaufskurven der männlichen und weiblichen Probanden unter Erhalt der Einteilung der Gruppen in „ohne Karte“ und „mit Karte zeilenweise“ und „mit Karte spaltenweise“ dargestellt.

In Abbildung 3.8 wird nur noch nach „ohne Karte“ und „mit Karte“ unterschieden; die Unterteilung nach der Leserichtung wird nicht mehr berücksichtigt.

Als zentrales Resultat stellt Abbildung 3.9 eine Zusammenfassung aller SRT - Ver-

3 Ergebnisse

laufdaten dar. Die Unterscheidung nach der Vorlage der Karte wird jetzt ebenfalls nicht mehr berücksichtigt.

In den beiden Abbildungen 3.5 und 3.6 kann trotz kleinerer Variationen auf den ersten Blick in jeder der drei Linien ein Abwärtstrend ausgemacht werden. Diese Tendenz unterliegt teils mehr oder weniger starken individuellen Schwankungen. Im Gesamtkontext betrachtet zeigt sich hier der erste Anhaltspunkt auf einen Lerneffekt bei mehrfacher Testwiederholung. Beim Vergleich von Männern und Frauen zeigte sich kein offensichtlicher Unterschied. Diese Vermutung bestätigt sich in den Varianzanalysen.

Die niedrigsten SRT-Schallpegel erreichten die meisten Männer im vierten oder fünften Testdurchlauf der dritten Sitzung: 8 von 16 in Sitzung 3 - Durchgang 5 und 3 von 16 in Sitzung 3 - Durchgang 4. Somit lieferten 11 von 16 Männern (68,75%) in den letzten beiden Testdurchläufen ihr bestes Ergebnis ab. 4 von 16 (25,00%) waren im zweiten Testdurchlauf und 1 von 16 (6,25%) im ersten am besten.

Die meisten weiblichen Probanden erreichten ihre niedrigsten SRT-Schallpegel im dritten, vierten oder fünften Testdurchlauf der dritten Sitzung: 7 von 16 in Sitzung 3 - Durchgang 5, 2 von 16 in Sitzung 3 - Durchgang 4 und 4 von 16 in Sitzung 3 - Durchgang 3. Somit lieferten 13 von 16 Frauen (81,25%) in den letzten drei Testdurchläufen ihr bestes Ergebnis ab; 9 von 16 Frauen (56,25%) waren in den letzten beiden Testdurchläufen am besten. 14 von 16 (87,50%) waren insgesamt im dritten Testdurchlauf und 2 von 16 (12,50%) waren im zweiten am besten.

Zusammenfassend verteilen sich unter Nichtbeachtung der Geschlechtertrennung die niedrigsten SRT-Werte auf folgende Testdurchläufe: 1 von 32 Probanden (3,125%) erreichte seinen niedrigsten SRT-Wert in der ersten Testsitzung. 6 von 32 Probanden (18,75%) erreichten ihren niedrigsten SRT-Wert in der zweiten Testsitzung. 25 von 32 Probanden (78,125%) erreichten ihren niedrigsten SRT-Wert in der dritten Testsitzung. 15 von 32 Probanden (46,875%) erreichten ihren niedrigsten SRT-Wert in Sitzung 3 - Durchgang 5. 20 von 32 Probanden (62,5%) erreichten ihren niedrigsten SRT-Wert in Sitzung 3 - Durchgang 4 oder 5. 24 von 32 Probanden (75%) erreichten ihren niedrigsten SRT-Wert in Sitzung 3 - Durchgang 3, 4 oder 5. Der niedrigste SRT-Wert wurde nie in einem ersten Durchlauf erreicht (Sitzung 1 - Durchgang 1, Sitzung 2 - Durchgang 1, Sitzung 3 - Durchgang 1).

In den beiden Zusammenfassungen, in denen das Geschlecht nicht mehr, aber die Vorlage der Karte noch weiter berücksichtigt wird (3.7 und 3.8) können kontinuierlich abfallende SRT-Verlaufswerte beobachtet werden. Nur in der Gruppe „mit Karte spaltenweise“ ändert der Kurvenverlauf in jeder der drei Kurven zweimal die Richtung (ein Wendepunkt). In Abbildung 3.8 existieren keine Wendepunkte mehr. Die drei Kurven

überschneiden sich in keiner der Grafiken. Die niedrigsten erreichten SRT-Werte werden durchschnittlich in letzten Durchgang der letzten Sitzung (Sitzung 3 - Durchgang 5) erreicht. Dabei ist es irrelevant, ob und in welcher Form die Karte verwendet wird. Die Steilheit des Abfalls der SRT-Verlaufskurven wird sichtbar flacher und die Kurven rücken näher zusammen. (Der mittlere Abstand der Kurven ist zwischen der schwarzen und dunkelgrauen Kurve enger als zwischen der hellgrauen und dunkelgrauen; Ausnahme: „mit Karte spaltenweise“.) Im Laufe der 5 Testdurchläufe (in jeder Testsitzung) ist auch im Verlauf der letzten Testsitzung keine Sättigung der Verbesserung bzw. eine asymptotische Annäherung an einen Wert erkennbar.

Absolutgrößen des Lerneffekts: In der Zusammenfassung aller selektierten Probanden 3.9 zeigen sich sämtliche, in den vorherigen Abbildungen ansatzweise ersichtliche Effekte am deutlichsten: Die drei SRT-Verlaufskurven verlaufen kontinuierlich, aber mit unterschiedlichem Stärkegrad abwärts (unterschiedliche, aber immer negative Steigungen), wobei die Werte der ersten Testsitzung (über die ersten fünf Durchläufe) insgesamt einen steileren Verlauf nehmen; die Werte in der zweiten und dritten Testsitzung (über die zweiten und dritten fünf Durchläufe) laufen zwar mit kleineren negativen Steigungen, aber dennoch kontinuierlich abwärts. Es existieren keine Wendepunkte. Im Laufe der 5 Testdurchläufe und in jeder Testsitzung ist auch hier keine Stagnation der Verbesserung erkennbar. Die Werte nähern sich nur bedingt einem Minimalwert an. Auch hier werden die niedrigsten SRT-Werte im Durchschnitt im letzten Durchgang der letzten Sitzung (Sitzung 3 - Durchgang 5) erreicht. Die abfallenden SRT-Werte stellen eine Verbesserung des Hörverständnisses dar. Im Einzelnen belaufen sich die errechneten SRT-Werte und Differenzen auf folgende Größen: Die SRT-Werte des ersten Durchgangs (hellgrau) fallen von -5,07 auf -7,68 dB SNR (Verbesserung um 2,61 dB). Die SRT-Werte des zweiten Durchgangs (dunkelgrau) fallen von -6,70 auf -8,47 dB SNR (Verbesserung um 1,77 dB). Die SRT-Werte des dritten Durchgangs (schwarz) fallen von -7,56 auf -9,21 dB SNR (Verbesserung um 1,65 dB). Der absolute Lerneffekt von Beginn der 1. Sitzung bis zum Ende der 3. Sitzung beläuft sich auf 4,13 dB. Bei dem ersten Testdurchlauf der Testsitzungen 2 und 3 war im Vergleich zu dem letzten Testdurchlauf der 1. und 2. Sitzung wieder eine Erhöhung des Schalldruckpegels um 0,98 bzw. 0,91 dB SPL nötig.

3 Ergebnisse

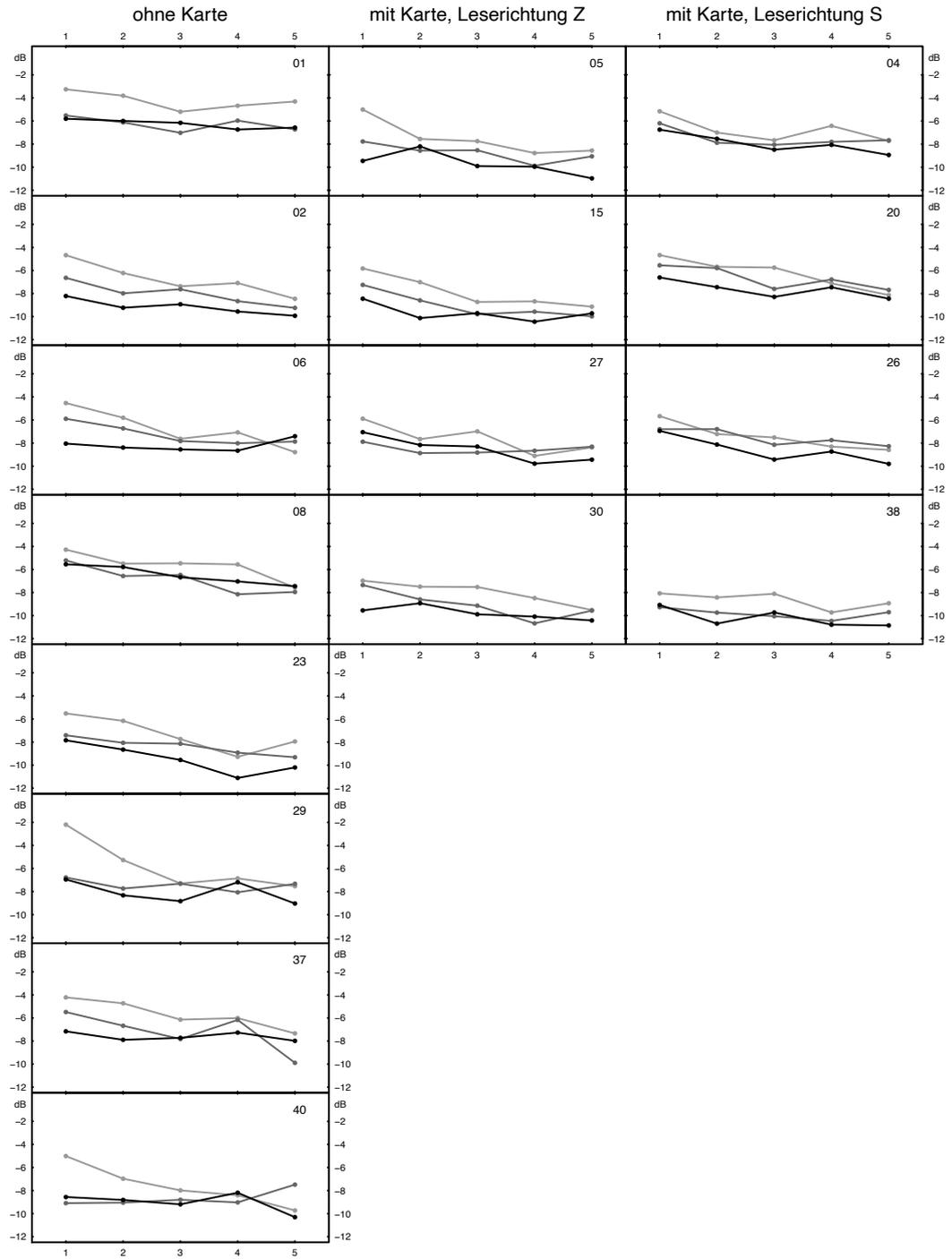


Abbildung 3.5: SRT-Verlaufswerte der selektierten männlichen Probanden

3.3 Einflussfaktoren auf das Sprachverständnis

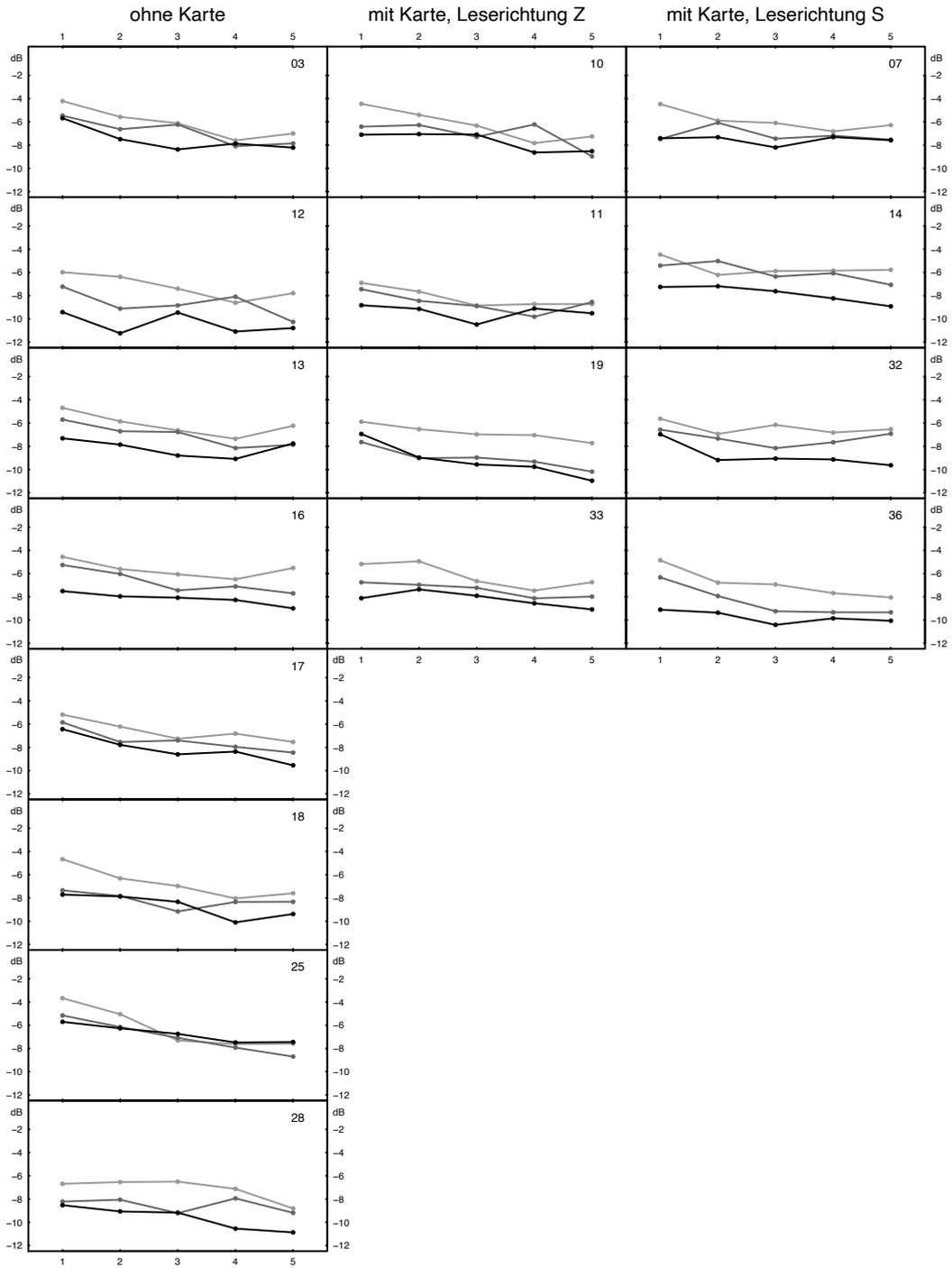


Abbildung 3.6: SRT-Verlaufswerte der selektierten weiblichen Probanden

3 Ergebnisse

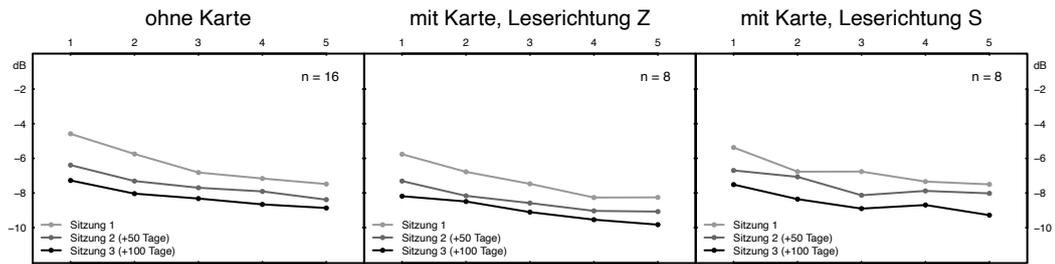


Abbildung 3.7: SRT-Verlaufswerte der selektierten Probanden bei Erhalt der 3 Gruppen

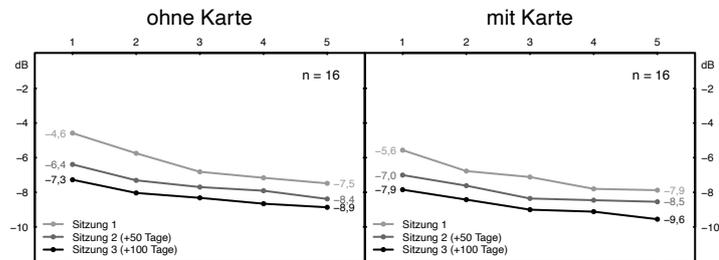


Abbildung 3.8: SRT-Verlaufswerte der selektierten Probanden bei Erstellung von 2 Gruppen

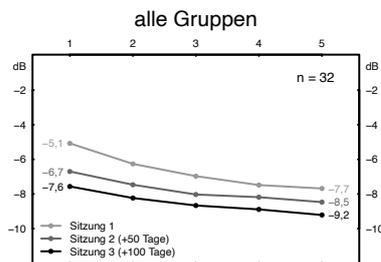


Abbildung 3.9: SRT-Verlaufswerte der selektierten Probanden bei Zusammenfassung der 3 Gruppen

Tabelle 3.1: ANOVA-Ergebnistabelle: Abhängige Variable: SRT (in dB SNR), Faktoren: Geschlecht, Vorlagenkarte, Testsitzung, Testlauf und Interaktionen (selektierte Hauptgruppe, $n = 32$). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Geschlecht (G)	1	7,6000	0,6600	0,4248	0,0152
Karte (C)	1	37,8500	3,2700	0,0815	0,0712
Sitzung (S)	2	267,3600	189,4200	< 0,001	0,3514
Testlauf (R)	4	244,4900	138,9200	< 0,001	0,3313
G × C	1	27,3400	2,3600	0,1358	0,0525
G × S	2	2,9700	2,1100	0,1313	0,0060
C × S	2	0,8900	0,6300	0,5366	0,0018
G × R	4	0,6900	0,3900	0,8157	0,0014
C × R	4	1,1600	0,6600	0,6202	0,0024
S × R	8	14,0000	4,8900	< 0,001	0,0276
G × C × S	2	1,1300	0,8000	0,4539	0,0023
G × C × R	4	1,8500	1,0500	0,3836	0,0037
G × S × R	8	5,0600	1,7700	0,0844	0,0102
C × S × R	8	4,3800	1,5300	0,1484	0,0088
G × C × S × R	8	1,5400	0,5400	0,8275	0,0031

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,014$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 4 von 60

Varianzhomogenität (Levene-Test): $p = 0,1809$

Resultate der Varianzanalysen

Die Varianzanalyse geht der zentralen Frage nach der Signifikanz der untersuchten Einflussfaktoren nach.

Tab. 3.1 zeigt die Resultate der vier untersuchten Einflussfaktoren (Geschlecht, Karte, Sitzung, Testlauf) sowie deren Interaktionen untereinander. In der Varianzanalyse werden die Freiheitsgrade, Quadratsummen, F-Werte, p-Werte und das deskriptive Maßes für die Varianzaufklärung bestimmt. Hauptbeachtung findet dabei der p-Wert.

Im Vorfeld zu dieser Varianzanalyse (3.1) wurden die Daten mithilfe einer ersten Varianzanalyse untersucht. In dieser ersten Varianzanalyse mit $n = 16$ (B.1 auf Seite 160) wurden die vier Faktoren Geschlecht, Leserichtung, Testsitzung und Testdurchgang be-

3 Ergebnisse

trachtet. Da sich die Leserichtung in der Untergruppe als nicht signifikant erwies, gilt der oben beschriebenen, zusammenfassenden Varianzanalyse mit $n = 32$ (3.1 auf Seite 49) das Hauptaugenmerk.

Alle weiteren Varianzanalysen finden sich aus Gründen der besseren Übersicht im Anhang.

Die beiden Voraussetzungen für die Anwendbarkeit einer Varianzanalyse sind die Normalverteilung und die Varianzhomogenität der zum Vergleich ausgewählten Daten. Wie in 2.6 erörtert, wurde in dieser Arbeit mit der in medizinischen Studien gebräuchlichen Signifikanzschwelle von 5 % gearbeitet. Die Normalverteilung wurde mit dem Shapiro-Wilk-Test geprüft. Bei einem $p > 0,05$ können die Daten als normalverteilt erachtet werden. Unterhalb der Tabelle der Varianzanalyse sind der kleinste, errechnete p-Wert der Shapiro-Wilk-Tests sowie die Anzahl der nicht normalverteilten Teilmengen dargestellt. Diese liegen mit 4 von 60 in einem akzeptablen Rahmen, da die Varianzanalyse bei Abweichungen von der Normalverteilung bei einer großen Anzahl an Stichproben weiterhin robust bleibt. Die Testung der Varianzhomogenität erfolgte mit dem Levene-Test. Mit 0,1809 liegt der p-Wert über 0,05 und erfüllt damit das Kriterium der Varianzhomogenität.

Die p-Werte sind für die Faktoren Sitzung, Testlauf und deren Interaktion kleiner als 0,05 und somit signifikant. Das deskriptive Maß für die Varianzaufklärung beläuft sich bei den signifikanten Einflussfaktoren auf 0,7103. Damit kann mit 71,03% ein sehr großer Anteil des Effekts erklärt werden. Die Einflussfaktoren Geschlecht, Karte sowie sämtliche Interaktionen, in denen diese beiden Faktoren mitspielen, stellen sich als nicht signifikant dar. Auch alle sonstigen Interaktionen sind mit p-Werten größer 0,05 nicht signifikant. Die Voraussetzungen zur Durchführung einer Varianzanalyse wurden mit dem Shapiro-Wilk-Test und dem Levene-Test geprüft (vgl. 2.6).

Da sich in diesen beiden Varianzanalysen nur die Faktoren Testsitzung und Testdurchgang als signifikant erwiesen, wurden nur mit diesen beiden Größen zwei weitere, korrigierte Varianzanalysen gerechnet.

Der Unterschied zu den vorherigen Analysen besteht darin, dass die beiden Zeitfaktoren Testsitzung und Testdurchgang, die sogenannten *within-Faktoren* zusammengefasst wurden. Diese neuen Varianzanalysen bestätigen die vorherigen Resultate. Alle Daten entstammen dabei der selektierten Hauptgruppe.

Tabelle 3.2: Post-hoc-Vergleich (Tukey HSD) der Testsitzungen. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	< 0,001	< 0,001
2		< 0,001

Tabelle 3.3: Post-hoc-Vergleich (Tukey HSD) der Testdurchläufe. Grau hinterlegt: $p < 0,05$

Testlauf	2	3	4	5
1	< 0,001	< 0,001	< 0,001	< 0,001
2		0,00461	< 0,001	< 0,001
3			0,34819	0,00463
4				0,46426

Resultate der Post-hoc-Vergleiche

Als post-hoc Testverfahren, welches der genaueren Untersuchung der signifikanten Einflussfaktoren dient, wurde der Tukey HSD Test (Tukey's Honest Significant Difference Test) ausgewählt. Die Voraussetzungen für die Anwendbarkeit dieser Testmethode sind mit Varianzhomogenität und gleicher Gruppengröße gegeben.

Die beiden zusammenfassenden Post-hoc-Vergleiche 3.2 und 3.3 (Seite 51 und 51) untersuchen den Einfluss der signifikanten Faktoren Testsitzungen und Testdurchläufe. Grau hinterlegte Werte zeigen dabei signifikante Werte ($p < 0,05$).

Die drei Testsitzungen unterscheiden sich signifikant voneinander ($p < 0,001$). Bei genauerer Betrachtung der Signifikanz der einzelnen Testdurchläufe zueinander zeigt sich $p < 0,05$ bei dem Vergleich des ersten und zweiten Durchgangs mit allen anderen sowie des dritten mit dem fünften Durchgang. Alle p-Werte sind dabei $< 0,005$. Die Durchläufe drei und vier sowie vier und fünf unterscheiden sich nicht signifikant voneinander. Dieser Sachverhalt wird in den Überlegungen zur Eradikation des Lerneffekts 4.4 aufgegriffen.

Die Ergebnisse aller weiteren nachgeschalteten Post-hoc-Vergleiche (B.8, B.9, B.10, B.11, B.12, B.13, B.14 und B.15) sind im Anhang zu sehen.

3 Ergebnisse

Tabelle 3.4: SRT-Mittelwerte (mit Standardabweichung) bei Zusammenfassung aller Gruppen (N=32) (Werte in dB SNR)

Sitzung/ Durchgang	1	2	3	4	5
I	-5,07 ($\pm 1,14$)	-6,26 ($\pm 1,00$)	-6,97 ($\pm 0,90$)	-7,48 ($\pm 1,15$)	-7,68 ($\pm 1,20$)
II	-6,70 ($\pm 1,11$)	-7,46 ($\pm 1,17$)	-8,03 ($\pm 1,01$)	-8,18 ($\pm 1,21$)	-8,47 ($\pm 0,99$)
III	-7,56 ($\pm 1,14$)	-8,23 ($\pm 1,22$)	-8,66 ($\pm 1,07$)	-8,89 ($\pm 1,26$)	-9,21 ($\pm 1,21$)

Tabelle 3.5: SRT-Mittelwerte (mit Standardabweichung) bei der Gruppen OHNE Karte (N=16) (Werte in dB SNR)

Sitzung/ Durchgang	1	2	3	4	5
I	-4,58 ($\pm 1,05$)	-5,75 ($\pm 0,78$)	-6,82 ($\pm 0,83$)	-7,16 ($\pm 1,16$)	-7,48 ($\pm 1,30$)
II	-6,39 ($\pm 1,19$)	-7,31 ($\pm 1,00$)	-7,70 ($\pm 0,92$)	-7,91 ($\pm 0,84$)	-8,39 ($\pm 0,98$)
III	-7,28 ($\pm 1,18$)	-8,04 ($\pm 1,34$)	-8,32 ($\pm 1,02$)	-8,66 ($\pm 1,45$)	-8,87 ($\pm 1,35$)

Lerneffekt und Signifikanz der Testdurchläufe zueinander

Nachfolgend soll zunächst der Lerneffekt im Laufe der Testungen und deren Signifikanz zueinander genauer betrachtet werden. Die drei wichtigsten Abbildungen 3.7, 3.8 und 3.9 sind hier in drei Tabellen (3.4, 3.5 und 3.6) zusammengefasst. In drei weiteren Tabellen (3.7, 3.8 und 3.9) wird der Lerneffekt von jedem Durchgang (zu je 30 Sätzen) zu jedem darauffolgenden Durchgang numerisch dargestellt. Am Ende jeder Tabellenzeile findet sich der innerhalb der jeweiligen Sitzung erzielte Lerneffekt. Tabelle 3.10 zeigt die Signifikanz der Durchgänge zueinander. Zwei sich signifikant voneinander unterscheidende Durchgänge sind mit einem roten „S“ markiert. (Beispiele: Der erste Durchgang der ersten Sitzung (01) unterscheidet sich signifikant von allen folgenden Durchgängen. Der erste Durchgang der zweiten Sitzung (06) unterscheidet sich signifikant von 08, 09, 10, 12, 13, 14, 15. Der erste Durchgang der dritten Sitzung (11) unterscheidet sich signifikant von 13, 14, 15.) Diese Tabelle 3.10 zeigt, dass theoretisch vor jeder Sitzung zwei 30er Listen zu Übungszwecken präsentiert werden müssen, um valide Daten bei den eigentlichen Messungen zu erhalten. Unter 4.4 soll der hier erzielte Lerneffekt mit anderen Arbeiten verglichen werden.

Die Übersicht 3.11 (Seite 55) fasst die Vorgehensweise und Resultate bezüglich der Varianzanalysen und Post-hoc-Vergleiche zusammen.

Zusammenfassend kann nach Betrachtung aller bisherigen Resultate beobachtet werden, dass bei mehrfacher Wiederholung des Oldenburger Satztests zwischen allen drei Testsitzungen und vom ersten bis zu dritten Testdurchlauf ein signifikanter Ab-

Tabelle 3.6: SRT-Mittelwerte (mit Standardabweichung) bei der Gruppe MIT Karte (N=16) (Werte in dB SNR)

Sitzung/ Durchgang	1	2	3	4	5
I	-5,57 ($\pm 1,03$)	-6,77 ($\pm 0,94$)	-7,12 ($\pm 0,96$)	-7,80 ($\pm 1,08$)	-7,88 ($\pm 1,09$)
II	-7,01 ($\pm 0,96$)	-7,62 ($\pm 1,35$)	-8,36 ($\pm 1,01$)	-8,46 ($\pm 1,47$)	-8,55 ($\pm 1,03$)
III	-7,85 ($\pm 1,06$)	-8,42 ($\pm 1,10$)	-9,00 ($\pm 1,03$)	-9,12 ($\pm 1,04$)	-9,55 ($\pm 0,97$)

Tabelle 3.7: Lerneffektberechnung aus den SRT-Mittelwerten bei Zusammenfassung aller Gruppen (N=32) (Werte in dB SNR)

Sitzung/ Durchgang	Δ 1./2.	Δ 2./3.	Δ 3./4.	Δ 4./5.	Δ 1./5. bzw. Δ gesamt
I	-1,19	-0,71	-0,51	-0,20	-2,61
II	-0,76	-0,57	-0,15	-0,29	-1,77
III	-0,67	-0,43	-0,23	-0,32	-1,65
					-4,13

Tabelle 3.8: Lerneffektberechnung aus den SRT-Mittelwerten in der Gruppe OHNE Karte (N=16) (Werte in dB SNR)

Sitzung/ Durchgang	Δ 1./2.	Δ 2./3.	Δ 3./4.	Δ 4./5.	Δ 1./5. bzw. Δ gesamt
I	-1,17	-1,07	-0,34	-0,32	-2,90
II	-0,92	-0,39	-0,21	-0,48	-2,00
III	-0,76	-0,28	-0,34	-0,21	-1,59
					-4,29

Tabelle 3.9: Lerneffektberechnung aus den SRT-Mittelwerten in der Gruppe MIT Karte (N=16) (Werte in dB SNR)

Sitzung/ Durchgang	Δ 1./2.	Δ 2./3.	Δ 3./4.	Δ 4./5.	Δ 1./5. bzw. Δ gesamt
I	-1,20	-0,35	-0,68	-0,08	-2,31
II	-0,61	-0,74	-0,10	-0,09	-1,54
III	-0,57	-0,58	-0,12	-0,43	-1,70
					-3,98

Tabelle 3.10: Signifikanz der Durchgänge zueinander (aus Post-hoc-Analyse)

Durchgang / Durchgang	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
01 1. Sitzung	■	S	S	S	S	S	S	S	S	S	S	S	S	S	S
02		■		S	S		S	S	S	S	S	S	S	S	S
03			■					S	S	S		S	S	S	S
04				■						S			S	S	S
05					■	S							S	S	S
06 2. Sitzung						■		S	S	S		S	S	S	S
07							■			S			S	S	S
08								■							S
09									■						S
10										■					
11 3. Sitzung											■		S	S	S
12												■			S
13													■		
14														■	
15															■

S Signifikanter Unterschied (Anm. : kein Unterschied bzgl. der Anw. GCSR bzw. GCT)

Tabelle 3.11: Praktisches Vorgehen bei der Varianzanalyse

Potentielle Einflussfaktoren (auf den SRT -Wert) G ender C ard D irection S ession R un T ime (global) <small>Time (global) = Zusammenfassung von Session u. Run</small>	GCSR (n=32)	GDSR (n=16)	GCT (n=32)	GDT (n=16)
Reduktion auf signifikante Einflussfaktoren <small>1. Elimination D (nicht signifikant) 2. Elimination G u. C (nicht signifikant)</small>				
Berücksichtigung nur signifikanter Einflussfaktoren	SR	SR	T	T
Post-hoc-Vergleich paarweise Mittelwertvergleiche der signifikanten Einflussfaktoren Procedere: (Anwendung des Tukey HSD der Gruppe GCSR) 1. Post-hoc Vergleich der Testsitzungen zueinander (Session) 2. Post-hoc Vergleich der Testdurchläufe zueinander (Run) 3. Post-hoc Vergleich aller Testsitzungen und Testdurchläufe zueinander (manuelle Grafik)	Tukey HSD	Tukey HSD	Tukey HSD	Tukey HSD

wärtstrend des SRT-Werts auftritt. Diese Verbesserung kann als Lerneffekt verstanden werden. Der Lerneffekt ist dabei von den zeitabhängigen Faktoren Testsitzung und Testdurchlauf abhängig, jedoch unabhängig vom Geschlecht und offenem oder geschlossenem Testdesign. Er tritt kontinuierlich in jeder Sitzung und zwischen dem ersten bis zum dritten Durchgang auf. Zwischen dem dritten und vierten sowie dem vierten und fünften Durchgang findet kein signifikanter Abwärtstrend mehr statt. Es kann abgewogen werden, ob die fehlende Signifikanz der Resultate im Verlauf einer tatsächlichen Lernsättigung oder einer vom Testleiter beobachteten Ermüdung der Probanden mit einem Nachlassen der Aufmerksamkeit zuzuschreiben ist.

3.4 Darstellung der Absolutwerte der Schalldruckpegel des Sprachsignals

Die drei Grafiken 3.10, 3.11 und 3.12 zeigen die Absolutwerte der Schalldruckpegel des Sprachsignals in der jeweiligen Sitzung. Jede der drei Grafiken besteht aus 2x5 Kacheln. In der oberen Zeile wurden die Werte derjenigen Probanden erfasst, welchen die Vorlagenkarte vorenthalten wurde. In der unteren Zeile sind die Werte derjenigen Probanden festgehalten, welchen die Vorlagenkarte mit der Instruktion diese spalten- bzw. zeilenweise zu lesen dargeboten wurde. Die Darstellung jeder Kachel zeigt: Im oberen Teil sind auf der Abszisse die 30 Sätze eines jeden Durchgang dargestellt. Auf der Ordinate sind die Schalldruckpegel der dargebotenen Sätze (in dB SPL) abgebildet. Jede Kachel repräsentiert also einen aus 30 aufgesprochenen Sätzen bestehenden Testdurchlauf. Die schwarzen Punkte mit den zugehörigen Balken der Standardabweichung sind für jeden Satz auf Höhe des entsprechenden Schalldruckpegels eingezeichnet. Der erste Punkt beginnt immer bei einem Schalldruckpegel von 65 dB SPL, da der erste Satz immer in dergleichen Lautstärke wie das Störgeräusch präsentiert wird. Der 31. grau markierte Wert stellt den errechneten Schalldruckpegel des 31. Satzes dar, welcher aber nicht mehr vorgespielt wird. In allen Kurven ist zwischen dem ersten und zweiten dargebotenen Satz eine absteigende Richtung ersichtlich. Auch im weiteren Verlauf erreicht das Sprachsignal nie Werte von 65 dB SPL oder höher. Weiterhin ist in allen Kurven erst ein Absteigen der Werte und anschließend ein Einpendeln im Rahmen von Schalldruckpegeln um den grauen errechneten Wert zu beobachten. Dabei fällt die Ausgangsmittelwertkurve in Testsitzung 2 stärker als in 1, in Testsitzung 3 stärker als in 2. Das Einpendeln geschieht insgesamt grob nach etwa 3-15 Sätzen. Zu Beginn einer Sitzung (insbesondere der ersten Sitzung) benötigen die Probanden mehr Sät-

3.5 Korrelation der Testsitzungen und -durchläufe hinsichtlich des SRT-Werts

ze, um im weiteren Verlauf Werte ohne größere Schwankungen zu erhalten. Bei den folgenden Durchgängen pendelt sich bei den meisten Probanden der Wert ziemlich schnell auf konstante Werte ein. Schon innerhalb der ersten 3 bis 5 Sätze bleiben die einzelnen Messwerte relativ gleich und liegen beinahe auf einer Horizontalen. Im unteren Teil der Kacheln ist die Anzahl der Wendepunkte (zugeordnet zu dem jeweiligen Satz auf der Abszisse) aufgetragen. Wendepunkte entstehen, wenn ein Proband bei einem Satz viel verstanden hat (3-5 Wörter), woraufhin der nächste Satz leiser präsentiert wird; jetzt versteht der Proband allerdings recht schlecht (0-2 Wörter), was dazu führt, dass der folgende Satz wieder lauter abgespielt wird (v-förmiger Wendepunkt). Ein Wendepunkt kann auch im umgekehrten Fall entstehen: Wenn bei einem Satz wenig verstanden wird (0-2 Wörter), wird der nächste Satz lauter präsentiert; dieser wird recht gut verstanden (3-5 Wörter); daraufhin wird der folgende Satz wieder leiser abgespielt (bogen-förmiger Wendepunkt). Die Anzahl der Wendepunkte geht exponentiell in die Formel der SRT-Wert-Berechnung ein und nimmt so Einfluss auf die Resultate eines Durchgangs (2.5.3).

Insgesamt sei an dieser Stelle darauf hingewiesen, dass die Abbildungen aus Mittelwerten bestehen und dadurch die Kurven modifiziert und geebnet werden, was auch die Interpretation limitieren könnte. Individuelle Schwankungen der Werte mit Ausreißern in beide Richtungen können in den Übersichtstabellen der einzelnen Probanden im Anhang betrachtet werden.

Die individuellen, absoluten Schalldruckpegel des Satzes variierten zwischen 65 und etwa 50 bis 55 dB SPL. In seltenen Fällen, in denen der Proband beim ersten Satz einer Liste lediglich zwei oder weniger Wörter korrekt verstand, erhöhte sich der Schalldruckpegel auf Werte knapp unter 70 dB SPL.

3.5 Korrelation der Testsitzungen und -durchläufe hinsichtlich des SRT-Werts

Die beiden Grafiken 3.13 und 3.14 beschreiben die Korrelation der drei Testsitzungen hinsichtlich des SRT-Werts. Die beiden Abbildungen 3.15 und 3.16 zeigen die Korrelation der fünf Testdurchläufe hinsichtlich des SRT-Werts. Es wird untersucht, in welcher Stärke die SRT-Werte der einzelnen Testsitzungen und -durchläufe in Wechselwirkung zueinander stehen, also ob und wie beispielsweise mittels des SRT-Werts der ersten Sitzung schon vorausgesagt werden kann, wie das Resultat in der zweiten und dritten Sitzung ausfällt.

3 Ergebnisse

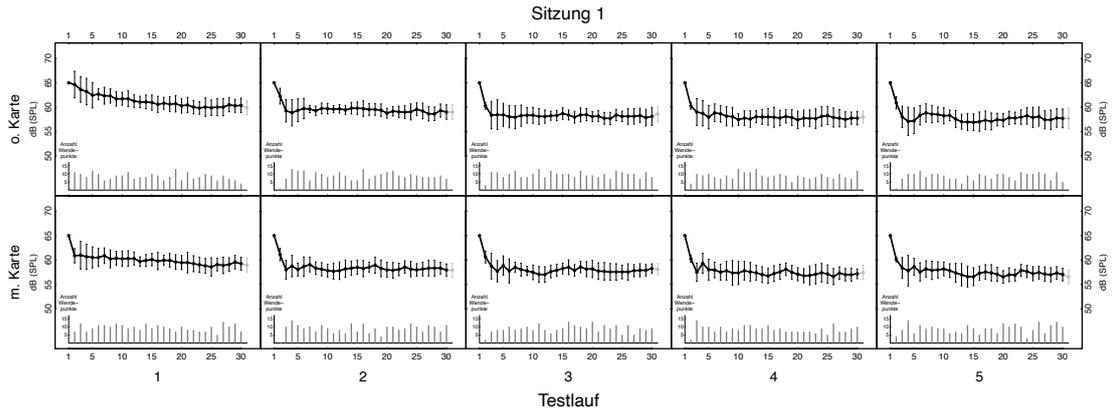


Abbildung 3.10: Mittelwertverlaufskurve der Absolutwerte der Schalldruckpegel des Sprachsignals der Testsitzung 1

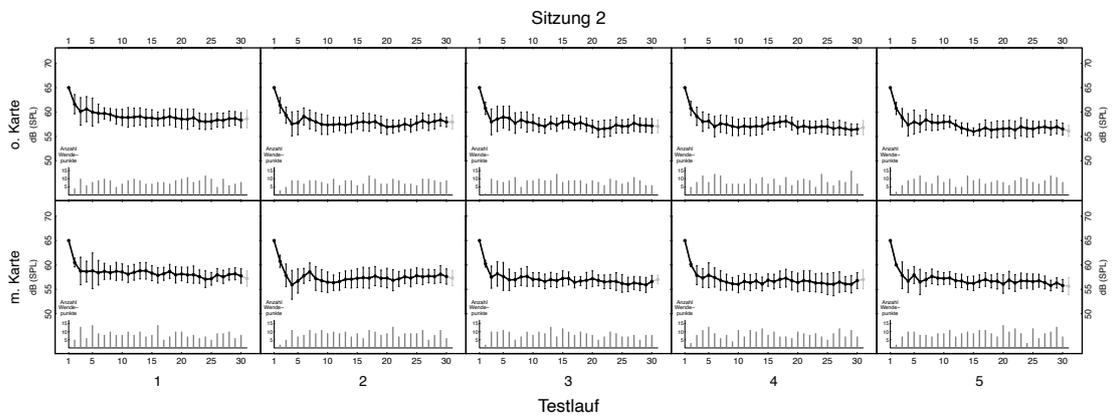


Abbildung 3.11: Mittelwertverlaufskurve der Absolutwerte der Schalldruckpegel des Sprachsignals der Testsitzung 2

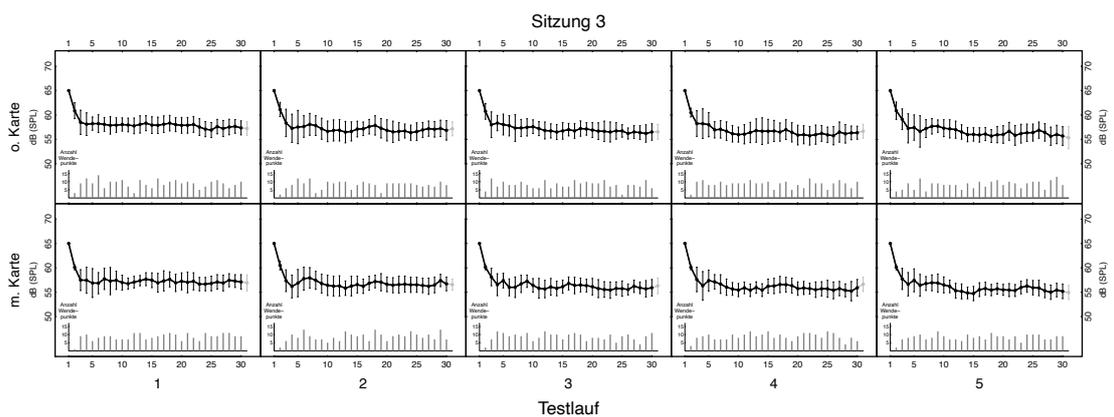


Abbildung 3.12: Mittelwertverlaufskurve der Absolutwerte der Schalldruckpegel des Sprachsignals der Testsitzung 3

Dabei wird in 3.13 und 3.15 die Gruppe dargestellt, der keine Karte vorgezeigt wurde und in 3.14 und 3.16 die Gruppe, die mit der Vorlagenkarte arbeiten durfte.

In den 9 bzw. 15 Kacheln sind verschiedene Werte aufgezeichnet: Die Zahlen 1 bis 3 bzw. 1 bis 5 repräsentieren die Sitzung bzw. den Durchlauf in vertikaler und horizontaler Richtung. Die Korrelation wird mittels Punktwolke mit errechneter, roter Korrelationslinie, mittels Histogramm und mittels Korrelationskoeffizient dargestellt. In den Punktdiagrammen sind die jeweiligen SRT-Werte (in dB SNR) gegeneinander aufgetragen. In den Histogrammen wird die Gesamtverteilung der erzielten SRT-Werte dargestellt. Die Zahlenwerte stellen die Korrelationskoeffizienten nach Pearson als ein Maß für die Übereinstimmung bzw. die Ähnlichkeit dar. Je gerader dabei die rote Ausgleichskurve und je höher der Korrelationskoeffizient ist, desto größer ist die Korrelation zwischen den Sitzungen bzw. Durchläufen. Ein Wert von 1,00 entspräche dabei vollkommener Voraussagbarkeit. Alle errechneten, roten Korrelationslinien in den vier Punktwolkendiagrammen ergeben annähernd gerade Diagonalen mit positiver Steigung. Im unteren Bereich sind die p-Werte, welche in allen Kacheln aller vier Abbildungen unter 0,001 liegen zu sehen. (Bei allen Ergebnissen der Sitzungen bzw. der Durchläufe ist die jeweils andere Dimension außer Acht gelassen.)

In den Diagrammen 3.13 (Gruppe ohne Karte) und 3.14 (Gruppe mit Karte) zeigen alle Korrelationskoeffizienten Werte über 0,80. Benachbarte Sitzungen ähneln sich dabei mehr als nicht aufeinander folgende Sitzungen. Die höchste Korrelation ist dabei nahezu unabhängig von der Vorlage einer Karte.

Auch in den Diagrammen 3.15 (Gruppe ohne Karte) und 3.16 (Gruppe mit Karte) sind dauerhaft Werte von mindestens 0,75 sichtbar. Wieder ähneln sich dabei benachbarte Durchläufe stärker. Hier ist in der Gruppe mit Karte der Korrelationszusammenhang insgesamt etwas stärker ausgeprägt.

3.6 Gesamtdauer der Testdurchführung und Antwortzeit

Im folgenden wird die Gesamtdauer der Testdurchführung und die Antwortzeit genauer unter die Lupe genommen. Betrachtet werden die Zeiten der 32 Probanden der Hauptgruppe. Die Testzeit ist die Zeit, die für die Durchführung eines Testdurchlaufes (eine Liste mit 30 Sätzen) benötigt wird. Diese betrug in dieser Arbeit zwischen 245 und 873 Sekunden. (Dabei gab es insgesamt drei Ausreißer über 600 Sekunden: Proband 30 benötigte im ersten Durchgang 873 Sekunden, Proband 36 benötigte im zweiten Durchgang 755 Sekunden und Proband 28 benötigte im ersten Durchgang 744 Sekunden.)

3 Ergebnisse

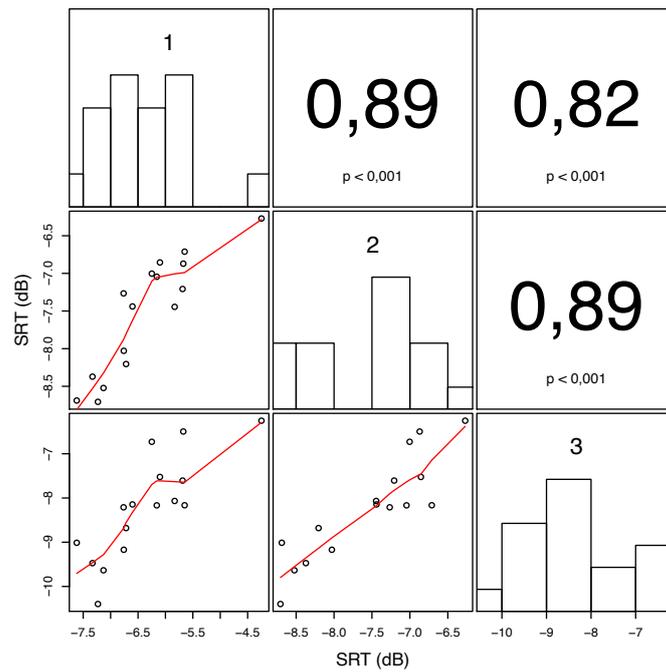


Abbildung 3.13: Korrelation der Testsitzungen zueinander in den Gruppen 1 und 4

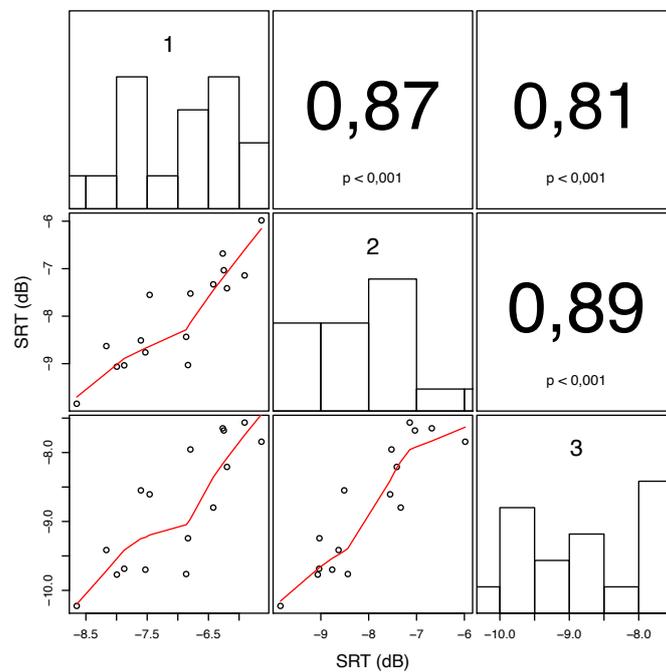


Abbildung 3.14: Korrelation der Testsitzungen zueinander in den Gruppen 2 und 3

3.6 Gesamtdauer der Testdurchführung und Antwortzeit

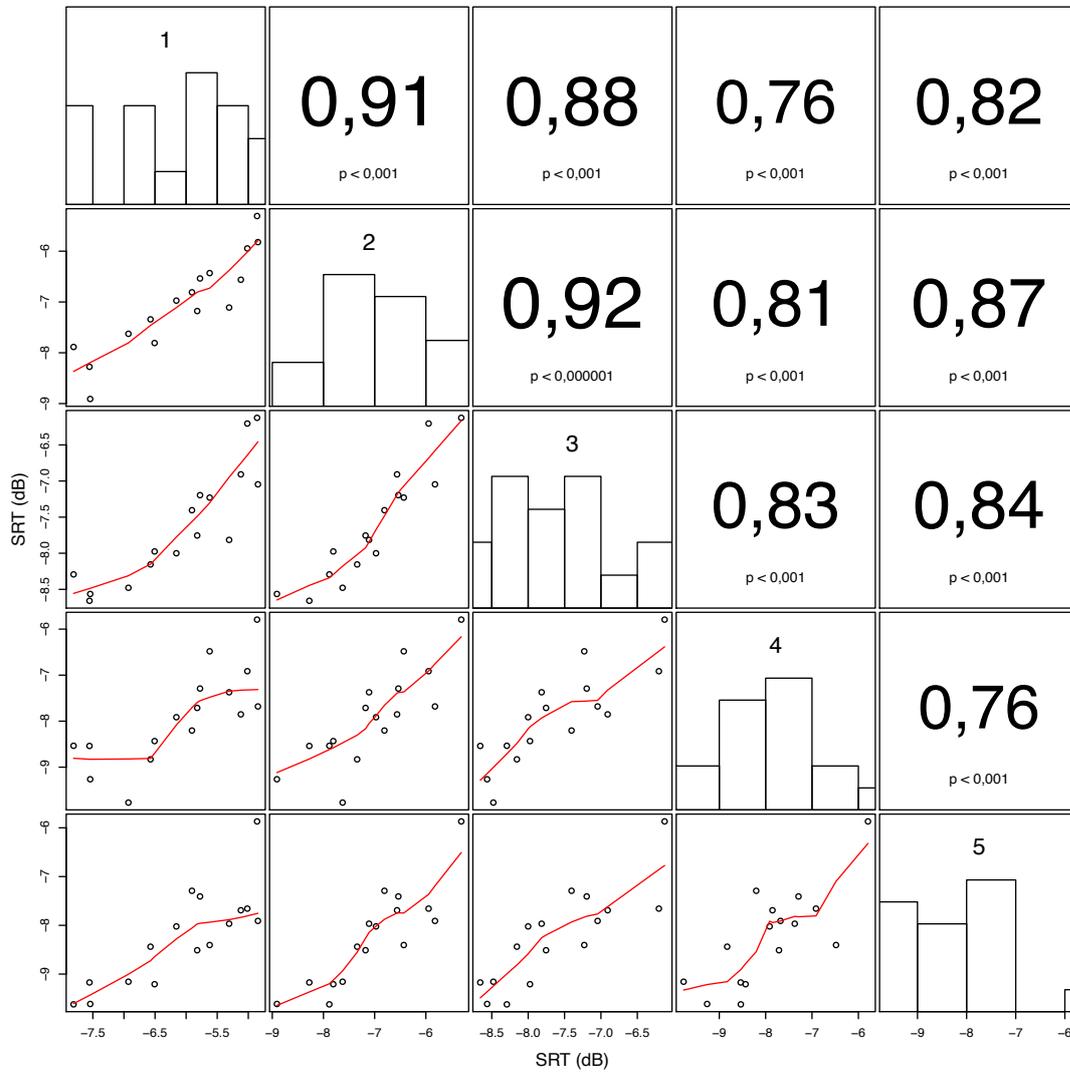


Abbildung 3.15: Korrelation der Testdurchläufe zueinander in den Gruppen 1 und 4

3 Ergebnisse

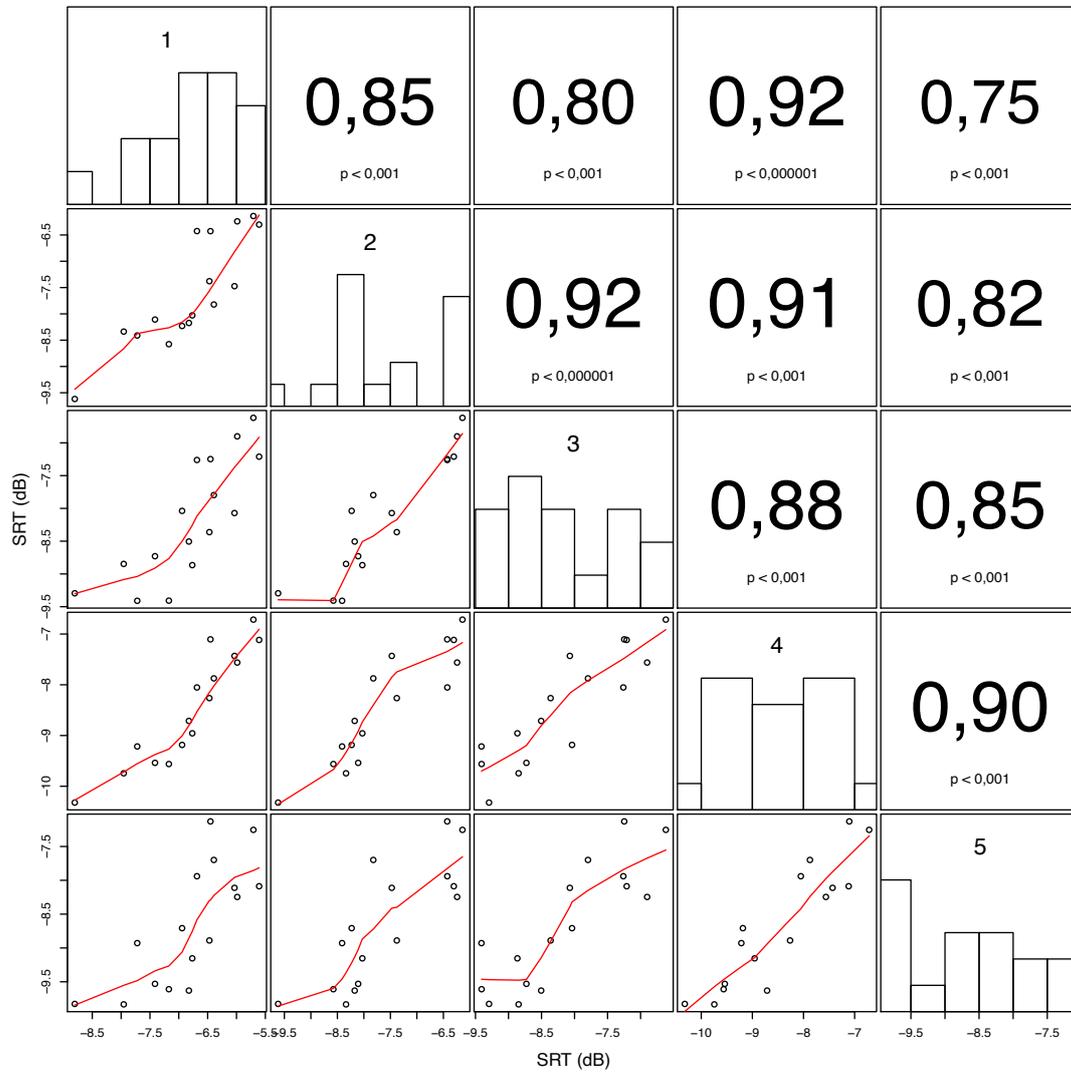


Abbildung 3.16: Korrelation der Testdurchläufe zueinander in den Gruppen 2 und 3

Die durchschnittliche Testzeit betrug mit 373,5 (\pm 69,3) Sekunden etwa sechs Minuten.

Insgesamt zeigte sich ein individuell sehr unterschiedliches Antwortverhalten: Manche Probanden antworteten schnell, anderen überlegten länger. Einige füllten nicht verstandene Lücken in den Sätzen kreativ oder mit ähnlich klingenden Worten auf, während andere nur die Worte wiederholten, die sie sicher verstanden hatten.

Grob überschlagen dauerte eine gesamte Sitzung (je nach individuellem Antwortverhalten) zwischen 45 und 60 Minuten.

Von der Testzeit kann die Antwortzeit abgegrenzt werden. In der Testzeit ist die Antwortzeit enthalten, es können Rückschlüsse gezogen werden.

Die Abbildungen 3.17, 3.18, 3.19 und 3.20 zeigen die durchschnittliche Antwortzeit der männlichen und weiblichen Probanden je nach Gruppenzugehörigkeit.

Die Antwortzeit (*response time*) wird durch die LOST-Software erfasst, protokolliert und gespeichert. Im Vergleich zu der Reaktionszeit des Menschen von etwa 200 Millisekunden geschieht dies mit einer Präzision von etwa 10 Millisekunden sehr genau. Definiert wird die *response time* als die Zeit zwischen Ende des Stimulus (bestehend aus Sprachsignal und Störgeräuschen) und dem ersten Antwortklick durch den Versuchsleiter (Grimm (2014, Abb. 3.7)). Pro Liste werden die für jeden der 30 Sätze nach diesem Prinzip ermittelten *response time* - Werte addiert und ergeben die Antwortzeit. Dementsprechend wird die Antwortzeit maßgeblich durch Versuchsleiter, Proband und LOST-Software beeinflusst. Durch die Variabilität dieser drei beeinflussenden Faktoren wird deutlich, dass die dokumentierte Antwortzeit nicht unbedingt der realen Antwortzeit entsprechen muss. Die Diskrepanz kann wie folgt erklärt werden: Der Versuchsleiter verzögert die Antwortzeit durch seine Reaktionszeit und durch seinen ersten Antwortklick, der entweder zu Beginn oder am Ende der kompletten Antwort eines Probanden getätigt werden kann. Andererseits kann durch diese Protokollierung der Antwortzeit die Fehleranfälligkeit durch Rückfragen oder zusätzliche Äußerungen des Probanden bzw. Erläuterungen des Versuchsleiters minimiert werden.

Die Grafiken 3.21, 3.22 und 3.23 fassen die Abbildungen 3.17, 3.18, 3.19 und 3.20 zusammen.

Mit sinkenden SRT-Werten sinkt auch die Antwortzeit kontinuierlich innerhalb der Sitzungen und Durchgänge.

Die individuellen Antwortzeiten der einzelnen Probanden lassen sich aus den SRT-Verlaufskurven im Anhang A.1 ab S. 107 ablesen.

3 Ergebnisse

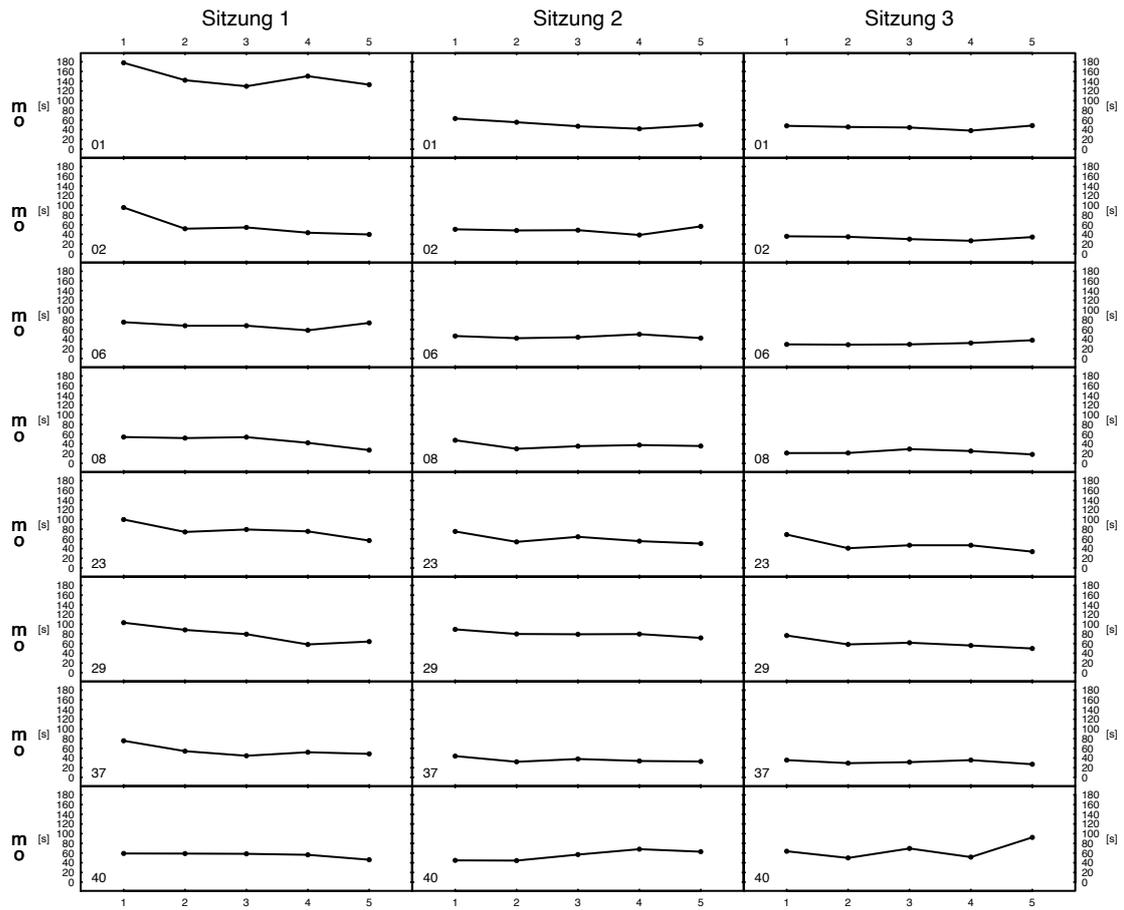


Abbildung 3.17: Antwortzeit (in Sekunden) der Gruppe OHNE Karte, männliche Probanden

3.6 Gesamtdauer der Testdurchführung und Antwortzeit

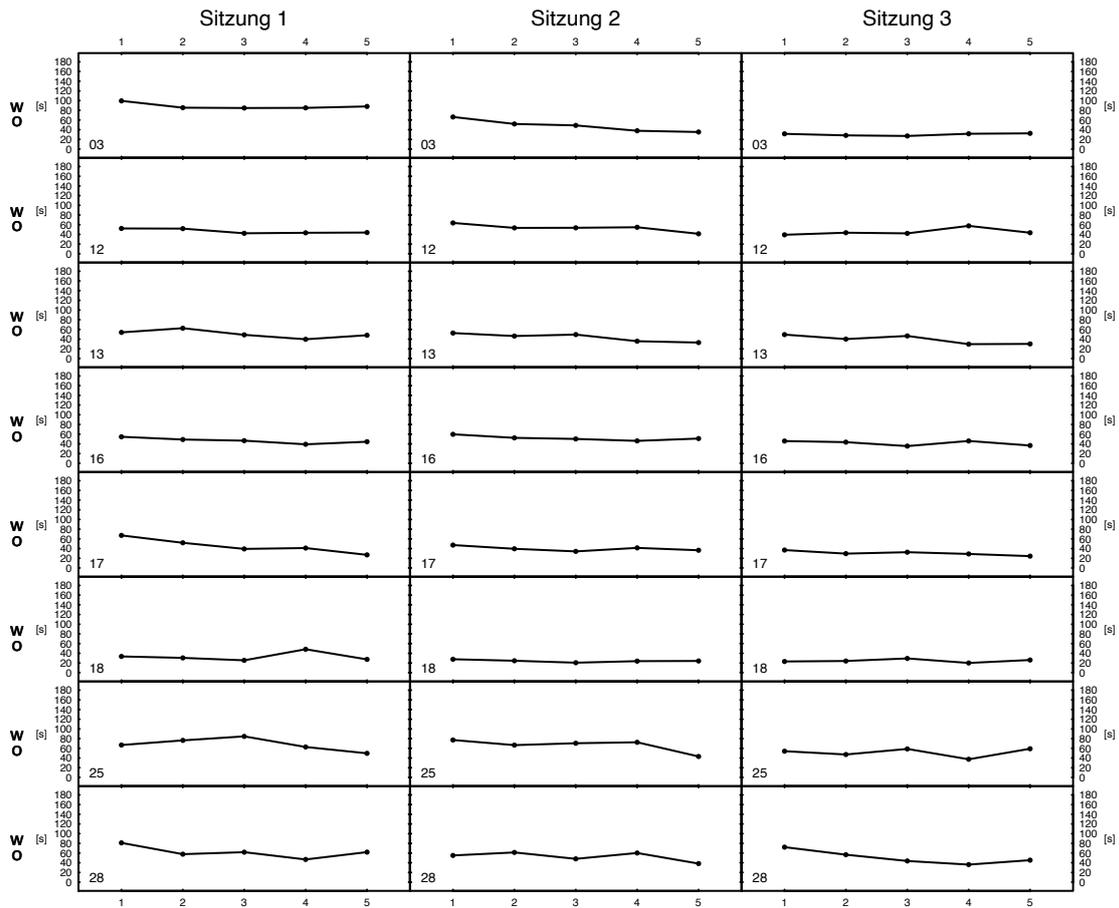


Abbildung 3.18: Antwortzeit (in Sekunden) der Gruppe OHNE Karte, weibliche Probanden

3 Ergebnisse

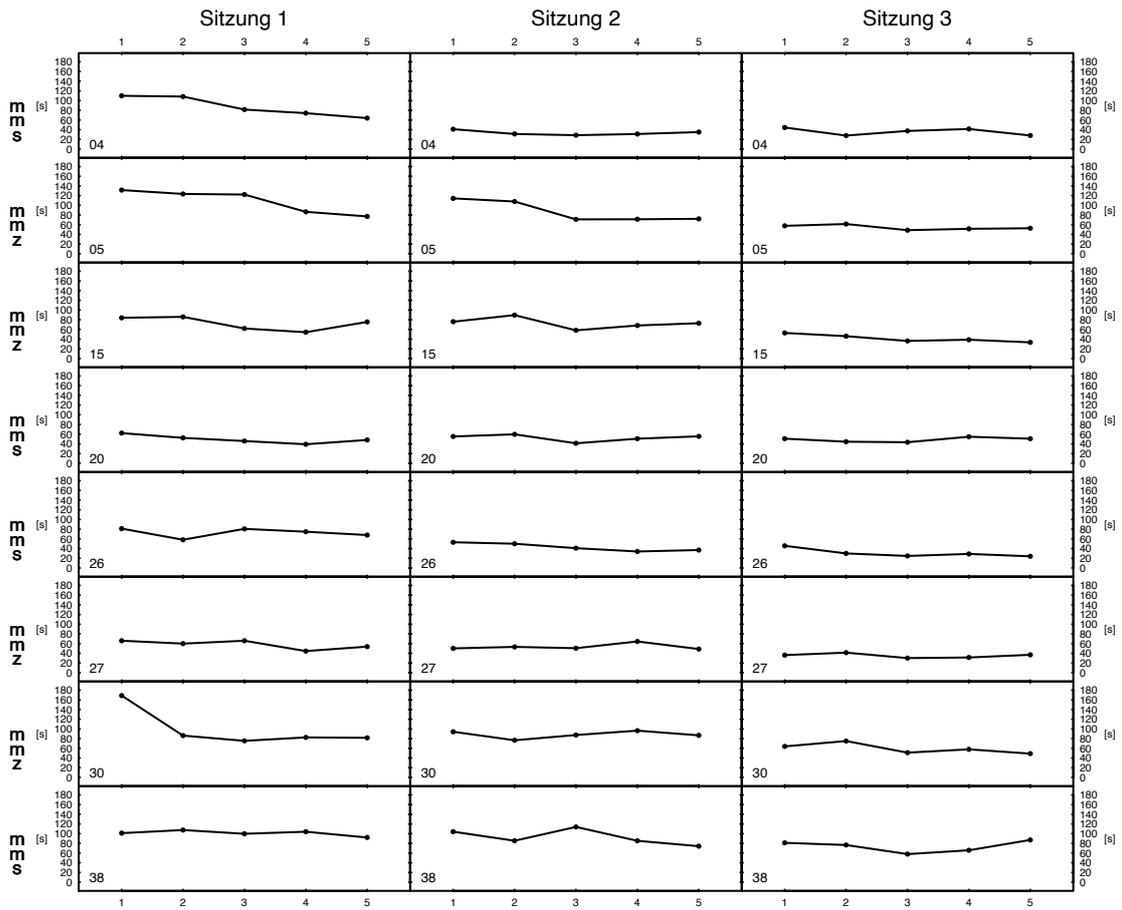


Abbildung 3.19: Antwortzeit (in Sekunden) der Gruppe MIT Karte, männliche Probanden

3.6 Gesamtdauer der Testdurchführung und Antwortzeit

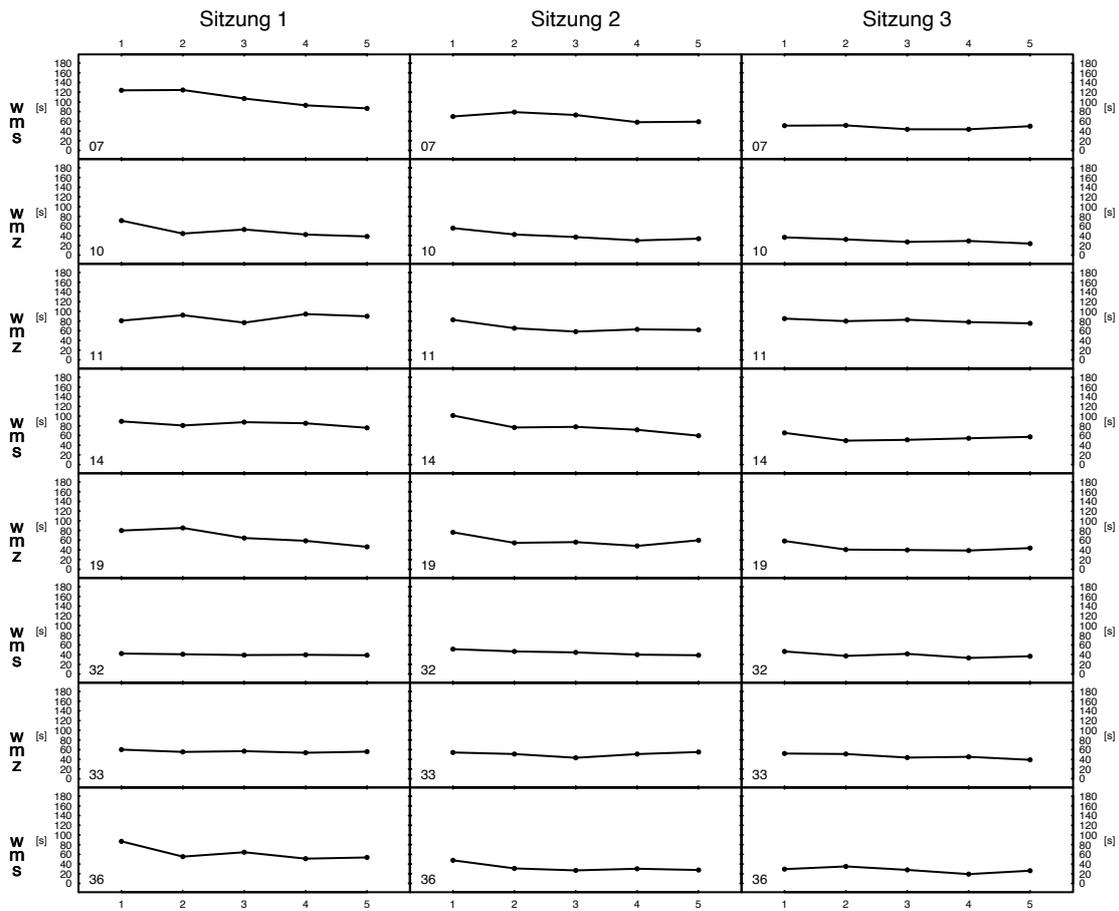


Abbildung 3.20: Antwortzeit (in Sekunden) der Gruppe MIT Karte, weibliche Probanden

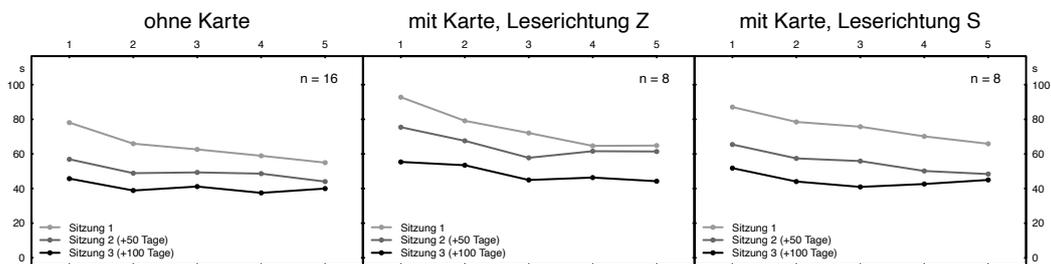


Abbildung 3.21: Durchschnittliche Antwortzeit (in Sekunden) bei Unterteilung in 3 Gruppen

3 Ergebnisse

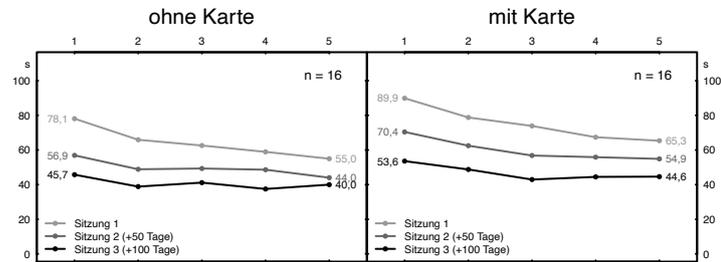


Abbildung 3.22: Durchschnittliche Antwortzeit (in Sekunden) bei Unterteilung in 2 Gruppen

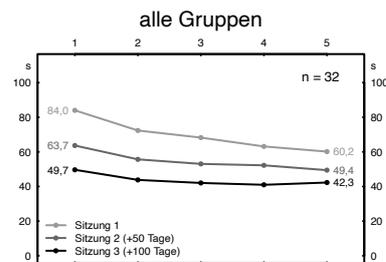


Abbildung 3.23: Durchschnittliche Antwortzeit (in Sekunden) bei Zusammenfassung aller Gruppen

4 Diskussion

Diese Arbeit beschäftigt sich mit dem Oldenburger Satztest, einem deutschsprachigen Sprachverständlichkeitstest, welcher an einem hörgesunden, jungen Probandenkollektiv durchgeführt wurde. Gemessen, dokumentiert und bewertet wurden das Sprachverstehen und die Antwortzeit. Der Fokus der Diskussion liegt jedoch auf dem kurz- und langfristig auftretenden Lerneffekt sowie dessen Komponenten und potentiellen Einflussfaktoren. Weiterhin soll erörtert werden, ob und wie eine Eradikation dieses Lerneffekts angestrebt werden könnte. Abschließend wird der Stellenwert des Oldenburger Satztests im deutschsprachigen Raum diskutiert und ein Ausblick auf die weitere Optimierung dieses Satztests unternommen.

4.1 Reintonaudiometrie

Die mittels Reintonaudiometrie im Freifeld für jedes Ohr nachgewiesene Normalhörigkeit stellte ein zentrales Selektionskriterium bei der Probandenauswahl dar. Eine einheitliche Definition der Normalhörigkeit existiert nicht. Es sollte eine Abschätzung der Normalhörigkeit in Abhängigkeit vom Alter in Anlehnung an die (inzwischen reformierte) Norm [ISO 7029:2017 \(2017\)](#) erfolgen ([Michel \(2014b\)](#), [Michel \(2014a\)](#)). Die hier gewählte Definition in Anlehnung an den Standardkatalog [ISO 7029:2000 \(2000\)](#) und vorherige Dissertationen von [Domes \(2009\)](#) und [Qualen \(2010\)](#) war für die jungen, gesunden Probanden im Regelfall unabhängig von Geschlecht und Sitzung leicht zu erfüllen, da ein gewisser Hörverlust als physiologisch zu werten ist. Es ist jedoch davon auszugehen, dass insbesondere ältere Probanden und Patienten mit einem Lärmschaden in der Anamnese vor allem im Hochtonbereich größere Defizite aufweisen ([Schmidt und Lang, 2007](#)). In diesem Zusammenhang wäre auch ein anderes Resultat in Bezug auf die im Oldenburger Satztest erzielten Ergebnisse (ein schlechteres Abschneiden bei den SRT-Werten) denkbar. Eine Varianzanalyse, welche den Zusammenhang zwischen den Ergebnissen der Reintonaudiometrie und den Resultaten im Oldenburger Satztest untersucht, erwies sich jedoch als nicht signifikant. Dabei ist die Varianzanalyse nicht signifikant bezüglich Geschlecht und ist daher nicht weiter aufgesplittet. Auch besteht kein Unterschied zwischen dem rechten und linken Ohr. Der erzielte Lerneffekt

kann damit nicht durch Unterschiede in den Tonaudiogrammen begründet werden. Jedoch ist im Einzelfall nicht auszuschließen, dass Probanden mit besseren Werten im Tonaudiogramm ebenfalls bessere Werte beim Oldenburger Satztest erzielen können.

4.2 Kurz- und langfristige Lerneffekte beim Oldenburger Satztest

Bisher noch nicht untersuchte Effekte und klinische Relevanz dieser Dissertation:

In bisherigen Studien zum Oldenburger Satztest lag der Fokus nicht auf der Betrachtung des Lerneffekts (Wagener et al., 1999a,c). Die Konditionen des Studiendesigns und der Probandenanzahl und -auswahl waren dadurch nicht ideal ausgewählt: Eine inhomogene, zwölköpfige Probandenklientel unterschiedlichen Alters und mit teils vorbestehenden Kenntnissen über die Testprozedur machen die Ergebnisse anfechtbar. Die so ermittelten Resultate zum Lerneffekt müssen daher nicht unbedingt die Realität wiedergeben. Die Angaben um einen fraglich signifikanten Lerneffekt von etwa 1 dB zwischen den ersten beiden Messungen bzw. von kleiner als 0,5 dB im Verlauf der Evaluationsmessungen (6 Messungen zu je 20 Sätzen) erscheinen relativ vage.

Der in dieser Dissertation eindeutig nachgewiesene, signifikante Lerneffekt findet über die einzelnen Testdurchläufe (kurzfristiger Lerneffekt) und zusätzlich über die Testsitzungen hinweg (langfristiger Lerneffekt) statt. Dabei ist keine Sättigung oder asymptotische Annäherung an einen bestimmten SRT-Tiefstwert ersichtlich. Aus diesen beiden wesentlichen Punkten ergibt sich eine kritische Bewertung in der wiederholten und mehrmaligen Anwendbarkeit des Oldenburger Satztests im klinischen Alltag. Die Vergleichbarkeit und Validität der gemessenen Resultate ist dabei durch die exakte Umsetzung der im Oldenburger-Satztest-Handbuch (Version 1.0) der Firma HörTech (2011) beschriebenen Einstellungen gewährleistet.

4.2.1 Diskussion um den Begriff „Lerneffekt“

Ganz allgemein ist ein Lerneffekt ein unerwünschter Nebeneffekt, der bei Versuchen, Studien, Experimenten oder Testungen aufgrund von Wiederholungen bestimmter Aufgaben, welche in derselben oder in einer ähnlichen Form durchgeführt werden, entsteht. So ist es möglich, dass bei nach der ersten Testsitzung stattfindenden Versuchen bessere Resultate erzielt werden. Die Ätiologie der Verbesserung beruht dabei vermutlich auf multiplen Mechanismen, über die größtenteils gemutmaßt werden muss.

Wie in Kapitel 1.3.2 ab Seite 14 beschrieben wird, kann der Lerneffekt in unterschiedliche Anteile gesplittet werden: Trainings- oder Übungseffekt, Gewöhnungs-, Habituations-

oder Anpassungseffekt, Erinnerungs- oder Wiedererkennungseffekt, Handlungswissen, Vorwissen, *procedural learning*, *content learning* sowie Effekt des Übergangs von einem offenen zu einem geschlossenen Tests. Diese Komponenten überschneiden sich dabei zumindest teilweise oder interagieren miteinander.

Es bleibt somit sogar offen oder zumindest fraglich, ob der Lerneffekt allgemein umbenannt werden sollte, da die besseren Resultate nicht (oder möglicherweise nur teilweise) auf einer realen Leistungssteigerung beruhen. Erinnerungen an richtige Lösungen (Erinnerungseffekt) oder auf die Kenntnis über die Art der zu lösenden Aufgabe (offener/geschlossener Test) spielen eventuell eine größere Rolle. Aus der Kombination der wörtlichen Definitionen der Wörter *lernen* („sich/seinem Gedächtnis einprägen“), *Effekt* („(nicht) bezweckte (überraschende, beeindruckende) (Aus-)Wirkung“) = *Lerneffekt* („in Lernen bestehender Effekt“) aus dem Duden ergibt sich ein eher aktiver, bewusster Aufnahmeprozess. Bei dem Gewöhnungs-, Habituations- oder Anpassungseffekt geht es eher darum, dass ein Proband sich nach wiederholter Darbietung desselben Vorgangs auf eine Situation einstellt und sich anpasst. Der Effekt erscheint eher passiv und unbewusst. Im Gegensatz dazu wird der Trainingseffekt (in Anlehnung an die sportliche Trainingslehre) als physiologische, mit anatomischen Korrelaten einhergehende Anpassungsreaktion des Körpers betrachtet. Diese ist Resultat einer erhöhten, über eine größere Zeitspanne stattfindenden Belastung.

Das *skill-Lernen* und die Habituation (Gewöhnung) als Leistungen des impliziten (prozeduralen) Gedächtnisses sind unbewusste Lernvorgänge (Schmidt und Lang (2007, S. 224-238)). Beim Oldenburger Satztest kann im Rahmen dieser neurophysiologischen Betrachtung eine Anpassungsreaktion erfolgen: Das Störäuschen (als länger dargebotener Reiz ohne Bedeutung) kann mit der Zeit ausgeblendet werden, während konsekutiv die ganze Aufmerksamkeit auf das Sprachsignal (als wichtigeres Ereignis) gelenkt werden kann. Das *skill-Lernen* kann mit dem Erlernen der Testsituation und des Testablaufs (*procedural learning*) in Verbindung gebracht werden.

Neuronale Plastizität charakterisiert die Veränderung neuronaler Strukturen in Abhängigkeit von deren Aktivität und betrifft die Kopplung von Neuronen und damit die Effizienz der Synapsen. Das Ziel der neuronalen Plastizität ist die Erhaltung, Modifikation, Anpassung und Erweiterung der Funktionen des Nervensystems, in Abhängigkeit von sich ändernden Umweltbedingungen (Hebb (2002, Kap. 4), Schmidt und Lang (2007, S. 229), Dudel et al. (2013, S. 487-525), Pape et al. (2014, S. 123)). Die Effekte der neuronalen Plastizität sind demnach keine kurzfristigen Lerneffekte innerhalb von Minuten, sondern entwickeln sich über eine längere Zeitspanne. In der kurzen Zeit eines Testdurchgangs des Oldenburger Satztests ist folglich kein materielles Korrelat denk-

bar. Zwischen den Testsitzungen (je 50 Tage) wäre ein organischer Umbau zumindest theoretisch denkbar. Der in dieser Studie zwischen dem jeweils letzten Durchgang einer Sitzung und dem jeweils ersten Durchgang der folgenden Sitzung beobachtete Rückschritt des Lernvorgangs (3.9) wäre aber in diesem Zusammenhang nicht einleuchtend.

Das *content learning* (Kennenlernen der vorkommenden Wörter) sollte in der Gruppe ohne Karte initial am stärksten ausgeprägt sein. Die Vorlagenkarte und damit der Testmodus (offen versus geschlossen) stellt jedoch keinen signifikanten Einflussfaktor dar. Bei Betrachtung von Abbildung 3.8 könnte interpretiert werden, dass in der Gruppe mit Karte schneller bessere SRT-Werte erreicht werden, welche allerdings im Verlauf auch von der Gruppe ohne Karte erreicht werden.

Das Kurzzeitgedächtnis bietet maximal Speicherplatz für 7 ± 2 Informationseinheiten (= Ketten; chunks) (Schmidt und Lang (2007, S. 224-238)). In diesem Zusammenhang wird deutlich, dass beim Oldenburger Satztest mehr als nur das Hörverständnis getestet wird und ein gewisser Intellekt vorausgesetzt wird (Mrowinski (2006), Probst (2013)).

Bei einer großen Anzahl an Wiederholungen des Oldenburger Satztests innerhalb von Monaten und Jahren könnten sämtliche Inhalte ins Langzeitgedächtnis übernommen worden sein und ein sofortiges Erinnern an sämtliche Wörter selbst bei offenem Testmodus wäre denkbar, sodass von Anfang an keine Verwechslungen der Wörter passieren.

Aufgrund des Testdesigns des Oldenburger Satztests kann davon ausgegangen werden, dass kein aktives Auswendiglernen aller 30 Sätze der 40 verschiedenen Listen möglich ist und passiert. Ein Verknüpfen bestimmter Wörterkombinationen wäre aber durchaus denkbar. Tatsächlich sind manche Wörterkombinationen im Oldenburger Satztest häufiger vertreten als andere (Schraud, 2014). Ein Verknüpfen und eine damit erhöhte Ratewahrscheinlichkeit kann in Betracht gezogen werden.

4.2.2 Untersuchte Einflussfaktoren auf den Lerneffekt

Interpretation der Resultate der ANOVA

Wie in Kapitel 3 auf Seite 39 beschrieben, wurden fünf potentielle Einflussfaktoren auf die Sprachverständnisschwelle (SRT) untersucht: Geschlecht, Testsitzung, Testdurchgang, Karte und Leserichtung (im Falle des Vorlegens einer Karte).

Dabei stellten sich nur die Einflussfaktoren Testsitzung und Testdurchgang sowie deren Interaktion als signifikant heraus. Dies erscheint aus Gründen des Gewöhnens, des Eindenkens bzw. eines Lerneffekts schlüssig.

Betrachtet man die Abhängigkeit der Sprachverständnisschwellen von dem Testmodus (offen versus geschlossen), so stellt sich kein signifikanter Unterschied dar. Vor der Testdurchführung wurde überlegt, ob sich das Vorhandensein einer Karte positiv auf die SRT-Werte auswirken könnte: Erstens käme es zu keinen Verwechslungen der Wörter. Zweitens erhöhe sich die Ratewahrscheinlichkeit durch die Karte, indem bei dem Verstehen einzelner Silben auf ein Wort aus der Liste geschlossen werden könnte. Des Weiteren könnte auch bei komplettem Nichtverständnis mit einer a priori höheren Trefferquote geraten werden. In den praktischen Durchführungen konnten diese theoretischen Überlegungen beobachtet werden: Probanden mit Karte zeigen eine größere Entscheidungsfreudigkeit. Es wurden häufig fünf Wörter genannt, auch wenn nichts verstanden wurde. Andererseits muss auch erwähnt werden, dass Probanden mit der Karte häufiger Rückfragen stellten, zumindest zeitweise überfordert wirkten und konsequent mehr Zeit benötigten.

Im Gegensatz dazu verwechselten Probanden der Gruppe ohne Karte häufig bestimmte Wörter. Beispielsweise wurde statt des Namens *Britta* häufig *Jutta* oder anstelle von *kauft* wurde *klaut* verstanden. Diese einmal falsch wahrgenommenen Wörter wurden in dieser Gruppe nicht selten über alle Sitzungen hinweg beibehalten. Der Proband erlernte quasi ein falsches Wort und wurde auf diesen nicht in der Liste vorkommenden Begriff geprägt. Probanden mit Karte passieren derartige Fehler nicht oder äußerst selten bei nachlassender Konzentration. In der Praxis zeigte sich, dass Probanden ohne Karte weniger raten. Wenn nichts verstanden wurde, wird ohne zu Raten zum nächsten Satz fortgeschritten. Bezüglich der Handhabung der Karte wäre es möglich, dass eine unterschiedliche Herangehensweise an die Sätze oder die Wörtergruppen und deren systematische Interpretation die Aufgabe beeinflusst. Tatsächlich konnte keine dieser theoretischen Überlegungen bestätigt werden. Karte und Leserichtung stellten in der Arbeit nicht signifikante Faktoren dar. Die Bedeutung der Instruktion der Leserichtung wurde nach Bestätigung der nicht vorhandenen Signifikanz als irrelevant bewertet und in den weiteren Untersuchungen als Einflussfaktor ausgeschlossen.

Der Lerneffekt ist dementsprechend in der Gruppe mit Karte genauso vorhanden wie in der Gruppe ohne Karte. Trotz der Kenntnis über alle Wörter findet eine vergleichbare, signifikante Verbesserung statt. Diese Resultate kongruieren mit der eingangs vorgestellten Studie von [Brand et al. \(2004\)](#), die zeigte, dass es keine Rolle spielt, ob der Test als offene oder geschlossene Version präsentiert wird. Im Gegensatz zu den hier vorliegenden Ergebnissen beschränkt sich der Lerneffekt in der Arbeit von [Brand et al. \(2004\)](#) im geschlossenen Format auf die ersten drei Durchgänge.

Betrachtet man die Abhängigkeit der Sprachverständnisschwelle vom Geschlecht, so

stellt dieses (ebenso wie in der Studie von Grimm (2014)) keinen signifikanten Einflussfaktor dar.

Interpretation der SRT-Werte in Abhängigkeit von Testsitzung und Testdurchlauf

Die Verlaufskurven der SRT-Werte mit ihren negativen Steigungen (insbesondere von 3.9, aber auch von 3.7 und 3.8) können als Lernkurven interpretiert werden. Der Lerneffekt besagt, dass in jeder Sitzung im Laufe der Durchläufe 1-5 immer niedrigere Schalldruckpegel der Sprache toleriert werden und trotzdem genauso viel verstanden wird, wie bei der Präsentation der Sprache mit lauterem Schalldruckpegel. Aus 3.9 wird ersichtlich: Während der ersten Testsitzung verbesserte sich jeder Proband im Durchschnitt um 2,61 dB, im zweiten um 1,77 dB und im dritten um 1,65 dB. Das ergibt einen absoluten Lernwert von Beginn der 1. Sitzung bis zum Ende der 3. Sitzung von 4,13 dB. Dadurch, dass keine Wendepunkte und keine asymptotische Annäherung an einen Wert erkennbar sind, kann nicht von einer Sättigung der Lernkurven gesprochen werden. Es wäre möglich, dass der Lerneffekt im Laufe weiterer Testsitzungen weiter zunimmt. Eine kleine zusätzliche Verbesserung und eine asymptotische Annäherung an einen Wert nicht weit unter -9,5 dB wäre durchaus denkbar.

Ein Vergleich der hier ermittelten Resultate mit dem Lerneffekt aus den Originalpublikationen von Wagener et al. fällt bei näherer Betrachtung nicht unbedingt leicht: Hauptgründe sind das offene Testkonzept (Fehlen einer Testkarte), die Auswahl von Testlisten zu je 20 Sätzen und die Selektion von Probanden, welchen der Oldenburger Satztest bereits bekannt war. Abbildung 1 aus der Studie von Wagener et al. (1999b) kann hierbei am ehesten mit der Tabelle 3.8 verglichen werden: Der hier erzielte Lerneffekt von 1,17 dB zwischen dem ersten und zweiten Durchgang entspricht annähernd dem Wert von etwa 1 dB bei Wagener et al.. Diese Studie bezieht sich dabei genau wie diese Arbeit auf junge, hörgesunde Probanden.

Dass auch bei schwerhörigen Patienten mit Cochlea-Implantat im Laufe der Messungen ein signifikanter Lerneffekt von 0,4 dB stattfindet, zeigte Müller-Deile (2009, Abb. 9) in Studien an insgesamt 46 Patienten.

Zusammenfassend muss in Anbetracht der Studienlage davon ausgegangen werden, dass auch bei schwerhörigen und Cochlea-Implantat-Patienten ein Lerneffekt auftritt. Spekulativ beläuft sich dieser auf ein geringeres Ausmaß als bei normalhörenden Probanden.

Jedoch muss insgesamt in diesem Zusammenhang betont werden, dass sich das Ausmaß der Lernkurven auf die Durchschnittswerte der selektierten Hauptgruppe der

Probanden bezieht. Einzelne Probanden zeigen zum Teil relativ große, in beide Richtungen weisende Abweichungen von den Mittelwerten. Daher muss der Lerneffekt im klinischen Alltag unter Berücksichtigung weiterer Faktoren (wie Art einer vorliegenden Pathologie des Gehörs, Alter, mentaler Status oder Vorgeschichte) teilweise neu und individualisiert beachtet und reevaluiert werden.

Interpretation der Absolutwerte der Schalldruckpegel des Sprachsignals

Die Tatsache, dass die Schalldruckpegel des Sprachsignals in jedem Testdurchlauf absteigen (siehe 3.10, 3.11 und 3.12), lässt die Schlussfolgerung zu, dass alle Probanden mit der Zeit besser hören bzw. „akustisch dazulernen“. Da sich in Testsitzung 2 und 3 der Wert auf niedrigere Schalldruckpegel einpendelt als in Testsitzung 1 und 2, kann von einem weiteren Lerneffekt über die Sitzungen ausgegangen werden. Es wird besser hörverstanden. Auch innerhalb der 5 Testdurchläufe kann man jeweils im Verlauf von 1-5 ein Einpendeln auf niedrigere Schalldruckpegel registrieren: Dies lässt die Schlussfolgerung der Existenz eines Lerneffekts innerhalb einer Sitzung zu. Das relativ schnelle Einpendeln um den errechneten Wert (nach etwa 3-15 Sätzen, ohne starke Schwankungen) spricht für die Genauigkeit des Tests. Bezüglich der Differenz der Schalldruckpegel zwischen Sprache und Störrauschen wird folgender Zusammenhang sichtbar: Im Durchschnitt sind bei den Schalldruckpegeln der Sprache nie Werte über 65 dB SPL nötig. Lediglich beim ersten und zweiten Satz eines Durchgangs wird gelegentlich einmalig ein höherer Schalldruckpegel als 65 dB SPL benötigt. So werden von den fünf Wörtern eines jeden ersten Satzes zumindest 60%, sprich 3 von 5 Wörtern richtig verstanden. Bei der Betrachtung der Wendepunkte ist nicht wirklich ein Muster erkennbar. Die Anzahl der Wendepunkte häuft sich in keiner Sitzung bzw. in keinem Durchgang ersichtlich in einem bestimmten Bereich oder bei einem bestimmten Satz. Es ist also nicht so, dass beispielsweise der 3. Satz besonders oft einen Wendepunkt darstellt.

Interpretation der Korrelation der Testsitzungen zueinander

In beiden Diagrammen (ohne und mit Karte) 3.13 und 3.14 ähneln sich die benachbarten Sitzungen (1 und 2 bzw. 2 und 3) mehr als die nicht benachbarte Sitzung (1 und 3). Dies wird auch in der numerischen Darstellung der Werte deutlich. Die Punktwolken mit roter Verbindungslinie verlaufen bei benachbarten Sitzungen „linearer“ bzw. „gerader“ ohne größeres Abknicken. Der Knick in der Verbindungslinie der Korrelation der Sitzung

1 und 3 ist stärker ausgeprägt. Aus den Balkendiagrammen in den mittleren Kacheln ist keine genauere Korrelation zu entnehmen.

Interpretation der Korrelation der Testdurchläufe zueinander

In dem Diagramm 3.15 (Gruppe ohne Karte) besteht ein Gefälle von links nach rechts bzw. von oben nach unten. In dem Diagramm 3.16 (Gruppe mit Karte) ist diese Tendenz eher nicht zu sehen; stattdessen korrelieren insgesamt benachbarte Durchläufe eher miteinander (Gefälle von der Diagonale nach außen zur Ecke). Die Punktwolken mit der roten Verbindungslinie und die Balkendiagrammen verdeutlichen diesen Effekt.

4.2.3 Nicht untersuchte Einflussfaktoren auf den Lerneffekt

Im folgenden werden potentielle Einflussfaktoren auf den Lerneffekt genannt, welche aber hier nicht explizit untersucht wurden. Es kann eine grobe Einteilung in *probandenbezogene Faktoren* und *unabhängige Faktoren* vorgenommen werden. In weiteren Studien wäre es eventuell sinnvoll, diese Faktoren zumindest teilweise zu beachten oder von Vorhinein im Studiendesign zu berücksichtigen.

Probandenbezogene Faktoren:

- Definierte Altersgruppe
- Bildungsniveau und sozialer Status
- Ausbildungsstatus des Probanden, Studentenstatus
- Eigeninteresse, wissenschaftliches Interesse, Motivation, Recherche im Internet, Austausch der Probanden untereinander
- Ehrgeiz, Leistungsorientierung
- Zeit, Termindruck, Arbeitsstress, -belastung
- Schlaf(-mangel), Wachheit, Konzentration, Kaffeegenuss, Aufregung, Nervosität, Hunger und Durst
- Hörbelastung vor der Testsitzung bzw. am Tag davor
- Herkunft innerhalb Deutschlands (aufgrund der regionalen Bekanntheit der Namen)
- Unbewusstes Üben in einer ähnlichen Hörsituation z. B. durch Lärm in der Wohnung, Hintergrundmusik im Restaurant etc.
- subjektives Wohlbefinden, Gemütszustand, leichtes Krankheitsgefühl

Probanden-unabhängige Faktoren:

- Tageszeit
- Motivation durch den Testleiter
- Sauerstoffgehalt und Temperatur der Kabine C

Bezugnehmend auf einige der oben genannten Faktoren kann die Arbeit wie folgt bewertet werden:

Es ist festzuhalten, dass alle 38 Probanden (mit Abitur als Schulabschluss) als Studenten (29 Humanmedizin, 3 Lehramt, 2 BWL, 1 Sozialpädagogik, 1 Zahnmedizin, 1 Informatik, 1 Journalismus; homogene Gruppe) ein eher hohes Bildungsniveau (und einen eher hohen sozialen Status) aufweisen.

Im Bezug auf die Erkenntnis, dass bei Sprachverständnistests Sensorik, Perzeption und Kognition getestet werden (Ptok und Kießling, 2004), könnte der eher hohe akademische Level trotz des eher simplen Wortmaterials eine Rolle spielen.

Auch das im Vorfeld im Studiendesign festgelegte Alter (Einschlusskriterium: 20-29 Jahre bei der ersten Sitzung) könnte einen signifikanten Einflussfaktor darstellen.

Die meisten Testungen wurden unter der Woche (an Arbeitstagen) am Nachmittag nach 16 Uhr durchgeführt, da die Kabine C davor im klinischen Alltag die meiste Zeit belegt war. Seltener wurde an Wochenenden getestet. Die daraus folgenden Umstände (zum einen Tageszeit, zum anderen Schlaf, Wachheit, Konzentration, Kaffeegenuss, Aufregung, Nervosität, Hunger und Durst sowie Hörbelastung vor der Testsitzung bzw. am Tag davor) spielen eventuell mit in die Resultate ein. Die Konzentrationsfähigkeit (und Wachheit) hätte evtl. mit einem Konzentrationstest (z. B. D2-Konzentrationstest) im Voraus ermittelt werden können.

Sowohl die Eigenmotivation (inklusive Nachlesen und Informieren im Internet und Austausch der Probanden untereinander), als auch die Motivation durch den Testleiter sind als psychologisch, pädagogisch relevante Faktoren in Betracht zu ziehen.

Bedeutung der Sprachperzeption: Sprachperzeption ist die Basis für sprachaudiometrische Untersuchungen (Ptok und Kießling, 2004). Da bei der Sprachaudiometrie neben sensorischen auch perzeptive und kognitive Fähigkeiten des Probanden oder Patienten geprüft werden, deutet ein schlechtes Ergebnis in einem Sprachverständlichkeitstest nicht zwangsläufig auf einen sensorischen Schaden hin. Defiziten im perzeptiven und kognitiven Bereich wird zu Gunsten von sensorischen Einbußen in der klinischen Praxis häufig zu wenig Bedeutung zugesprochen. Sprachaudiometrische Resultate sind somit mit kognitiven Funktionen und Leistungen eng verknüpft (Meister et al. (2011), Meister (2017), Knopke und Olze (2017)). In diesem Zusammenhang kann auch

erwähnt werden, dass durch eine Hörverbesserung mithilfe von Cochlea-Implantaten auch kognitive Funktionen positiv beeinflusst werden können (Knopke und Olze, 2017).

Viele Sprachverständnistests scheinen unterschiedliche Wahrnehmungsprozesse zu untersuchen. Das Wissen darum und die korrekte Deutung stellen eine wichtige Weiche in der Praxis bezüglich der effizienten Indikationsstellung von Hörgeräten dar.

Die Tatsache, dass verschiedene Wörter (unabhängig oder nur mäßig abhängig von der Länge bzw. Anzahl der Buchstaben) unterschiedlich schnell verstanden werden, wirft Fragen nach der Art der Speicherung der Sprache im Gehirn und nach Sprachproduktion, -wiedergabe und -verständnis auf.

Die Art der Organisation des sprachlichen Wissens erfolgt in Form eines *mentalen Lexikons* und einer *mentalen Grammatik*, in denen sprachliche Einheiten und deren Kombinationsmöglichkeiten systematisch organisiert aufbewahrt werden. Hier besteht ein Anknüpfungspunkt zum Oldenburger Satztest: Der vorgegebene Satzbau kann die Verständlichkeit einzelner Wörter vereinfachen.

Bedeutung der Vorhersagbarkeit: Satztests mit niedriger Vorhersagbarkeit sollten theoretisch bei der Schätzung der SRT-Werte und der Steigungen der Diskriminationsfunktionen effizienter sein als Satztests mit niedriger Vorhersagbarkeit. Ein hohes j heißt niedrige Vorhersagbarkeit.

Der Göttinger Satztest zeichnet sich durch eine hohe Vorhersagbarkeit ($j=1,95$ bei -8 dB SNR; $j=2,38$ bei -4 dB SNR) aus. Der Oldenburger Satztest besitzt eine niedrige Vorhersagbarkeit ($j=3,18$ bei -9 dB SNR; $j=4,29$ bei -5 dB SNR) (Brand und Kollmeier, 2002).

Allerdings konnte in Messungen an hörgesunden und hörgeschädigten Probanden dieser erwartete Vorteil des Oldenburger Satztests nicht beobachtet werden, was möglicherweise einem Lerneffekt zuzuschreiben ist (Brand und Kollmeier, 2002).

4.3 Kurz- und langfristige Lerneffekte bei Sprachverständnistests

Ein Lerneffekt kann ebenso bei der wiederholten Durchführung anderer, ähnlich aufgebauter oder fremdsprachiger Matrixtests beobachtet oder ausgeschlossen werden. Im Folgenden werden weitere Tests, deren Prinzip und deren Lerneffekt mit Ausmaß und Auswirkungen vorgestellt.

Matrixtests leiten sich nach dem Funktionsprinzip des Satztests nach Hagerman (1982) ab und existieren in vielen verschiedenen Sprachen (Brand und Wagener, 2017),

beispielsweise auf schwedisch (Hagerman (1982)), deutsch (Wagener et al. (1999b)), dänisch (Wagener et al. (2003)), polnisch (Ozimek et al. (2010)), französisch (Jansen et al. (2012)), spanisch (Hochmuth et al. (2012)), italienisch (Puglisi et al. (2014)), finnisch (Dietz et al. (2014)), türkisch (Zokoll et al. (2015)), russisch (Warzybok et al. (2015)) und niederländisch (Houben und Dreschler (2015)).

Der Einsatz von Matrixtests in ihrer geschlossenen Form beschränkt sich dabei nicht auf Muttersprachler (Brand und Wagener, 2017). Dies passt zu den Resultaten des spanischsprachigen Probanden, dessen SRT-Werte nicht wesentlich von denen der deutschsprachigen Probanden abweichen.

Abgesehen von kleineren landesspezifischen Unterschieden können die Eigenschaften und Resultate dieser Matrixtests gut miteinander verglichen werden (Wagener et al., 2007, Studie an Satztests auf deutsch, englisch, niederländisch und schwedisch). Insgesamt besteht eine „hohe internationale Vergleichbarkeit“ dieser Matrixtests (Brand und Wagener, 2017).

Außerdem existieren für „Personen mit eingeschränkter Hörmerkspanne (z. B. Kinder)“ verkürzte Formen von Matrixtests.

Hey et al. (2014) argumentieren, dass aufgrund der geringen Anzahl an verwendeten Wörtern bei Matrixtests ein ausgeprägter Lerneffekt bestehe, welcher dem Kennenlernen der getesteten Wörter und der Testprozedur zuzuschreiben sei. Dem ist entgegenzusetzen, dass in diesem Falle größere Unterschiede bei der Verfahrenswahl der Testmodi (offen versus geschlossen) nachweisbar sein müssten.

Die Aussage von Brand und Wagener (2017), „Anschließend liefern Matrixtests auch bei wiederholter Verwendung gleichwertige Ergebnisse.“, kann in Anbetracht der Resultate dieser Arbeit zumindest als fragwürdig betrachtet werden.

Mithilfe von Matrixtests gelingt es allerdings, den Erinnerungseffekt an komplette Sätze aus einem Satztest zu minimieren (Brand und Wagener, 2017). Im Vergleich zu den Matrixtests ist der prozedurale Lerneffekt bei Satztest mit Alltagssätzen lediglich in geringem Ausmaß vorhanden. Hier genügen wenige Übungssätze vor der eigentlichen Messung (Sukowski et al., 2009).

Die Indikation für Matrixtests beschreiben Brand und Wagener (2017) folgendermaßen: „Matrixtests sind für Hörgeräte- und Cochlea-Implantat-Träger insbesondere für die begleitende Überprüfung der Anpassung und des Versorgungserfolgs geeignet.“

Tabelle 4.1: Übersicht: Lerneffekt bei Sprachverständlichkeitstests in Satzform

Satztest (Name und Autor)	Probandenzahl (und ggf. Testmodus)	SRT	Lerneffekt
Schwedischer Satztest (Hagerman, 1982)	20	-7,3 ($\pm 0,2$) dB SNR	2,2 dB bei der Testung von 28 Listen zu je 10 Sätzen allgemein: 0,1 dB pro Liste zu 10 Sätzen
Oldenburger Satztest (Wagener et al., 1999b)	20	-7,1 ($\pm 1,1$) dB SNR (Ursprungsarbeit) -8,4 dB SNR (Evaluationsarbeit)	1 bis maximal 2 dB bei der Testung von 6 Listen zu je 20 Sätzen
DANTALE II (Wagener et al., 2003)	16 (Trainingsmessung) 60 (Evaluationsmessung)	-8,43 dB SNR	2,2 dB bei der Testung von 8 Listen zu je 20 Sätzen 1,4 dB innerhalb der ersten beiden Listen 0,3 dB zw. der letzten Trainingsmessung u. der Evaluationsmessung
Polish Matrix Test (Ozimek et al., 2010)		-8,0 ($\pm 1,3$) dB SNR	
FrMatrix (Jansen et al., 2012)	10	-6,0 ($\pm 0,6$) dB SNR	1,8 dB bei der Testung von 6 Listen zu je 20 Sätzen 1,0 dB zw. der 1. u. 2. Liste
Spanish matrix sentence test (Hochmuth et al., 2012)	33 (offen) 10 (geschlossen)	-6,2 ($\pm 0,8$) dB SNR -7,2 ($\pm 0,7$) dB SNR	1,6 dB bei der Testung von 6 Listen zu je 20 Sätzen (offen) 1,1 dB zw. 1. u. 2. Liste (offen) 0,8 dB zw. 1. u. 2. Liste (geschlossen)
Italian matrix sentence test (Puglisi et al., 2014)	26 (offen)	-6,7 ($\pm 0,7$) dB SNR	2,5 dB bei der Testung von 6 Listen zu je 20 Sätzen 1,5 dB zw. der 1. u. 2. Liste
Finnish Matrix Test (Dietz et al., 2014)	21	-10,1 ($\pm 0,1$) dB SNR	2,3 dB bei der Testung von 7 Listen zu je 20 Sätzen 1,1 dB zw. 1. u. 2. Liste 1,8 dB zw. 1. u. 3. Liste
Durch matrix sentence test (Houben und Dreschler, 2015)	45	-8,4 ($\pm 0,2$) dB SNR	(keine explizite Untersuchung vorgenommen)
Russian Matrix Test (Warzybok et al., 2015)	77	-8,8 ($\pm 0,8$) dB SNR	
Turkish Matrix Test (Zokoll et al., 2015)	26	-7,2 ($\pm 0,8$) dB SNR	2,0 dB bei der Testung von 6 Listen zu je 20 Sätzen (offen) 1,4 dB bei der Testung von 6 Listen zu je 20 Sätzen (geschlossen)
Oldenburger Satztest (in dieser Arbeit)	32		4,13 dB bei der Testung von 15 Listen zu je 30 Sätzen 2,61 dB innerhalb der ersten Testsitzung 1,77 dB innerhalb der zweiten Testsitzung 1,65 dB innerhalb der dritten Testsitzung

Anmerkung: Bei der Optimierung und Evaluation der Satztests wurden teilweise unterschiedliche Voreinstellungen verwendet. Die Tabelle zeigt nur eine Auswahl der existierenden Matrixtests.

Lerneffekte bei Sprachverständnistests in Satzform

Nachfolgend soll der bei internationalen Sprachverständlichkeitstests in Satzform aufgetretene Lerneffekt vergleichend betrachtet werden. Tabelle 4.1 auf Seite 80 zeigt die große Spannweite des Lerneffekts bei verschiedensprachigen Tests.

Anmerkungen zu Tabelle 4.1: Der größte Lerneffekt findet immer zwischen der ersten und zweiten und zwischen der zweiten und dritten Liste statt.

Zusammenfassend zeigt sich bei allen nach dem Grundprinzip von Hagerman (1982) basierenden Satztests bei der Präsentation von 120 bis 280 Sätzen ein kurzfristiger Lerneffekt im Bereich zwischen 1 bis maximal 2,5 dB.

Vorstellung internationaler Studien: Hagerman (1982) beschreibt bei dem von ihm konzipierten schwedischen Satztest einen Kurzzeit-Lerneffekt von 0,1 dB pro Liste (à 10

Sätze) sowie einen zusätzlichen Lerneffekt zwischen der 1. und 2. Liste. Der Lerneffekt betrug zwischen Liste Nummer 1 und 28 etwa 2,2 dB bzw. zwischen Liste Nummer 2 und 28 etwa 1,7 dB (Testkonditionen: 20 hörgesunden Probanden; offene Version; 28x 10 Sätze). [Hagerman](#) unterteilt diesen Lerneffekt in zwei Teile: Das zentrale Erlernen der Worte und den eher peripheren Informationsfluss. Durch das Kennenlernen des Sprachmaterials werden die Worte bekannt und somit vorhersagbarer. Zugeschrieben wird dieser Teil der zentralen Verarbeitung der Sprache. Dieses Phänomen erklärt allerdings nur einen kleinen Teil des kompletten Lerneffekts. Es wird angenommen, dass der größere restliche Teil des Lerneffekts aufgrund des Informationsflusses/-gehalts, der durch das Spektrum des Sprachsignals und des Störrauschens zustande kommt, entsteht. Das Störrauschen wurde aus dem Sprachmaterial generiert, wodurch es aus demselben Spektrum wie das Sprachsignal besteht. Dieser Teil basiert weniger auf zentralen, als auf peripheren Prozessen. Diese Ergebnisse passen gut zu denen von [Plomp und Mimpen \(1979\)](#), die verschiedene Wörter in ihren unterschiedlichen Listen benutzen, was die Annahme unterstützt, dass nur ein kleiner Teil des Lerneffekts von dem Erlernen des Wortmaterials herrührt. Es wird darauf hingewiesen, dass sich der Lerneffekt verändern könnte, wenn das Sprachsignal ohne Rauschen präsentiert werden würde, da er vermutlich von dem Zusammenhang zwischen Sprache und Rauschen, welches durch spektrale Gemeinsamkeiten entsteht, abhängig ist.

Der Oldenburger Kinder-Satztest (kurz: OLKiSA, verkürzte Form des OLSA) besteht aus 3 Wörtern. Bei Test-Retest-Messungen an 43 jungen, normalhörenden Versuchspersonen zeigte sich ein signifikanter Lerneffekt ([Ebner et al., 2008](#)). Dieser war in der Gruppe, welcher keine initiale Übung in Ruhe präsentiert wurde größer (2,0 dB statt 1,3 dB). Der OLKiSA wurde (unter verschiedenen Bedingungen) an insgesamt 43 20-26-jährigen, normalhörenden Versuchspersonen getestet. Den Probanden wurden 4 Listen (zu je 14 Sätzen) im Freifeld unter konstantem Sprachschallpegel (65 dB SPL) dargeboten. Einer Testgruppe wurde der OLKiSA mit einem Probedurchgang in Ruhe präsentiert; eine Untergruppe hatte nach der ersten Testliste eine Pause von 4 Tagen. In allen Gruppen zeigte sich ein signifikanter Lerneffekt.

In dem dänischen Satztest Dantale II zeigte sich ein Lerneffekt von 1,4 dB innerhalb der ersten beiden Messungen bzw. 2,2 dB zwischen der 1. und 8. Messung (über 160 Sätze hinweg; Testkonditionen: 16 hörgesunden Probanden, offene Version, 8x 20 Sätze) ([Wagener et al., 2003](#)). In der Disserationsarbeit von [Wagener \(2003\)](#) betrug der Trainingseffekt beim Dantale II nach acht 20er Listen 0,3 dB (Evaluationsmessung, 60 Normalhörende). [Hernvig und Olsen \(2005\)](#) untersuchten den Lerneffekt bei Dantale II an 34 hörgeschädigten Probanden (mittleres Alter: 68 Jahre). Hier wurden folgende

4 Diskussion

Testkonditionen gewählt: SRT-Bestimmung, konstantes Störgeräusch bei 65 dB SPL; Training: 3 Listen (30 Sätze); Messungen: 9, 12 bzw. 15 Listen (90, 120 bzw. 150 Sätze) je nach Gruppe; Zusammenfassung der Listen in 30-Satz-Blöcke; SRT-Bestimmung aus den letzten 20 Sätzen eines Blocks. Der Lerneffekt in dieser Studie von [Hernvig und Olsen \(2005\)](#) zeigte mit rund 2,7 dB (abgelesen, von dem 2. bis zum 6. 30-Satz-Block) ein ähnliches Ausmaß wie bei [Wagener et al. \(2003\)](#). Aufgrund des unterschiedlichen Alters der Probandenkollektive in den beiden Studien, aber ähnlicher Resultate, liegt der Verdacht nahe, dass der Lerneffekt unabhängig von Alter ist.

In dem französischen Satztest FrMatrix zeigte sich ein Lerneffekt von 1,0 dB (Evaluation-sentence-scoring) zwischen der 1. und 2. Messung bzw. von 1,8 dB zwischen der 1. und 6. Messung ([Jansen et al., 2012](#)). (Dabei galten diese Testkonditionen: 10 hörgesunde Probanden; 6x 20 Sätze.)

Im Spanish matrix sentence test ergab sich ein Lerneffekt von 1,1 dB (in der offenen Version) bzw. 0,8 dB (in der geschlossenen Version) zwischen der 1. und 2. Messung bzw. von 0,5 dB (in der offenen Version) zwischen der 3. und 6. Messung ([Hochmuth et al., 2012](#)). (Folgende Testkonditionen wurden gewählt: hörgesunde Probanden; offene (33) und geschlossene (10) Version; 6x 20 Sätze.)

Der Italian matrix sentence test wies einen Lerneffekt von 1,5 dB zwischen der 1. und 2. Messung bzw. von 1 dB zwischen der 3. und 6. Messung ([Puglisi et al., 2014](#)) auf. (Dabei galten nachstehende Testkonditionen: 26 hörgesunde Probanden; offene Version; 6x 20 Sätze.) [Puglisi et al.](#) begründen die besseren Ergebnisse in der geschlossenen Version dadurch, dass der Proband durch die visuelle Präsentation der Antwortmöglichkeiten das zuvor akustisch Aufgenommene besser zu benennen vermag.

[Wilson et al. \(2003\)](#) zeigten bei der Durchführung eines 3-Wort-Satztests (mit 7-9 Silben) einen signifikanten Lerneffekt je nach Altersgruppe (<30, 40-60, >65) von 6,5/6,8 dB (<30) bzw. 8,4/8,1 dB (40-60) bzw. 7,7/8,4 dB (>65) über 5 Sitzungen hinweg. Der Lerneffekt wird dem *procedural learning* zugeschrieben (u. a. Gewöhnung an Testumfeld, Stimme des Sprechers) und nicht dem *content learning* (Lernen der Wörter bzw. Sätze des Tests). (Es galten folgende Testkonditionen: 3-Wort-Sätze (mit 7-9 Silben); Probanden: 30 Probanden (10 pro Altersgruppen: <30, 40-60, >65) (von normalhörend bis hörgeschädigt); Sitzungen: 5 1h-Sitzungen in einem Zeitraum von 5 bis 10 Tagen; Ermittlung von 8 SRT-Werte pro Sitzung.)

[Yund und Woods \(2010\)](#) teilen den Lerneffekt in einen durch den Inhalt verursachten Lerneffekt (*content learning*) und einen durch das Testprocedere entstehenden Lerneffekt (*procedural learning*). Testmaterial sind zum einen der HINT (24 Listen à 10 Sätze),

zum anderen der QuickSIN (18 Listen).

Im ersten Experiment (HINT) an 8 hörgesunden Probanden zeigten **Yund und Woods (2010)** einen Lerneffekt aufgrund des *content learning* von 2,7 dB und einen nicht signifikanten Lerneffekt aufgrund des *procedural learning* von 0,3 dB. Ziel: Messung des Lerneffekts (content and procedural learning) bei Anwendung des HINT (im Störgeräusch), wenn einige Sätze wiederholt präsentiert werden und andere nicht. Test: HINT Probanden: 8 hörgesunde Probanden. Sitzungen: 5 1h-Sitzungen in einem Zeitraum von 5 bis 13 Tagen und einem Abstand von 1 bis 6 Tagen. 24 Listen in Sitzung 2 - 4: 12 wiederholt (2x in jeder Sitzung, einmal vorwärts, einmal rückwärts). 30 Listen in Sitzung 1 + 5: 12 wiederholt (2x in jeder Sitzung, einmal vorwärts, einmal rückwärts) + 12 einfach. Sitzung 1 - 5: alle Sätze doppelt zur Messung des Lerneffekt (content learning). Sitzung 1 + 5: zusätzlich nur einfach vorkommende Sätze zur Messung des Lerneffekt (procedural learning). Ergebnisse der wiederholten Sätze (content learning): SRT Verbesserung um 2,7 dB. Ergebnisse der einfachen Sätze (procedural learning): SRT Verbesserung um 0,3 dB (nicht signifikant!). Fazit: Lerneffekt aufgrund des content learning, minimaler Lerneffekt aufgrund des procedural learning.

In einem zweiten Experiment an 23 jungen, hörgesunden Probanden zeigten **Yund und Woods (2010)** im HINT einen Lerneffekt aufgrund des *procedural learning* von 0,2 dB pro Sitzung und im QuickSIN eine nicht signifikante Verbesserung um 0,1 dB pro Sitzung. Ziel: Messung/Abschätzung des Lerneffekts (procedural learning) bei Anwendung des HINT und des QuickSIN an einer größeren Probandengruppe (im Störgeräusch) ohne eine Wiederholung der Sätze über die Sitzungen hinweg. Test: HINT und QuickSIN. Probanden: 23 junge, hörgesunde Probanden. Sitzungen: 3 innerhalb eines Zeitraumes von 2 Wochen. 4 HINT 20er Listen pro Sitzung (keine Wiederholungen). 6 QuickSIN Listen pro Sitzung (keine Wiederholungen). Sitzung 1 - 3: nur einfach vorkommende Sätze/Listen zur Messung des Lerneffekt (procedural learning). Ergebnisse der wiederholten Sätze HINT (procedural learning): SRT Verbesserung um 0,2 dB pro Sitzung. Ergebnisse der wiederholten Sätze QuickSIN (procedural learning): SRT Verbesserung um 0,1 dB pro Sitzung (nicht signifikant!).

Das dritte Experiment von **Yund und Woods (2010)** diente der Quantifizierung des Lerneffekts *content and procedural learning* bei Anwendung des HINT und des QuickSIN, wenn diese Satztests in einem Intervall von mehreren Monaten wiederholt werden. Die Tests an 11 älteren (60 -79 Jahre), hörgesunden Probanden ergab in beiden Test einen kleinen, aber signifikanten Lerneffekt aufgrund des *procedural learning*. Die SRT-Verbesserung betrug bei dem HINT 0,5 dB und bei dem QuickSIN 0,4 dB. (Ziel: Messung/Beurteilung des Lerneffekts (content and procedural learning) bei Anwendung

des HINT und des QuickSIN), wenn Satztests in einem Intervall von mehreren Monaten wiederholt werden. Test: HINT und QuickSIN. Probanden: 11 ältere (60 -79 Jahre), hörgesunde Probanden. Sitzungen: 3 + 1 (=Wiederholung nach 3 - 6 Monaten. 4 HINT 20er Listen pro Sitzung (mit Wiederholungen). 6 QuickSIN Listen pro Sitzung (mit Wiederholungen). 3 - 6 Monate vorher an 3 verschiedenen Tagen mit einem Abstand von 12 Tagen. Wiederholung in einer 1h - Sitzung. Sitzung 1 - 3 + Retest: mehrfach vorkommende Sätze/Listen zur Messung des Lerneffekt (procedural and content learning?). Retest: nach einem Intervall (3 - 6 Monaten) erneute Testung mit denselben Sätze/Listen (content learning). Ergebnisse HINT: SRT Verbesserung um 0,5 dB (bei wiederholten Messungen). Ergebnisse QuickSIN: SRT Verbesserung um 0,4 dB.)

Bedeutung von Hörtraining: [Burk und Humes \(2007\)](#) zeigten, dass sich nach einem Hörtraining die Leistung der Probanden sowohl bei bekannten, als auch bei unbekanntem Wörtern signifikant verbessert. Durchgeführt wurde das Hörtraining mit 9 jungen Normalhörenden in einem 5h-Trainingsprotokoll (short-term) sowie 8 jungen Normalhörenden in einem 15h-Trainingsprotokoll (long-term). Weiterhin zeigten [Burk und Humes \(2008\)](#), dass die wiederholte Präsentation von bekannten Wörtern im Störgeräusch zu einer besseren Hörleistung auch bei älteren Probanden führt. (Hier wurden folgende Testkonditionen gewählt: 8 hörgeschädigte Probanden, Training für 12 Wochen, Retest für 14 Wochen.)

4.4 Lösungsvorschläge zur Eradikation des Lerneffekts

Prinzipiell ist das Ziel, einen Satztest zu konstruieren, welcher keinen Lerneffekt aufweist und somit auch die Resultate des tatsächlichen Hörverstehens nicht negativ beeinflusst oder tangiert.

Wiedererkennungs- und Lerneffekte entstehen bei Sprachverständlichkeitstests, wenn innerhalb kurzer Zeit mit exakt denselben Testlisten gearbeitet wird. Die Aussage „Hier könnten neuere Testverfahren wie z. B. der Oldenburger Satztest Abhilfe schaffen“ ([von Wedel, 2001](#)) darf in Anbetracht dieser und anderer Studien, welche deutlich zeigen, dass auch beim Oldenburger Satztest ein Lerneffekt nicht verhindert werden kann, angezweifelt werden.

[Löhler et al. \(2014\)](#) betonen die Notwendigkeit eines Trainings, damit „am Ende der Lernkurve“ valide Resultate erzielt werden. [von Wedel \(2001\)](#) hebt hervor, dass die Trainingsphase bei psychoakustischen Tests je nach Schwierigkeit modifiziert werden

muss. Dabei erfordern anspruchsvolle Aufgaben eine längere Übungsphase.

Lösungsvorschläge: Zur Eradikation bzw. Verringerung des kurzfristigen Lerneffekts wird von nahezu allen Autoren der Sprachverständnistests auf schwedisch, deutsch, dänisch, französisch, spanisch und italienisch der folgende Lösungsansatz vorgeschlagen (Hagerman (1982), Wagener et al. (1999b), Wagener et al. (2003), Jansen et al. (2012), Hochmuth et al. (2012) und Puglisi et al. (2014), Hey et al. (2014)): Zu Übungszwecken, zum *Einhören* und zur Gewöhnung an die Testsituation sollte der Test vor der eigentlichen Datenerhebung in Form von zwei Doppel-Listen zu je 20 Sätzen (insgesamt 40 Sätze) präsentiert werden (Brand und Wagener, 2017). Auf diese Weise kann der Lerneffekt (*residual training effect*) unter vertretbarem zeitlichen Aufwand auf weniger als 1 dB verringert werden (Wagener et al. (1999b), Wagener et al. (2003), Hochmuth et al. (2012), Jansen et al. (2012)). Die erste Übungsliste sollte in einer gut verständlichen Lautstärke dargeboten werden (konstanter SRT, welcher klar über der Sprachverständlichkeitsschwelle liegt). Die zweite Übungsliste sollte mit dem üblichen adaptiven Verfahren präsentiert werden. Diese Anweisung gilt für folgende Matrixtests: American English, Arabic, Finnish, French, German, Italian, Polish, Russian, Spanish, Turkish (HörTech, 2014).

Pro Testung einer Doppel-Liste rechnen Wagener et al. (1999b) dabei mit einem zusätzlichen Zeitaufwand von 5 Minuten. Müller-Deile (2009) schlägt zur Eradikation bzw. Verringerung des langfristigen Lerneffekts neben der Präsentation von 140 Sätzen in Ruhe vor der allerersten Durchführung des Oldenburger Satztests vor, vor jeder neuen Sitzung eine Liste in Ruhe zu messen, um den Probanden bzw. den Patienten wieder mit der Situation und dem Testprinzip vertraut zu machen. Es bleibt bei Müller-Deile (2009) die Frage bezüglich der praktischen Durchführung, beispielsweise die Frage nach der offenen oder geschlossenen Durchführung offen. Schlüter et al. (2012) zeigten, dass der intrasessionale Lerneffekt/Kurzzeit-Lerneffekt durch 2 Übungslisten auf 0,5 dB (oder weniger) sinkt, sodass anschließend valide Ergebnisse entstehen. Bezogen auf die eigenen Ergebnisse erscheint die Präsentation von nur zwei Übungslisten nicht ausreichend zu sein. Drei bis vier Listen wären zur Verringerung des Lerneffekts unter eine signifikante Grenze wünschenswert, da sich erst die Durchläufe drei und vier sowie vier und fünf nicht mehr signifikant voneinander unterscheiden. Beachtet werden muss in diesem Zusammenhang, dass ein zu langes Hörtraining eventuell aufgrund von nachlassender Konzentration wieder zu schlechteren Ergebnissen führen könnte. Das Hörtraining vor der Datenerhebung dürfte damit nicht beliebig verlängert werden. Eine definitive Aussage über die zur endgültigen Eradikation des Lerneffekts nötigen Anzahl

der Übungssätze kann nicht mit letzter Sicherheit getroffen werden. Dazu muss auch der Faktor Testzeit in einem für die Tests angemessenen Rahmen bleiben. Die Ausprägung des Lerneffekts in dieser Arbeit ist verglichen mit den hier aufgeführten Studien deutlich höher. Ob dies an der Vorselektion der Probandenklientel (junges Alter, gute Motivation) liegen kann, muss in Betracht gezogen werden. In welchem Ausmaß verschiedene Hörpathologien und das Alter der Testpersonen eine Rolle spielen können, müsste in anderen Studien analysiert werden. Ältere Patienten könnten mit der Aufgabe überfordert sein, im Verlauf Konzentrationsstörungen zeigen oder angesichts vermeintlich schlechter Resultate resignieren. Mit diesem Hintergrund wäre eine individuelle Anpassung des Hörtrainings abhängig von Alter, Vorkenntnissen über den Test, mentalem Status und Hörpathologie denkbar.

Es soll darauf hingewiesen sein, dass auch bei anderen Sprach- und Satztests Lerneffekte, die in der Praxis zugunsten valider Resultate möglichst gering gehalten werden sollten, existieren.

Yund und Woods (2010) schlagen bei Hörtests die Präsentation von mindestens zwei Übungslisten zur Elimination des anfänglichen Lerneffekts aufgrund des *procedural learning* vor. Sowohl der HINT, als auch der QuickSIN liefern haltbare und sensitive Ergebnisse auch bei mehrfachen Messungen des Sprachverständnisses, vorausgesetzt die Listen/Sätze werden nicht wiederholt. Die Resultate bei der Durchführung des HINT und des QuickSIN im Störgeräusch weichen von bisherigen Ergebnissen, die bei der Durchführung des Satzverständnistests in Ruhe erhoben wurden, insofern ab, als in Ruhe eine SRT-Wert-Verbesserung von 6 bis 9 dB über 5 Sitzungen hinweg, erzielt werden konnte, unabhängig davon, ob die Sätze/Listen einfach oder mehrfach präsentiert wurden. Unterschiede zwischen der Testung im Störgeräusch und in Ruhe scheinen für den Unterschied in der Stabilität der Messungen von Satztests verantwortlich zu sein.

Somit gestalten sich die Lösungsansätze bei anderen Sprach- und Satztests ähnlich wie die zum Oldenburger Satztest. Übungsphasen sind aufgrund der Lerneffekte fast immer nötig. Der relative Nutzen muss im klinisch praktischen Alltag in Anbetracht des erforderlichen, teils erheblichen, zeitlichen Mehraufwands genau abgewogen werden.

Der Göttinger Satztest zeigte in der Studie von Brand und Kollmeier (2002) keinen signifikanten Lerneffekt. Laut Wagener und Brand (2005) kann eine Versuchsperson diesen Test dennoch für gewöhnlich nur einmal durchführen, da die bedeutungsvollen Sätze leicht erinnerlich sind und die Wörter im Kontext erschlossen werden können. Eine wiederholte Messung mit derselben Liste ist erst nach ausreichend vergangener Zeit, etwa einem halben Jahr oder sogar länger wieder möglich.

Korrekturvorschrift: Theoretisch wäre eine mathematische Korrekturvorschrift, im Sinne eines Herausrechnen des Lerneffekts denkbar. Das Wissen um die Hintergründe und Komponenten des Lerneffekts ist allerdings noch nicht soweit erforscht, dass eine sinnvolle Berechnung entstehen kann.

4.5 Interpretation der Berechnungsvorschrift des SRT-Werts

Zur Berechnung des SRT-Werts werden in der Originalpublikation nicht alle SNR-Werte berücksichtigt, sondern nur die Werte vom 12. bis zum errechneten, aber nicht mehr abgespielten 31. Satz (Wagener et al. (1999c), HörTech (2011)). Die Auswahl aller SNR-Werte wäre aufgrund relativ großer Einpendelabweichungen ungünstig. Die Auswahl zu weniger Sätzen wäre aufgrund zu weniger Werte zu ungenau.

Ob die originale Berechnungsvorschrift tatsächlich der optimalen Auswahl der SNR-Werte zutrifft kann anhand der Abbildung 4.2 untersucht und diskutiert werden.

Abbildung 4.1 zeigt eine 3D-Darstellung der SRT-Berechnung mit variierender Anzahl der einbezogenen SNR-Werte gemittelt über alle Testsitzungen und alle Testdurchläufe. Zugehörig sind in Abbildung 4.2 in dreidimensionaler Ansicht die Standardfehler des arithmetischen Mittels bei der jeweiligen Berechnungsvorschrift gemittelt über alle Testsitzungen und alle Testdurchläufe dargestellt.

Am genauesten ist die Messung, wenn der Standardfehler möglichst gering ist. Dieser kann auf zwei verschiedene Arten möglichst klein gehalten werden: Zum einen durch das Berücksichtigen so vieler SNR-Werte wie möglich, zum anderen durch das Nicht-Berücksichtigen von Schwankungen der SNR-Werte zu Beginn einer Testliste. Es muss daher ein Kompromiss zwischen diesen beiden gegensätzlichen Anforderungen gefunden werden.

Abbildung 4.1 zeigt ein räumliches kartesisches Koordinatensystem. Auf der Abszisse findet eine Variation der Berücksichtigung der SNR-Werte (in dB) (2 - 31) statt. Auf der Ordinate wird der letzte für die SRT-Berechnung berücksichtigte SNR-Wert (in dB) variiert. Auf der Applikate ist der errechnete, über die SNR-Werte aller Probanden gemittelte SRT-Wert aufgezeichnet. Es resultiert eine Darstellung aller möglichen SRT-Werte in Abhängigkeit von der Berechnungsvorschrift. Es entsteht eine Art *Wertelandschaft*. Dabei entsprechen Gipfelwerte weniger negativen SRT-Werten und damit einem schlechteren Sprachverständnis. Dies geschieht beispielsweise bei der Berücksichtigung aller Werte: Wie erwartet verschlechtert die Beachtung des anfänglichen Einjustierens der Werte die Resultate.

In der Grafik 4.2 ist auf der Applikate der Standardfehler des arithmetischen Mittels bei der jeweiligen Berechnungsvorschrift gemittelt über alle Testsitzungen und alle Testdurchläufe dargestellt. Auf diese Weise entsteht wieder eine *Wertelandschaft*, welche nach vorne hin und nach rechts abfällt. Zur Verdeutlichung der Höhe der Werte wurden zusätzlich in Farbabstufungen verwendet: Wie auf einer geographischen Landschaftskarte repräsentieren rot-braune Farbtöne hohe, unerwünschte Werte (Berge) und grüne Farbtöne niedrige, gewünschte Werte (Täler) für den Standardfehler. Der tiefste Punkt der *Wertelandschaft* ist durch die dunkelgrüne Farbe gekennzeichnet und befindet sich bei Satz Nummer 8. Dieser Satz wäre somit der am besten geeignete Ausgangswert für die Auswertung. Tatsächlich beginnt die Auswertung bei dem rot umkreisten Satz Nummer 12. Dies entspricht der Berechnungsvorschrift wie in den Originalpublikationen angegeben, mit der auch in dieser Studie gearbeitet wurde.

Einen Grund, warum die Auswertung trotzdem weiterhin bei Satz Nummer 12 beginnen sollte, liefert die Tatsache, dass das Wortmaterial innerhalb der jeweils ersten, zweiten und dritten 10 Sätze einer 30er-Liste ausbalanciert ist (Schraud, 2014).

Aus Übersichtsgründen sind in diesem Kapitel lediglich diese beiden zusammenfassenden Grafiken enthalten. Die Abbildungen der 3D-Darstellung der SRT-Berechnung mit variierender Anzahl der einbezogenen SNR-Werte gemittelt über die jeweils genannte Testsitzungen bzw. den genannten Testdurchlauf sowie deren Standardabweichungen und Standardfehlern finden sich im Anhang.

4.6 Untersuchung der Antwortzeit und Testzeit

Die Abbildungen 3.21, 3.22 und 3.23 (auf Seite 67 und 68) fassen 3.17, 3.18, 3.19 und 3.20 (auf Seite 64, 65, 66 und 67) zusammen und erinnern dabei an 3.7, 3.8 und 3.9 (auf Seite 48).

Mit sinkenden SRT-Werten sinkt auch die Antwortzeit kontinuierlich innerhalb der Sitzungen und Durchgänge.

Abbildung 3.22 zeigt, dass die Antwortzeit der Probanden in einem geschlossenen Testformat je nach Sitzung und Durchgang zwischen 4,6 und 13,5 Sekunden länger ist als im offenen Format. Ursächlich liegt der Grund des Suchens nach der richtigen Lösung auf der Vorlagenkarte und die größere Bereitschaft zu Raten auf der Hand.

Die verlängerte Antwortzeit im geschlossenen Testformat in dieser Arbeit passt zu der verlängerten Testzeit in der Studie von Brand et al. (2004): Diese betrug bei der Testung von einer Liste zu 20 Sätzen in der offenen Version 3,1 Minuten und in der

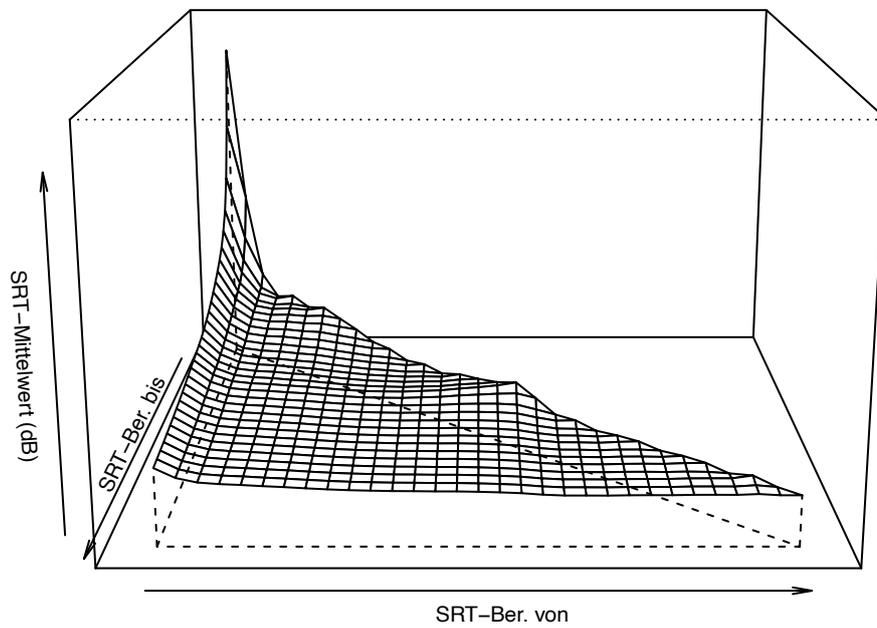


Abbildung 4.1: 3D-Darstellung der SRT-Berechnung mit variierender Anzahl der einbezogenen SNR-Werte gemittelt über alle Testsitzungen und alle Testdurchläufe

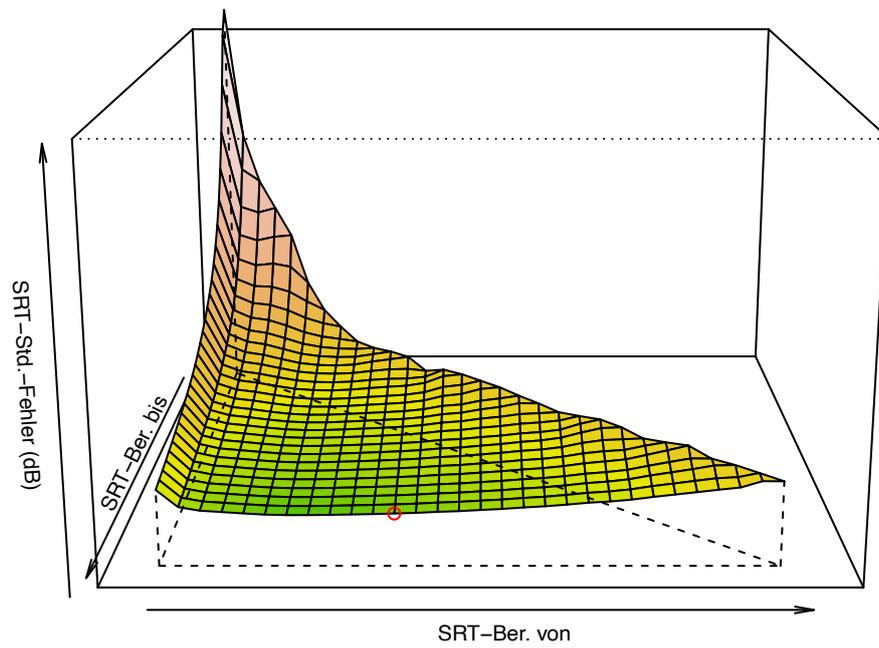


Abbildung 4.2: Standardfehler des arithmetischen Mittels bei der jeweiligen Berechnungsvorschrift gemittelt über alle Testsitzungen und alle Testdurchläufe

geschlossenen 4,9 Minuten. Das Testformat wurde dabei als „Maß für die Schwierigkeit der Aufgabe“ angesehen. In diesem Zusammenhang muss daran erinnert werden, dass die Testzeit nicht nur die Antwortzeit, sondern auch unter anderem die Abspielzeit und die vom Probanden für das Markieren, ggf. Korrigieren der Wörter auf einem Touchscreen benötigten Zeit beinhaltet (Brand et al. (2004, Abbildung 1)). In dieser Arbeit wurde daher nicht die Testzeit, sondern die Antwortzeit herangezogen, um die Komponenten der Einflussfaktoren auf ein Minimum zu reduzieren.

Die Vermutung, dass die Vorlagenkarte Einfluss auf die Testergebnisse nimmt, bestätigt sich nicht. Die Lernkurve unterscheidet sich nicht signifikant im offenen und geschlossenen Format. Im Sinne einer Ablenkung durch die Karte mit einer Notwendigkeit der Konzentration auf akustische *und* optische Reize könnte die geschlossene Testversion als anspruchsvoller bezeichnet werden. Probanden ohne Karte könnten sich ohne visuelle Distraktoren ggf. auch mit geschlossenen Augen auf die Sprachsignale fokussieren. Die Karte könnte auf der anderen Seite natürlich dafür sorgen, dass weniger Verwechslungsfehler entstehen, da nach den Wörtern gesucht werden kann.

Es darf nicht außer Acht gelassen werden, dass in dieser Arbeit junge, normalhörende Probanden untersucht wurden. Dass die Antwortzeit beim Oldenburger Satztest bei älteren, schlechthörenden Patienten signifikant verlängert sein könnte, lässt sich mit der Arbeit von Tun et al. (2010) belegen. Dies könnte vor allem durch eine etwaige Überforderung mit der Testsituation und eine reduzierte kognitive Leistung begründet werden.

Abbildung 3.23 zeigt, dass sich die durchschnittliche Antwortzeit von der ersten bis zur fünfzehnten Liste in etwa halbiert (von 84,0 auf 42,3 Sekunden). Gründe hierfür könnten neben Faktoren wie der Anpassung an die Hörsituation auch die nachlassende Aufmerksamkeit und gesunkene Motivation durch das Verstehen von lediglich etwa der Hälfte des Gehörten sein. Weiterhin ist davon auszugehen, dass die Probanden mit der Karte als Vorlage auch deshalb schneller werden, da der Inhalt der Karte inzwischen komplett gelernt worden ist und der Begriff nicht mehr gesucht werden muss.

4.7 Überlegungen zum Oldenburger Satztest

Unter dem Punkt 1.3 wurden die Anforderungen an einen modernen Sprachtest postuliert. Im Vergleich zu anderen gebräuchlichen Sprachverständnistests werden die Bedingungen beim Oldenburger Satztest besser zusammengebracht. Dennoch birgt die Testkonstruktion ein Dilemma: Während auf der einen Seite die Äquivalenz der Stimuli

für eine steile Diskriminationsfunktion und somit eine hohe Messgenauigkeit und Effizienz sorgt, muss auf der anderen Seite darauf geachtet werden, dass die Äquivalenz der Listen nicht durch Repetition von identischen Signalen garantiert wird, da im anderen Fall Lerneffekte die Korrektheit der Resultate verzerren können. Die derzeit bestmögliche Lösung in der Kompromissfindung zwischen Listenäquivalenz und Umgehung von Lerneffekten scheint die Konstruktion von Matrix-Satztests zu sein. Die hypothetischen Überlegungen konnten allerdings in der Praxis nicht bestätigt werden. Ein Lerneffekt konnte hier nachgewiesen werden und findet sich auch in der Literatur (Wagener et al. (1999c), Müller-Deile (2009), Grimm (2014)).

Der Gebrauch derselben Wörter in einer vorgegebenen Satzstruktur sowie eine Pegelkorrektur von maximal 2 dB SPL nach oben bei schwerer verständlichen Wörtern und nach unten bei akustisch leichter aufgenommenen Wörtern garantiert die Äquivalenz der Testlisten (Wagener et al., 1999c). Der Erfolg des Wort-spezifischen Pegelausgleichs kann in Anbetracht der Resultate aus der Arbeit von Grimm (2014) zumindest teilweise angezweifelt werden.

Eine Verstärkung des Lerneffekts könnte durch die nicht vollends ausgenutzte Möglichkeit der Kombination der Wörter bei der Erstellung der insgesamt 40 Testlisten entstanden sein. Wie eingangs beschrieben, wäre in der Theorie ein Repertoire aus 100.000 verschiedenen Sätzen möglich. Die tatsächliche Anzahl der aufgesprochenen Sätze beläuft sich aber auf lediglich 100. Eine Verringerung des Lerneffekts durch maximale Ausschöpfung aller Möglichkeiten wäre denkbar. Zusätzlich wäre eine ausgeglichene Verteilung der Wörterkombinationen wünschenswert (Schraud, 2014). „Die Test-Retest-Reliabilität kann auch durch das Erinnern und Wiedererkennen von Testelementen bei Mehrfachtestung beeinträchtigt werden, wenn für umfangreiche Untersuchungen nicht genügend Testlisten zur Verfügung stehen.“, konstatiert Steffens (2017b).

Die 40 Testlisten des Oldenburger Satztests (zu je 30 Sätzen) bestehen insgesamt aus 1.200 Sätzen, was bedeutet, dass im gesamten Listenformat jeder der 100 Sätze insgesamt zwölfmal abgespielt wird. In dieser Dissertation wurden jedem Probanden 15 Listen präsentiert. Bei 450 vorgespielten Sätzen hört jeder Proband 50 Sätze fünfmal und 50 Sätze viermal. Aufgrund dieser Wiederholung der Listen könnten vor allem im dritten Testdurchgang deutliche Erinnerungseffekte entstehen. Andererseits ist es fraglich, ob bei einer Vergrößerung des Listenrepertoires um den Faktor 1.000 die im aktuellen Format sehr hohe Äquivalenz der Testlisten aufrecht erhalten werden kann. Bei einer etwaigen Weiterentwicklung oder Modifikation des Oldenburger Satztests könnte hier angeknüpft werden.

Insgesamt stellt die bei der Konstruktion des Oldenburger Satztests angewendete

Methode des Satzbaus nach dem Matrix-Prinzip den derzeit besten Kompromiss zwischen hoher Äquivalenz der Testlisten und Reduktion des Lerneffekts dar.

Der in dieser Arbeit ermittelte Lerneffekt ist unabhängig vom Testformat und größer als eingangs angenommen. Plausible Gründe hierfür sind die Gewöhnung an die Testsituation, die Adaption des Gehörs, das Erlernen des Wortrepertoires und auch die Verwendung identischer Sätze innerhalb der Listen. Diese Effekte reduzieren sich im Laufe der Sitzungen und Durchläufe. Im Laufe der ersten Testsitzung sinkt der Wert des Lerneffekts erst nach der vierten Liste auf unter 0,5 dB, bei der zweiten Sitzung nach der dritten Liste und bei der dritten Sitzung nach der zweiten Liste. Eine völlige Elimination des Lerneffekts wird jedoch zu keinem Zeitpunkt erreicht.

4.8 Stellenwert des Oldenburger Satztests

Der Stellenwert des Oldenburger Satztests kann für verschiedene Einsatzgebiete unterschiedlich betrachtet werden:

Forschung: Zu Forschungszwecken ist der Oldenburger Satztest sehr gut geeignet. Die wiederholte Durchführbarkeit ist aufgrund der Tatsache, dass die 40 Listen praktisch nicht auswendig gelernt oder aktiv im Gedächtnis behalten werden können, garantiert.

HNO-Praxis und klinischer Alltag: In der klinischen Praxis (z. B. im Rahmen von Routineuntersuchungen oder bei der Anpassung von Hörgeräten) birgt der Oldenburger Satztest gewisse Nachteile, durch die sein Einsatzprofil kritisch evaluiert werden muss. Der existierende kurz- und langfristige Lerneffekt muss bei der Wiederholbarkeit der Messungen beachtet und ggf. abgezogen werden. Weiterhin wird durch den kurzfristigen Lerneffekt die Durchführung von Übungslisten nötig, wodurch der Zeitverbrauch in einem inadäquaten Rahmen ansteigt. Wenn ein Training aufgrund nur weniger benötigter Messungen nicht lohnenswert erscheint, wird von [Brand et al. \(2004\)](#) der Göttinger Satztest als eine geeignete Alternative zum Oldenburger Satztest empfohlen.

Medizin-Politik: Ob der Oldenburger Satztest, wie derzeit in der medizin-politischen Diskussion gefordert verpflichtender Bestandteil in der standardisierten Hördiagnostik werden sollte, bleibt v. a. aufgrund des Zeitaufwandes und des beschriebenen Lerneffekts zumindest zu diskutieren. Dabei soll der hohe Nutzen, die gelungene authentische Hörsituation und die klinische Relevanz dieses Tests jedoch in keinster Weise in Frage

gestellt werden. Bei der Neufassung der Hörgeräteverordnung wurde die Alleinstellung des Oldenburger Satztests abgelehnt ([Gemeinsamer Bundesausschuss, 2017](#)).

4.9 Abschluss und Ausblick

In dieser Arbeit wird der Frage nach der Existenz eines kurz- und langfristigen Lerneffekts planmäßig nachgegangen, allerdings beschränkt auf hörgesunde Probanden in der Altersgruppe zwischen 20 und 30 Jahren und nur für einen zeitlichen Abstand von 2 x 8 Wochen. Die Existenz eines intrasessionalen und intersessionalen Lerneffekts konnte nachgewiesen werden. Gründe dafür können diskutiert, allerdings nicht abschließend eindeutig kausal bestätigt werden. Multiple akustische, neurologische, kognitive, psychologische und von der Testkonstruktion bedingte Faktoren scheinen dabei eine Rolle zu spielen. Zur Reduktion und Kompensation des Lerneffekts sind Übungslisten nötig.

Die eingangs als hilfreich angedachte Hinzunahme einer Karte aller Testwörter brachten weder einen Benefit bezüglich der Resultate, noch bei der Testzeit.

Es ergeben sich zusätzliche, auch für die klinische Arbeit relevante Fragestellungen, beispielsweise ob und wie sich der Lerneffekt verändert, wenn für die Messung eine andere Altersgruppen und/oder hörgeschädigte Probanden herangezogen werden und der zeitliche Abstand zwischen den Messungen mehr als zwei Monate bzw. mehrere Jahre beträgt. Eine solche Studie wäre insofern näher an der Realität gelegen, als die Patientenklientel, welcher der Oldenburger Satztest zugeführt wird, in der Regel älter ist und zur Indikationsstellung, Verlaufsbeurteilung oder Befundkontrolle mehrfach innerhalb mehrerer Jahre in der Klinik vorstellig wird. Die Grundsatzfrage, ob auch bei nicht normalhörenden Probanden ein Lerneffekt zu beobachten ist, bestätigt [Müller-Deile \(2009\)](#).

Außerdem gilt, dass eine größere Probandenzahlen die statistische Reliabilität erhöht.

Weiterhin wäre zur Verringerung des Lerneffekts eine Modifikation des Oldenburger Satztests hinsichtlich Wörter-, Satz- und Listenanzahl denkbar und könnte sich positiv (aber auch negativ) auswirken. Darüber hinaus könnte eine Optimierung hinsichtlich des zeitlichen und personellen Aufwandes angestrebt werden. In diesem Zusammenhang wären eine völlig computergesteuerte Testung mit automatischer Instruktion der Probanden und Patienten sowie eine etwaige Verkürzung der Übungsphase zu nennen.

5 Zusammenfassung

Das Heraushören von Stimmen aus einer lauten Umgebung mit Störgeräuschen und Umgebungslärm jeglicher Art ist eine zentrale Anforderung an das menschliche Gehör. Sprachaudiometrische Tests quantifizieren das Sprachverständnis. Es existieren verschiedene Sprachtests in deutscher Sprache mit unterschiedlichem Profil. Bei einigen wird die Filterfunktion des Gehörs getestet, indem zu einer aufgesprochenen Stimme zusätzlich ein sprachähnliches Störgeräusch präsentiert wird.

Ein solcher sprachaudiometrischer Test ist der Oldenburger Satztest, entwickelt von [Wagener et al. \(1999c\)](#) nach dem Vorbild des Satztests nach [Hagerman \(1982\)](#). Jeder Satz des Oldenburger Satztests setzt sich aus fünf Wörtern zusammen. Die Satzstruktur mit einem Namen als Subjekt, einem Verb als Prädikat und einem dreiteiligen Objekt aus Zahlwort, Adjektiv und Substantiv, wird durchweg beibehalten. Für jede der fünf Satzpositionen stehen zehn verschiedene Wörter zur Verfügung, die beliebig miteinander kombiniert werden können. Dargeboten werden die Sätze parallel zu einem Störrauschen. Der Einstellungsmodus der Satztestentwickler wurde in dieser Arbeit beibehalten: Während das Rauschen konstant bei 65 dB SPL gehalten wurde, veränderte sich der Sprachsignalschallpegel zur Bestimmung der Fünfzig-Prozent-Sprachverständlichkeitsschwelle des Signal/Rausch-Verhältnisses (SRT-Wert) adaptiv.

Das Ziel dieser Dissertation stellte das strukturierte Erfassen des kurz- und langfristigen Lerneffekts bei der mehrfachen Durchführung des Oldenburger Satztests in seiner computerisierten Form dar. Der kurzfristige Lerneffekt findet intrasessional, der langfristige Lerneffekt intersessional statt. Speziell der langfristige Lerneffekt (über mehrere Testsitzungen hinweg) wurde bisher nicht genauer untersucht. Das untersuchte Versuchspersonenkollektiv (32 Probanden) bestand aus 20- bis 29-jährigen, gemäß Tonaudiogramm normalhörenden, muttersprachlich deutschen Probanden. Die Versuchspersonen wurden dreimal im Abstand von jeweils 7-9 Wochen getestet (3 Testsitzungen à 5 Testdurchläufe à 30 Sätze). Untersucht wurde der SRT_{50} -Wert, als der Wert, bei dem 50% des Sprachsignals korrekt verstanden wird sowie der daraus ableitbare Lerneffekt. Als potentielle Einflussfaktoren auf die Ergebnisse des Oldenburger Satztests wurden das Geschlecht, die Testsitzung, der Testdurchlauf, der Testmodus (offen versus geschlossen durch Vorlage einer Karte mit dem gesamten Wortrepertoire) und

5 Zusammenfassung

im geschlossenen Testdesign die Leserichtung in Betracht gezogen und untersucht. Dementsprechend erfolgte eine Einteilung in Gruppen und Untergruppen.

Nur die beiden zeitabhängigen Faktoren Testsitzung und Testdurchlauf konnten als signifikante Einflussfaktoren bestätigt werden. Bei der Auswertung der SRT-Werte zeigte sich unabhängig von der Gruppenzugehörigkeit ein signifikanter Lerneffekt sowohl innerhalb einer Testsitzung, als auch über die Testsitzungen hinweg: Lerneffekt von 2,61 dB innerhalb des ersten Durchgangs, 1,77 dB innerhalb des zweiten Durchgangs und 1,65 dB innerhalb des dritten Durchgangs. Insgesamt beträgt der Lerneffekt von Beginn der 1. Sitzung bis zum Ende der 3. Sitzung 4,13 dB. Eine komplette Sättigung des Lerneffekts konnte nicht beobachtet werden. Einhergehend mit der Verbesserung zeigte sich auch eine zunehmend schnellere Antwortzeit. Eine vollständige Klärung der zugrunde liegenden Mechanismen kann dabei nicht gegeben werden.

Die hier erarbeiteten Resultate bestätigen somit den kurzfristigen Lerneffekt und zeigen eindeutig einen langfristigen Lerneffekt, der in der Forschung und im klinischen Alltag beachtet werden sollte.

6 Literaturverzeichnis

- Ahlich G. 1985. Anmerkungen zum Freiburger Sprachverständnistest. *Sprache Stimme Gehör*, 9:1–6.
- Akeroyd M. A. *The psychoacoustics of binaural hearing*. International Journal Of Audiology; 45 (Supplement 1):S25–S33, 2006.
- Arons B. 1992. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12(7):35–50.
- Baljić I., Winkler A., Schmidt T., und Holube I. 2016. Untersuchungen zur perzeptiven äquivalenz der Testlisten im Freiburger Einsilbertest. *HNO*, 64(8):572–583.
- Batsoulis C. und Lesinski-Schiedat A. 2017. Sprachaudiometrie in der Begutachtung des Hörvermögens. *HNO*, 65(3):203–210.
- Behrends J. C. *Physiologie: 93 Tabellen*. Duale Reihe. Thieme, 2010. URL <https://books.google.de/books?id=LHyjLfia7HIC>.
- Benninghoff A. und Drenckhahn D. *Taschenbuch Anatomie*. Elsevier, Urban & Fischer, 2008. URL <https://books.google.de/books?id=Wwt5LwuRIsAC>.
- Blauert J. *Räumliches Hören*. S. Hirzel Verlag Stuttgart, 1972.
- Bortz J. und Schuster C. *Statistik für Human- und Sozialwissenschaftler. Lehrbuch mit Online-Materialien*. Springer-Verlag, 2010.
- Brand T. und Kollmeier B. 2002. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America*, 111(6):2801–2810.
- Brand T. und Wagener K. C. 2017. Eigenschaften, Leistungen und Grenzen von Matrixtests. *HNO*, 65(3):182–188.
- Brand T., Wittkop T., Wagener K. C., und Kollmeier B. Vergleich von Oldenburger Satztest und Freiburger Wörtertest als geschlossene Versionen. In *Proceedings of the 7th annual meeting of the Deutsche Gesellschaft für Audiologie (DGA), Leipzig*, 2004.

- Bronkhorst A. W., Brand T., und Wagener K. C. 2002. Evaluation of context effects in sentence recognition. *The Journal of the Acoustical Society of America*, 111(6): 2874–2886.
- Burk M. H. und Humes L. E. 2007. Effects of training on speech recognition performance in noise using lexically hard words. *Journal of Speech, Language, and Hearing Research*, 50(1):25–40.
- Burk M. H. und Humes L. E. 2008. Effects of long-term training on aided speech-recognition performance in noise in older adults. *Journal of Speech, Language, and Hearing Research*, 51(3):759–771.
- Cherry E. C. *Some experiments on the recognition of speech with one and two ears*, Bd. 25. The Journal Of The Acoustical Society Of America, 1953.
- Claßen J. und Schnitzler A. *Interventionelle Neurophysiologie: Grundlagen und therapeutische Anwendungen*. Thieme, 2012. URL <https://books.google.de/books?id=s2Q2qy0g1ToC>.
- Dietz A., Buschermöhle M., Aarnisalo A. A., Vanhanen A., Hyyrynen T., Aaltonen O., Löppönen H., Zokoll M. A., und Kollmeier B. 2014. The development and evaluation of the Finnish Matrix Sentence Test for speech intelligibility assessment. *Acta otolaryngologica*, 134(7):728–737.
- Domes C. C. *Evaluierung und Weiterentwicklung eines Prosodieverständnistests*. Dissertation, Julius-Maximilians-Universität Würzburg, 2009.
- Dudel J., Menzel R., und Schmidt R. F. *Neurowissenschaft: Vom Molekül zur Kognition*. Springer-Verlag, 2013.
- Ebner K., T. Steffens, und J. Hellbrück. 2008. Sprachverstehen in Ruhe und im Störgeräusch und Lerneffekte bei normalhörenden und unilateral sowie sequentiell bilateral kochleaimplantierten Kindern. *Zeitschrift für Audiologie*, 47 (3), 100-110., 47 (3):100–110.
- Feldmann H. 2004. 200 Jahre Hörprüfungen mit Sprache, 50 Jahre deutsche Sprachaudiometrie – ein Rückblick. *Laryngo-Rhino-Otologie*, 83(11):735–742.
- Friesecke A. *Die Audio-Enzyklopadie: Ein Nachschlagewerk für Tontechniker*. De Gruyter Reference. De Gruyter, 2014. URL <https://books.google.es/books?id=iqboBQAAQBAJ>.

- Gekle M. *Physiologie*. Thieme, 2010. URL <https://books.google.de/books?id=tiRctmU2KzkC>.
- Gemeinsamer Bundesausschuss . 2017. Hilfsmittel-Richtlinie des Gemeinsamen Bundesausschusses über die Verordnung von Hilfsmitteln in der vertragsärztlichen Versorgung (Stand: 17. Februar 2017). *Hilfsmittel-Richtlinie*.
- Green D. M. 1976. An Introduction to Hearing.
- Grimm K. *Lern- und Habituationseffekte beim Einsatz des Oldenburger Satzverständnistestes bei jungen Normalhörenden*. Dissertation, Julius-Maximilians-Universität Würzburg, 2014.
- Hagerman B. 1982. Sentences for testing speech intelligibility in noise. *Scandinavian audiology*, 11(2):79–87.
- Hahlbrock K.-H. Juli 1953. Über Sprachaudiometrie und neue Wörtertreste. *Archiv für Ohren-, Nasen- und Kehlkopfheilkunde*, 162(5):394. ISSN 1434-4726. URL <http://dx.doi.org/10.1007/BF02105664>.
- Hebb D. O. 2002. The organization of behavior. 1949. *New York Wiley*.
- Hernvig L. H. und Olsen S. O. 2005. Learning effect when using the Danish Hagerman sentences (Dantale II) to determine speech reception threshold: El efecto de aprendizaje al utilizar oraciones danesas Hagerman (Dantale II) para determinar el umbral de recepcion del habla. *International journal of audiology*, 44(9):509–512.
- Hey M., Hocke T., Hedderich J., und Müller-Deile J. 2014. Investigation of a matrix sentence test in noise: reproducibility and discrimination function in cochlear implant patients. *International journal of audiology*, 53(12):895–902.
- Hochmuth S., Brand T., Zokoll M. A., Castro F. Z., Wardenga N., und Kollmeier B. 2012. A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International journal of audiology*, 51(7):536–544.
- Holube I. und Schlüter A. Lerneffekte des Oldenburger Satztests: Die Einflüsse von Schwerhörigkeit und Zeitkompression auf Kurz- und Langzeiterneffekte. Diplomarbeit, Institut für Hörtechnik und Audiologie Jade Hochschule Oldenburg / Elsfleth / Wilhelmshaven, 2012.

- Hoppe U. und Hast A. 2017. Sprachaudiometrie bei der Indikation von Hörhilfen und Hörimplantaten. *HNO*, 65(3):195–202.
- Hoth S. 2016. The Freiburg speech intelligibility test: A pillar of speech audiometry in German-speaking countries. *HNO*, 64:540–548.
- Houben R. und Dreschler W. A. 2015. Optimization of the Dutch matrix test by random selection of sentences from a preselected subset. *Trends in hearing*, 19: 2331216515583138.
- Huppelsberg J. und Walter K. *Kurzlehrbuch Physiologie*. KURZLEHRBUCH. Thieme, 4. Ausgabe, 2013. URL <https://books.google.de/books?id=6YhrAAAAQBAJ>.
- HörTech . *Oldenburger Satztest Bedienungsanleitung*. HörTech, 1.0 Ausgabe, 2011.
- HörTech . 2014. International Matrix Tests. *HörTech*.
- ISO 7029:2000 . 2000. Acoustics – Statistical distribution of hearing thresholds as a function of age. *International Organization for Standardization, Standards catalogue, Acoustics*.
- ISO 7029:2017 . 2017. Acoustics – Statistical distribution of hearing thresholds related to age and gender. *International Organization for Standardization, Standards catalogue, Acoustics*.
- Jansen S., Luts H., Wagener K. C., Kollmeier B., Del Rio M., Dauman R., James C., Fraysse B., Vormè E., s, und Frachet B. 2012. Comparison of three types of French speech-in-noise tests: A multi-center study. *International journal of audiology*, 51(3): 164–173.
- Kießling J. *Endlich wieder besser hören: aktiv dabei: mehr Lebensfreude mit Hörgeräten ; wie Sie damit optimal klarkommen ; alles über die neuen digitalen Hör-Systeme*. TRIAS ärztlicher Rat. TRIAS, 2002. URL <https://books.google.de/books?id=I1NWIjo5LmMC>.
- Kießling J. 2010. Moderne Sprachtests auf dem Weg von der Forschung in die Praxis. *HNO*, 58(6):595–596.
- Kießling J., Kollmeier B., und Diller G. *Versorgung und Rehabilitation mit Hörgeräten*. Thieme, 2008. URL <https://books.google.com/books?id=ryM5-9Mj2yMC>.

- Klockgether S., Epp B., und Verhey J. L. 2011. Einfluss von auditorischer Bewegung auf Lokalisation unter Freifeldbedingungen. *Fortschritte der Akustik*.
- Knopke S. und Olze H. September 2017. Hörrehabilitation mithilfe von Cochleaimplantaten und kognitive Fähigkeiten. *HNO*, Seite 1. ISSN 1433-0458. URL <http://dx.doi.org/10.1007/s00106-017-0423-z>.
- Kollmeier B. *Moderne Verfahren der Sprachaudiometrie*. Buchreihe audiologische Akustik. Median-Verlag von Killisch-Horn, 1992. URL <https://books.google.de/books?id=3ugWAwAACAAJ>.
- Kollmeier B., Lenarz T., Kießling J., Müller-Deile J., Steffens T., Wedel v. H., Döring W. H., Buschermöhle M., Wagener K. C., und Brand T. 2014. Zur Diskussion um den Freiburger Einsilbertest. *HNO*, 62:49–56.
- 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. *HNO*, 59(10):1012.
- Kuyper P. 1972. The cocktail party effect. *International Journal of Audiology*, 11(5-6): 277–282.
- König E. R. 1979. Kurzer Historischer Rückblick auf die Entwicklung der Hörhilfemittel. *6. Audio-Symposium*. Brommer AG. Zürich, 6.
- Lehnhardt E. und Laszig R. *Praxis der Audiometrie*. Thieme, 2009. URL https://books.google.de/books?id=bJnR_YQ3gYEC.
- Lenarz T. und Boenninghaus H.-G. *Hals-Nasen-Ohren-Heilkunde für Studierende der Medizin*. Springer Verlag, 13. Ausgabe, 2007.
- Lewald J. und Getzmann S. 2015. Electrophysiological correlates of cocktail-party listening. *Behav. Brain Res.*, 292:157–166.
- Löhler J., Akcicek B., Pilnik M., Saager-Post K., Dazert S., Biedron S., Oeken J., Mürbe D., Löbert J., Laszig R., Wesarg T., Langer C., Plontke S., Rahne T., Machate U., Noppeney R., Schultz K., Plinkert P., Hoth S., Praetorius M., Schlattmann P., Meister E.F., Pau H.W., Ehrt K., Hagen R., Shehata-Dieler W., Cebulla M., Walther L.E., und Ernst A. 2013. Evaluation des Freiburger Einsilbertests im Störschall. *HNO*, 61(7): 586–591.

- Löhler J., Akcicek B., Wollenberg B., und R. Schönweiler. 2014. Umsetzung der neuen Qualitätssicherungsvereinbarung zur Hörgeräteversorgung im Praxisalltag. *HNO*, 62 (9):667–682.
- Meister H. März 2017. Sprachaudiometrie, Sprachwahrnehmung und kognitive Funktionen. *HNO*, 65(3):189. URL <http://dx.doi.org/10.1007/s00106-016-0229-4>.
- Meister H., Schreitmüller S., Grugel L., Landwehr M., Wedel H., Walger M., und Meister I. Juli 2011. Untersuchungen zum Sprachverstehen und zu kognitiven Fähigkeiten im Alter. *HNO*, 59(7):689. URL <http://dx.doi.org/10.1007/s00106-011-2262-7>.
- Michel O. 2014a. Die Beurteilung der Normalhörigkeit in der Hals-Nasen-Ohrenärztlichen Begutachtung. *HNO*, 62(5):382–384.
- Michel O. 2014b. Grade der Normalhörigkeit. *HNO*, 62(9):664. ISSN 1433-0458. URL <http://dx.doi.org/10.1007/s00106-014-2889-2>.
- Mrowinski D. *Audiometrie: eine Anleitung für die praktische Hörprüfung ; 10 Tabellen ; mit CD-ROM AUDIOSIM-Lernprogramm für die Tonschwellenaudiometrie mit Vertäufung*. Thieme Flexible Taschenbücher. Thieme, 2006. URL https://books.google.de/books?id=AKoAVPa5c_EC.
- Müller-Deile J. 2009. Sprachverständlichkeitsuntersuchungen bei Cochleaimplantatpatienten. *HNO*, 57(6):580–592.
- Nagel A. Auswirkung eines Hörtrainings auf das Sprachverstehen im Störlärm. Diplomarbeit, Hochschule Aalen für Technik und Wirtschaft, 2011.
- Ozimek E., Warzybok A., und Kutzner D. 2010. Polish sentence matrix test for speech intelligibility measurement in noise. *International journal of audiology*, 49(6):444–454.
- Pape H. C., Kurtz A., Silbernagl S., Persson A. B., Brenner B., Burckhardt G., Draguhn A., Ehmke H., Eysel U., und Fandrey J. *Physiologie*. DeL. Thieme, 2014. URL <https://books.google.de/books?id=l6bTBAAQBAJ>.
- Papke R. *Sprachverstehen im Störgeräusch und Lokalisationsfähigkeit Normalhörender in Abhängigkeit vom Lebensalter*. Dissertation, Julius-Maximilians-Universität Würzburg, 2013.
- Paulsen F. und Waschke J. *Sobotta, Atlas der Anatomie des Menschen Band 3: Kopf, Hals und Neuroanatomie*. Atlas der Anatomie des Menschen: [mit Online-Zugang zu

- www.e-sobotta.de]. Elsevier Health Sciences Germany, 2011. URL <https://books.google.de/books?id=JXLt7zjcJAwC>.
- Plomp R. und Mimpen A. M. 1979. Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology*, 18(1):43–52.
- Probst R. Januar 2013. Moderne Sprachaudiometrie in der Begutachtung? *HNO*, 61(1): 12. ISSN 1433-0458. URL <http://dx.doi.org/10.1007/s00106-012-2620-0>.
- Pschyrembel W. *Pschyrembel Klinisches Wörterbuch*. de Gruyter, 261. Ausgabe, 2007. URL <https://books.google.de/books?id=HUbhGAAACAAJ>.
- Ptok M. und Kießling J. 2004. Sprachperzeption. *HNO*, 52(9):824–830.
- Puglisi G. E., Warzybok A., Hochmuth S., Astolfi A., Prodi N., Visentin C., und Kollmeier B. Construction and first evaluation of the italian matrix sentence test for the assessment of speech intelligibility in noise. In *The Proceedings of Forum Acusticum*, 2014.
- Qualen J. F. *Evaluation des Einsilber-Sprachmaterials M-2007 und Entwurf einer Methodik für die Zusammenstellung gleichwertiger Listen*. Dissertation, Julius-Maximilians-Universität Würzburg, 2010.
- R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL <https://www.R-project.org>.
- Rahne T. März 2017. Zielgenaue Anwendung der Sprachaudiometrie. *HNO*, 65(3):180. ISSN 1433-0458. URL <http://dx.doi.org/10.1007/s00106-016-0299-3>.
- Reiß M. *Facharztwissen HNO-Heilkunde: Differenzierte Diagnostik und Therapie*. Springer, 2009. URL <https://books.google.de/books?id=VTKiTs1k0JYC>.
- Schlüter A., Holube I., und Lemke U. 2012. Trainingseffekte bei normaler und schneller Sprache. *15. Jahrestagung der Deutschen Gesellschaft für Audiologie*.
- Schmidt R. F. und Lang F. *Physiologie des Menschen: mit Pathophysiologie*. Springer-Lehrbuch. Springer London, Limited, 30. Ausgabe, 2007. URL <https://books.google.de/books?id=k5sfBAAAQBAJ>.
- Schmidt T. und Baljić I. 2016. Untersuchung zum Trainingseffekt des Freiburger Einsilbertests. *HNO*, 64(8):584–588.

- Schorn K. 2004. Hörgeräteüberprüfung in der Praxis. *HNO*, 52(10):875–885.
- Schorn K. März 2006. Die Aufgaben des Hals-Nasen-Ohren-Arztes bei der Hörgeräteversorgung. *HNO*, 54(3):233. URL <http://dx.doi.org/10.1007/s00106-006-1378-7>.
- Schraud T. Detaileigenschaften des Oldenburger Satztests und Ansätze zur Weiterentwicklung zu einem deutschsprachigen Matrix-Test. Diplomarbeit, Fachhochschule Lübeck, 2014.
- Schweizer K. *Leistung und Leistungsdiagnostik*. Springer, 2006. URL <https://books.google.de/books?id=vXQyjEn1Rf0C>.
- Schünke M., Schulte E., und Schumacher U. *Kopf, Hals und Neuroanatomie*. Prometheus. Thieme, 2. Ausgabe, 2009. URL <https://books.google.de/books?id=oNQhARc1340C>.
- Silbernagl S. *Taschenatlas Physiologie*. DeL. Thieme, 2012. URL https://books.google.de/books?id=nvff2_nmqn4C.
- Steffens T. 2015. Der Freiburger Sprachtest: ein Klassiker mit Stärken und Schwächen. *HNO Nachrichten*, 45(4):22–25.
- Steffens T. jun 2016. Verwendungshäufigkeit der Freiburger Einsilber in der Gegenwartssprache. *HNO*, 64(8):549–556.
- Steffens T. März 2017a. Die systematische Auswahl von sprachaudiometrischen Verfahren. *HNO*, 65(3):219. ISSN 1433-0458. URL <http://dx.doi.org/10.1007/s00106-016-0249-0>.
- Steffens T. 2017b. Die systematische Auswahl von sprachaudiometrischen Verfahren. *HNO*, 65(3):219–227.
- Sukowski H., Brand T., Wagener K. C., und Kollmeier B. 2009. Untersuchung zur Vergleichbarkeit des Freiburger Sprachtests mit dem Göttinger Satztest und dem Einsilber-Reimtest nach von Wallenberg und Kollmeier. *HNO*, 57(3):239.
- Tun P. A., Benichov J., und Wingfield A. 2010. Response latencies in auditory sentence comprehension: effects of linguistic versus perceptual challenge. *Psychology and aging*, 25(3):730.

- Wedel von H. 2001. Fehlermöglichkeiten in der Ton- und Sprachaudiometrie. *HNO*, 49 (11):939. URL <http://dx.doi.org/10.1007/s001060170025>.
- 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. *International journal of audiology*, 44(3):144–156.
- Wagener K. C., Brand T., und Kollmeier B. 2006. The role of silent intervals for sentence intelligibility in fluctuating noise in hearing-impaired listeners. *International Journal of Audiology*, 45(1):26–33.
- Wagener K. C., Brand T., und Kollmeier B. 2007. International cross-validation of sentence intelligibility tests. *8th EFAS Congress / 10th Congress of the German Society of Audiology*.
- Wagener K. C., Josvassen J. L., und Ardenkjær R. 2003. Design, optimization and evaluation of a Danish sentence test in noise. *International journal of audiology*, 42 (1):10–17.
- Wagener K. C., T. Brand, und Kollmeier B. 1999a. Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics* 38, 44-56., 38:44–56.
- Wagener K. C., T. Brand, und Kollmeier B. 1999b. Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics* 38, 86-95., 38:86–95.
- Wagener K. C., V. Kühnel, und Kollmeier B. 1999c. Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Zeitschrift für Audiologie/Audiological Acoustics* 38, 4-15., 38:4–15.
- Wagener K.C. *Factors influencing sentence intelligibility in noise*. Dissertation, Universität Oldenburg, 2003.
- Warzybok A., Zokoll M., Wardenga N., Ozimek E., Boboshko M., und Kollmeier B. 2015. Development of the russian matrix sentence test. *International journal of audiology*, Seiten 1–9.
- Wilson R. H., Bell T. S., und Koslowski J. A. 2003. Learning effects associated with repeated word-recognition measures using sentence material. *J Rehabil Res Dev.*, 40.

Yund E. W. und Woods D. L. 2010. Content and procedural learning in repeated sentence tests of speech perception. *Ear and hearing*, 31(6):769–778.

Zervos-Kopp J. *Anatomie, Biologie und Physiologie: Ergotherapie Prüfungswissen*. Thieme, 2013. URL https://books.google.de/books?id=_Rw7itfVgMcC.

Zokoll M. A., Fidan D., Türkyılmaz D., Hochmuth S., Ergenç İ., Sennaroğlu G., und Kollmeier B. 2015. Development and evaluation of the Turkish matrix sentence test. *International journal of audiology*, 54(sup2):51–61.

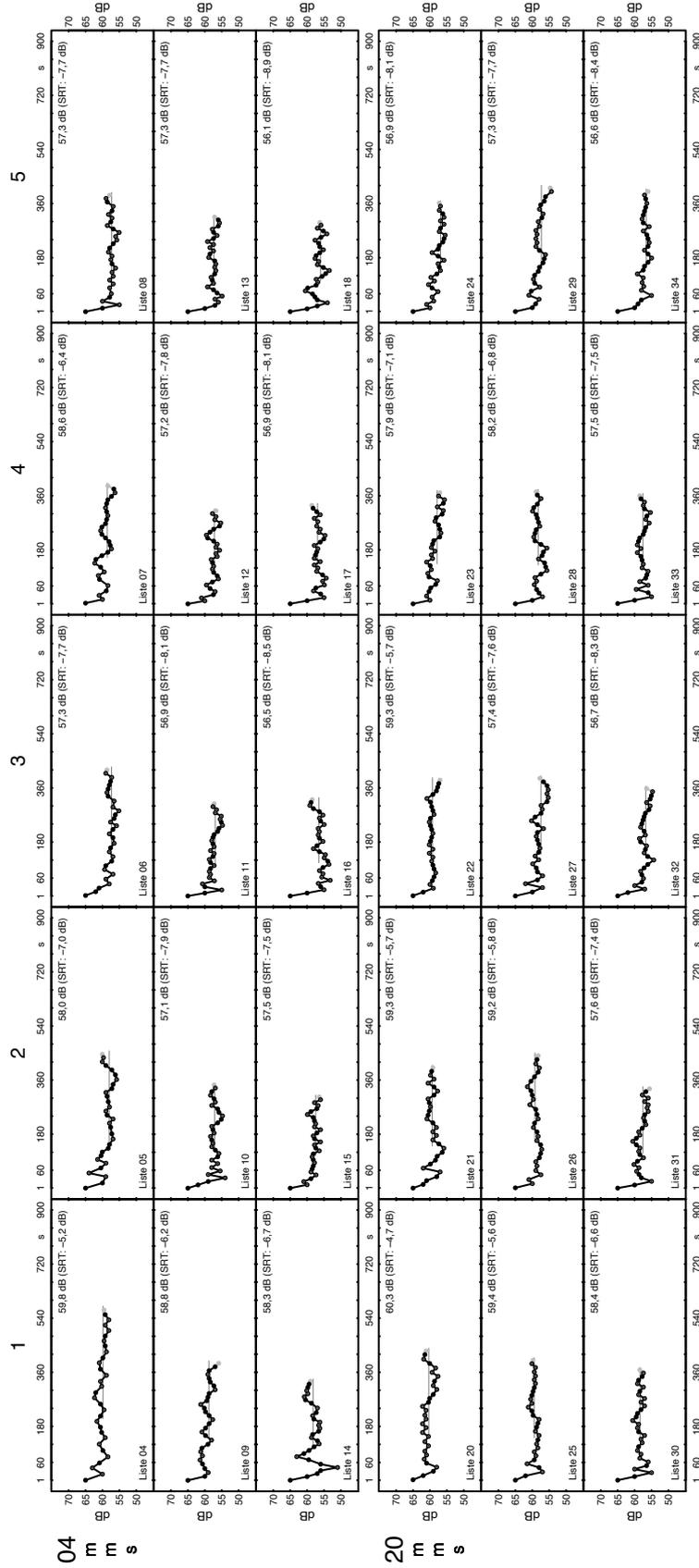
A Anhang: Übersichtstabellen von Rohdaten

A.1 SRT-Werte aller Probanden aller Testsitzungen aller Durchläufe

Dargestellt sind die zeitabhängigen SRT-Verlaufskurven aller 38 Probanden. Ein Proband füllt dabei insgesamt 15 Kacheln (3 Zeilen und 5 Spalten): Die 3 Zeilen repräsentieren die 3 Sitzungen, die 5 Spalten alle dazugehörigen 5 Testdurchläufe. Geordnet wurde nach Geschlecht (m = männlich; w = weiblich), Gruppenzugehörigkeit (m = mit Karte; o = ohne Karte) und der etwaigen Vorlage einer Karte (s = spaltenweise; z = zeilenweise; n = keine Karte und daher auch keine Vorlage). Daraus ergibt sich die Beschriftung *mms*, *mmz*, *mon* und *wms*, *wmz*, *won*. Am linken äußeren Rand steht die Nummer des Probanden. Die Nummern der 32 Probanden der Hauptgruppe sind schwarz, die der nicht in den Auswertungen berücksichtigten sind grau dargestellt. Am oberen Rand sind die einzelnen Testdurchgänge von 1 bis 5 durchnummeriert. In jeder Kachel sind die jeweiligen abgespielten 30 SNR-Werte in schwarz und der errechnete 31. SNR-Wert in grau in ein Koordinatensystem eingetragen (Abszisse: Zeit (in Sekunden); Ordinate: Schalldruckpegel (in dB SPL)). Oben rechts in jeder Kachel steht der SNR- und SRT-Wert des Durchgangs. Unten links wurde die jeweilige Listennummer notiert. (Die Listennummerierung wurde wie in [2](#) beschrieben durchgeführt; lediglich bei Proband Nummer 2 wurde in der dritten Testsitzung im fünften Durchgang nicht wie geplant die Liste 16, sondern die Liste 14 abgespielt.)

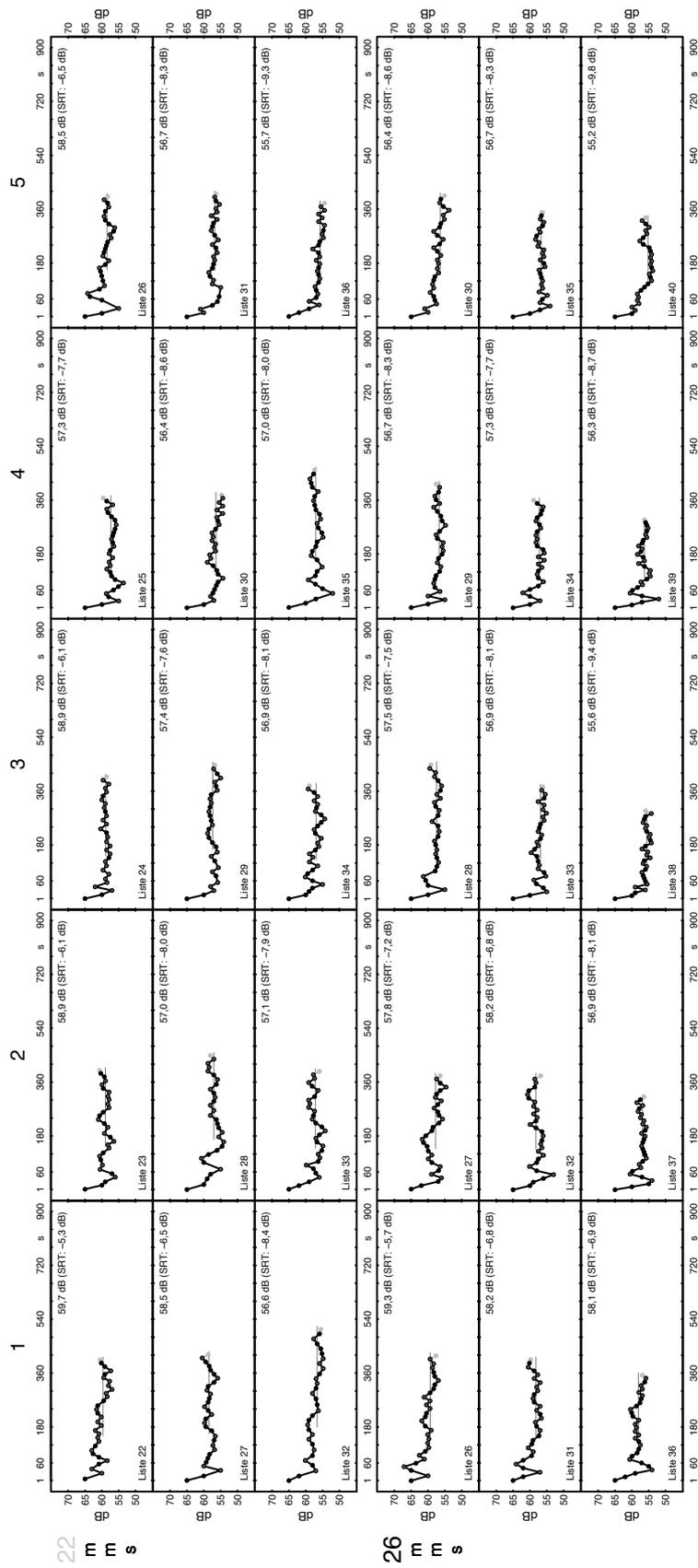
Aus Gründen der besseren Übersicht sind in Kapitel [3](#) auf den Seiten [156](#) und [157](#) lediglich Zusammenfassungen der SRT-Verlaufsdaten zu sehen.

Probanden 04 , 20

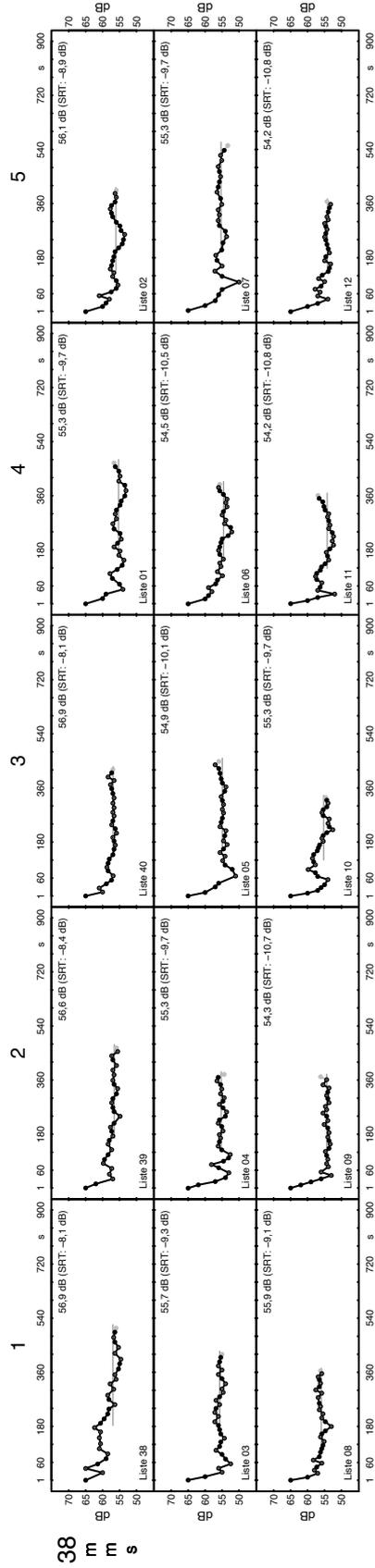


A.1 SRT-Werte aller Probanden aller Testsituationen aller Durchläufe

Probanden 22, 26

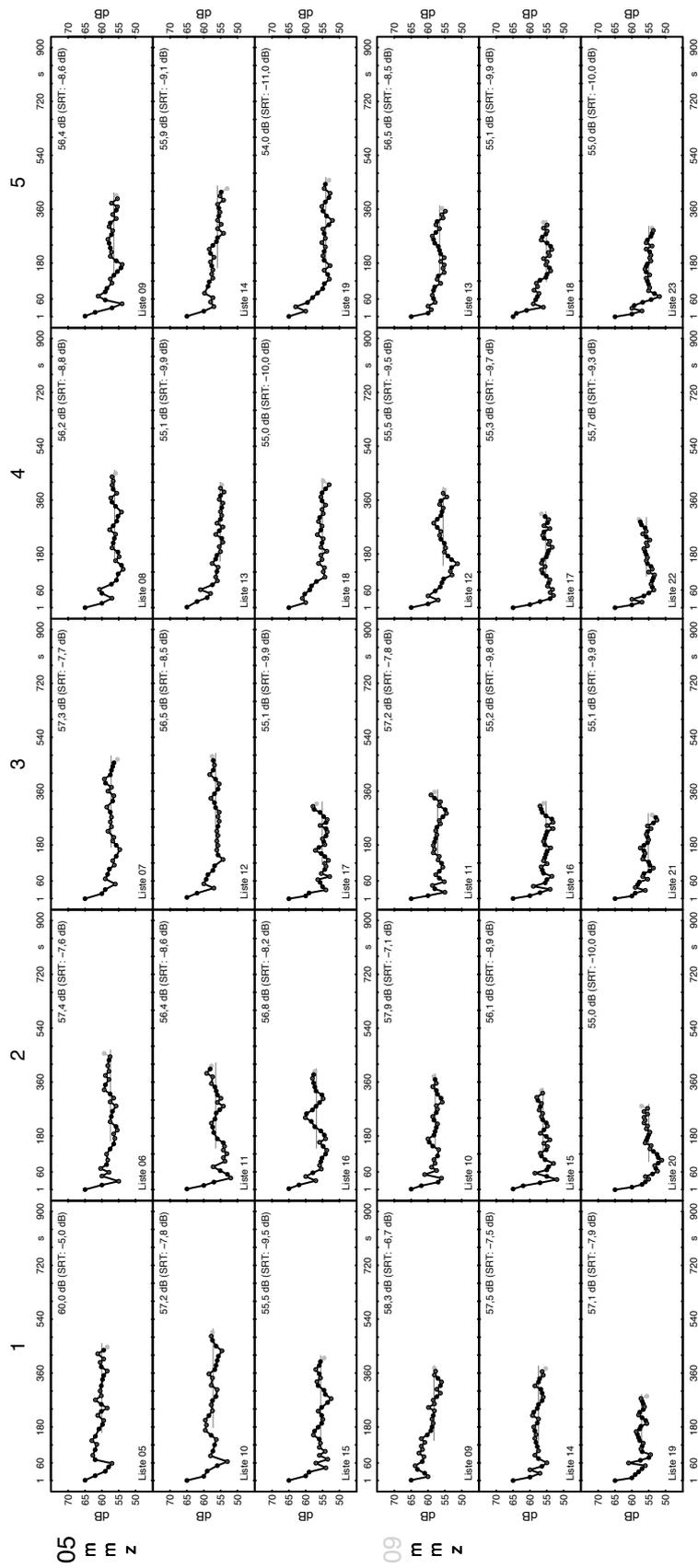


Proband 38

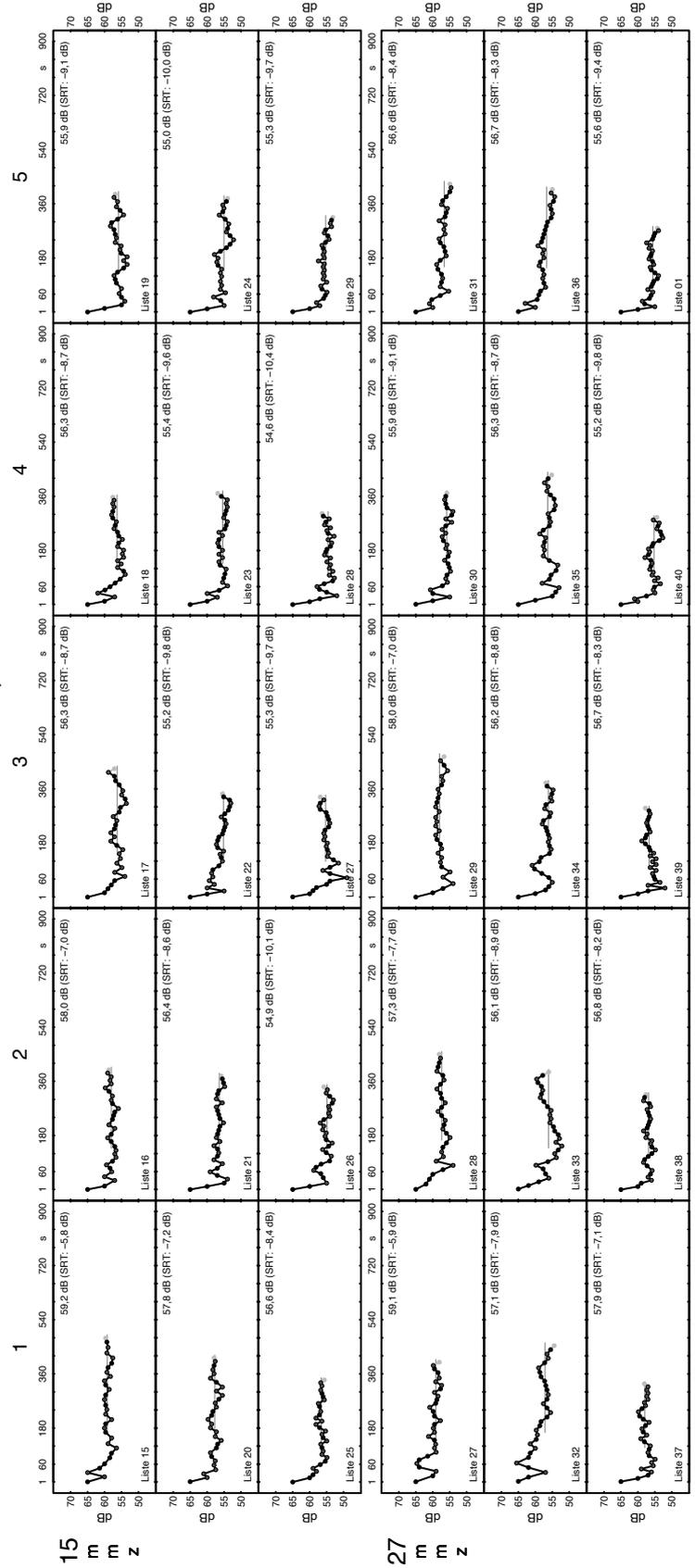


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

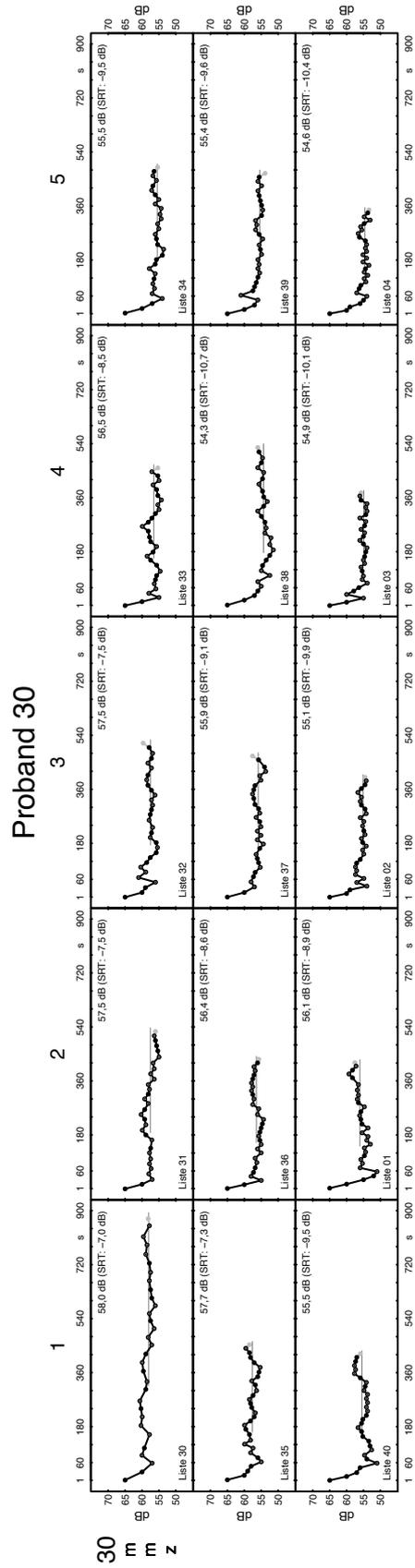
Probanden 05 , 09



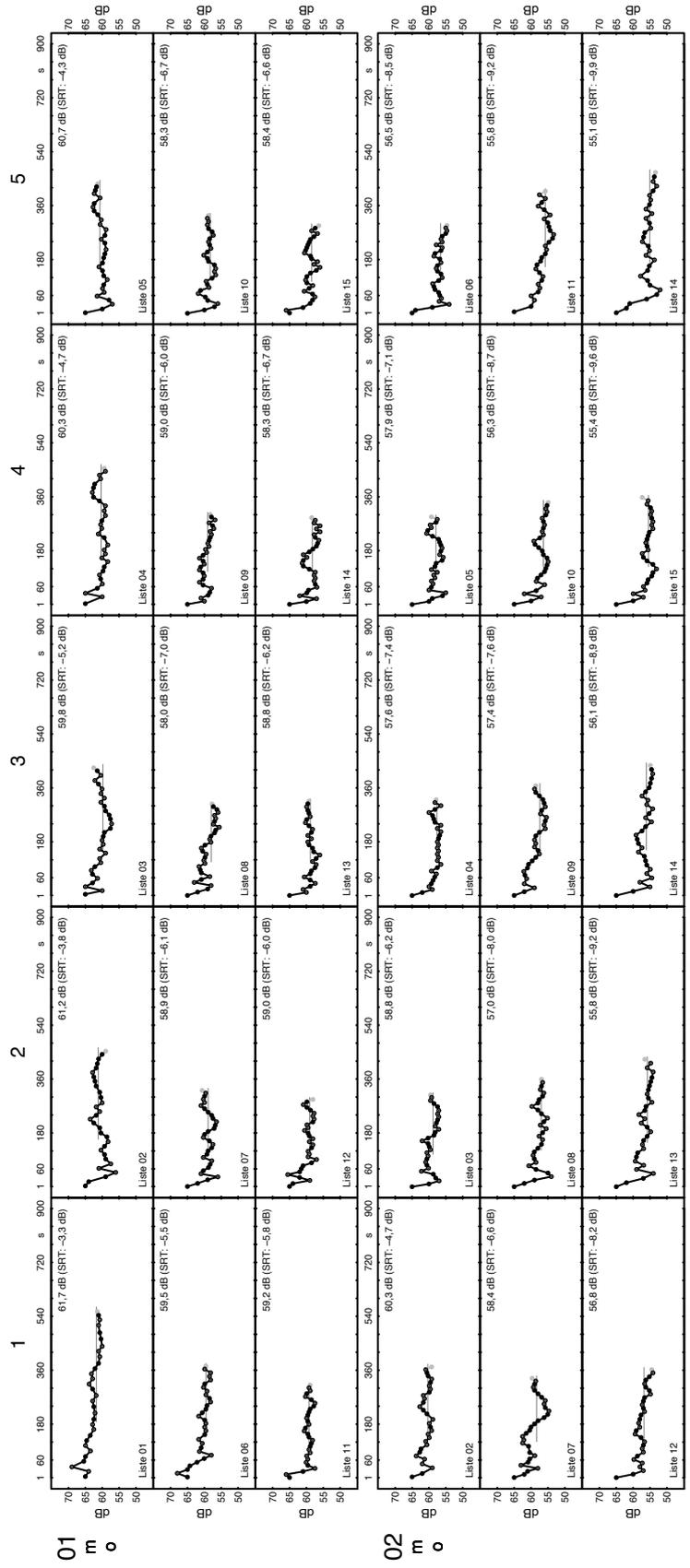
Probanden 15, 27



A.1 SRT-Werte aller Probanden aller Testsitzungen aller Durchläufe

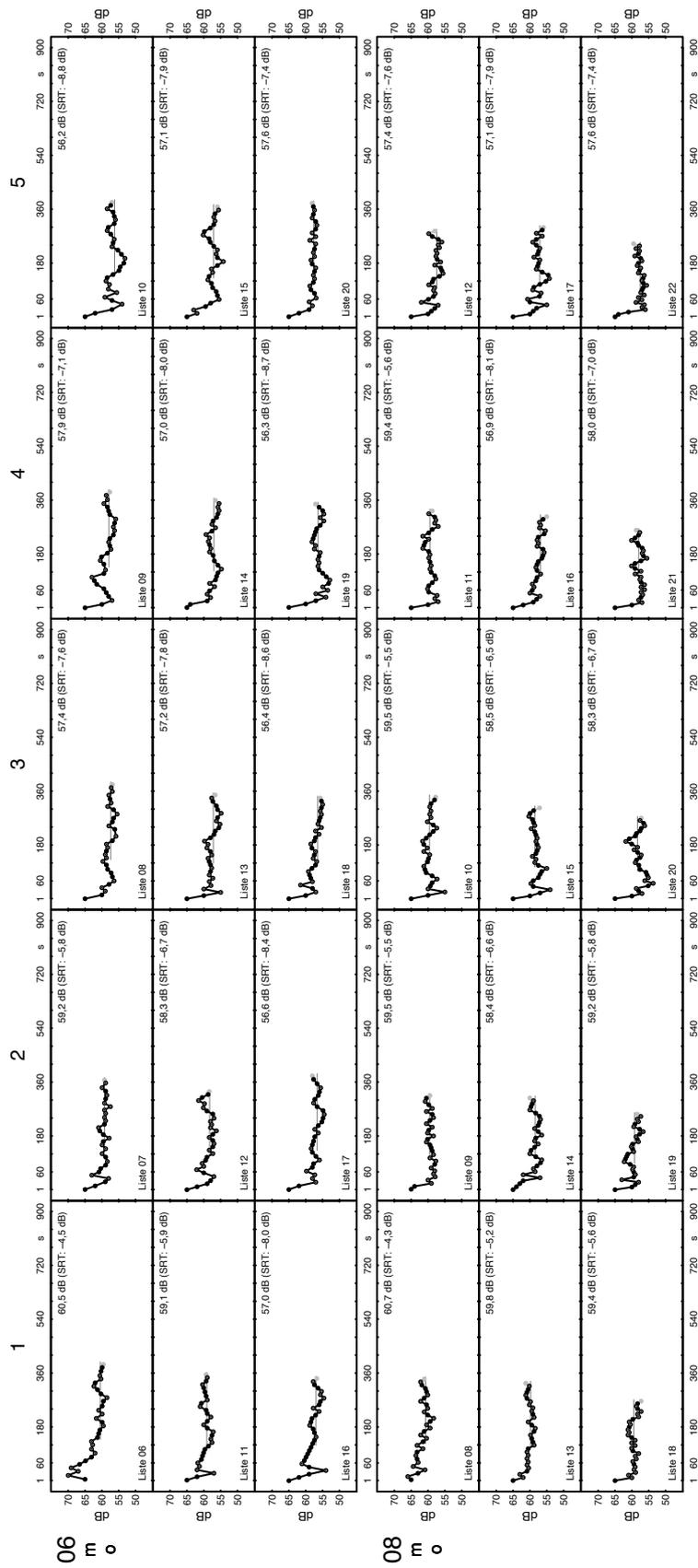


Probanden 01 , 02

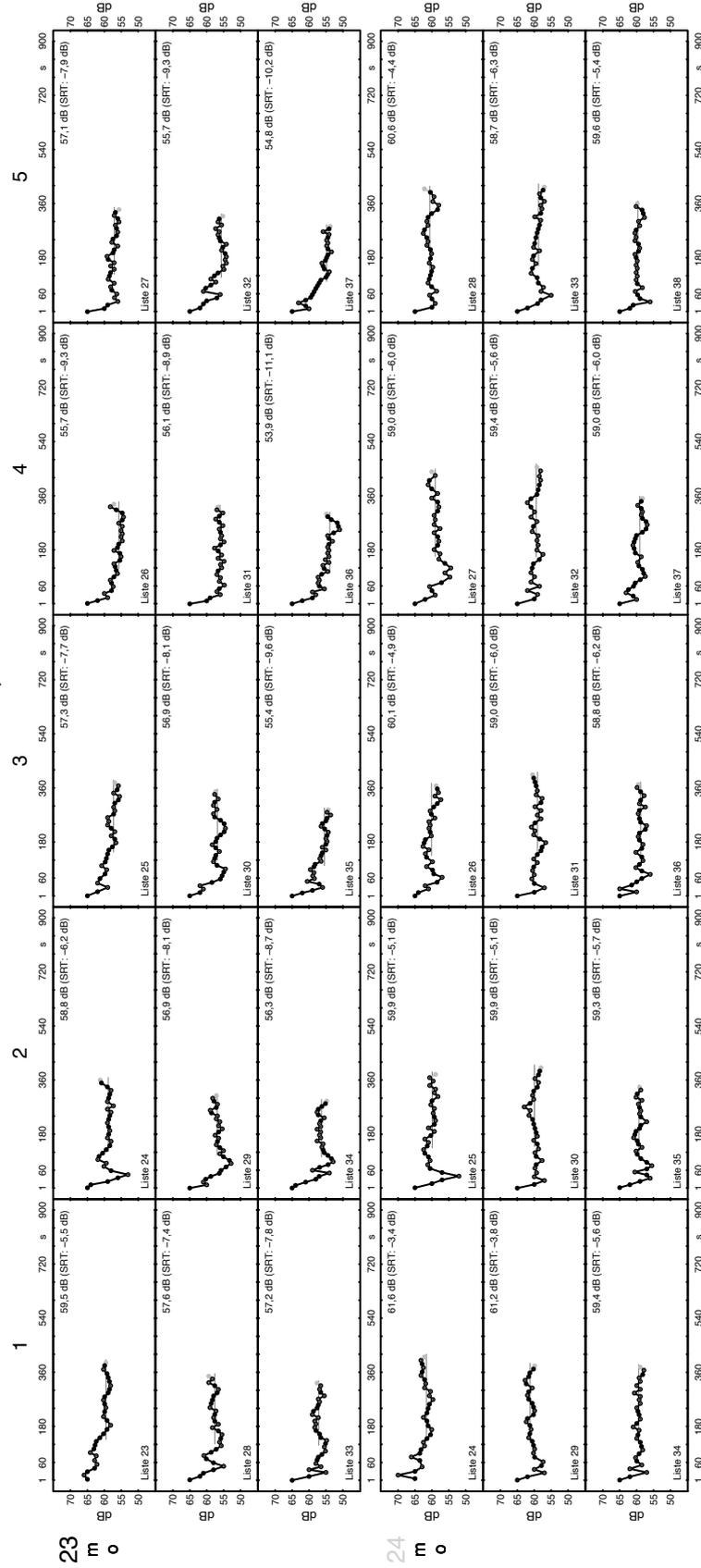


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 06 , 08

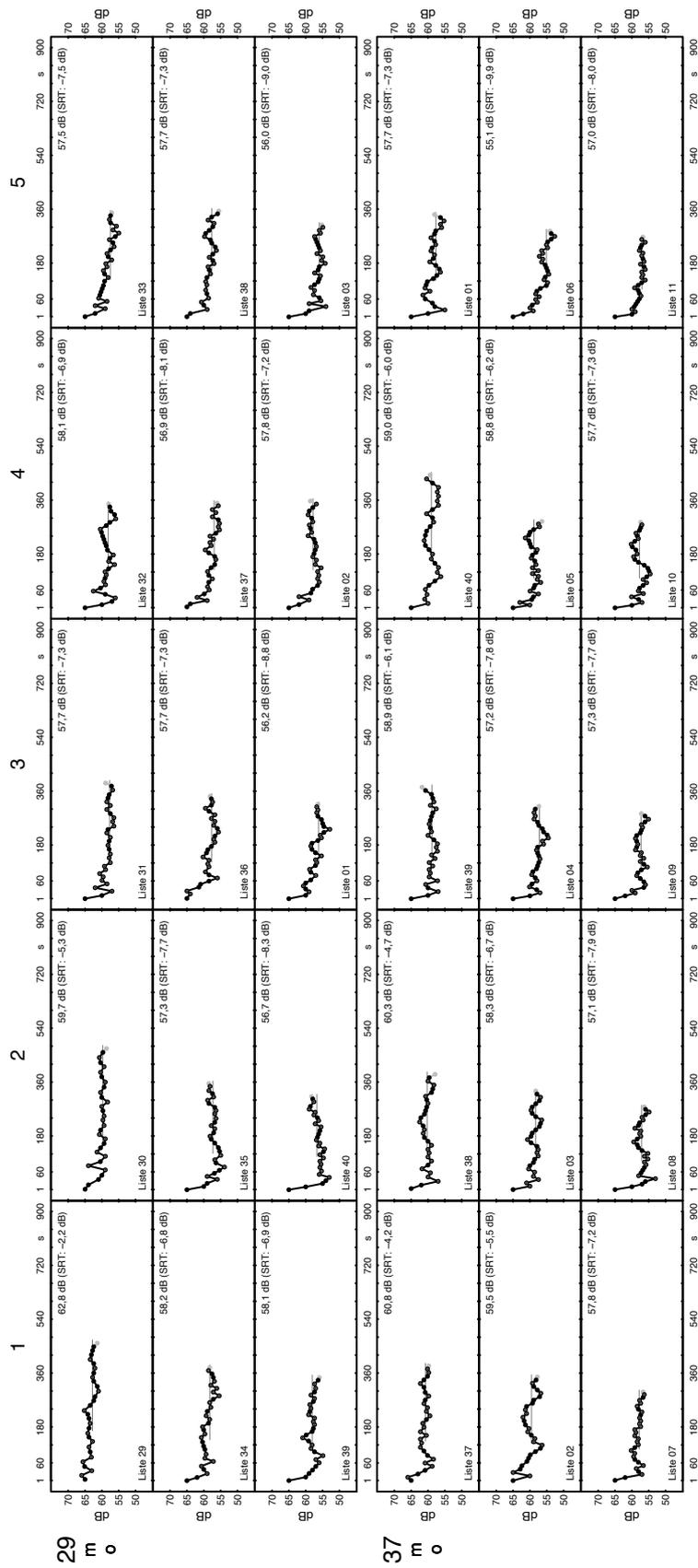


Probanden 23, 24

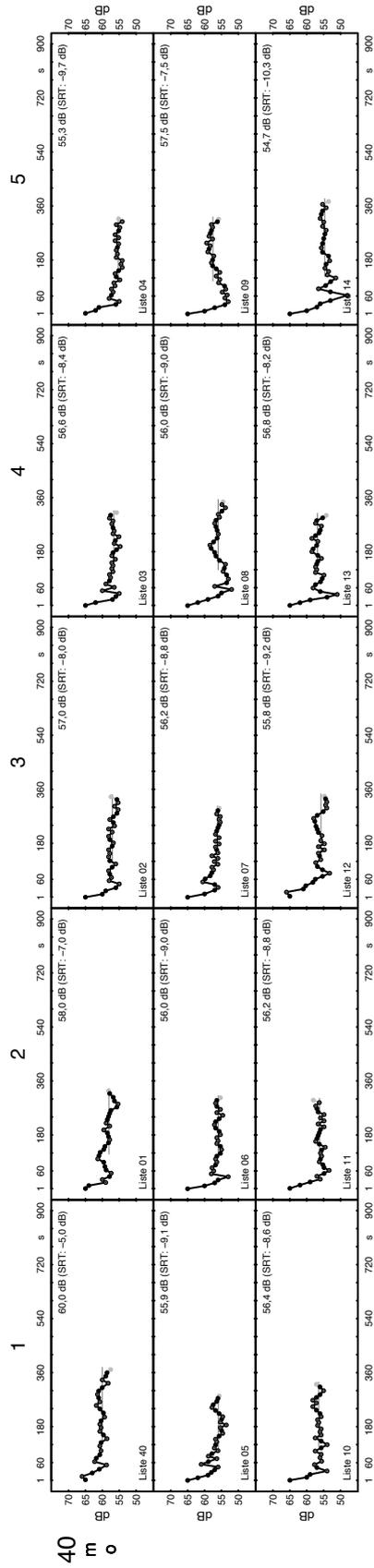


A.1 SRT-Werte aller Probanden aller Testsitzungen aller Durchläufe

Probanden 29, 37

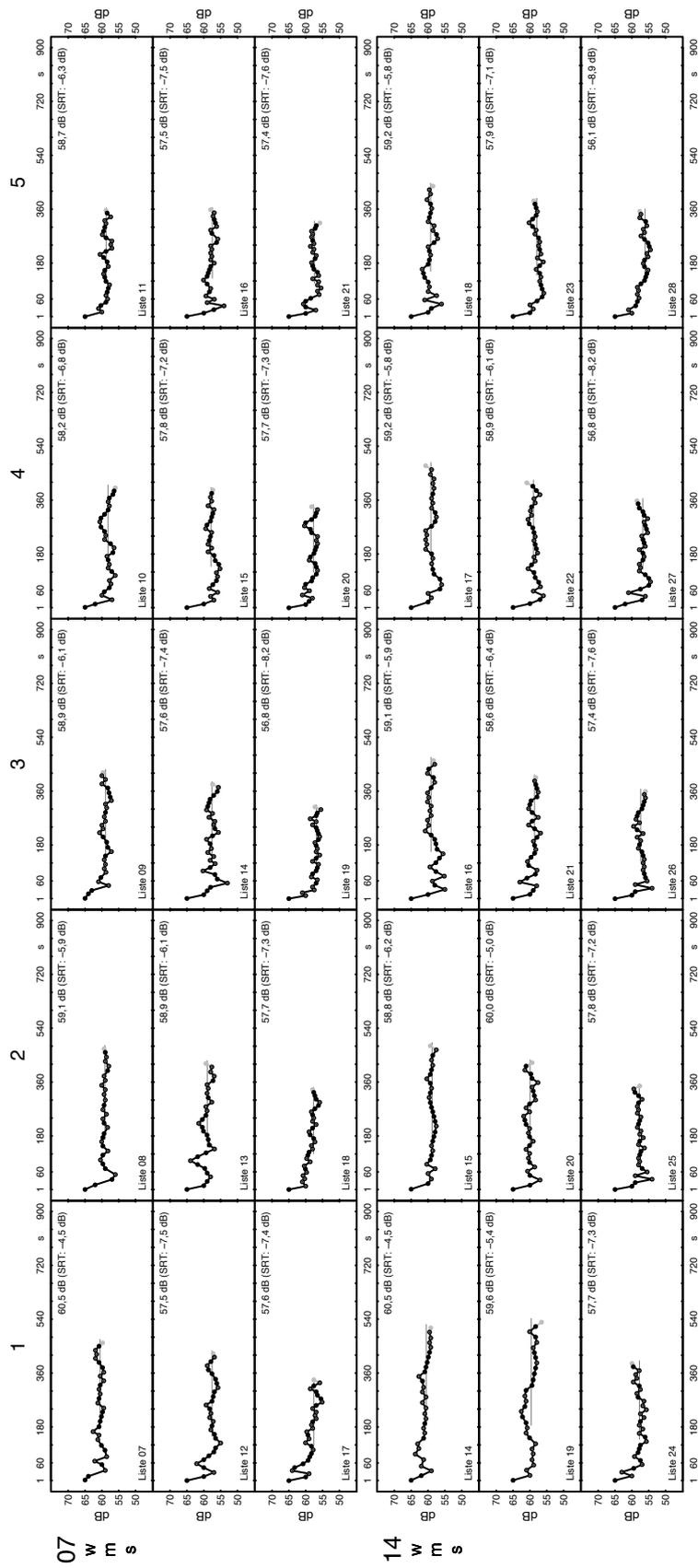


Proband 40

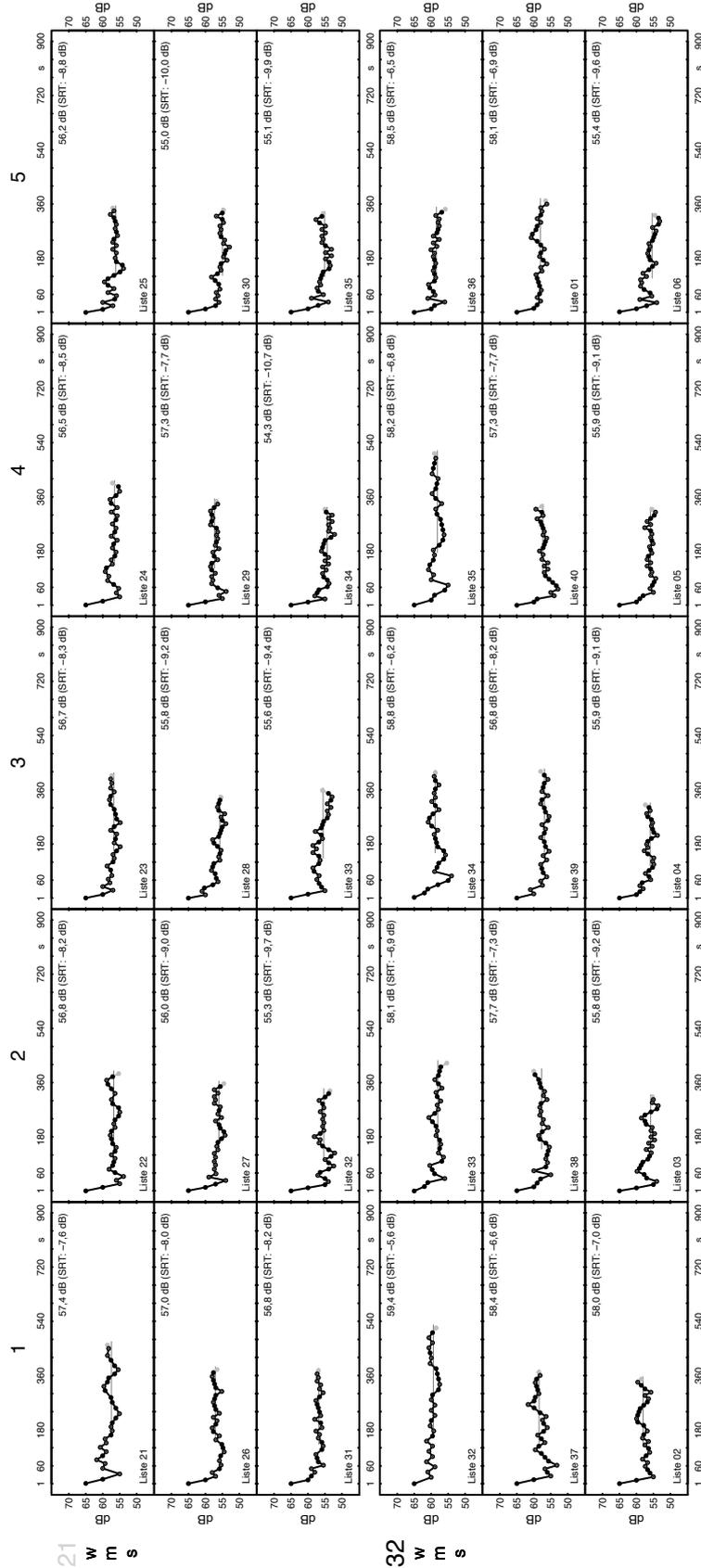


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 07, 14

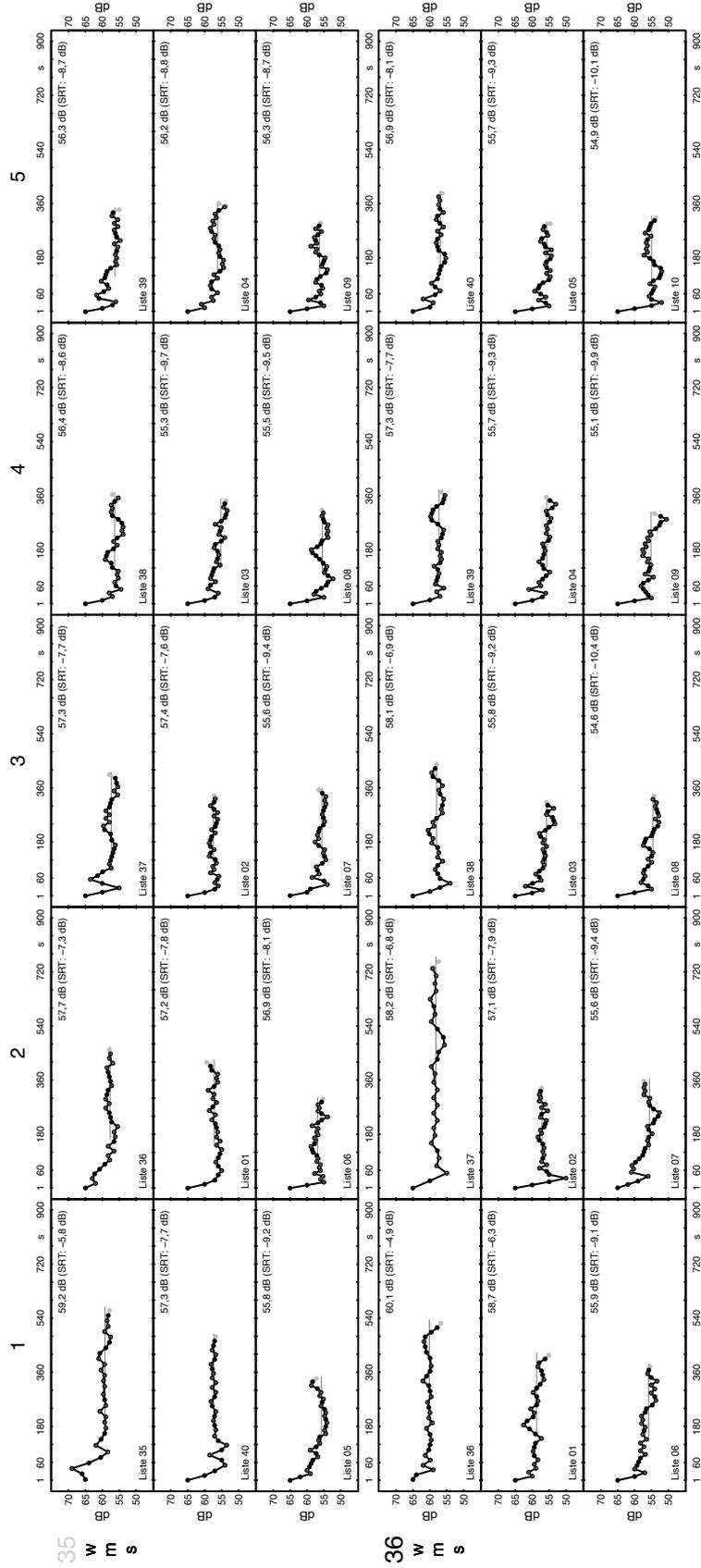


Probanden 21, 32

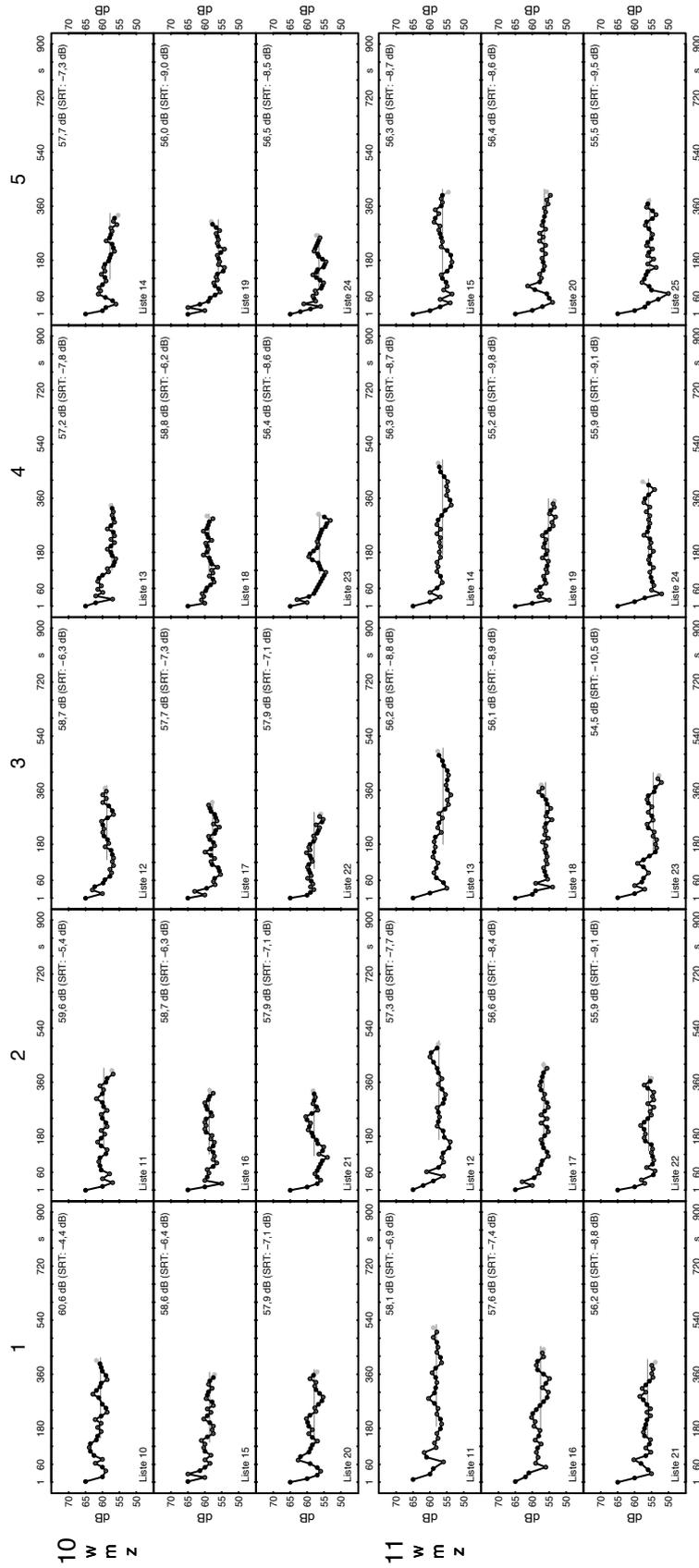


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 35, 36

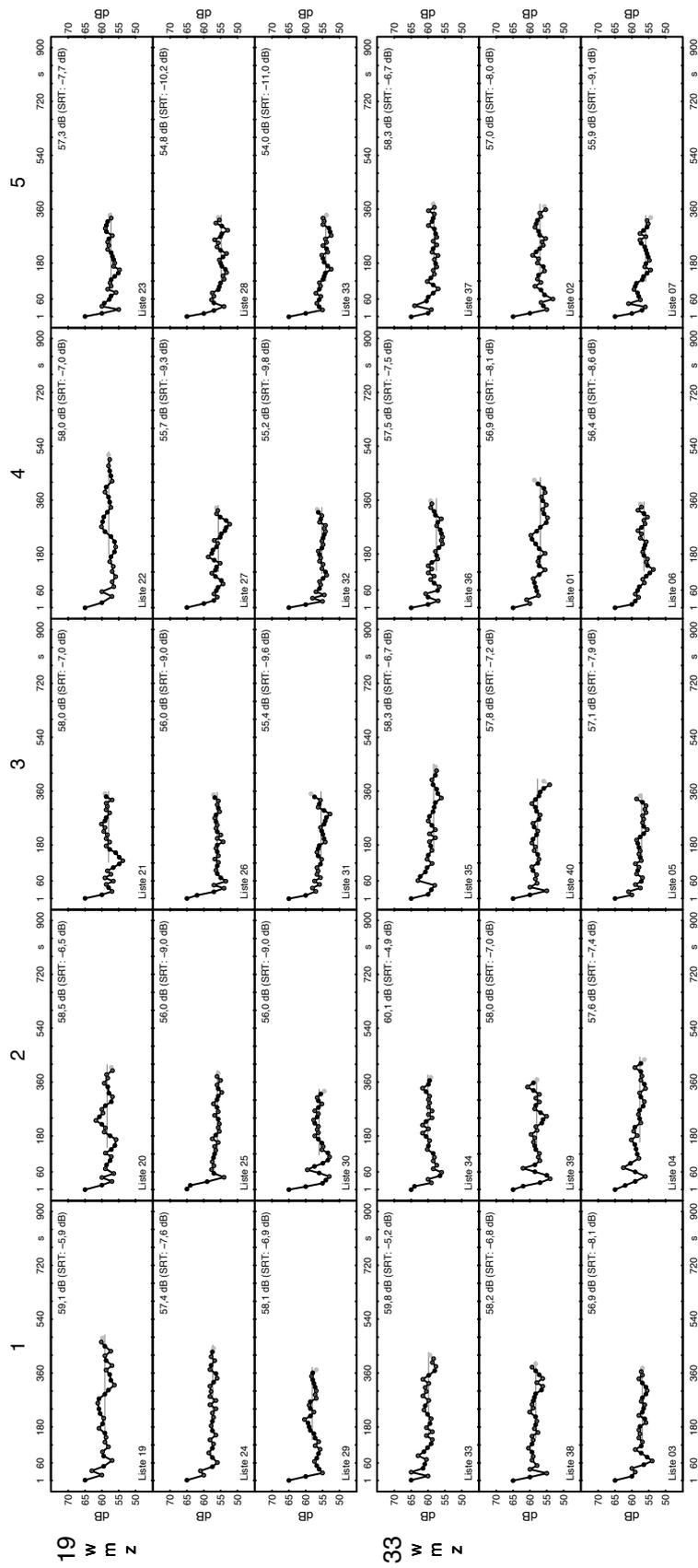


Probanden 10, 11

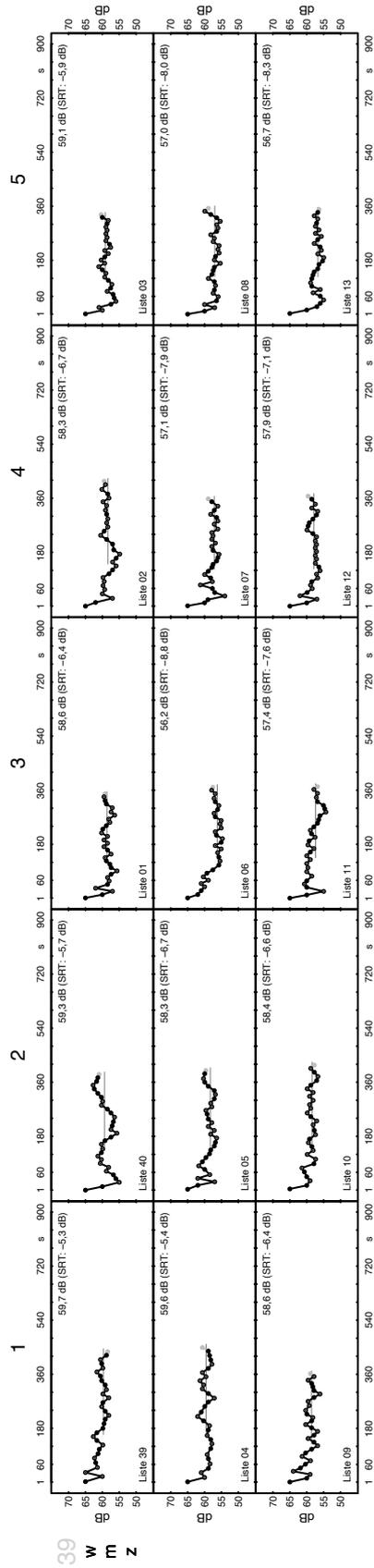


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 19, 33

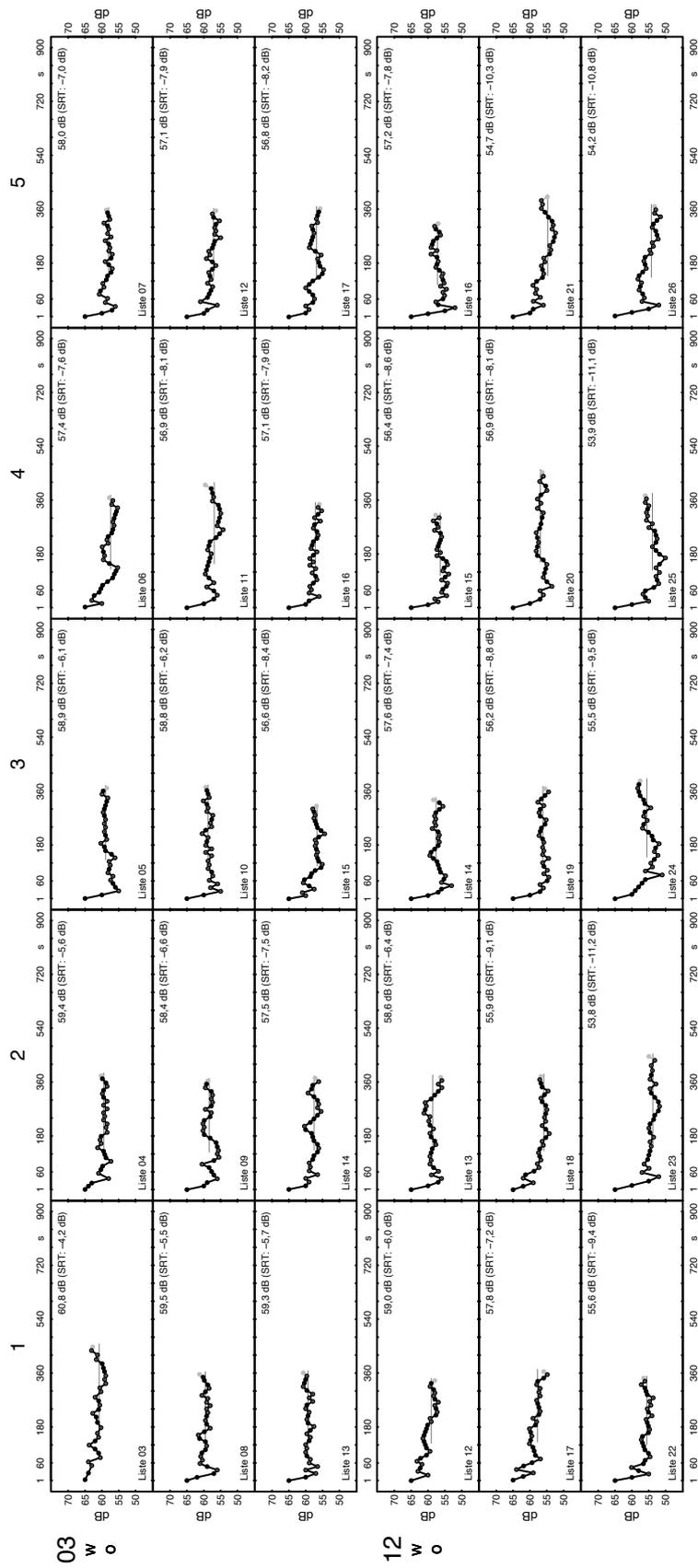


Proband 39

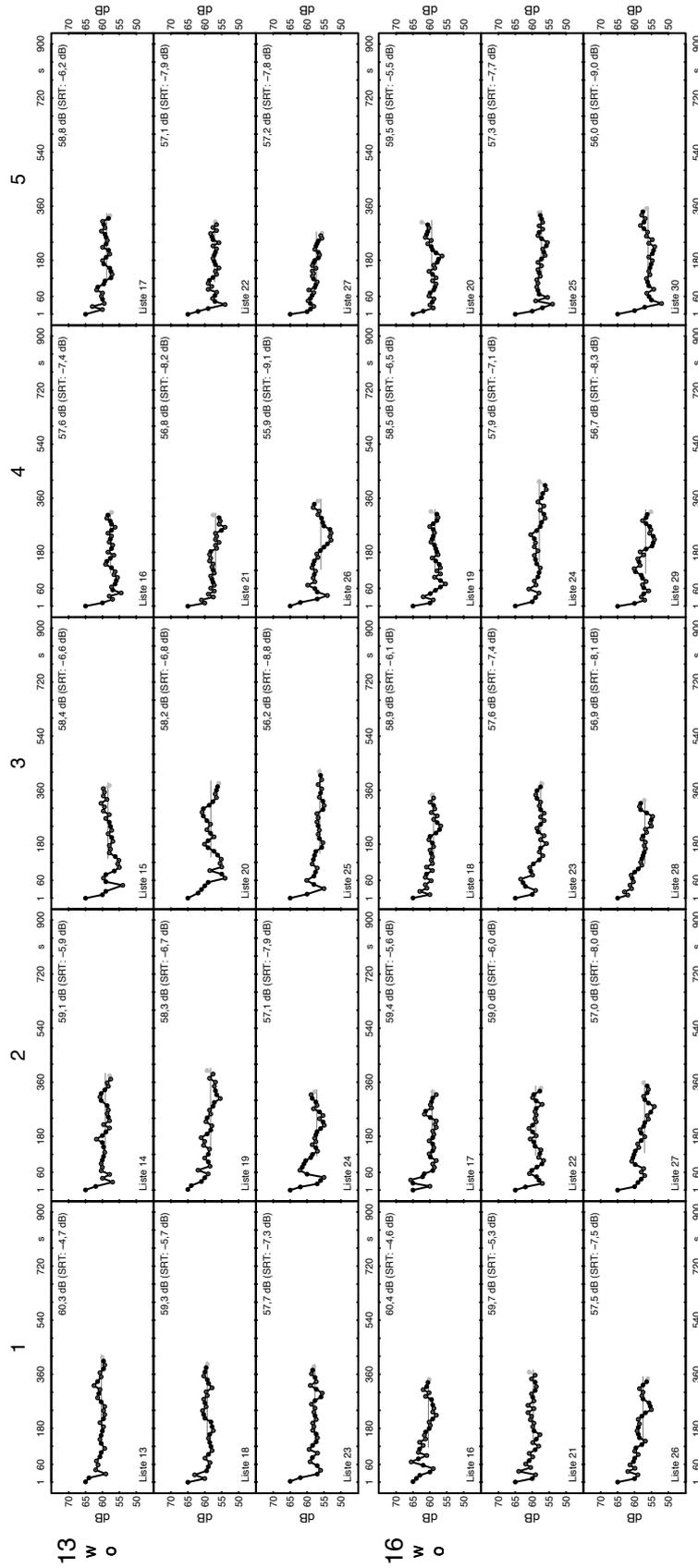


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 03, 12

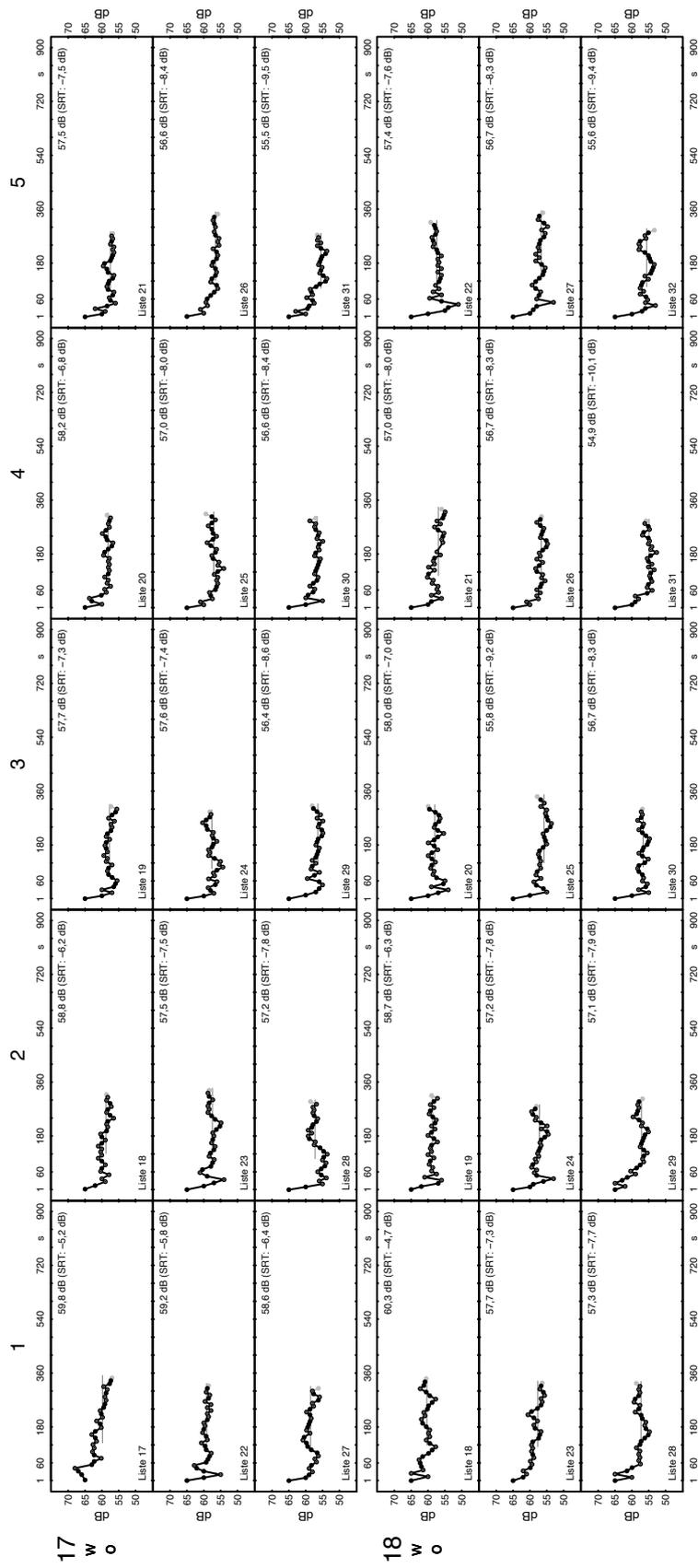


Probanden 13, 16

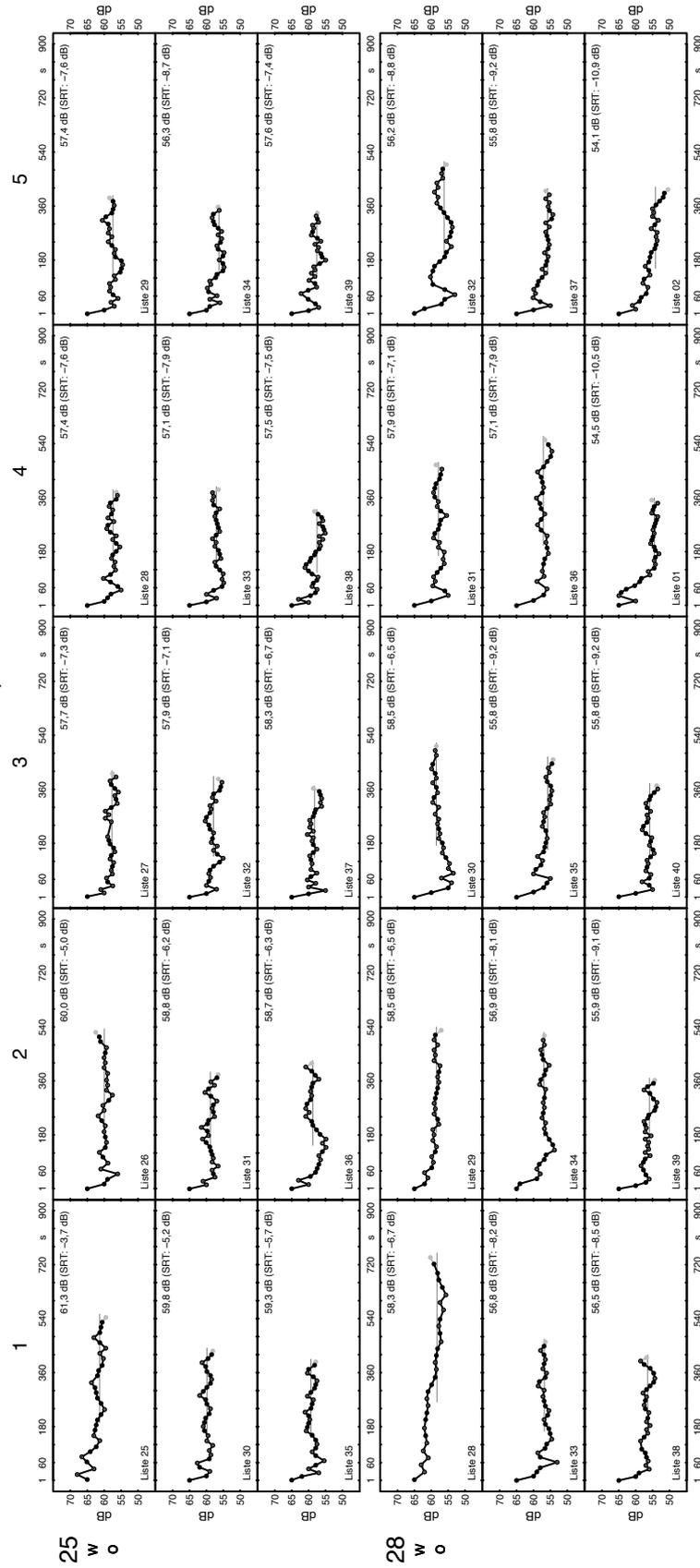


A.1 SRT-Werte aller Probanden aller Testsitzenngen aller Durchläufe

Probanden 17, 18



Probanden 25, 28



A.2 3D-Darstellung der Mittelwerte der SRT-Berechnung mit Standardabweichungen und Standardfehlern

In Kapitel 4 sind die beiden wichtigsten dreidimensionalen Darstellungen der SRT-Berechnung mit variierender Anzahl der einbezogenen SNR-Werte (4.1) und der Standardfehler des arithmetischen Mittels bei der jeweiligen Berechnungsvorschrift (4.2) dargestellt.

Aus Übersichtsgründen sind in Kapitel 4 lediglich diese beiden zusammenfassenden Grafiken enthalten.

Die hier folgenden Abbildungen zeigen eine 3D-Darstellung der SRT-Berechnung mit variierender Anzahl der einbezogenen SNR-Werte gemittelt über die jeweils genannte Testsitzungen bzw. den genannten Testdurchlauf sowie deren Standardabweichungen.

Weiterhin folgt eine 3D-Darstellung des Standardfehlers des arithmetischen Mittels bei der jeweiligen Berechnungsvorschrift gemittelt über die entsprechende Testsitzungen bzw. den entsprechenden Testdurchlauf.

Die Beschriftung verhält sich dabei wie in Kapitel 4 beschrieben.

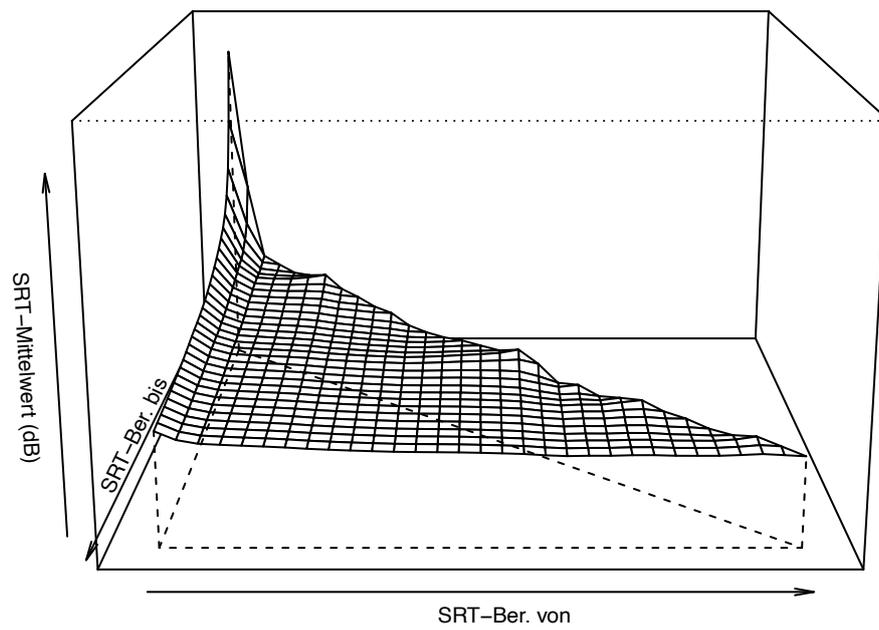


Abbildung A.1: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte aller Testdurchläufe der 1. Testsitzung

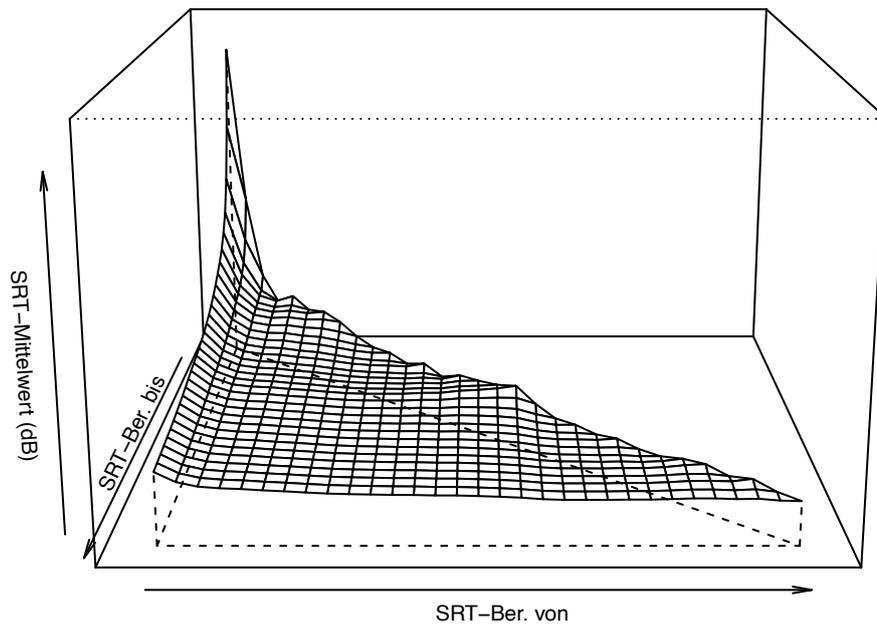


Abbildung A.2: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte aller Testdurchläufe der 2. Testsitzung

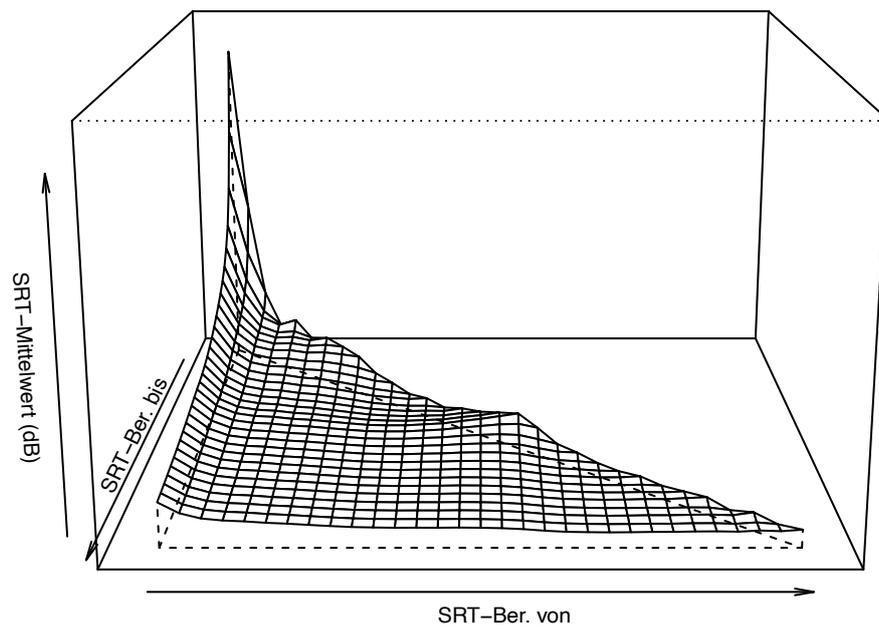


Abbildung A.3: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte aller Testdurchläufe der 3. Testsitzung

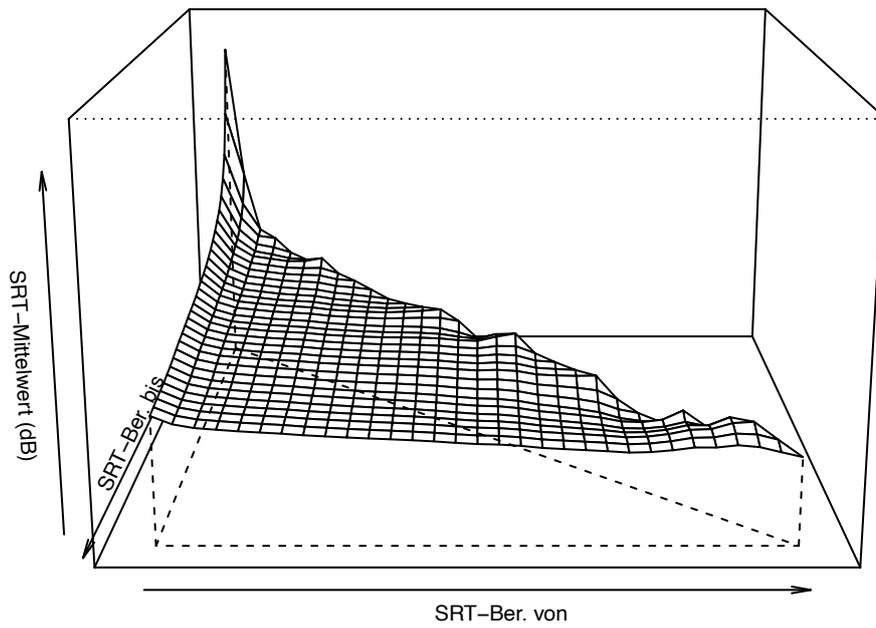


Abbildung A.4: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 1. Testdurchläufe aller Testsitzungen

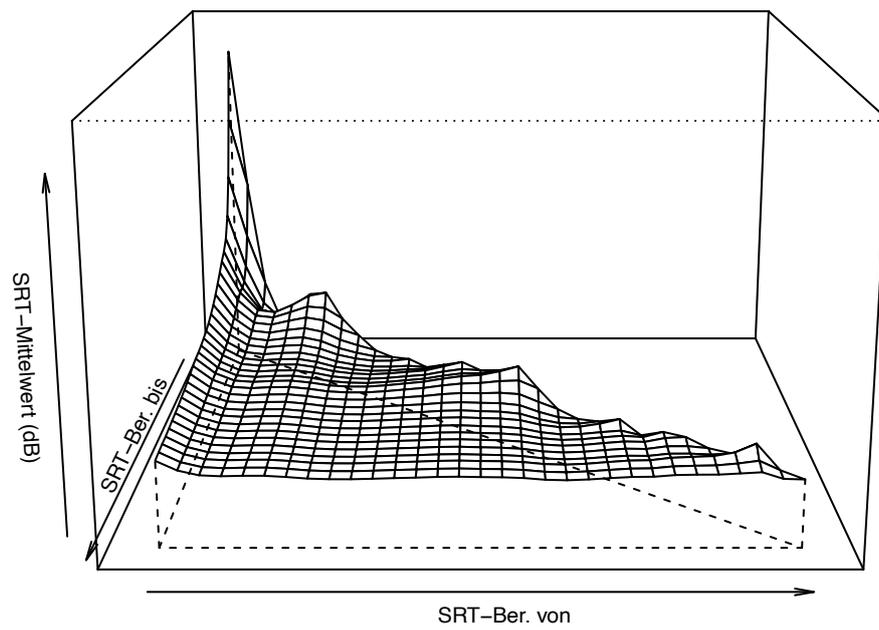


Abbildung A.5: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 2. Testdurchläufe aller Testsitzungen

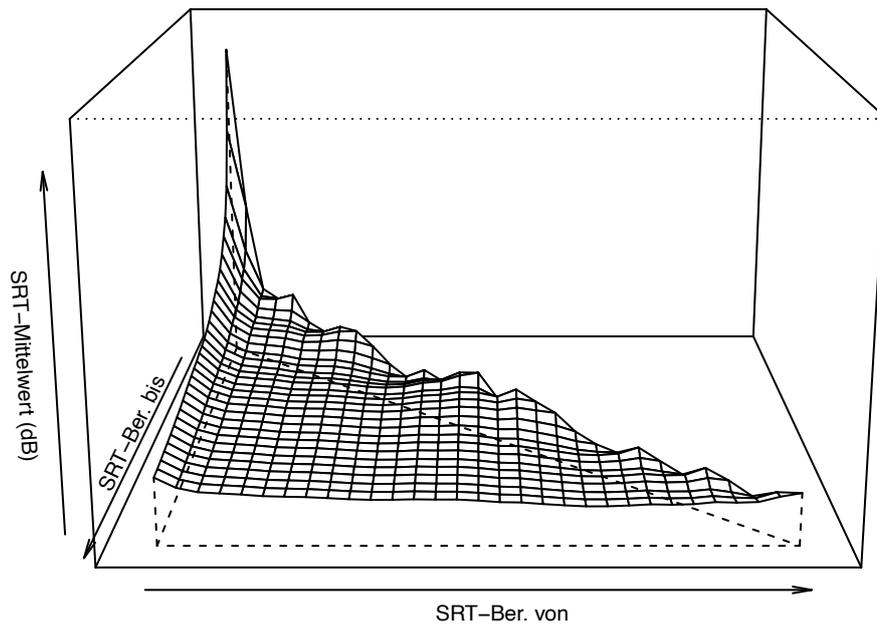


Abbildung A.6: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 3. Testdurchläufe aller Testsitzungen

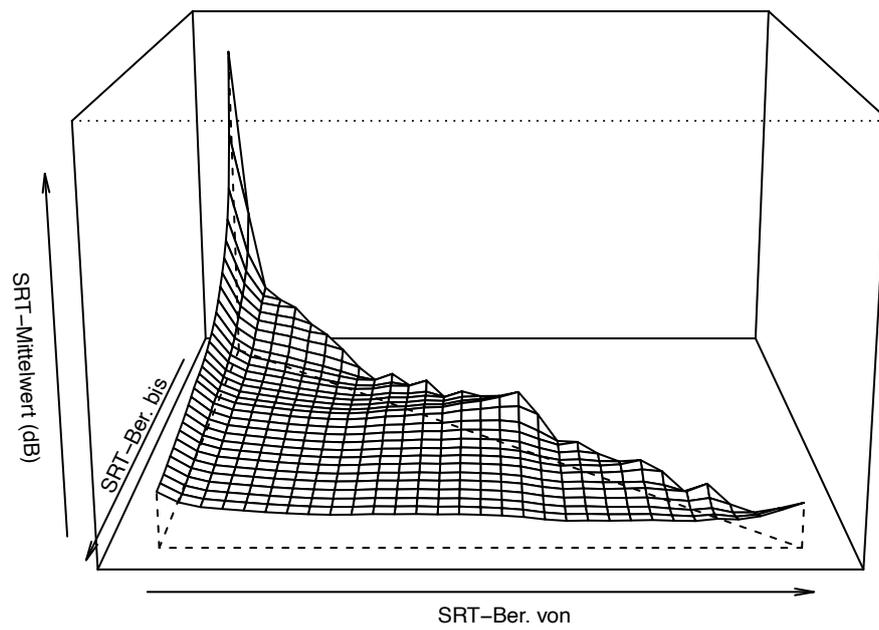


Abbildung A.7: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 4. Testdurchläufe aller Testsitzungen

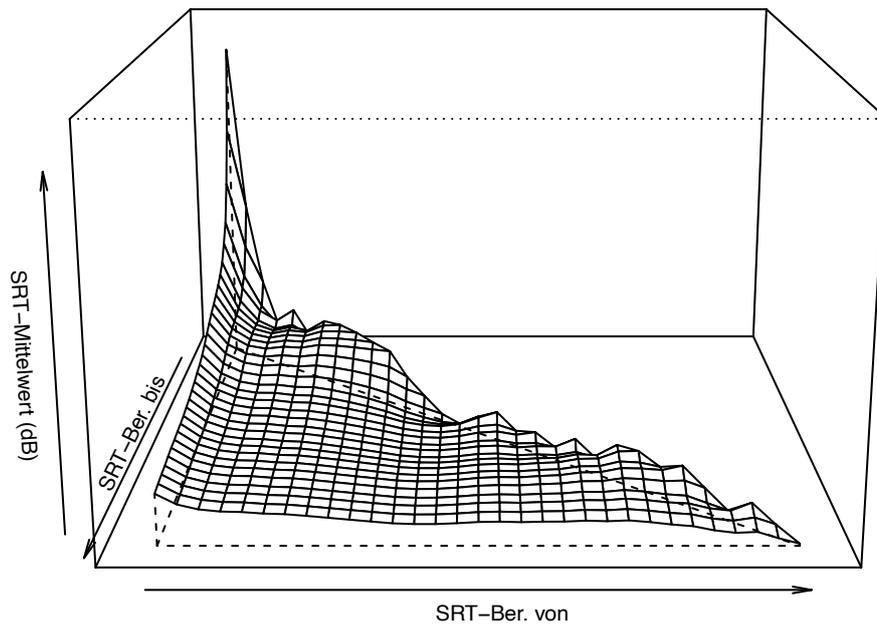


Abbildung A.8: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 5. Testdurchläufe aller Testsitzungen

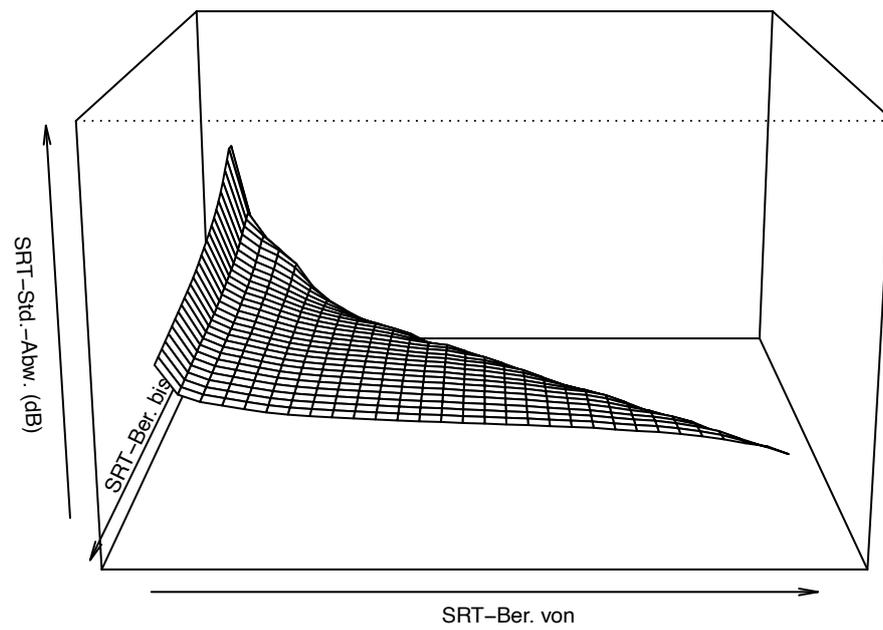


Abbildung A.9: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte aller Testdurchläufe aller Testsitzungen

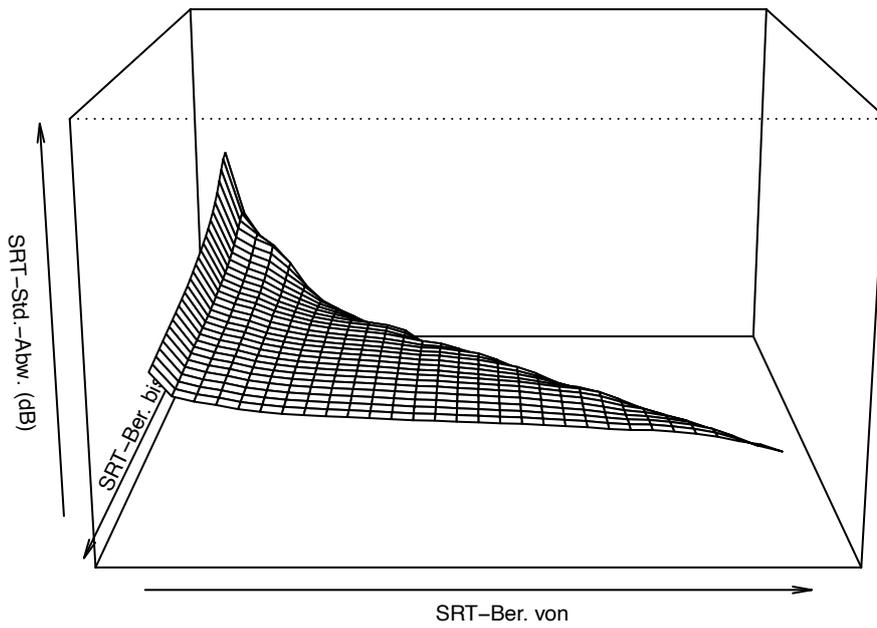


Abbildung A.10: 3D-Darstellung der berechneten Standardabweichung der SRT-Werte aller Testdurchläufe der 1. Testsitzung

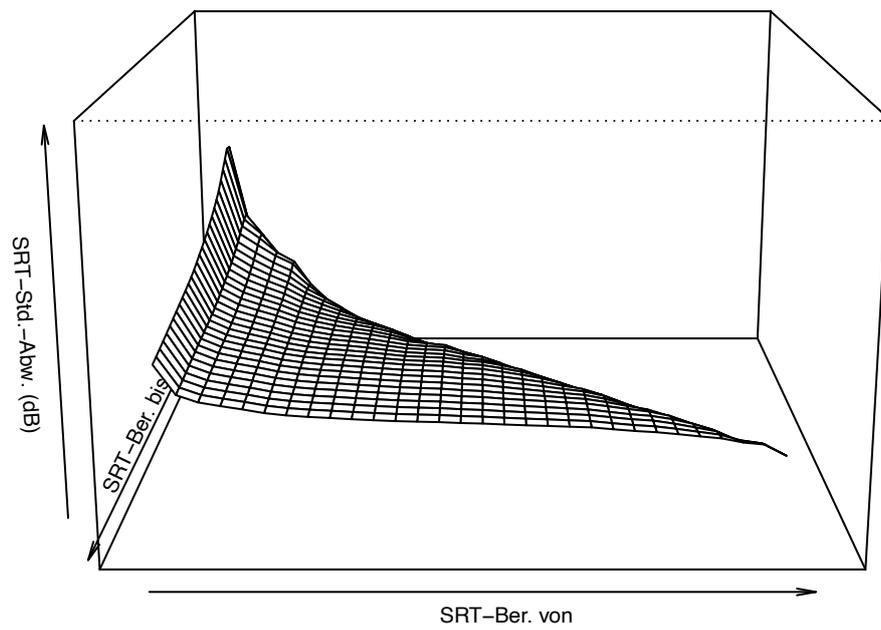


Abbildung A.11: 3D-Darstellung der berechneten Standardabweichung der SRT-Werte aller Testdurchläufe der 2. Testsitzung

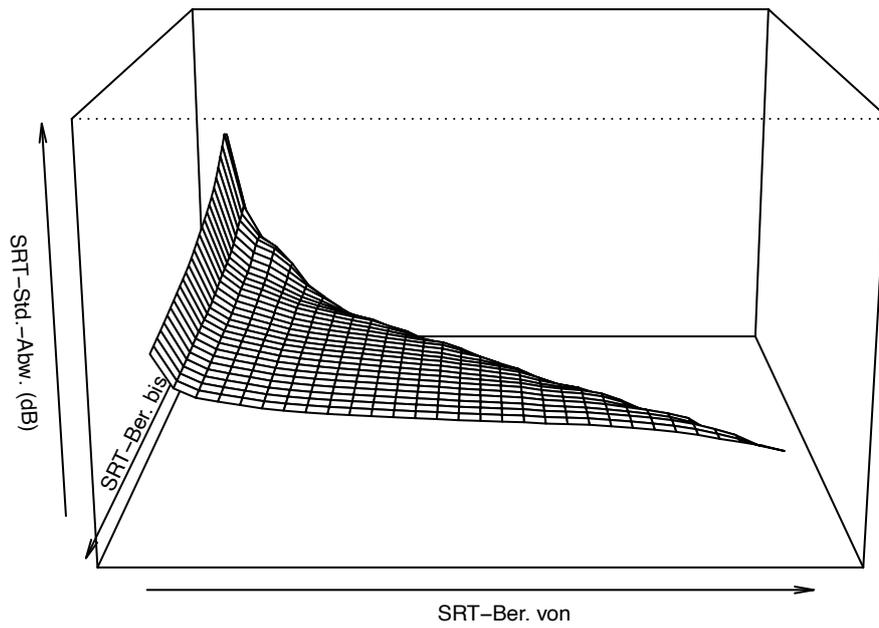


Abbildung A.12: 3D-Darstellung der berechneten Standardabweichung der SRT-Werte aller Testdurchläufe der 3. Testsitzung

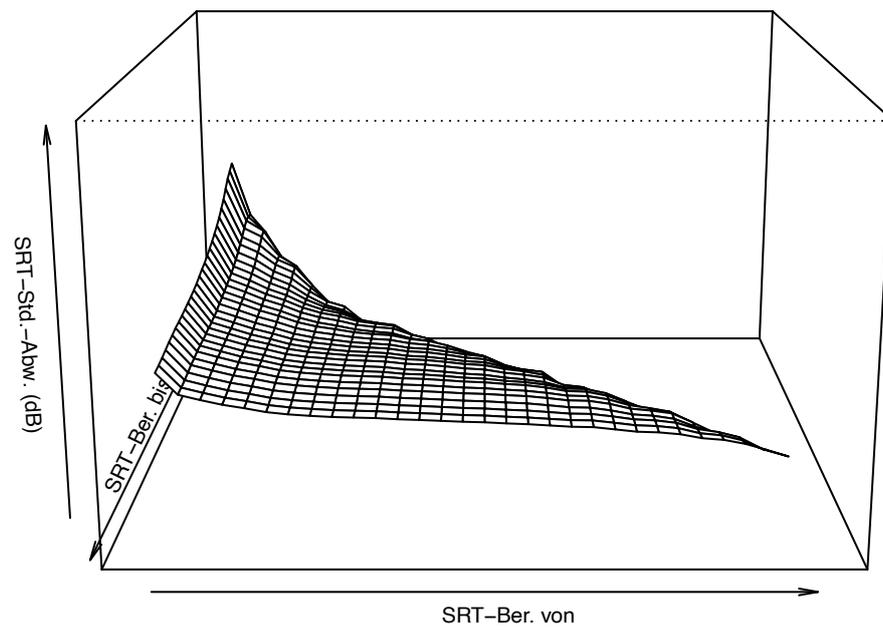


Abbildung A.13: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 1. Testdurchläufe aller Testsitzungen

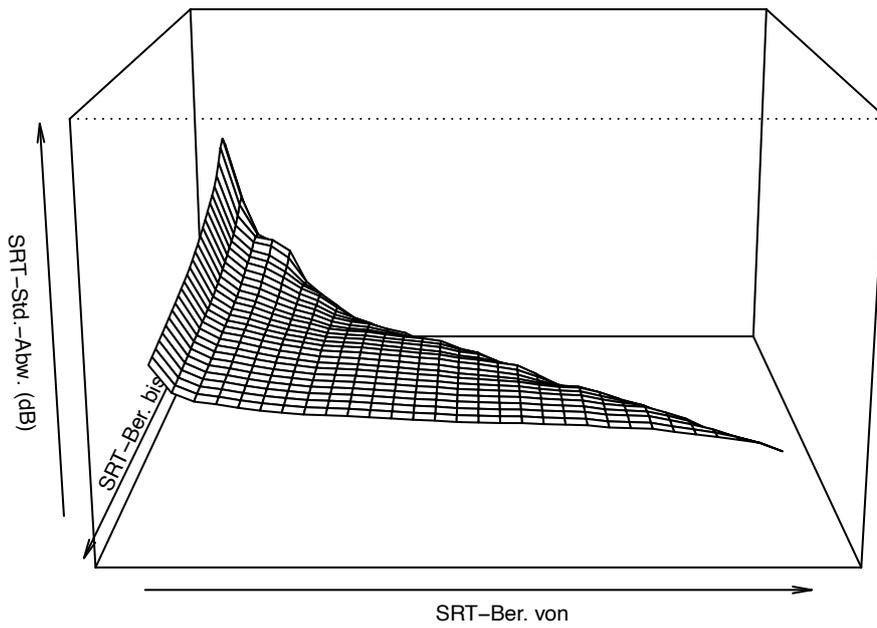


Abbildung A.14: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 2. Testdurchläufe aller Testsitzungen

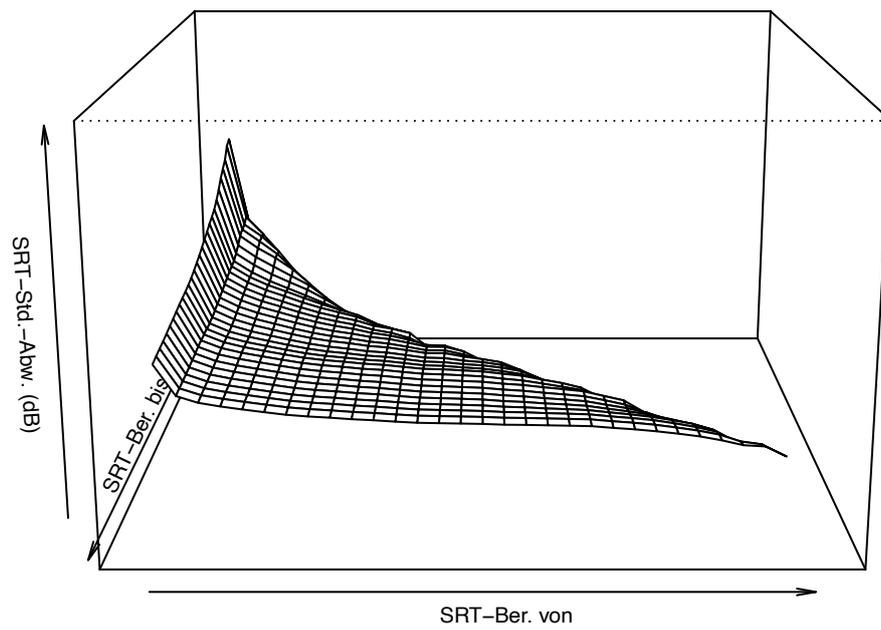


Abbildung A.15: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 3. Testdurchläufe aller Testsitzungen

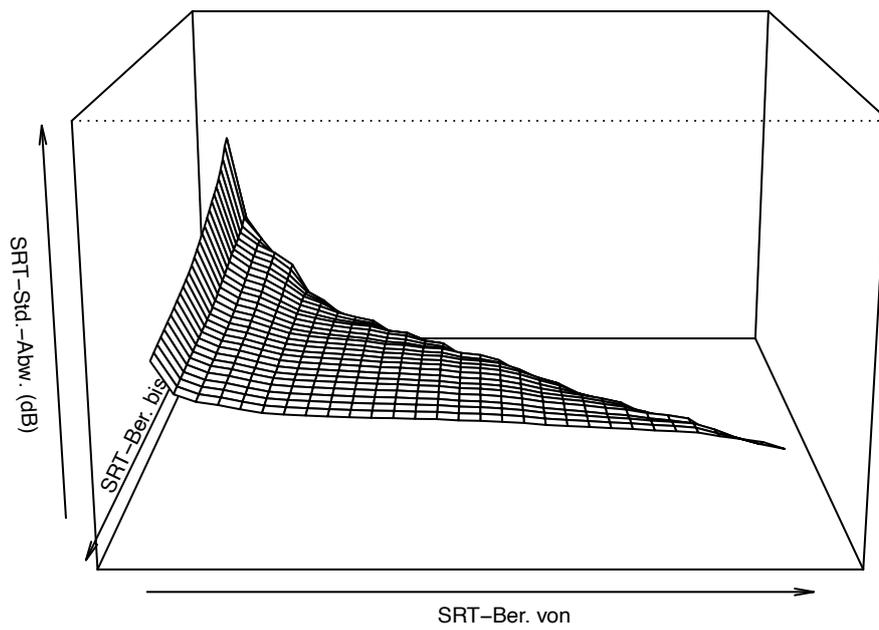


Abbildung A.16: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 4. Testdurchläufe aller Testsitzungen

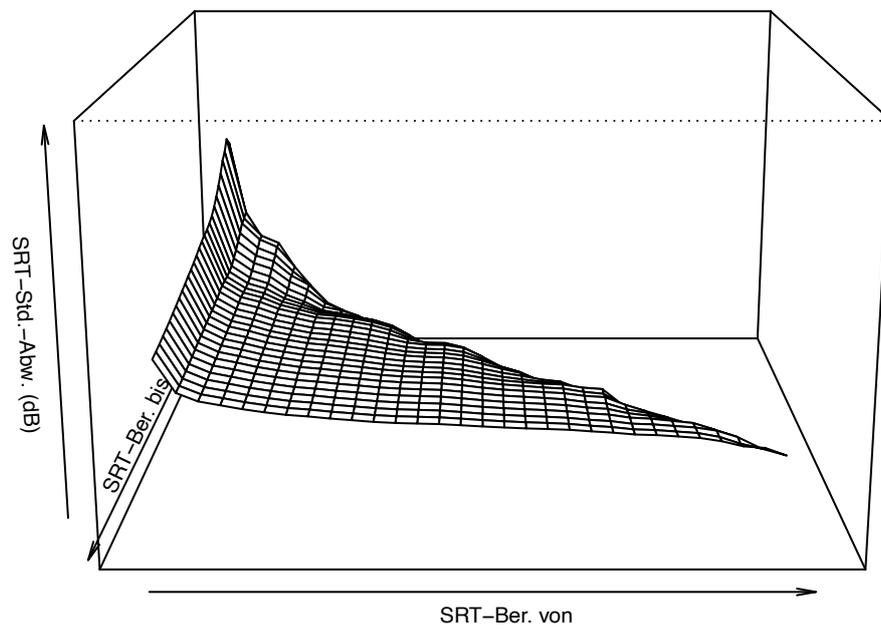


Abbildung A.17: 3D-Darstellung der berechneten Mittelwerte der SRT-Werte der 5. Testdurchläufe aller Testsitzungen

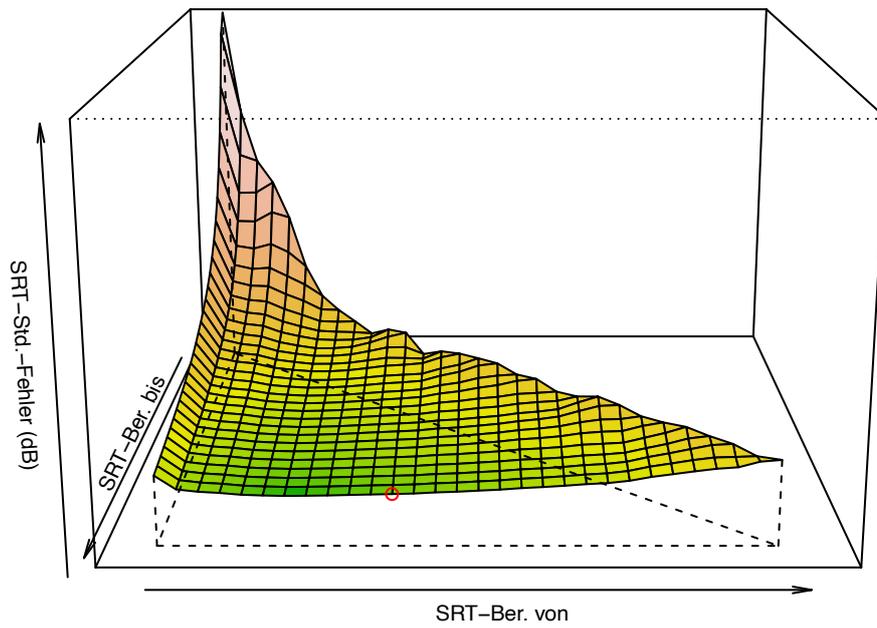


Abbildung A.18: 3D-Darstellung der berechneten Standardfehler der SRT-Werte aller Testdurchläufe der 1. Testsitzung

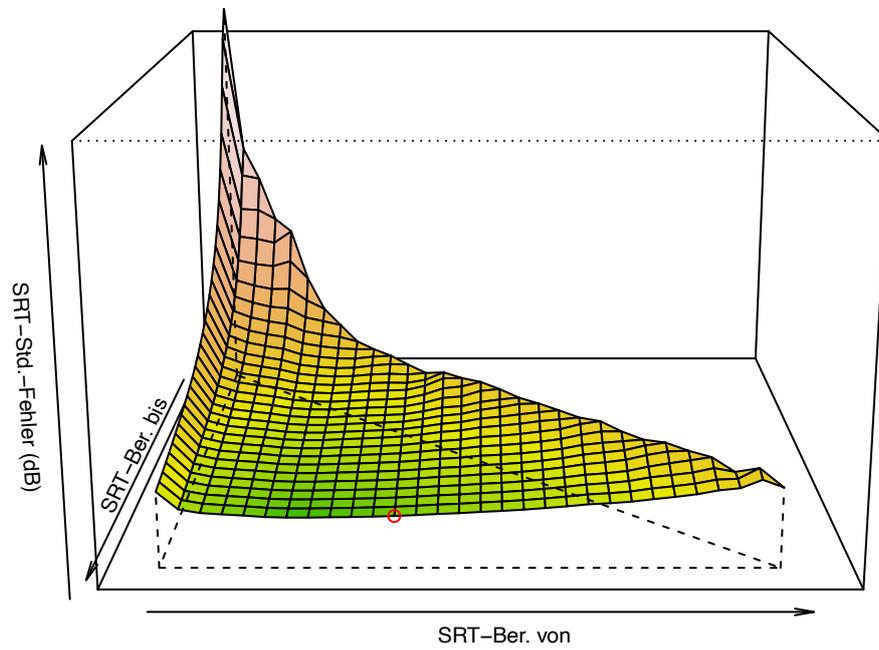


Abbildung A.19: 3D-Darstellung der berechneten Standardfehler der SRT-Werte aller Testdurchläufe der 2. Testsitzung

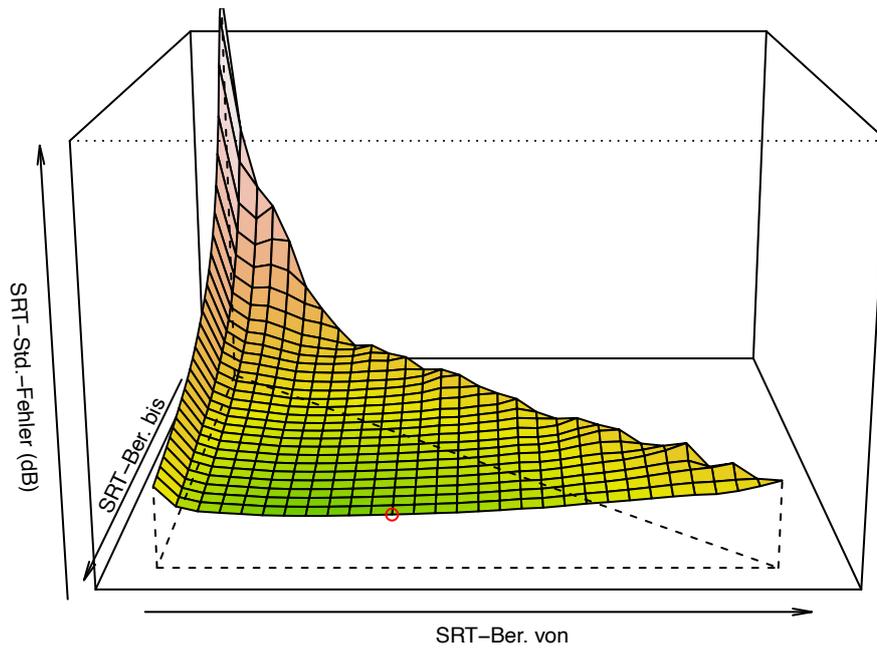


Abbildung A.20: 3D-Darstellung der berechneten Standardfehler der SRT-Werte aller Testdurchläufe der 3. Testsitzung

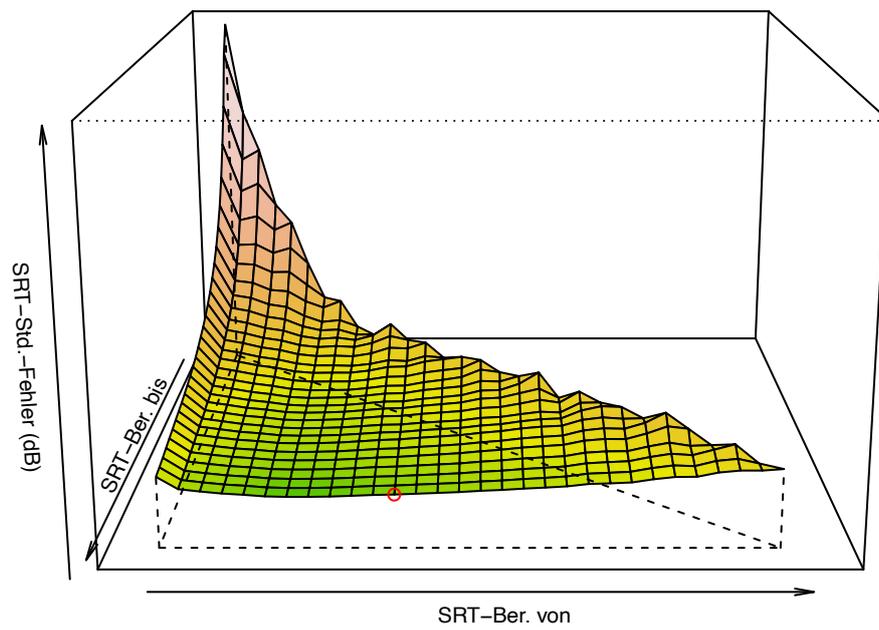


Abbildung A.21: 3D-Darstellung der berechneten Standardfehler der SRT-Werte der 1. Testdurchläufe aller Testsitzungen

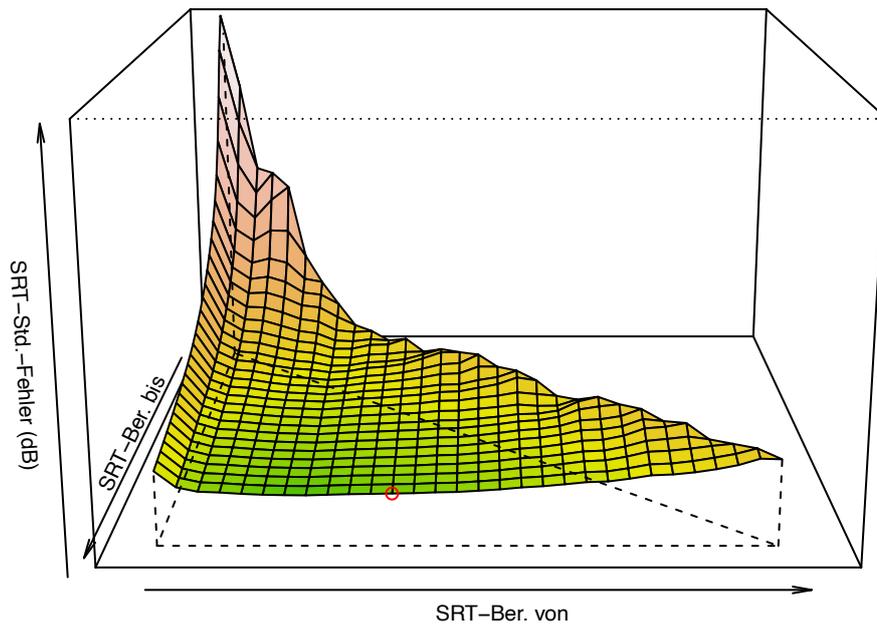


Abbildung A.22: 3D-Darstellung der berechneten Standardfehler der SRT-Werte der 2. Testdurchläufe aller Testsitzungen

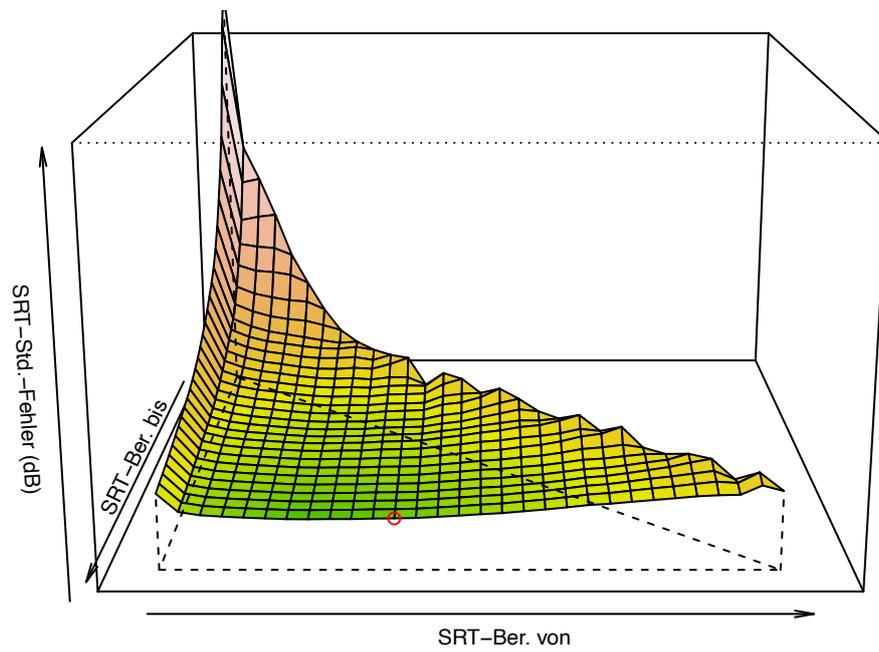


Abbildung A.23: 3D-Darstellung der berechneten Standardfehler der SRT-Werte der 3. Testdurchläufe aller Testsitzungen

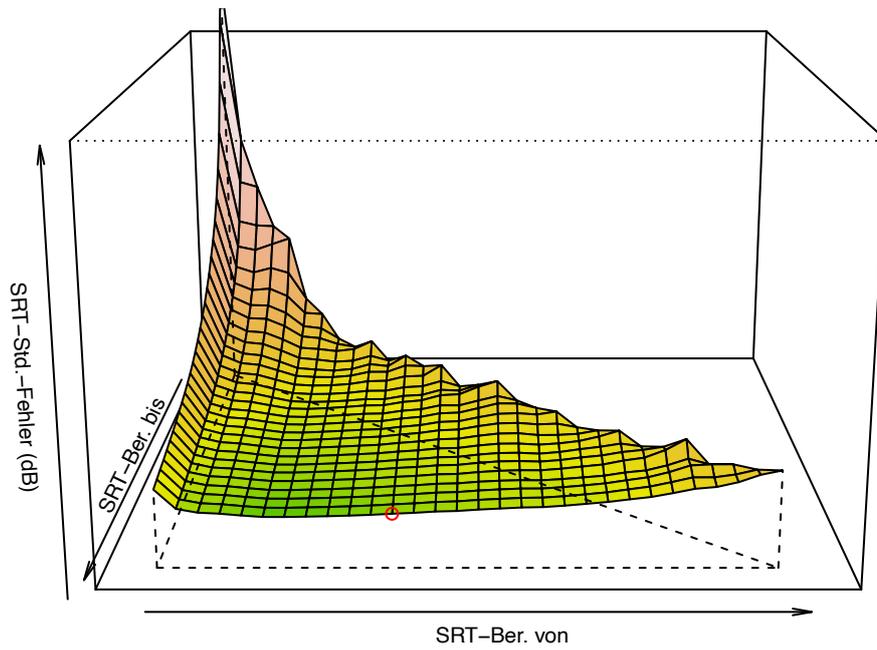


Abbildung A.24: 3D-Darstellung der berechneten Standardfehler der SRT-Werte der 4. Testdurchläufe aller Testsitzungen

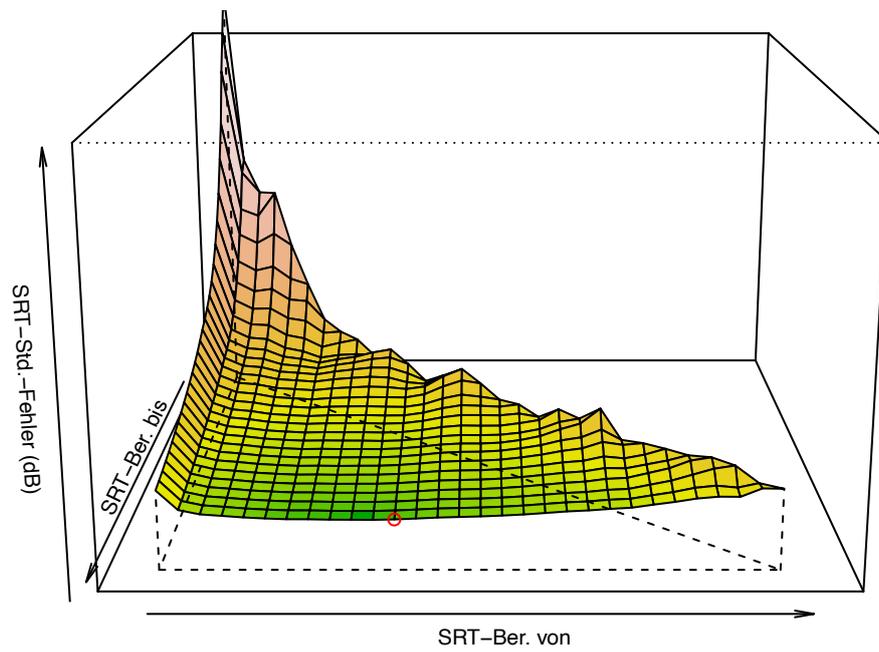


Abbildung A.25: 3D-Darstellung der berechneten Standardfehler der SRT-Werte der 5. Testdurchläufe aller Testsitzungen

A.3 SRT-Verlaufswerte der nicht selektierten Probanden

A Rohdaten

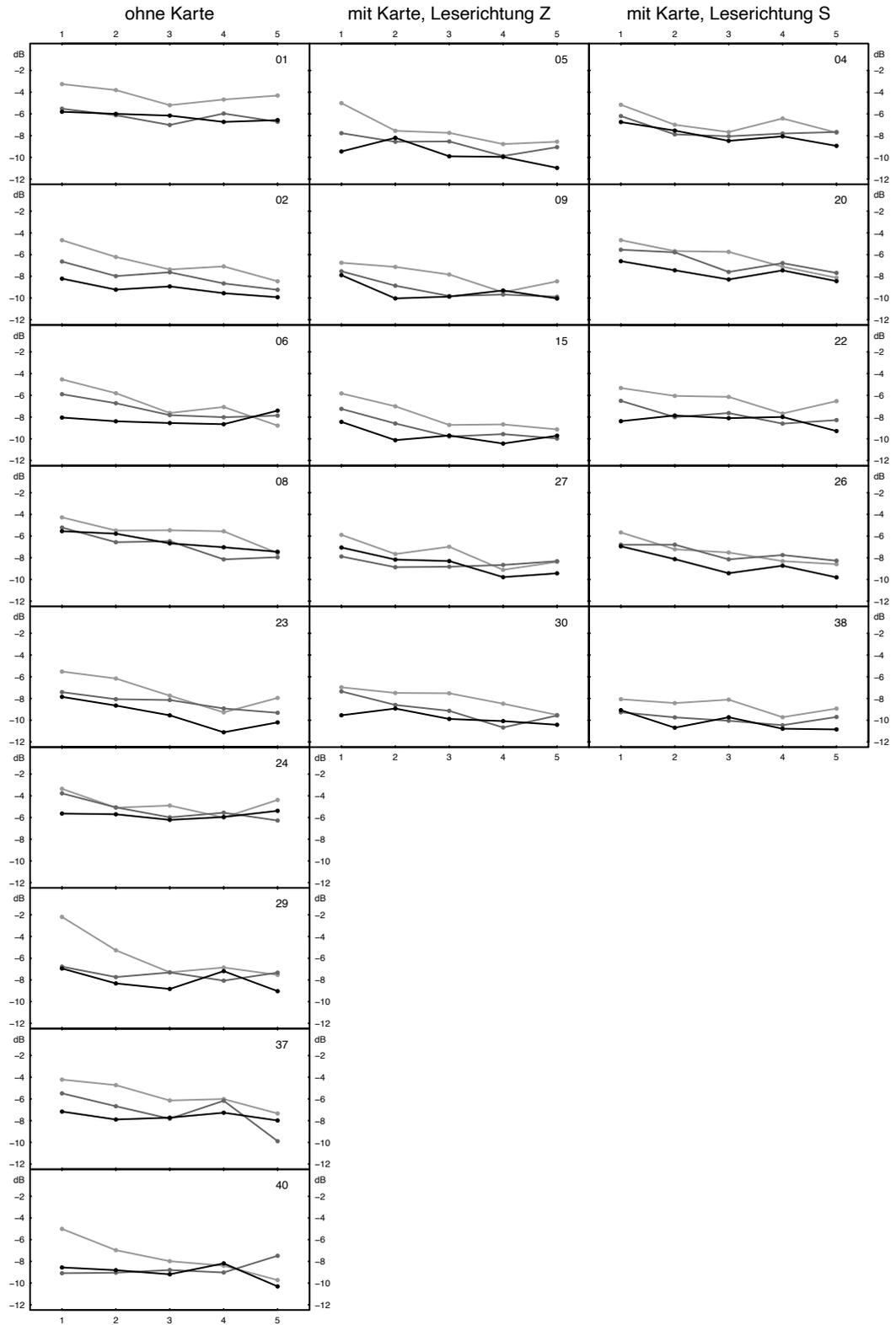


Abbildung A.26: SRT-Verlaufswerte der nicht selektierten, männlichen Probanden

A.3 SRT-Verlaufswerte der nicht selektierten Probanden

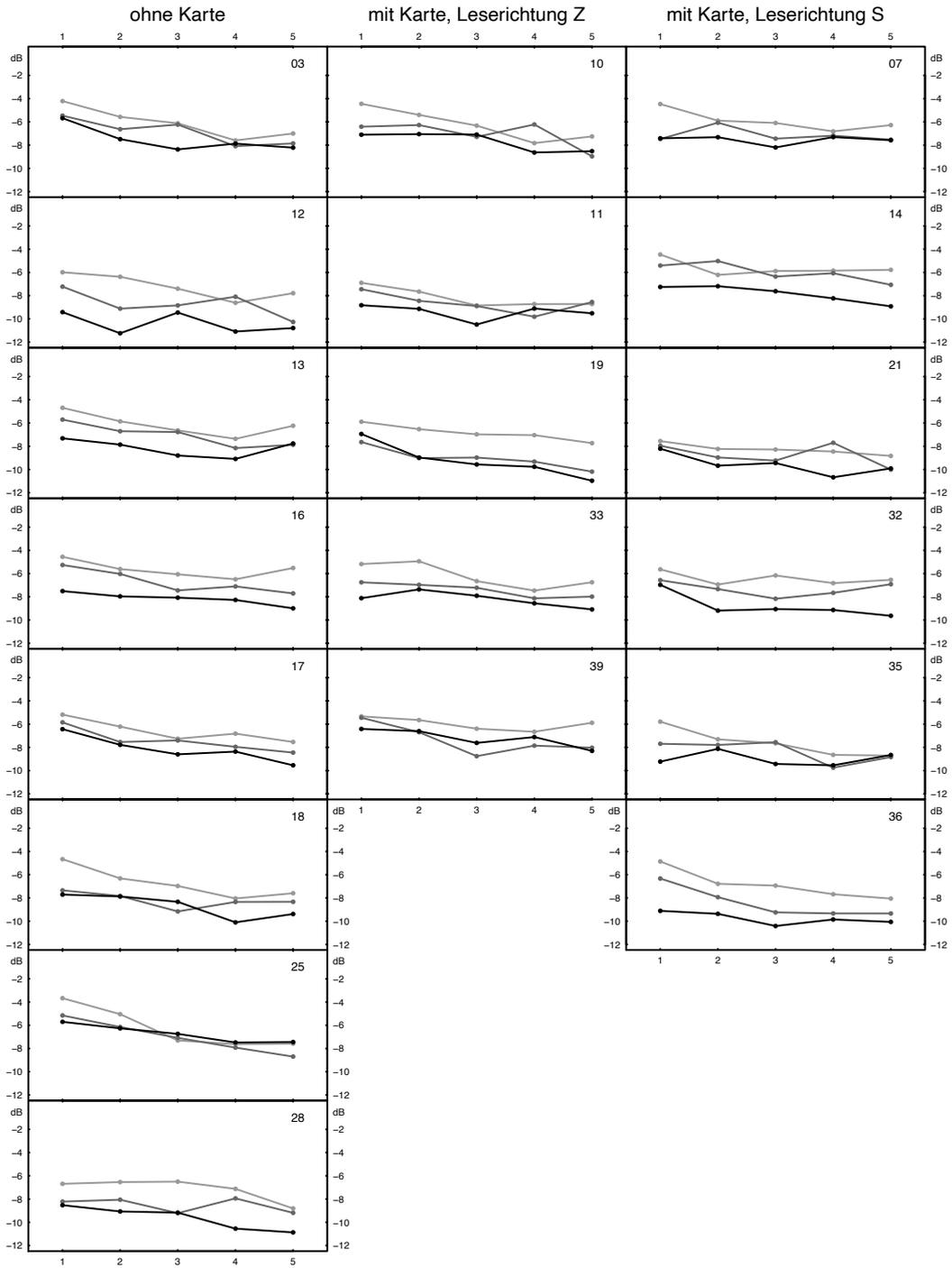


Abbildung A.27: SRT-Verlaufswerte der nicht selektierten, weiblichen Probanden

B Vollständige Darstellung der Varianzanalysen

Tabelle B.1: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Geschlecht, Leserichtung, Testsitzung, Testlauf und Interaktionen (selektierte Hauptgruppe, nur Probanden mit Vorlagenkarte, n = 16). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Geschlecht (G)	1	31,8900	2,9900	0,1096	0,1432
Leserichtung (D)	1	24,3400	2,2800	0,1570	0,1132
Sitzung (S)	2	124,6700	112,8700	< 0,001	0,3953
Testlauf (R)	4	107,0800	66,2000	< 0,001	0,3596
G × D	1	0,6300	0,0600	0,8116	0,0033
G × S	2	2,8600	2,5900	0,0961	0,0148
D × S	2	1,7300	1,5700	0,2287	0,0090
G × R	4	1,1800	0,7300	0,5763	0,0061
D × R	4	2,6900	1,6700	0,1734	0,0139
S × R	8	5,8200	2,3400	0,0245	0,0296
G × D × S	2	1,9200	1,7400	0,1966	0,0100
G × D × R	4	0,8700	0,5400	0,7091	0,0045
G × S × R	8	1,6300	0,6500	0,7295	0,0085
D × S × R	8	2,4900	1,0000	0,4414	0,0129
G × D × S × R	8	1,5100	0,6100	0,7703	0,0079

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,017$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 3 von 60

Varianzhomogenität (Levene-Test): $p = 0,1308$

Tabelle B.2: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Testsit-
zung, Testlauf und Interaktion (selektierte Hauptgruppe, n = 32). df = Frei-
heitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives
Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Sitzung (S)	2	267,3600	186,2000	< 0,001	0,3134
Testlauf (R)	4	244,4900	143,0600	< 0,001	0,2944
S × R	8	14,0000	4,7600	< 0,001	0,0233

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,105$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 0 von 15

Varianzhomogenität (Levene-Test): $p = 0,8880$

Tabelle B.3: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Test-
situation, Testlauf und Interaktion (selektierte Hauptgruppe, nur Probanden
mit Vorlagenkarte, n = 16). df = Freiheitsgrad, QS = Quadratsumme, F = F-
Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Sitzung (S)	2	124,6700	94,5900	< 0,001	0,3204
Testlauf (R)	4	107,0800	66,5000	< 0,001	0,2882
S × R	8	5,8200	2,4600	0,0169	0,0215

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,020$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 1 von 15

Varianzhomogenität (Levene-Test): $p = 0,4123$

Tabelle B.4: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Geschlecht, Vorlagenkarte, Testlauf (global) und Interaktionen (selektierte Hauptgruppe, n = 32). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Geschlecht (G)	1	7,6000	0,6600	0,4248	0,0152
Karte (C)	1	37,8500	3,2700	0,0815	0,0712
Test-global (T)	14	525,8500	87,1400	< 0,001	0,5159
G × C	1	27,3400	2,3600	0,1358	0,0525
G × T	14	8,7200	1,4500	0,1293	0,0174
C × T	14	6,4300	1,0700	0,3878	0,0129
G × C × T	14	4,5200	0,7500	0,7240	0,0091

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,014$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 4 von 60

Varianzhomogenität (Levene-Test): $p = 0,1809$

Tabelle B.5: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Geschlecht, Leserichtung, Testlauf (global) und Interaktionen (selektierte Hauptgruppe, nur Probanden mit Vorlagenkarte, n = 16). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Geschlecht (G)	1	31,8900	2,9900	0,1096	0,1432
Leserichtung (D)	1	24,3400	2,2800	0,1570	0,1132
Test-global (T)	14	237,5700	45,5500	< 0,001	0,5547
G × D	1	0,6300	0,0600	0,8116	0,0033
G × T	14	5,6700	1,0900	0,3725	0,0289
D × T	14	6,9200	1,3300	0,1960	0,0350
G × D × T	14	4,3000	0,8300	0,6408	0,0221

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,017$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 3 von 60

Varianzhomogenität (Levene-Test): $p = 0,1308$

Tabelle B.6: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Testlauf (global), (selektierte Hauptgruppe, n = 32). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Test-global (T)	14	525,8500	86,4200	< 0,001	0,4730

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,105$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 0 von 15

Varianzhomogenität (Levene-Test): $p = 0,8880$

Tabelle B.7: ANOVA-Ergebnistabelle: Abhängige Variable: SRT in dB, Faktoren: Testlauf (global), (selektierte Hauptgruppe, nur Probanden mit Vorlagenkarte, n = 16). df = Freiheitsgrad, QS = Quadratsumme, F = F-Wert, p = p-Wert, η^2 = deskriptives Maß für die Varianzaufklärung

Faktor	df	QS	F	p	η^2
Test-global (T)	14	237,5700	44,8400	< 0,001	0,4732

Prüfung der Testvoraussetzungen:

Normalverteilung (p_{\min} aller Shapiro-Wilk-Tests): $p > 0,105$

Anzahl nicht normalverteilter Teilmengen (laut Shapiro-Wilk-Tests): 0 von 15

Varianzhomogenität (Levene-Test): $p = 0,4123$

Tabelle B.8: Post-hoc-Vergleich (Tukey HSD) der Testdurchläufe: Sitzung 1. Grau hinterlegt: $p < 0,05$

Testlauf	2	3	4	5
1	< 0,001	< 0,001	< 0,001	< 0,001
2		0,0731	< 0,001	< 0,001
3			0,3184	0,0684
4				0,9487

Tabelle B.9: Post-hoc-Vergleich (Tukey HSD) der Testdurchläufe: Sitzung 2. Grau hinterlegt: $p < 0,05$

Testlauf	2	3	4	5
1	0,0476	< 0,001	< 0,001	< 0,001
2		0,2491	0,0734	0,0034
3			0,9800	0,5043
4				0,8418

Tabelle B.10: Post-hoc-Vergleich (Tukey HSD) der Testdurchläufe: Sitzung 3. Grau hinterlegt: $p < 0,05$

Testlauf	2	3	4	5
1	0,1634	0,0025	< 0,001	< 0,001
2		0,5897	0,1751	0,010
3			0,9396	0,3446
4				0,8111

Tabelle B.11: Post-hoc-Vergleich (Tukey HSD) der Sitzungen: Testdurchlauf 1. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	< 0,001	< 0,001
2		0,0079

Tabelle B.12: Post-hoc-Vergleich (Tukey HSD) der Sitzungen: Testdurchlauf 2. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	< 0,001	< 0,001
2		0,0221

Tabelle B.13: Post-hoc-Vergleich (Tukey HSD) der Sitzungen: Testdurchlauf 3. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	< 0,001	< 0,001
2		0,0325

Tabelle B.14: Post-hoc-Vergleich (Tukey HSD) der Sitzungen: Testdurchlauf 4. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	0,0578	< 0,001
2		0,0558

Tabelle B.15: Post-hoc-Vergleich (Tukey HSD) der Sitzungen: Testdurchlauf 5. Grau hinterlegt: $p < 0,05$

Sitzung	2	3
1	0,0186	< 0,001
2		0,0277

Danksagung

Herrn Univ.-Prof. Dr. med. Dr. h. c. Rudolf Hagen, Direktor der Klinik und Poliklinik für Hals-, Nasen- und Ohrenkrankheiten, plastische und ästhetische Operationen der Universität Würzburg danke ich herzlichst für die Übernahme des Referates.

Weiterhin gilt mein Dank Frau Priv.-Doz. Dr. med. Regina Taurines für die Übernahme des Koreferates.

Einen besonderen Dank möchte ich Herrn Dipl.-Ing. Stefan Brill für seine herausragende Betreuung und fachliche Hilfe aussprechen.

Außerordentlicher Dank gebührt meinen Freunden und Kommilitonen, die als motivierte Probanden maßgeblich zum Gelingen der Arbeit beigetragen haben und durch die die Arbeit an der Dissertation zu einer großen Freude wurde.

Weiterhin danke ich sämtlichen Mitarbeitern der HNO-Abteilung der Universität Würzburg.

Vor allem danke ich meinen Eltern und meinen beiden Schwestern sowie meinen Freunden und Wegbegleitern für deren fortwährenden Zuspruch, deren kraftvolle Motivation, deren stetige Geduld und deren hervorragende Unterstützung, auf die ich immer bauen durfte.

Besten Dank Ihnen und Euch allen!