

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

Direktor: Prof. Dr. med. Rudolf Hagen

**Die Methode der Urteilsabgabe  
als Einflussfaktor auf Lautheitsurteil  
und Vorgängerabhängigkeit  
bei kategorialer Lautheitsskalierung**

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

Roland Haaga

aus Rottenacker

Hamburg, September 2009



Referent: Prof. Dr. med. Rudolf Hagen  
Korreferent: Priv.-Doz. Dr.-Ing. habil. Mario Cebulla  
Dekan: Prof. Dr. med. Matthias Frosch

Tag der mündlichen Prüfung:

Der Promovend ist Arzt



Gewidmet meinen Eltern, die mich immer unterstützt haben.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Einführung . . . . .	1
1.2	Probleme der Hörgeräteakustik . . . . .	2
1.2.1	Die Problematik der Lautheitsmessung . . . . .	3
1.3	Das Würzburger Hörfeld (WHF) . . . . .	6
1.3.1	Das Skaliertablett . . . . .	8
<b>2</b>	<b>Problemstellung</b>	<b>11</b>
2.1	Kontexteffekte . . . . .	11
2.1.1	Einfluss der Urteilsmethode auf die Skalierung . . . . .	12
2.1.2	Sequenzeffekte . . . . .	13
2.1.3	Einfluß der Zeitposition auf die Lautheitsskalierung . . . . .	14
<b>3</b>	<b>Versuchsplanung und Methodik</b>	<b>15</b>
3.1	Auswahl der Probanden . . . . .	15
3.2	Tonaudiogramm . . . . .	16
3.3	Voruntersuchungen zum Testablauf . . . . .	17
3.4	Reihenfolge der verschiedenen Urteilsmethoden . . . . .	17
3.5	Urteilsabgabe mit dem Skaliertablett . . . . .	18
3.6	Mündliche Abgabe des Urteils . . . . .	19
3.7	Urteilsabgabe mit dem Laserpointer . . . . .	20
3.8	Registrierung der abgegebenen Urteile . . . . .	21
3.9	Aufbau der Stimulussequenzen . . . . .	21
3.10	Das CCITT-Rauschen . . . . .	23
3.11	Probandeninstruktion . . . . .	24
<b>4</b>	<b>Versuchsdurchführung</b>	<b>27</b>
4.1	Camera silens - schalltoter Raum . . . . .	27
4.2	Aufstellung der Geräte in der Kammer . . . . .	29
4.3	Kontrolle der Schallpegel . . . . .	29

4.4	Versuchsablauf	31
<b>5</b>	<b>Technischer Aufbau</b>	<b>33</b>
5.1	Computer	33
5.2	Videokamera	33
5.3	USB-Soundkarte	33
5.4	Lautsprecherauswahl	34
5.5	Laserpointer	34
5.6	Skaliertablett	34
5.7	Skalierungsprogramm	35
5.8	Patientendatenbank	36
<b>6</b>	<b>Ergebnisse</b>	<b>37</b>
6.1	Statistische Auswertung	37
6.1.1	Verwendete statistische Tests und Verfahren	38
6.1.2	Umgang mit fehlerhaften Werten	39
6.2	Visuelle Betrachtung der Rohdaten	41
6.3	Visuelle Betrachtung der Urteilsdichte	43
6.3.1	Festlegung des Kerndichteschätzers	45
6.4	Urteilsunterschiede der Probanden	46
6.5	Urteil in Abhängigkeit von der Urteilsmethode	51
6.6	Urteil in Abhängigkeit vom Pegel	51
6.7	Abhängigkeit von der Zeitposition	57
6.8	Vorgängerpegel und -urteilsabhängigkeit	58
6.8.1	Vorgängereffekte in Abhängigkeit vom Probanden	64
6.8.2	Vorgängereffekte in Abhängigkeit von der Urteilsmethode	65
6.8.3	Vorgängereffekte in Abhängigkeit vom Pegel	68
6.8.4	Effekt der Medianlage der Urteile	76
6.9	Überblick über Vorgängereffekte	76
6.10	Verhalten und Aussagen der Probanden	77
<b>7</b>	<b>Diskussion</b>	<b>81</b>
7.1	Visuelle Betrachtung des Urteilsbereichs und der Urteilsdichte	81
7.2	Urteilsunterschiede der Probanden	83
7.3	Auswirkungen der Urteilsmethoden auf das Urteil	83
7.4	Auswirkungen des Pegels auf das Urteil	85



7.5	Auswirkungen der Zeitposition der Urteilmethoden . . . . .	86
7.6	Vorgängerpegel und -urteilsabhängigkeit . . . . .	87
7.6.1	Probandenabhängigkeit . . . . .	88
7.6.2	Abhängigkeit von der Urteilmethode . . . . .	88
7.6.3	Abhängigkeit vom Pegel . . . . .	90
7.6.4	Abhängigkeit von der Medianlage der Urteile . . . . .	90
7.7	Mögliche unberücksichtigte Einflussfaktoren und Fehler . . . . .	91
<b>8</b>	<b>Ausblick</b>	<b>93</b>
<b>9</b>	<b>Zusammenfassung</b>	<b>95</b>
<b>A</b>	<b>Übersichtstafeln Lautheitsskalierung und Urteilsdichte</b>	<b>99</b>
<b>B</b>	<b>Skalierungsdaten und Vorgängerabhängigkeiten</b>	<b>105</b>
	<b>Literaturverzeichnis</b>	<b>133</b>
	<b>Danksagung</b>	
	<b>Lebenslauf</b>	



# 1 Einleitung

## 1.1 Einführung

Wie kein anderer Sinn beeinflusst die Funktionstüchtigkeit des Gehörs die zwischenmenschliche Kommunikation. Trotzdem wird die Fähigkeit des Hörens und Verstehens von den Allermeisten als selbstverständlich angesehen. Die Komplexität des Hörvorganges ist dabei den wenigsten bewusst. Hellbrück (1993) bezeichnet das Gehör als das intellektuellste aller menschlichen Sinnesorgane. Diese Leistungsfähigkeit des menschlichen Gehörs ermöglicht ein äußerst feines Differenzierungsvermögen in Lautstärke, Tonhöhe, Färbung, Rauigkeit etc. der wahrgenommenen Geräusche. Je komplexer ein System aber ist, desto schwerer wird es auch, dieses System wissenschaftlich zu analysieren, zu verstehen und zu rekonstruieren. Obwohl laut Hellbrück (1993) „[...] die Akustik - genauer die Psychoakustik - als die erste *experimentelle* Wissenschaft [...]“ der Weltgeschichte gilt, sind auch noch heute bei weitem nicht alle Aspekte des Vorganges „Hören“ verstanden. Crane (1983) schreibt:

„The auditory system, like the nervous system as a whole, is so complex that one can find instances and evidence to support almost any theory.“

Dies liegt auch daran, dass man das Gehör nur schwer in leicht zu untersuchende Einzelteile oder Funktionseinheiten zerlegen kann, die man nur aneinanderzureihen braucht um das Gehör als Ganzes zu verstehen. Nichtsdestotrotz konnte man durch intensive Forschung die einzelnen Vorgänge der Schallübertragung, -rezeption und -verarbeitung recht exakt darlegen. Einer der wichtigsten Forscher war Georg von Békésy (1899 - 1972), welcher 1961 den Nobelpreis für Medizin für seine Forschung zur Wanderwellen-Theorie erhielt. Als sehr nützlich im Rahmen der Untersuchung von einzelnen Parametern haben sich objektive Testverfahren (z.B. Hirnstammaudiometrie (BERA) oder Stapediusreflexmessung) erwiesen. Jedoch können auch diese Tests nicht die zentrale Verarbeitung im Gehirn objektiv darstellen.

## 1 Einleitung

Anders als die Physiologie oder die Anatomie betrachtet die Psychoakustik ein umfassenderes Konzept des Gehörs. Das heißt, sie greift nicht einzelne Aspekte des Hörvorganges heraus, sondern sieht das Gehör als Ganzes, also inklusive der Integrationsleistung des zentralen Nervensystems. Die Psychoakustik ist ein Teilbereich der von Gustav Theodor Fechner (1801 - 1887) begründeten Psychophysik. In der Psychophysik steht die Idee der quantitativen Beschreibung von Empfindungen und Wahrnehmungen im Mittelpunkt.

„Da hierzu die physikochemische Welt der Reize mit der subjektiven Welt der Wahrnehmungen in Beziehung gesetzt werden, nennt man diese Forschungsrichtung Psychophysik.“ (Schmidt und Schai-ble, 2006)

Hierdurch wird der Mensch zwangsläufig selbst zum Messgerät, indem er vorgegebene Stimuli nach bestimmten Kriterien bewertet. Solche Tests werden als subjektive Tests zusammengefasst. Ein sehr einfacher subjektiver Test ist zum Beispiel der Stimmgabel-Test nach Weber und Rinne, welcher eine differentialdiagnostische Unterscheidung zwischen Innenohr- und Mittelohrschädigung erlaubt. Dass der Mensch durchaus als zuverlässiges Messinstrument funktionieren kann zeigte z. B. Nowak (1980).

Ein in der Psychoakustik häufig angewandtes Verfahren ist die Einordnung der Lautheit eines Stimulus. Mittels der Lautheitsskalierung versuchen Mediziner und Hörgeräteakustiker zusammen mit schwerhörigen Patienten Hörhilfen individuell optimal anzupassen. Ein bestimmtes Verfahren der Lautheitsskalierung wurde nach dem Entwicklungsort als „Würzburger Hörfeld“ bezeichnet. Darauf soll in Kapitel 1.3 (Das Würzburger Hörfeld (WHF), S. 6) näher eingegangen werden.

### 1.2 Probleme der Hörgeräteakustik

Die Hörgeräteanpassung ist ein äußerst komplizierter Prozess, der einer interdisziplinären Zusammenarbeit von Audiologen, Pädagogen, Sprachtherapeuten, Ingenieuren und Ärzten bedarf. Hörhilfen sollen in erster Linie ein freies Sprachverständnis, d.h. das Verstehen von Sprache ohne visuelle Unterstützung wie Lippenlesen oder Gebärdensprache ermöglichen und dadurch die Kommunikationsfähigkeit des Schwerhörigen wiederherstellen. Deshalb muss vor allem der

Schallbereich verstehbar gemacht werden, der der menschlichen Umgangssprache entspricht. Die Frequenz dieses sogenannten Hauptsprachbereiches liegt zwischen 250 Hz und 4000 Hz und der Schalldruck zwischen 40 dB und 80 dB. Graphisch lässt sich der Hauptsprachbereich deutlich übersichtlicher darstellen (vgl. grün hinterlegte Fläche in Abb. 1.1). Aufgrund seiner hier erkennbaren Form wird er auch als „Sprachbanane“ bezeichnet.

In erster Linie wird durch das Hörgerät das ankommende Signal so verstärkt, dass es für den Schwerhörigen hörbar wird.

„The first order of business in the audiological management of a hearing-impaired person must be to increase the intensities of sounds so they become audible.“ (Gelfand, 2001, S. 443)

Dies beschreibt jedoch bei weitem nicht die Möglichkeiten eines modernen digitalen Hörgerätes. Denn nicht nur die Lautheit spielt eine entscheidende Rolle im Sprachverstehen. Auch andere Parameter wie z.B. die Verzerrung oder das Eigenrauschen des Hörgerätes beeinflussen das Hörerleben der Patienten. „Allerdings stellt die Normalisierung der Lautheitsempfindung [...] eine notwendige, nicht aber hinreichende Bedingung dar, d.h. eine sprachaudiometrische und subjektive Überprüfung des Versorgungserfolgs ist unverzichtbar.“ (Kießling, 1995)

Im Weiteren soll entsprechend des Themas der vorliegenden Arbeit jedoch nur auf die Problematik der Lautheit eingegangen werden.

### 1.2.1 Die Problematik der Lautheitsmessung

Die subjektive Lautheitsempfindung hängt neben dem Schallpegel auch von der Frequenz ab. Ein Stimulus mit immer gleichem Schalldruckpegel wird mit ansteigender Frequenz erst lauter und ab ca. 5000 Hz wieder leiser empfunden, bis er schließlich nicht mehr zu hören ist (Hellbrück und Heller, 1980). Stanley Smith Stevens (1906 - 1973) führte deshalb die Einheit „phon“ als Maß für die subjektiv empfundene Lautheit ein. Die subjektive Lautheit wird durch Vergleich mit der Lautheit und dem Schalldruckpegel eines 1000 Hz Referenztones bestimmt. Daraus ergeben sich Kurven gleich empfundener Lautstärken, sogenannte Isophone. Abbildung 1.1 zeigt den Verlauf der Isophone und den Hauptsprachbereich (grün hinterlegt). Ein weiterer Faktor, der die Lautheit komplexer Schalle beeinflussen kann ist die Bandbreite des Stimulus. Breitbandige Stimuli, d.h. Stimuli die mehrere Frequenzgruppen umfassen, werden statistisch signifikant

# 1 Einleitung

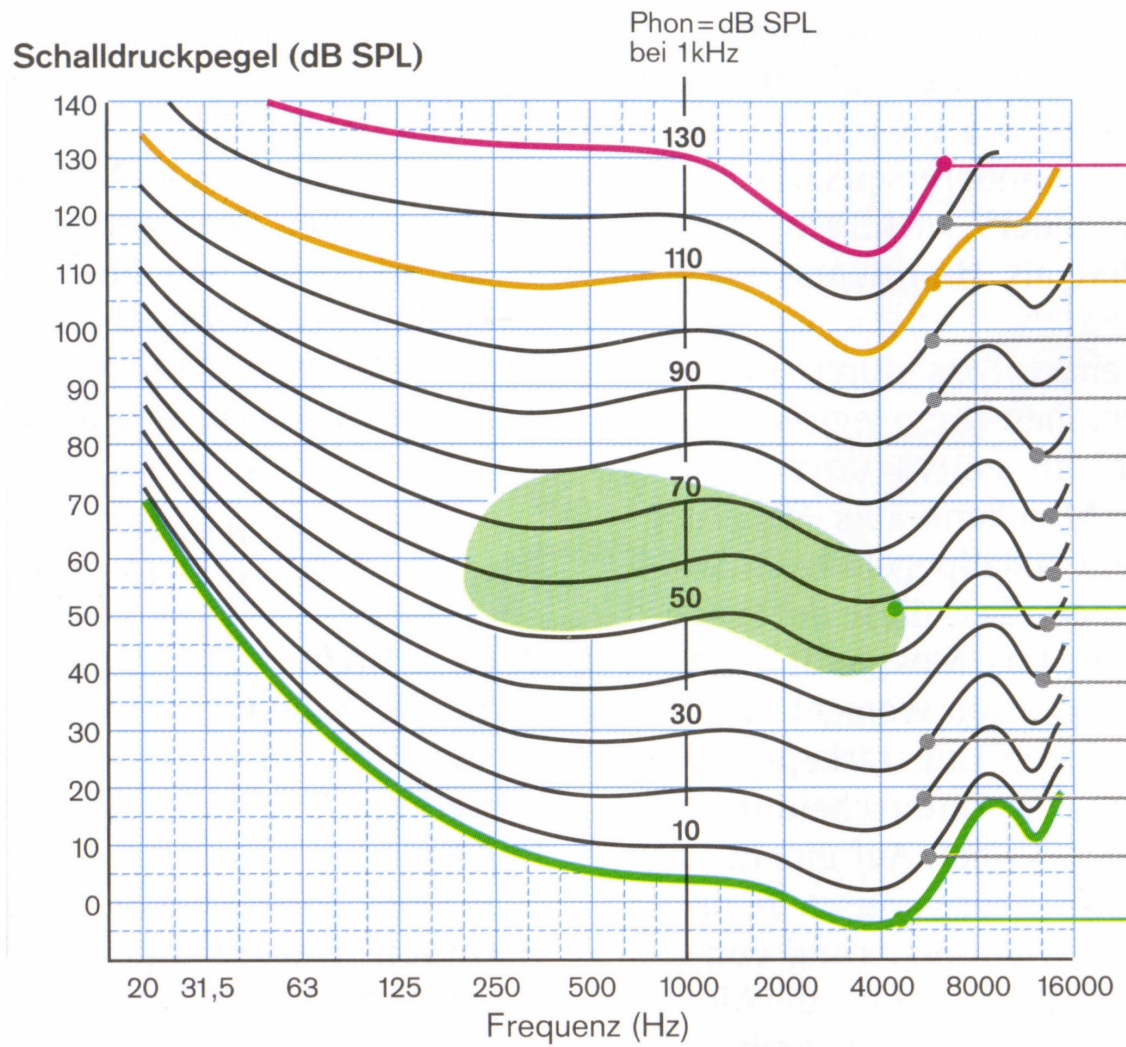


Abbildung 1.1: Verlauf der Isophone und Hauptsprachbereich

lauter beurteilt als schmalbandige Stimuli (Hubner und Ellermeier, 1993; Nägle, 2004; Ricketts und Bentler, 1996; Zwicker und Fastl, 1990). Auch die Dauer des Stimulus ist wichtig; allerdings nur bis zu einer Länge von 300 Millisekunden. Ist der Stimulus länger als 300 ms, spielt die Präsentationsdauer keine Rolle mehr (Zwicker und Fastl, 1990). Auch physikalisch schlecht fassbare Eigenschaften, wie z.B. das Empfinden des Stimulus als unangenehm bzw. angenehm können das Lautheitsempfinden beeinflussen (Fucci et al., 1993). Man sieht also, dass das subjektive Empfinden von Lautheit vielen verschiedenen Parametern unterworfen ist, die in der Audiologie berücksichtigt werden müssen.

Für eine optimale Einstellung der Hörhilfe reicht es nicht aus, nur die Hörschwelle, den Pegel des angenehmen Hörens (most comfortable loudness, MCL) oder die Unbehaglichkeitsschwelle (uncomfortable loudness, UCL) des Patienten zu bestimmen. Aus diesen einzelnen Parametern lässt sich der überschwellige Lautheitsanstieg des Patienten nämlich nicht ableiten (Hellbrück und Moser, 1985; Hohmann und Kollmeier, 1995; Kießling, 1995; Launer et al., 1996; Moser, 1987).

„Insofern bestätigt sich, daß hörschwellenorientierte Formeln, wie z.B. Berger, POGO oder NAL, allenfalls für eine Voreinstellung der Hörgeräte geeignet sind, nicht aber für die endgültige Wahl der Anpaßparameter.“ (Kießling, 1995)

Auch Schorn und Baumann (1999) messen der Lautheitsskalierung unter Berücksichtigung der individuellen Restdynamik eine besondere Bedeutung zur Feinanpassung zu.

Das Hörvermögen des Patienten muss also im ganzen Schallpegelumfang und insbesondere im Hauptsprachbereich analysiert werden. S. S. Stevens führte dazu in seiner „neuen Psychophysik“ das Verfahren der direkten Skalierung ein (Stevens, 1955, 1956, 1957). Darauf aufbauend entstanden das ebenfalls von Stevens entwickelte Verfahren der absoluten Größenschätzung (Absolute Magnitude Estimation, AME) und der kreuzmodale Größenabgleich (Cross Modality Matching, CMM). Diese beiden Verfahren haben sich besonders in der amerikanischen Audiometrie durchgesetzt. Sie arbeiten mit der Methode der Verhältnisskalierung.

Auf einer Partitions- bzw. Kategorienskalierung beruhen dagegen die verschiedenen Verfahren der Kategorienunterteilung (KU). Dadurch ergeben sich konzeptionelle Unterschiede zwischen dem Verfahren der Größenschätzung und

## 1 Einleitung

den kategorienbasierten Verfahren. Deshalb herrscht

„ [...] Uneinigkeit hinsichtlich der Frage [...] wie Empfindungsgrößen adäquat *beurteilt* werden sollen. Unter den Skalenforschern und den Anwendern von psychophysischen Skalen in der Akustik gibt es Befürworter der Verhältnismethoden und die “Advokaten“ der Kategorienskalierung, und dies ist oft mit der Ablehnung der jeweils anderen Methode verbunden.“ (Hellbrück, 1993, S. 211)

In Hellbrück (1993) werden die entscheidenden Unterschiede der beiden Skalierungsverfahren mit ihren Konsequenzen für den Audiologen ausführlich dargestellt.

Die ersten praktisch anwendbaren Verfahren, die zur Hörgeräteanpassung die kategoriale Lautheitsskalierung nutzten wurden bereits von Pascoe (1978, 1986) und Allen et al. (1990) entwickelt. Dennoch konnte sich die kategoriale Lautheitsskalierung längere Zeit kaum durchsetzen. Kollmeier (1997) vermutet die Ursachen darin, dass das Verfahren „nur von wenigen Audiometer-Herstellern kommerziell angeboten wurde“ und das „positive Aufwand-Nutzenverhältnis nicht in ausreichender Deutlichkeit erbracht wurde.“ Erst in den 90er Jahren gewannen auf Kategorienskalen basierende audiometrische Verfahren im englischsprachigen Raum an Boden. Heute liegen zahlreiche Arbeiten vor, die die Lautheitsskalierung als solides Verfahren zur Anpassung von Hörhilfen und zur audiologischen (Topo-)Diagnostik sowohl bei Erwachsenen als auch bei Kindern anerkennen. Einige Beispiele sind: Gocht (2002); Hellbrück und Thomamüller (1997); Heller (1991); Hohmann und Kollmeier (1995); Kießling (2000); Kießling et al. (1996, 1994, 1993); Kugler und Moser (1995); McColl und Fucci (1999); Meister und von Wedel (1995); Moser (1987); Robinson und Gatehouse (1996); Völter et al. (2001)

### 1.3 Das Würzburger Hörfeld (WHF)

Man unterscheidet in der Psychophysik zwischen absoluten und vergleichenden Urteilen. „Als “absolut“ bezeichnet man Urteile wie “schwer“ und “leicht“, “laut“ und “leise“, “hell“ und “dunkel“ oder “groß“ und “klein“.“ (Haubensak, 1985). Unter vergleichenden Urteilen versteht man dagegen Urteile wie „lauter als ...“ oder „leiser als ...“. Das von Heller (1985), Hellbrück und Moser (1985) und Moser



Tabelle 1.1: Kategorien und Skala der Lautheitsskalierung nach Heller (1985)

»Sehr leise«	1 – 10	hel
»Leise«	11 – 20	hel
»Mittel«	21 – 30	hel
»Laut«	31 – 40	hel
»Sehr laut«	41 – 50	hel

(1987) entwickelte Verfahren des „Würzburger Hörfeld“ arbeitet mit Absoluturteilen. Es ist „[...] ein psycho-akustisches Hörtestverfahren, das auf subjektiver Lautheitsbeurteilung von Testsignalen [...] beruht.“ (Kugler und Moser, 1995).

Dem Patienten bzw. der Versuchsperson wird ein bestimmter Stimulus vorgespielt, dessen Lautheit er dann einstufen soll. Der Stimulus kann sowohl im Schallpegel als auch in der Frequenz variiert werden. Es können auch die verschiedensten Arten von Tönen und Geräuschen als Stimulus verwendet werden (s. Abschnitt 3.10, Das CCITT-Rauschen, S. 23). Damit die Versuchsperson die empfundene Lautheit angeben kann, wird ihr eine Skala vorgegeben. Diese ist in fünf Kategorien unterteilt, welche verbal mit »sehr laut«, »laut«, »mittellaut«, »leise« und »sehr leise« bezeichnet sind. Jede dieser fünf Kategorien ist nochmals zehnfach unterteilt und durchnummeriert. Somit entsteht eine von 1 bis 50 abgestufte Skala. Dieser Unterteilung in 50 Schritte wurde - benannt nach Otto Heller - die Bezeichnung „hel“ zugeteilt. Andere Arbeiten benutzen auch die Bezeichnung „KU-Einheiten“ (KU = Kategorien-Unterteilung) oder schlicht „cu“. Die verbale Benennung der fünf Kategorien gibt dem Probanden die Möglichkeit, die künstlichen Stimuli des Laborversuches mit Abstufungen aus dem Alltag zu vergleichen. „Die Verbalisierung ist notwendig, um dem Probanden die Urteilssicherheit zu geben, die in seinem individuellen Bezugssystem für Lautstärken begründet ist.“ (Hellbrück und Moser, 1985). Durch die Feinunterteilung der Kategorien wird dem „wissenschaftlichen Genauigkeitsanspruch“ (Hellbrück et al., 1995) Rechnung getragen. Außerdem ist es von großer Bedeutung, dass die Skala mindestens gleich viel Antwortmöglichkeiten zulässt wie verschiedene Stimuli vorhanden sind. Bei einer Anzahl von insgesamt nur fünf Kategorien wäre die maximal verwendbare Zahl von fünf verschiedenen Stimuli schnell überschritten. Eine Übersicht über die Kategorien und die Zuordnung der Zahlenwerte zeigt Tabelle 1.1. Als Bezugssystem bezeichnet man einen internen Skalierungsmechanismus, der bis zum ca. sechsten Lebensjahr ausgebildet ist

## 1 Einleitung

und welcher lebenslang stabil bleibt (Zwislocki und Goodman, 1980). Trotzdem ist jedes Bezugssystem variabel. Je nach Vergleich mit einer Maus oder einem Hochhaus kann ein Elefant rießig groß aber auch sehr klein sein. Deshalb sollte der Proband vor Beginn des Versuches sein Bezugssystem erst „eichen“. Eine Übungssequenz mit allen im Versuch vorkommenden Schallpegeln in zufälliger Reihenfolge soll dem Probanden den Umfang der verwendeten Stimuli aufzeigen. Diese Übungssequenz wird dem Probanden vor Beginn der eigentlichen Messung so lange vorgespielt bis er sich sicher fühlt. Diese „Orientiertheit“ des Probanden über die vorkommenden Stimuli bezeichnet Heller (1980) „als Prinzip jedes Skalierungsexperimentes“. In Heller (1991) verweist er nochmals auf die Wichtigkeit der Orientiertheit der Versuchsperson. Mit der Problematik der Bezugssysteme und absoluten Urteilen beschäftigen sich Haubensak (1985, 1992); Lauterbach und Sarris (1980); Witte (1960); Zwislocki (1983); Zwislocki und Goodman (1980) eingehend.

Auch die Anzahl der Kategorien ist ein häufig diskutiertes Thema. Nach Haubensak (1985) entsprechen fünf Kategorien einer Einteilung, wie sie häufig in vielen Bereichen des Alltags vorgenommen wird. Dies passt zu den Vorgaben von Hellbrück und Moser (1985) ein möglichst realitäts- und alltagsnahes Skalierungssystem zu verwenden. In Lauterbach und Sarris (1980) gibt es mehrere Artikel, die sich mit der Bedeutung der Kategorienanzahl beschäftigen.

Im ursprünglichen Verfahren des WHF bekommt der Proband den Stimulus zweimal zu hören. Beim ersten Mal soll er die Lautheit des Stimulus nur in eine der fünf verbal benannten Kategorien einteilen, während er nach Präsentation des zweiten Stimulus dann die Feinunterteilung innerhalb der Kategorie vornehmen soll. Weiterführende Untersuchungen von Kießling et al. (1993) und Hohmann und Kollmeier (1995) haben gezeigt, dass auch ein einstufiges Verfahren, bei dem der Proband gleich beim ersten Mal die Feinunterteilung vornimmt, ausreichend zuverlässige Ergebnisse liefert. Dadurch wird bei gleicher Leistung der Untersuchung eine enorme Zeitersparnis erreicht. Besonders in der audiologischen Praxis ist deshalb das einstufige Vorgehen von großem Vorteil.

### 1.3.1 Das Skaliertablett

Ein Test, der im klinischen Alltag zur Diagnostik eingesetzt wird, sollte für den Patienten leicht verständlich und gut handhabbar, sowie effizient und robust gegenüber Störungen sein. Er muss aber auch eine hohe Genauigkeit und Repro-

duzierbarkeit aufweisen und einfach zu interpretierende Resultate liefern (Kollmeier, 1997; Ricketts, 1997). Diese allgemeingültigen Kriterien wurden bei der Entwicklung des WHF bestmöglich umgesetzt. Die Urteilsabgabe erfolgt beim WHF mit einem elektronischen Skaliertablett, das an einen PC angeschlossen ist. In der Mitte des Tablett befindet sich ein Drucksensorstreifen, der mit der oben erwähnten Skala nach Heller beschriftet ist. Die Beschriftung des Skalier-tablettes ist in Abbildung 1.2 wiedergegeben. Der Proband drückt seinem Urteil entsprechend mit dem Finger auf den Sensorstreifen und der zugehörige Wert wird automatisch registriert. Die Beschriftung ist beliebig austauschbar und lässt sich somit an die Fähigkeiten der Versuchsperson anpassen. Bei Kindern werden z.B. anstatt der Beschriftung Symbole verwendet. Meister et al. (2001) verwenden größer werdende Elephanten und Ellis und Wynne (1999) skizzierten Gesichter zur Bezeichnung der Kategorien.

# 1 Einleitung

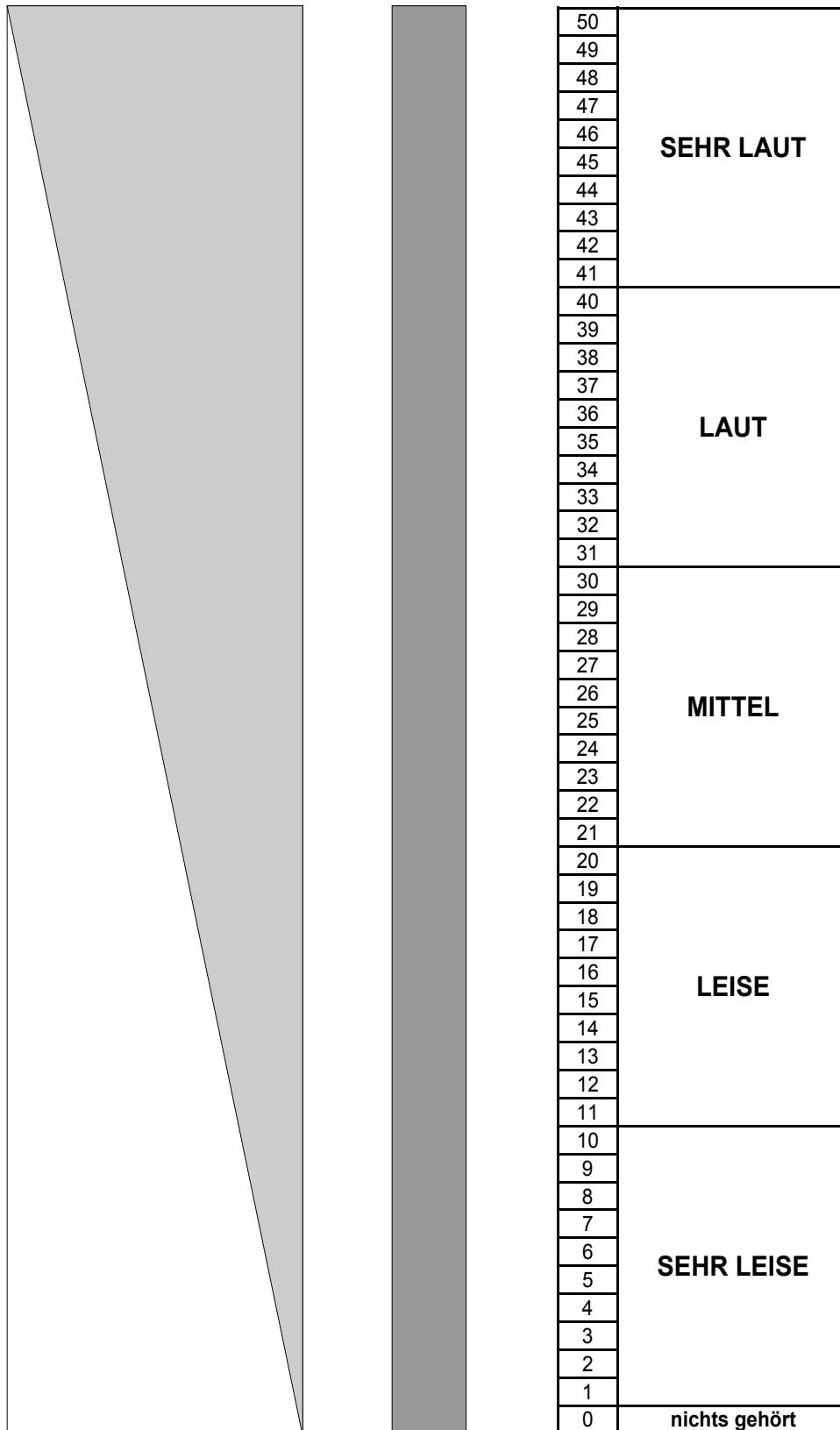


Abbildung 1.2: Skala und Beschriftung des Skaliertabletts

## 2 Problemstellung

Die ständige technische Weiterentwicklung der Hörhilfen gestattet immer komplexere Einstellungsmöglichkeiten. Besonders die Einführung der Digitaltechnik hat einen enormen Sprung nach vorne gebracht. Bei der Entwicklung des Würzburger Hörfeldes hatten **Hellbrück und Moser (1985)** noch Bedenken, den Informationsgewinn mit handelsüblichen Hörgeräten gar nicht voll nutzen zu können. Heute ist es so, dass z.B. manche Cochlea-Implantate bereits zu Techniken befähigt sind, die noch gar nicht benutzt werden.

### 2.1 Kontexteffekte

Um die neuen Möglichkeiten auch voll ausnutzen zu können, müssen in den spezifischen Hörtests erst einmal verlässliche Daten ermittelt werden, die dann in der Konfiguration der Hörhilfe berücksichtigt werden. Dazu müssen die verwendeten Tests selbst korrekt angewendet werden. Auch müssen Effekte ermittelt werden, die zwar nicht im Interesse der Untersuchung liegen, aber dennoch Einfluss auf das Urteil des Patienten haben und somit die Ergebnisse verändern bzw. verfälschen können. Solche experimentellen Randbedingungen werden unter dem Begriff „*Kontexteffekt*“ zusammengefasst. Zu den sehr vielfältigen Kontexteffekten zählen unter anderem:

- Reizverteilung
- Reizumfang
- Häufigkeit der Reize
- Länge des Stimulus
- zeitlicher Abstand der Stimuli = Interstimulusintervall
- (Vor)Erfahrung des Probanden
- Reihenfolge der Reize
- Methode der Antwortabgabe

## 2 Problemstellung

In der vorliegenden Arbeit soll besonders auf die beiden letztgenannten Effekte eingegangen werden. Kontexteffekte, die durch die Reihenfolge der Stimuli bedingt sind, werden in der Psychophysik als „Sequenzeffekte“ bezeichnet (s. Abschnitt 2.1.2, [Sequenzeffekte](#), S. 13).

Zur Bedeutung und Auswirkung von Kontexteffekten gibt es einige Untersuchungen. Erwähnt seien hier [Cox et al. \(1997\)](#); [Fucci et al. \(1986, 1997\)](#); [Gescheider und Hughson \(1991\)](#); [Hellstrom \(1979, 2003\)](#); [Marks \(1993, 1994\)](#) und [Marshall und Brand \(1980\)](#).

### 2.1.1 Einfluss der Urteilsmethode auf die Skalierung

Bei der Weiterentwicklung der kategorialen Lautheitsskalierung wurden verschiedene Verfahren der Urteilsabgabe ersonnen. Die sogenannte „Oldenburger Hörflächenskalierung“ benutzt „als Eingabemedium für die Versuchsperson oder den Versuchsleiter [...] ein berührungsempfindliches LCD-Display“ ([Kollmeier, 1997](#); [Launer et al., 1996](#)). Im Rahmen der Literaturrecherche konnte allerdings nur eine Untersuchung gefunden werden, die die Auswirkungen der unterschiedlichen Arten der Urteilsabgabe auch systematisch untersucht hat. [Kießling et al. \(1993\)](#) verglich die Eingabe der Urteile am Touch-Screen mit der mündlichen Übermittlung an den Versuchsleiter bei Vorlage der Skala auf Papier. Er zeigte, dass „die Abweichungen der Skalierung bei direkter (Versuchsperson) bzw. indirekter Antworteingabe (Untersuchungsleiter) in der Größenordnung der altersbezogenen Test-Retest-Streuung“ lagen. Somit existierte zwischen den beiden Methoden kein signifikanter Unterschied. [Kießling et al. \(1993\)](#) setzten die Untersuchung von Effekten der Urteilsmethode damit nicht weiter fort.

In der vorliegenden Arbeit wurden Effekte der Urteilsmethoden in einem größeren Umfang untersucht als dies bei Kießling et al. der Fall war. Dabei konnten existente Auswirkungen verschiedener Urteilsmethoden auf die Lautheitsskalierung nachgewiesen und qualifiziert werden.

Es wurden fünf verschiedene Arten der Urteilsabgabe entwickelt und die jeweiligen Auswirkungen auf die Skalierung ermittelt. Kapitel 3 ([Versuchsplanung und Methodik](#), S. 15) beschreibt die verschiedenen Urteilsmethoden und geht auf deren Eigenschaften ein.

### 2.1.2 Sequenzeffekte

Auch Sequenzeffekte wurden schon mehrfach in der Literatur dargestellt und analysiert. Jedoch sind die vorliegenden Arbeiten mit dem Verfahren der (Absolute) Magnitude Estimation durchgeführt worden (DeCarlo und Cross, 1990; Jesteadt et al., 1977; Lockhead, 1986; Staddon et al., 1980; Ward, 1979; Ward und Lockhead, 1970, 1971).

Nägle (2004) dagegen untersuchte in ihrer Dissertation Sequenzeffekte mit der kategorialen Lautheitsskalierung. Sie betrachtete den Einfluss von Geschlecht, Stimulustyp und Vorgängerstimulus auf die Skalierung. Es zeigte sich, dass sowohl der Pegel des Vorgängerstimulus als auch das Urteil über den Vorgängerstimulus einen signifikanten Einfluss auf den nachfolgend zu beurteilenden Pegel nehmen. Ursprünglich ging Nägle von einer *negativen* Abhängigkeit vom Vorgängerstimulus aus. Bei erster Überlegung erschien es wahrscheinlicher, dass ein mittellauter Stimulus mit lautem Vorgänger aufgrund des großen Lautstärkekontrastes eher leiser beurteilt wird als mit einem Vorgänger ähnlicher Lautstärke. Diese Annahme traf allerdings nicht zu. Im Gegenteil wurde ein *positiver* Zusammenhang zwischen Vorgänger- und Nachfolgerstimulus nachgewiesen. Das heißt, dass bei einem Vorgänger mit höherem Schallpegel der nachfolgende Stimulus relativ lauter und bei einem Vorgänger mit niedrigerem Schallpegel der nachfolgende Stimulus relativ leiser beurteilt wird. Auch Cross (1973) fand positive Abhängigkeiten vom Vorgängerpegel. Allerdings führte auch er seine Untersuchungen mit dem Verfahren der Magnitude Estimation durch. Die Ergebnisse waren aber mit denen von Nägle (2004) durchaus vergleichbar. Dieser Umstand der Vorgängerabhängigkeit gehört wie oben beschrieben zu den Sequenzeffekten. In dieser Arbeit sollen die Erkenntnisse Nägles ihre logische Fortführung finden. Sie spekulierte, dass eventuell doch eine negative Abhängigkeit bestünde, welche aber durch andere Einflüsse überdeckt werden würde.

Ihre Idee war, dass die Urteilsabgabe mit dem Skaliertablett quasi durch die „Faulheit“ des Probanden, seinen Arm zu bewegen, beeinflusst wird. Anders ausgedrückt wäre es möglich, dass wenn der Proband einen hohen Wert gedrückt hat, der nachfolgende Stimulus relativ lauter beurteilt wird, weil der Proband unbewusst zu träge ist, die Hand auf der Skala weiter nach unten zu bewegen. Natürlich soll diese Annahme, dass der Proband unbewusst einen kürzeren Weg wählt auch umgekehrt für einen leisen Vorgänger mit niedrigem Urteil an-

genommen werden. Dieser vermutete Effekt wird im Weiteren mit dem Begriff „Armträgheit“ beschrieben. Noch zu erwähnen ist, dass diese Armträgheit nicht zwingend durch subjektive, unbewusste psychologische Vorgänge bedingt sein muss, sondern durchaus auch als ein rein physikalischer Effekt gesehen werden kann.

Zur Evaluierung dieser potenziellen Armträgheit wurden die bereits in Abschnitt 2.1.1 (Einfluss der Urteilsmethode auf die Skalierung, S. 12) erwähnten fünf Arten der Urteilsabgabe so konstruiert, dass diese Fehlerquelle schrittweise ausgeschaltet wurde. Die verschiedenen Urteilsmethoden werden in Kapitel 3 (Versuchsplanung und Methodik, S. 15) eingehend beschrieben.

### 2.1.3 Einfluß der Zeitposition auf die Lautheitsskalierung

Ein weiterer Effekt, der die Lautheitsskalierung der Probanden beeinflussen könnte, ist in der zeitlichen Position der Urteilsmethoden zu suchen. Es ist denkbar, dass die Position der Urteilsmethode im Ablauf des Versuches Auswirkungen hat. Laut Kollmeier (1997) „[...] kann die Skalierungsqualität durch Lerneffekte offenbar deutlich gesteigert werden, [...]“. Andererseits konnten Hohmann und Kollmeier (1995) in ihrer Untersuchung keinen signifikanten Lerneffekt nachweisen. Vorstellbar ist auch, dass bei längeren Versuchen eine zunehmende Ermüdung des Probanden eintritt, welche eine verschlechterte Skalierung zur Folge hätte. Ward (1987) verfolgt einen weiteren Ansatz, der für das Vorhandensein von Lerneffekten spricht. Er postuliert, dass sich die Skalierung zusammensetzt aus im Kurzzeitgedächtnis abgespeicherten Kontrasteffekten von kurz zuvor Gehörtem und im Langzeitgedächtnis verankerten Erfahrungen. Bei längerer Versuchsdauer könne durch die zunehmende „Datenmenge“ die richtige Einschätzung quasi erlernt werden. Der wahre Sachverhalt scheint also nicht völlig geklärt zu sein.



### 3 Versuchsplanung und Methodik

Da die vorliegende Arbeit an die Dissertation von Therese Nägle (Nägle, 2004) anknüpft, wurde der Versuchsaufbau weitestgehend übernommen. So konnte eine bestmögliche Vergleichbarkeit gewährleistet werden.

#### 3.1 Auswahl der Probanden

Als Probanden dienten 25 normalhörende Personen, an welche mehrere Anforderungen gestellt wurden. In Analogie zu Nägle (2004) wurden nur Personen mit einer Hörschwelle niedriger 20 dB HL zwischen 125 Hz und 4 kHz zu den Versuchen zugelassen. Die Normalhörigkeit wurde vor den eigentlichen Versuchen kontrolliert. Dazu wurden die Probanden in einer Audiometrikammer der HNO-Universitätsklinik Würzburg untersucht (s. Abschnitt 3.2, Tonaudiogramm, S. 16). Somit konnte sichergestellt werden, dass die Probanden auch den leisesten Ton von 30 dB im Versuch sicher hören konnten.

Weiterhin wurden nur Personen zwischen 20 und 30 Jahren zugelassen. Ältere Personen wurden bewusst ausgeschlossen, da sich im Alter sowohl die absolute Hörschwelle als auch die Lautheitswahrnehmung verändern. Unter anderem haben Heller (1985) und Hellbrück und Moser (1985) auf den altersbedingten Hochttonverlust hingewiesen, der in Kombination mit einer Lautheitszunahme im mittleren Frequenzbereich auftritt.

Es nahmen insgesamt 12 weibliche und 13 männliche Versuchspersonen teil. Auf eine Gleichverteilung der Geschlechter wurde bei der Auswahl der Probanden nicht geachtet, da mehrere Arbeiten bereits nachweisen konnten, dass das Geschlecht bei Lautsprecherdarbietung der Stimuli keine signifikanten Auswirkungen auf das abgegebene Urteil hat (Hellbrück, 1983; Hellbrück et al., 1984, 1995; Nägle, 2004; Nowak, 1980). Dass die Geschlechter trotzdem annähernd gleichverteilt sind - soweit das bei einer ungeraden Anzahl von Probanden möglich ist - ist reiner Zufall. Im Gegensatz zur Freifelddarbietung mit Lautsprechern würde das Geschlecht bei Kopfhörerdarbietung sehr wohl eine Rolle spielen. Frauen haben nämlich durchschnittlich ein kleineres Gehörgangsvolumen als

Männer. Daraus resultiert trotz gleicher Schallenergie bei Frauen ein höherer Schalldruck am Trommelfell als bei Männern. Deshalb skalieren Frauen grundsätzlich etwas lauter. Die relative Empfindlichkeit ist bei beiden Geschlechtern jedoch gleich, d.h. ein Lautheitszuwachs wird von beiden Geschlechtern gleich skaliert (Hellbrück et al., 1980; Hellbrück, 1993). Ein weiterer Nachteil der Kopfhörerdarbietung wäre, dass diese wie Bandpassfilter funktionieren können und außerdem Lautheit und Tonfärbung beeinflussen (Zwicker und Fastl, 1990). Weiterhin kann alleine die unterschiedliche Positionierung des Kopfhörers, sprich dessen Auflagefläche das Lautheitsempfinden beeinflussen (Lehnhardt und Laszig, 2001). Deshalb kommen auch sie zu dem Schluss, dass es besser ist „das Hörvermögen mit Tönen zu messen, die von einem Lautsprecher abgestrahlt werden“. Allerdings wird auch eingeräumt, dass der technische Mehraufwand der Lautsprecherdarbietung in der alltäglichen Audiometrie zu groß ist, um diese generell zu bevorzugen.

## 3.2 Tonaudiogramm

Das Tonaudiogramm wurde mit dem Audiometer CAD03/1 der Firma Westra Electronic durchgeführt. Die folgenden Angaben wurden aus der Gerätebeschreibung (Westra CAD03/1) übernommen. Das Klasse I Audiometer ist ein Reinton- und Sprachaudiometer nach IEC 645-1 und zur Reintonaudiometrie von 125 Hz bis 12,5 kHz bei einer Frequenzgenauigkeit von 1 % geeignet.

Der Lautstärkepegel reicht von -20 dB bis +120 dB und ist in 2,5 dB und 5 dB Stufen einstellbar. Die Töne wurden dem Probanden über den zugehörigen Luftleitungskopfhörer DT48 als gepulstes Signal dargeboten. Dabei wurde das Hörvermögen seitengetrent voneinander getestet.

Die Kammer, in welcher das Audiogramm erstellt wurde, wird auch im klinischen Alltag täglich genutzt und entspricht den Normen, die für die Erstellung eines Audiogrammes erfüllt sein müssen. Sie ist schallisoliert und die Versuchsperson ist vom Untersucher akustisch getrennt, es besteht aber Blickkontakt. Der Untersucher kann außerdem mit der Versuchsperson über Mikrofon kommunizieren. Die Bedieneinheit des Audiometers ist für die Versuchsperson nicht einsehbar.

### 3.3 Voruntersuchungen zum Testablauf

Vor der Untersuchung der eigentlichen Hauptgruppe wurden zwei Personen getestet, um eventuelle Probleme beim Ablauf und bei der Durchführung der Untersuchungen aufzudecken. Diese beiden Vorprobanden hatten keine erkennbaren Probleme mit den Urteilsmethoden und dem Testablauf, so dass an der geplanten Umsetzung keine Mängel festzustellen waren. Die einzige Änderung ergab sich im Versagen des Laserpointers während eines Durchlaufes. Daraufhin wurde den Probanden zwei Laserpointer zur Verfügung gestellt, zwischen denen sie im Notfall wechseln konnten; siehe Kapitel 3.7 (Urteilsabgabe mit dem Laserpointer, S. 20). Die Probandeninstruktion (s. Abschnitt 3.11, Probandeninstruktion, S. 24 ff.) war so formuliert, dass häufige Fragen der Probanden, die bereits bei Nägle (2004) auftraten, beantwortet wurden.

Wie sich bei Nägle (2004) herausstellte, war ein Interstimulusintervall von 4 Sekunden ausreichend, um den Probanden genügend Zeit zu lassen, ihr Urteil abzugeben. Deshalb, und um die Vergleichbarkeit der beiden Arbeiten sicherzustellen, wurden die Abstände auch in der vorliegenden Arbeit konstant auf vier Sekunden festgelegt. Dass die Zeit, die zwischen den aufeinanderfolgenden Stimuli liegt, durchaus eine Rolle spielt, wurde sowohl für die Verfahren AME und CMM gezeigt (DeCarlo und Lawrence, 1992; Hellstrom, 1979) als auch für die kategoriale Lautheitsskalierung (Ruscetta und Palmer, 2000). Letztere untersuchten den Effekt des Interstimulusintervalls (ISI) mit dem Contour Test, einem kategorialen Lautheitsskalierungsverfahren nach Cox et al. (1997). Sie empfehlen ein ISI von mindestens 2 Sekunden um Zeitfehler in den Antworten des Patienten zu vermeiden. Welche Auswirkung die Dauer des ISI auf das abgegebene Urteil im WHF hat, ist Thema einer derzeit noch laufenden Dissertation an der HNO-Universitätsklinik Würzburg (Körner).

### 3.4 Reihenfolge der verschiedenen Urteilsmethoden

Insgesamt gab es fünf verschiedene Urteilsmethoden, die mit Großbuchstaben von A bis E bezeichnet wurden.

- Urteilsmethode A: Urteilsabgabe mit dem Skaliertablett und Ruhenlassen der Hand über der gedrückten Stelle

- Urteilmethode B: Urteilsabgabe mit dem Skaliertablett und Zurücklegen der Hand auf eine definierte Ausgangsposition
- Urteilmethode C: Ansage des Urteils bei sichtbarer, auf ein Poster abgedruckter Skala
- Urteilmethode D: Ansage des Urteils ohne sichtbare Skala
- Urteilmethode E: Urteilsabgabe durch Anzeige auf dem Poster mit einem Laserpointer

Um systematische Fehler zu vermeiden, wurde die Reihenfolge der Urteilsmethoden so festgesetzt, dass jede Methode gleich oft an jeder Stelle des Gesamtversuches stand. Das heißt, dass bei insgesamt 25 Probanden Urteilmethode A im Ablauf fünf mal an erster Stelle, fünf mal an zweiter Stelle, fünf mal an dritter Stelle usw. stand. Dasselbe gilt für die anderen Urteilsmethoden. Somit betrafen etwaige Lerneffekte oder Anfangsschwierigkeiten alle fünf Urteilsmethoden in gleicher Weise und konnten durch Mittelung statistisch minimiert werden.

Mit den fünf Urteilsmethoden könnte man theoretisch 120 ( $n!$  : n-Fakultät;  $5! = 120$ ) verschiedene Anordnungen aufstellen. Da diese Zahl jedoch das sinnvolle Maß an Probanden überschreitet, wurde die systematisch durcheinandergewürfelte Anordnung der Urteilsmethoden gewählt.

### 3.5 Urteilsabgabe mit dem Skaliertablett

In der klinischen Praxis gibt der Proband beim WHF das Urteil mittels Fingerdruck auf den Drucksensorstreifen des Skaliertablettes ab. Um eine Standardisierung der Urteilsabgabe zu gewährleisten, wurde diese Methode geringfügig modifiziert. Im Versuch wurden die Probanden angewiesen, mit dem Radiergummiaufsatz eines umgedrehten Bleistiftes auf den Registrierstreifen zu drücken. Damit wurde erreicht, dass die Fläche auf die gedrückt wurde konstant groß war. Außerdem wurden sie gebeten immer möglichst gleich, d.h. möglichst senkrecht und mit gleichem Druck zu drücken. Somit sollten zwar unbedeutende, aber dennoch mögliche Fehlmessungen, wie sie bei Vorversuchen bei [Nägler \(2004\)](#) auftraten, verhindert werden.

Jeder Proband musste zwei Durchläufe mit dem Skaliertablett absolvieren. Bei einem Durchlauf sollte der Proband, nachdem er das Urteil abgegeben hatte, den Bleistift über der eben gedrückten Stelle schweben lassen bis er den

dann folgenden Stimulus bewertete. Durch diese Vorgehensweise wurde erreicht, dass die Distanz der aufeinanderfolgenden Urteile dem Weg entsprach, die die Hand des Probanden zurücklegen musste. Bei weit auseinanderliegenden Urteilen war der Weg maximal, während bei nahe beieinander liegender Urteilen auch der zurückzulegende Weg des Armes gering war. Diese Art der Urteilsabgabe wird im Folgenden mit *Urteilsmethode A* bezeichnet.

Beim anderen Durchlauf sollte der Proband nach jedem abgegebenen Urteil seine Hand zurück auf eine festgelegte Ausgangsposition legen. Diese Ausgangsposition war mit Klebeband auf der Lehne des Sessels markiert. Das Starten von der Ausgangsposition bewirkte, dass der Arm - egal welches Urteil abgegeben wurde - immer die annähernd gleiche Strecke zu überbrücken hatte. Die Armträgheit sollte somit keine Rolle spielen, da der Weg für alle Urteile der gleiche war. Diese Art der Urteilsabgabe wird mit *Urteilsmethode B* bezeichnet.

Die komplette Probandeninstruktion ist in Kapitel [3.11 \(Probandeninstruktion, S. 24 ff.\)](#) nachzulesen.

## 3.6 Mündliche Abgabe des Urteils

Es gab zwei Durchläufe, bei denen der Proband seine Lautheitsurteile dem Versuchsleiter nur verbal mitteilen sollte. Somit konnte die Lautheitsskalierung durchgeführt werden, ohne dass eine Bewegung des Armes notwendig war. Der Proband wurde instruiert, sein Urteil laut und deutlich dem Versuchsleiter zu sagen. Der Versuchsleiter notierte die Antworten in einer Excel-Tabelle.

Um sich immer über die vorgegebene Skala orientieren zu können, wurde bei einem der beiden Durchläufe vor dem Probanden ein Poster mit derselben Beschriftung wie die des Skaliertablettes aufgestellt (s. [Abb. 1.2](#)). Dieser Durchlauf wird *Urteilsmethode C* genannt.

Das Poster im Format 1,24 m auf 0,95 m stand im Abstand von 2 Metern vor dem Probanden. Die Ausmaße des Posters entsprechen der Skala auf dem Skaliertablett, vergrößert um den Faktor 4. Der Abstand von 2 Metern entspricht ebenfalls einem Vierfachen des Abstandes vom Kopf des Probanden zum Skaliertablett. Diese Werte wurden so gewählt um bei den verschiedenen Durchläufen die gleichen räumlichen Relationen zu erhalten.

Aufgrund der räumlichen Gegebenheiten konnte das Poster nicht in gerader Blickrichtung des Probanden aufgestellt werden, da dort der Lautsprecher sowie

das Pult des Versuchsleiters standen (s. Abschnitt 4.2, [Aufstellung der Geräte in der Kammer](#), S. 29). Deshalb stand das Poster vom Probanden aus gesehen um 0,8 m nach links versetzt neben dem Pult.

Bei dem anderen Durchlauf hatte der Proband das Poster nicht vor Augen. Auch das Skaliertablett war für den Probanden nicht einsehbar. Somit konnte er sich nicht an einer sichtbaren Skala orientieren, sondern konnte als Referenz nur das im Gedächtnis abgespeicherte Muster der Skala zu Rate ziehen. Dieser Skalierungsdurchgang wurde mit *Urteilsmethode D* bezeichnet. Es sollte untersucht werden, welchen Effekt die sichtbare Skala auf die Lautheitsbeurteilung hat bzw. wie stark eine sichtbare Skala zur Konsolidierung der Urteile beiträgt. Vermutet wurde, dass das Nichtvorhandensein einer sichtbaren Skala eher zu einer Verschlechterung der Skalierung führt. Es war jedoch nicht vorhersehbar in welcher Weise sich diese Verschlechterung genau niederschlägt. So könnte eine größere Streuung der Urteile über einen Pegel erwartet werden oder auch ein Beschränken auf wenige Zwischenschritte, z.B. eine Skalierung, die nur noch in Fünferschritten abgestuft wäre. Auch eine Beschränkung auf nur einen Teilbereich der gesamten Skala wäre denkbar, so dass z.B. nur noch Urteile zwischen 10 hel und 30 hel abgegeben werden.

### 3.7 Urteilsabgabe mit dem Laserpointer

Bei dieser Methode sollte der Proband seine Urteile - mittels eines Laserpointers - auf der Zahlenskala des auch bei Urteilsmethode C (s. Abschnitt 3.6 ([Mündliche Abgabe des Urteils](#), S. 19)) verwendeten Posters anzeigen. Dadurch konnte ein fast vollständiges Ausschalten der Armbewegung erreicht werden. Nur eine minimale Bewegung aus dem Handgelenk war zur Urteilsabgabe erforderlich. Diese Methode ist sozusagen der Mittelweg zwischen den Methoden mit dem Skaliertablett (s. Abschnitt 3.5, [Urteilsabgabe mit dem Skaliertablett](#), S. 18) und den in Abschnitt 3.6 ([Mündliche Abgabe des Urteils](#), S. 19) aufgezeigten Methoden der mündlichen Ansage der Urteile. Sie wird als *Urteilsmethode E* bezeichnet. Dazu wurde ein handelsüblicher Laserpointer mit einem starken Lichtpunkt verwendet. Da das auf der Skala angezeigte Urteil vom Versuchsleiter von der Rückseite des Posters abgelesen wurde, musste der Lichtpunkt stark genug sein um durch das Poster durchzuscheinen. Um einer Unterbrechung des Versuches durch einen möglichen Ausfall des Laserpointers, z.B. durch entladene Batterien

vorzubeugen, wurden dem Probanden zwei gleichartige Laserpointer gegeben. So konnte er bei einem Ausfall des einen Laserpointers schnell auf den anderen wechseln, ohne dass dabei unterbrochen werden musste. Die abgegebenen Urteile wurden vom Versuchsleiter wiederum in einer Excel-Tabelle notiert.

### 3.8 Registrierung der abgegebenen Urteile

Beim Würzburger Hörfeld werden die auf dem Skaliertablett gedrückten Urteile direkt vom Skalierprogramm erfasst und in einer Datenbank gespeichert (s. Kapitel 5.8, [Patientendatenbank](#), S. 36).

Bei den Urteilsmethoden, die nicht mit dem Skaliertablett durchgeführt wurden, konnten die Urteile nicht automatisiert vom Computer aufgezeichnet werden. Vielmehr musste der Versuchsleiter selbst die Urteile notieren. Der Versuch, die Urteile der Probanden durch den Versuchsleiter mit dem Skaliertablett eingeben zu lassen, scheiterte daran, dass die vier Sekunden Interstimulusabstand zu kurz waren. Die Zeit reichte nicht aus, um den Probanden sein Urteil fällen zu lassen und dieses dem Versuchsleiter anzusagen bzw. anzuzeigen, der dann wiederum das Urteil selbst erst registrieren und auf dem Skaliertablett eingeben musste. Das Skalierprogramm hat einen Sicherheitsmechanismus, welcher verhindert, dass zu einem Stimulus keine Angabe vom Skaliertablett aufgezeichnet wird. Dieser Sicherheitsmechanismus musste bei den Durchgängen C, D und E also ausgeschaltet werden, da das Skalierprogramm ja keine Angaben vom Skaliertablett erhielt. Zusätzlich zeichnete eine Videokamera diese Durchgänge auf, um bei etwaigen Fehlern in den Notizen des Versuchsleiters die abgegebenen Urteile rekonstruieren zu können.

### 3.9 Aufbau der Stimulussequenzen

Die Stimulussequenzen wurden von [Nägler \(2004\)](#) übernommen. Es wurden in Fünferschritten abgestufte Schallpegel von 30 dB bis 90 dB verwendet. Das ergab 13 verschiedene Pegel. Dies wird zwar der Forderung, dass bei Lautheitsbestimmungen immer der gesamte Hörbereich abgedeckt werden sollte ([Hohmann und Kollmeier, 1995](#)) nicht ganz gerecht. Um potenzielle Hörschäden zu vermeiden wurden jedoch bewusst keine Schalldrücke höher als 90 dB verwendet. Jedem der 13 verschiedenen Pegel wurde jeder andere vorkommende Pegel



einmal vorangestellt. „So konnte von jedem der Pegel die Lautheitsbeurteilung in Abhängigkeit von jedem möglichen Vorgänger erfasst werden“ (Nägler, 2004). Eine Sequenz setzte sich also aus  $13 \times 13 = 169$  Stimuli zusammen.

„Da der erste Ton der gesamten Sequenz auch einen Vorgänger brauchte, wurde diesem der noch fehlende Vorgängerpegel vorangestellt. Dieser eine Pegel kam damit 14-mal, statt wie alle anderen 13-mal in der Sequenz vor. Daraus resultieren insgesamt 170 Stimuli in einer Sequenz. Um den Probanden eine Eingewöhnungsphase zu geben, die ihnen jedoch nicht bewusst sein sollte, wurde der eigentlichen Hauptsequenz noch einmal eine Initialsequenz aller vorkommenden Lautstärkepegel, also 13 Stimuli, vorausgeschickt. Zusammen mit dieser bestand die Gesamtsequenz also aus 183 Stimuli. Mit einem speziell entwickelten Computerprogramm wurden Sequenzen erzeugt, die die genannten Anforderungen erfüllten. Während die Initialsequenz unverändert blieb, wurde die eigentliche Kernsequenz in mehreren, um einige Stimuli durchrotierten Varianten verwendet.“ (Nägler, 2004).

Unabhängig von der Initial- und der Kernsequenz wurde noch eine Übungssequenz erstellt. Diese Übungssequenz beinhaltete alle 13 vorkommenden Pegel in zufälliger Reihenfolge. Mittels dieser Übungssequenz konnte der Proband ein eigenes Bezugssystem entwickeln. Wichtig war dabei, dass - wie bei Heller (1985); Hohmann und Kollmeier (1995) erwähnt - der Proband über das gesamte Spektrum der Pegel orientiert ist, ohne jedoch den lautesten und den leisesten Pegel explizit vorgestellt zu bekommen. Die Lautheitsskalierung der Übungssequenz wurde immer mit dem Skaliertablett durchgeführt.

Nachfolgend werden die Pegel der jeweils konstanten Übungs- und Initialsequenz, sowie die Pegel einer der möglichen Hauptsequenzen dargestellt.



60 35 85 55 80 40 65 30 90 50 75 45 70 (Übungssequenz)

60 35 85 55 75 45 65 30 90 50 80 40 70 (Initialsequenz)

### 60

60 35 60 40 60 65 60 75 60 45 60 70 60

30 60 85 60 50 60 55 60 80 60 90 35 35

40 35 65 35 75 35 45 35 70 35 30 35 85

35 50 35 55 35 80 35 90 40 40 65 40 75

40 45 40 70 40 30 40 85 40 50 40 55 40

80 40 90 65 65 75 65 45 65 70 65 30 65

85 65 50 65 55 65 80 65 90 75 75 45 75

70 75 30 75 85 75 50 75 55 75 80 75 90

45 45 70 45 30 45 85 45 50 45 55 45 80

45 90 70 70 30 70 85 70 50 70 55 70 80

70 90 30 30 85 30 50 30 55 30 80 30 90

85 85 50 85 55 85 80 85 90 50 50 55 50

80 50 90 55 55 80 55 90 80 80 90 90 **60** (Hauptsequenz)

## 3.10 Das CCITT-Rauschen

In bisherigen Forschungsarbeiten wurden bereits eine Vielzahl verschiedener Stimuli gebraucht. Nach [Lehnhardt und Laszig \(2001\)](#) werden in der Praxis meist schmalbandige Rauschimpulse verwendet, z.B. bei [Hellbrück und Thomamüller \(1997\)](#), [Heller \(1985\)](#) und [Ellermeier et al. \(1985\)](#). Aber auch viele andere Stimuli kamen zum Einsatz, z.B. Sinustöne, Rockmusik, „babble speech (Stimmengewirr)“, Sprachstimuli oder Wobbel-Töne ([Fucci et al., 1997](#); [Hellbrück und Thomamüller, 1997](#); [Hellstrom, 2003](#); [Hosoi et al., 1996](#); [Nägler, 2004](#)).

In dieser Arbeit wurde, wie bei [Nägler \(2004\)](#) und [Hellbrück et al. \(1995\)](#) auch, das CCITT-Rauschen benutzt. Dieses Breitbandrauschen wurde nach dem „Comité Consultatif International Téléphonique et Télégraphique“ benannt. Dieses Gremium ist ein Zusammenschluss aller Fernmeldeverwaltungen zum Zweck der

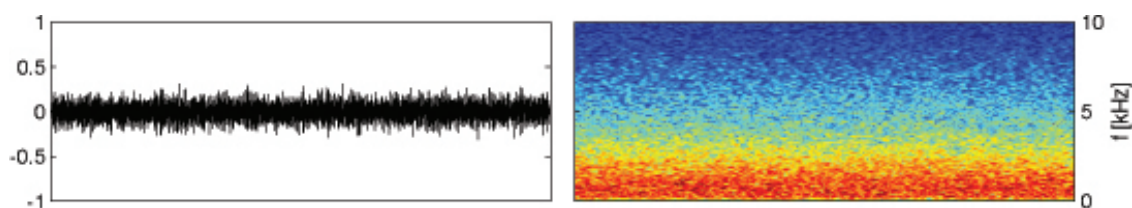


Abbildung 3.1: Zeitverlauf und Spektrogramm des CCITT-Rauschens.

Festlegung gemeinsamer technischer Normen. Ursprünglich für die Fernmelde-technik entwickelt, fand das CCITT-Rauschen später seinen Weg in die Audio-metrie und die Akustik.

Für diese Arbeit wurde das CCITT-Rauschen gewählt, weil es als breitbandi-ger Stimulus eine größere Vorgängerabhängigkeit aufweist als schmalbandige Stimuli (Nägle, 2004). Erwartet wurde, dass dadurch eine Veränderung der Ab-hängigkeiten durch die verschiedenen Urteilsmethoden besser erkennbar wür-de. Außerdem ist das CCITT-Rauschen in der akustischen Forschung gut eta-bliert.

Es ist so konfiguriert, dass es sich fließender Sprache gut annähert. Fastl (1987) benutzte das CCITT-Rauschen zur Vertäubung und erreichte eine ähnliche Re- duktion der Wortverständlichkeit wie mit Stimmengewirr. Außerdem wird das CCITT-Rauschen bei bilateral implantierten CI-Trägern als Reiz beim Richtungs- hörtest verwendet. Abbildung 3.1 zeigt den Zeitverlauf und das Spektrogramm des CCITT-Rauschens.

Das von Dipl.-Ing. S. Brill neu generierte CCITT-Rauschen wurde wiederum aus der Arbeit von Nägle (2004) übernommen. Es ist „genau eine Sekunde lang und wurden mit 16 Bit und einer Abtastfrequenz von 44,1 kHz aufgezeichnet. Es wurde eine verlustfreie digitale Kodierung in Microsoft WAV-Dateiformat verwen- det. [Das CCITT-Rauschen] wurde so erstellt, dass das Signal mit einer  $\sin^2$ - Rampe der Dauer von 5 ms begann und endete. So wurde ein Knacken am Anfang oder Ende des Stimulus vermieden.“ (Nägle, 2004).

#### 3.11 Probandeninstruktion

Vor Beginn des Versuches erhielt jeder Proband die gleiche schriftliche Aufklä- rung:

*"Der Zweck dieses Versuchs ist es, Ihre individuelle Lautheitsbeurteilung in Ab-*

*hängigkeit verschiedener Messmethoden zu erforschen. Dazu hören Sie verschieden laute Töne in Form eines Rauschens. Die Schallpegel bewegen sich abgestuft in Fünferschritten zwischen 30 dB und 90 dB. Sie müssen also keine Hörschädigung befürchten. Die Reihenfolge der verschiedenen Lautstärken ist willkürlich angeordnet.*

*Ihre Urteile werden immer mit der gleichen Skala erfasst. Diese Skala enthält fünf Hauptkategorien: »sehr leise«, »leise«, »mittellaut«, »laut« und »sehr laut«. Innerhalb dieser Kategorien sollen Sie nochmals eine Feinunterteilung und somit die Tendenz zur nächsten darunter oder darüber liegenden Kategorie angeben. Die Zahlen sind nur zur Orientierung angegeben. Benutzen Sie die Skala bitte mit allen Zwischenstufen.*

*Beachten Sie außerdem, dass Sie jeden Ton für sich beurteilen, wie Sie ihn in diesem Moment wahrnehmen. Sie sollen die Lautstärken nicht im Verhältnis zu den vorhergehenden Tönen beurteilen.*

*Innerhalb einer Testreihe werden 183 Stimuli unterschiedlicher Lautstärke präsentiert. Sie hören jeden Ton jeweils für eine Sekunde. Danach haben Sie vier Sekunden Zeit um Ihr Urteil abzugeben bevor der nächste Ton folgt.*

*Es werden fünf untereinander verschiedene Urteilsmethoden verwendet. Bei zwei Testreihen erfolgt Ihre Urteilsabgabe mit Hilfe eines Skaliertabletts, welches die Ergebnisse direkt an den angeschlossenen Computer übermittelt. Die verschiedenen Urteile sollen Sie mit dem Radiergummiaufsatz eines Bleistiftes abgeben. Drücken Sie dabei möglichst senkrecht und mit immer gleichem Druck auf das Skaliertablett. Stützen Sie sich jedoch nicht mit den Händen auf dem Tablett ab, da dadurch ein falscher Wert registriert werden könnte. Sollten Sie Ihr Urteil nicht innerhalb von 4 Sekunden abgeben, ist die Software so programmiert, dass sie automatisch um fünf Töne zurück springt um den fehlenden Wert zu ermitteln. Sie können also ganz normal im Test fortfahren.*

*Die beiden Testreihen mit dem Skaliertablett unterscheiden sich dadurch, dass Sie einmal Ihre Hand jeweils bis zur nächsten Urteilsabgabe über dem vorher gedrückten Urteil schweben lassen. Bei der anderen Testreihe sollen Sie Ihre Hand nach jeder Urteilsabgabe auf einen Ruhepunkt (auf der Armlehne des Sessels gekennzeichnet) legen.*

*Die dritte mögliche Urteilsabgabe erfolgt mittels eines Laserpointers. Zeigen Sie*

### 3 Versuchsplanung und Methodik

*mit dem Lichtpunkt auf der vor Ihnen aufgehängten Skala Ihr Urteil an. Lassen Sie dabei Ihren Arm auf der Armlehne liegen und bewegen den Laserpointer nur aus dem Handgelenk. Ich werde Ihr Urteil dann selbst in den Computer eingeben.*

*Bei den weiteren zwei Urteilsmethoden sollen Sie Ihr Urteil mündlich abgeben. Bei einer Testreihe werden Sie dabei die Skala vor sich haben. Bei der anderen Testreihe werden Sie die Skala nicht sehen können. Sagen Sie mir die Zahl laut und deutlich. Ich werde Ihr Urteil dann wiederum in den Computer eingeben.*

*Zu Beginn des Versuches wird Ihnen eine Übungssequenz angeboten, die alle im Versuch vorkommenden Lautstärken beinhaltet. Dies soll Sie mit dem Umgang der jeweiligen Urteilsmethode vertraut machen. Sie können die Übungssequenz so oft hören wie Sie wollen bis Sie sich mit der Methode vertraut fühlen.*

*Jede Testreihe dauert circa sechzehn Minuten. Diese Zeit wird Ihnen recht lange vorkommen. Damit Sie sich über das Fortschreiten des Tests orientieren können, werde ich Ihnen deshalb nach ungefähr acht Minuten ein Zeichen geben.*

*Während der Hauptsequenz der jeweiligen Testreihe dürfen keine Pausen gemacht werden. Außerdem dürfen weder Sie noch ich während des Versuchs sprechen. Ich werde Ihnen auch keine Rückmeldung über Ihre Urteile geben.*

*Versuchen Sie während des ganzen Versuchs möglichst konzentriert zu sein und die Lautstärke exakt einzuschätzen.*

*Sie können den Test jederzeit abbrechen, ohne dass Ihnen daraus irgendwelche Konsequenzen entstehen. Die bereits erhobenen Daten werden dann jedoch unbrauchbar und können nicht in die Arbeit miteinbezogen werden.*

*Während des Versuches zeichnet eine Videokamera die von Ihnen abgegebenen Urteile auf. Dies dient ausschließlich dazu, eventuelle Lücken in der Urteilsabgabe nachträglich in den Computer eingeben zu können. Sie werden auf der Aufnahme nicht zu sehen sein.*

*Die Daten werden zu wissenschaftlichen Zwecken erhoben und nur in anonymisierter Form ausgewertet. Die Daten werden verschlüsselt und wir verpflichten uns zur Einhaltung der datenschutzrechtlichen Bestimmungen.“*

## 4 Versuchsdurchführung

### 4.1 Camera silens - schalltoter Raum

Die Tests wurden in einem speziellen Raum - der Camera silens - im Keller der HNO-Klinik Würzburg durchgeführt. Ein Photo des experimentellen Aufbaus zeigt Abb. 4.1. Die Wände, sowie der Boden und die Decke dieses 5×5×4 Meter großen Raumes sind mit speziellen Schaumstoffkeilen und Glaswollewürfeln verkleidet, die so zueinander ausgerichtet sind, dass sie Schallwellen nicht reflektieren sondern schlucken. Dadurch werden mögliche Irritationen durch reflektierte oder stehende Schallwellen verhindert. Dies spielt speziell bei der Verwendung von Sinustönen eine entscheidende Rolle. Durch die dicke Isolierung wird auch Störschall von außen weitgehend vermieden. Damit werden die akustischen Randbedingungen möglichst gleichförmig gehalten, so dass von außen kommende Störungen vernachlässigbar waren. „Beleuchtung und eine Belüftungsanlage sind installiert. Damit sind für die Durchführung der Untersuchung eine akustisch und optisch reizarme Umgebung geschaffen, die eine bessere Konzentration der Probanden auf die Testsignale und eine Minderung der Umweltgeräusche gewährleistet.“ (Kugler und Moser, 1995). Während der Versuche wurde die Belüftungsanlage jedoch abgeschaltet, um sie als akustische Störquelle ausschließen zu können. Dass die akustischen Randbedingungen deutliche Auswirkungen auf die Lautheitsskalierung haben zeigten Hellbrück et al. (1995).

Die Camera silens wird auch für Versuche zum Richtungshören bei bilateral implantierten Cochlea-Implantat-Trägern benutzt. Die auf dem Bild (Abb. 4.1) sichtbaren, im Halbkreis aufgehängten Lautsprecher gehören zu diesen Tests und konnten für die Versuche der vorliegenden Arbeit nicht entfernt werden. Weiterhin sieht man den Sessel für die Probanden (grün), die verwendete Aktivbox (gegenüber dem Sessel im Regal), den Platz des Versuchsleiters sowie die Heller-Skala als Poster und auf dem Skaliertablett.

#### 4 Versuchsdurchführung



Abbildung 4.1: Der Versuchsaufbau in der Camera silens



## 4.2 Aufstellung der Geräte in der Kammer

Die Positionen der einzelnen Gegenstände sind in Abb. 4.1 sichtbar. Während des Versuches befanden sich sowohl der Proband als auch der Versuchsleiter in der Kammer. Auch das gesamte technische Equipment war in der Kammer installiert. Weil neben den hier dargestellten Versuchen auch noch andere Tests in der Camera silens durchgeführt wurden, wurde die Position aller beweglichen Teile auf dem Boden mit Klebeband markiert. Somit konnte der Versuchsaufbau auch nach Verschieben schnell und korrekt wiederhergestellt werden. Der Proband saß in einem Sessel gegenüber dem Versuchsleiter. Direkt vor dem Probanden befand sich die Aktivbox (s. Abschnitt 5.4, [Lautsprecherwahl](#), S. 34) auf ungefährer Kopfhöhe des Probanden. Der Sessel stand so, dass der Abstand vom Lautsprecher zur Kopfmitte des Probanden 1,5 Meter betrug. Als Kopfmitte wurde eine gedachte Frontalebene durch die Gehörgänge des Probanden definiert. Um schnell erkennen zu können, ob der Proband richtig sitzt, wurde im gewünschten Abstand eine Schraubenmutter mit einer Schnur an der Decke befestigt, die dann knapp über dem Kopf des Probanden hing. Das Poster mit der aufgedruckten Skala stand - bei Sicht des Probanden auf die Aktivbox - um 0,8 m nach links versetzt. Das hatte den Nachteil, dass der Proband sein rechtes Ohr etwas dem Lautsprecher zuwenden musste und dieses eventuell einem minimal größeren Schalldruckpegel ausgesetzt war als das linke Ohr. Selbstverständlich konnte das Poster aber nicht zwischen Proband und Lautsprecher stehen. Ebenso war die Aufstellung hinter dem Lautsprecher räumlich nicht möglich.

## 4.3 Kontrolle der Schallpegel

Nachdem der gesamte Versuchsaufbau festgelegt war, wurden die vom Computer vorgegebenen Pegel an der späteren Position des Kopfes des Probanden gemessen. Jeder im Versuch vorkommende Pegel wurde einzeln kontrolliert. Verwendet wurde der Schallpegelmesser „Investigator 2260“ der Firma Brüel & Kjær. Die Pegel wurden A-bewertet. Die maximal gemessene Abweichung betrug - 1,1 dB beim vorgegebenen Pegel von 30 dB. Bei allen anderen Pegeln lag die Abweichung jedoch unter  $\pm 1$  dB. In Tabelle 4.1 sind die Messwerte aufgeführt.

Da das für den Laserpointerversuch benötigte Poster als schallreflektierende Wand angesehen werden kann, wurde die Schallpegelmessung zweimal durchgeführt. Einmal mit dem Poster an seinem Platz in der Camera silens und einmal

Tabelle 4.1: Gemessene Schallpegel bei Pegelvorgabe vom Computer, gemessen an Position des Kopfes des Probanden, alle Angaben in dB

Pegel	CCITT
30	28,9
35	34,1
40	39,1
45	44,2
50	50,0
55	55,1
60	60,1
65	65,2
70	70,1
75	75,1
80	80,1
85	85,1
90	89,9

ohne das Poster. Beide Messungen ergaben jedoch die selben in Tabelle 4.1 angegebenen Pegel. Somit war gesichert, dass das Poster keinen Einfluss auf die Schallpegel hatte.

Auch bei absoluter Ruhe in der Camera silens konnte noch ein Schallpegel von etwas unter 20 dB gemessen werden. Dieser Schallpegel war aber subjektiv nicht zu hören. Die Herkunft dieses Hintergrundschallpegels ist nicht klar, könnte aber wie in [Nägle \(2004\)](#) vermutet, von tieffrequenten Maschinengeräuschen hervorgerufen werden.

Da die Videokamera nicht völlig geräuschlos lief, wurde der Schallpegel auch bei laufender Kamera gemessen. Der Hintergrundschallpegel am Platz des Probanden war auch dann noch unter 20 dB und stellte somit keine Beeinträchtigung des Probanden dar. Das im Versuchsaufbau verwendete Notebook lief subjektiv nahezu geräuschlos. Die Schallpegel wurden alle bei eingeschaltetem Notebook gemessen. Somit konnte ausgeschlossen werden, dass die Schallpegel durch den Computer beeinflusst wurden. Nähere Angaben zum verwendeten Notebook und den anderen technischen Geräten finden sich in Kapitel 5 ([Technischer Aufbau](#), S. 33).



## **4.4 Versuchsablauf**

Die freiwilligen Probanden wurden im Labor der Cochlea-Implantat-Abteilung begrüßt und erhielten dort die Probandeninstruktion und die Einverständniserklärung, die sie sich jeweils durchlesen und unterzeichnen mussten. Anschließend erfolgte die Kontrolle des Hörvermögens in der Audiometrie-Kammer. Bei erfolgreichem Bestehen wurden die Probanden dann in die Camera silens geführt. Vor Ort wurde ihnen nochmals der Versuchsablauf erklärt und eventuelle Fragen geklärt. Nachdem sich die Probanden anhand der Übungssequenz orientiert hatten und sich sicher fühlten, wurde mit den eigentlichen Versuchssequenzen gestartet. Nach Ablauf jeder Sequenz wurde den Probanden eine kurze Pause sowie die Möglichkeit zum Stellen von Fragen angeboten. Viele Probanden wollten zwischen den Durchläufen einmal für kurze Zeit die Camera silens verlassen.

#### 4 Versuchsdurchführung

## 5 Technischer Aufbau

Soweit möglich wurden dieselben technischen Geräte wie bei [Nägler \(2004\)](#) verwendet.

### 5.1 Computer

Das Skalierungsprogramm wurde auf einem Panasonic CF-37 Toughbook betrieben. Dieses Notebook läuft auf dem Betriebssystem Microsoft Windows 2000 Professional. Es hatte die folgende Ausstattung: Intel Pentium III Prozessor mit 700 MHz, 128 MB RAM, CD-Rom/DVD-Laufwerk. Der Lüfter des Notebooks war nicht zu hören. Trotzdem arbeitete das Notebook nicht völlig geräuschlos. Weitere Erläuterungen hierzu finden sich in Kapitel [4.3 \(Kontrolle der Schallpegel, S. 29\)](#).

### 5.2 Videokamera

Wie in [3.8](#) erwähnt, zeichnete bei den Durchgängen C, D und E eine Videokamera die abgegebenen Urteile auf. Es handelte sich dabei um eine Sony DCR-VX 1000 E. Dieser Camcorder zeichnet auf miniDV-Kassetten auf und hat ein fest eingebautes Elektret-Kondensator Stereo-Mikrofon. Die Kamera war auf einem Stativ befestigt und stand schräg links hinter dem Probanden. Das Objektiv war während des ganzen Versuches nur auf das Poster gerichtet, so dass der Proband nicht auf der Aufnahme zu sehen war. Somit war der Proband auf den Aufnahmen gut zu hören und der Laserpointer auf dem Poster gut zu erkennen, ohne dass die Kamera im Versuchsaufbau störte.

### 5.3 USB-Soundkarte

Wie auch bei [Nägler \(2004\)](#), wurde nicht die interne Soundkarte des Notebooks verwendet, sondern eine über eine USB-Schnittstelle angeschlossene externe

Soundkarte. „Die Messungen wurden mit einer an der Klinik entwickelten Audiokarte durchgeführt. Darin wurde ein D/A-Wandler (digital-analog-Wandler, Typ PCM 2707) mit USB-Schnittstelle der Firma Burr-Brown eingesetzt.“ (Nägler, 2004).

### 5.4 Lautsprecherwahl

Um eine saubere und verzerrungsfreie Wiedergabe der Stimuli im gesamten verwendeten Schallpegelbereich sicherzustellen, konnte nicht auf den schon in der Camera silens befindlichen Denon-HiFi-Verstärker mit angeschlossenem Lautsprecher zurückgegriffen werden. Stattdessen wurde die Audiometrieaktivbox „LAB 501“ verwendet, welche auch in den Audiometriekammern der Universitäts-HNO-Klinik zum Einsatz kommt. Außerdem gehört sie zur Standardausrüstung des von der Firma Westra-Electronic vertriebenen WHF. Die Daten sind aus der Gebrauchsanleitung entnommen (Westra CAD03/1). Die Freifeldbox hat die Abmessungen  $235 \times 335 \times 24$  mm. Sie beinhaltet einen vollkomplementären Gegenakt - Endverstärker mit hoher Linearität und geringen Verzerrungen. „Die Bauform bündelt den Schall stark auf den Bereich frontal zur Box.“ Der Klirrfaktor ist nach DIN 45 403  $< 0,1\%$ , der Gesamtklirrfaktor bei 90 dB  $< 0,3\%$ . Der maximale Schalldruckpegel beträgt im normalen Hörraum ca. 112 dB.

### 5.5 Laserpointer

Es wurden zwei handelsübliche Laserpointer verwendet. Es wurde darauf geachtet, dass sie möglichst handlich und leicht sind. Die Laserpointer sind zylindrisch mit einem Durchmesser von ca. 1,2 cm und einer Länge von ca. 6 cm. Um zu verhindern, dass sie aus Versehen durch den Gitterboden der Camera silens fallen, waren sie mit einer Schnur am Sessel des Probanden festgebunden.

### 5.6 Skaliertablett

Das verwendete Skaliertablett gehört zur Ausstattung des von der Firma Westra-Electronic vertriebenen Würzburger Hörfeldes. Es „nimmt das Lautheitsurteil des Probanden über einen Drucksensorstreifen auf und leitet es zur Auswertung an

den PC. Die Beschriftungsstreifen sind auswechselbar und erlauben die Gestaltung in unterschiedlichen Sprachen und mit beliebigen Symbolen.“ (s. Hersteller-Anleitung zum WHF, [Westra WHF](#)). Das Skaliertablett wurde über einen seriellen Anschluss an die COM-Schnittstelle des PC mit 9-poligem Stecker angeschlossen. Das Tablett kann insgesamt 256 ( $= 2^8$ , entsprechend einer 8-Bit-Verschlüsselung) verschiedene Werte registrieren. Diese werden dann vom Skalierungsprogramm auf die 50 Skalenteile der Skala nach Heller umgerechnet.

Der elektrisch leitende Drucksensorstreifen ist in einen Stromkreislauf eingebunden und dient als Ohmscher Widerstand. Je nachdem wo der Proband auf den Streifen drückt, ändert sich der Betrag dieses Widerstandes. Diese Änderung wird registriert und kann auf den zugehörigen Zahlenwert umgerechnet werden. Die registrierten Werte - sowohl der interne 8-Bit-Zahlenwert als auch der daraus abgeleitete Wert der Heller-Skala - werden automatisch in einer Datenbank gespeichert und außerdem dem Versuchsleiter graphisch präsentiert.

### 5.7 Skalierungsprogramm

Zum Abspielen der Stimuli wurde - entsprechend [Nägle \(2004\)](#) - das von Dipl.-Ing.(FH) Alexander Möltner modifizierte Skalierungsprogramm verwendet. In der Bedienoberfläche dieses Programms lassen sich die Optionen bezüglich des Ablaufs der Skalierung einstellen. Die Länge des Interstimulusintervalls lässt sich von einer bis zu acht Sekunden variieren. Weiterhin gibt es die Möglichkeit, zwischen verschiedenen Stimuli (unter anderem dem CCITT-Rauschen), die im .wav-Format auf dem Computer gespeichert sind, auszuwählen. Auch die gewünschte Dauer des Stimulus wird hier festgelegt. Die unterschiedlichen Stimulussequenzen (s. Kapitel [3.9, Aufbau der Stimulussequenzen](#), S. 21) sind ebenfalls auf dem Computer abgespeichert und lassen sich im Skalierungsprogramm auswählen. Das Programm registriert die Daten vom Skaliertablett und speichert sie in einer Datenbank ab (s. Kapitel [5.8, Patientendatenbank](#), S. 36). Weiterhin zeichnet die Software die Antworten direkt in ein Koordinatensystem ein. Auf der Abszisse sind die Pegel von 30 dB bis 90 dB aufgetragen und auf der Ordinate die Antworten von 0 hel bis 50 hel. Für jede registrierte Antwort wird an entsprechender Stelle ein grüner Punkt im Koordinatensystem eingezeichnet. Das Schaubild wird in Echtzeit erstellt und kann vom Versuchsleiter direkt eingesehen werden. Anhand der graphischen Darstellung kann der Versuchsleiter schnell

einen Eindruck über das Skalierungsverhalten des Probanden gewinnen und bei Auffälligkeiten entsprechend reagieren. Weiterhin lässt sich im Programm ablesen welcher Stimulus in der Sequenz mit welchem Pegel gerade abgespielt wird.

Erhält das Programm innerhalb des Interstimulusintervalls kein Urteil vom Probanden, wird beim zugehörigen Pegel ein roter Punkt auf der Abszisse aufgetragen. Um aber für jeden vorgespielten Pegel ein Urteil zu erhalten, beinhaltet das Programm eine Sicherheitsautomatik. Wenn keine Antwort registriert wird, springt die Software in der Stimulussequenz fünf Schritte zurück und spielt die Sequenz von dort an wieder normal ab. Dies geschieht, ohne dass der Ablauf gestört wird bzw. der Proband es merkt. In der Datenbank wird dann nur der fehlende Wert ersetzt, während die Antworten zu den anderen doppelt abgespielten Pegeln ignoriert werden. Diese Sicherung ist vom Versuchsleiter an- und ausschaltbar.

### 5.8 Patientendatenbank

Die ermittelten Daten aller untersuchten Probanden wurden in einer speziell konstruierten Datenbank festgehalten. Diese Datenbank ist Teil des von der Firma Westra-Electronic vertriebenen Skalierprogrammes zur Durchführung des WHF. Bei Anmeldung des Probanden im Skalierprogramm werden zunächst Nach- und Vorname, sowie Geburtsdatum und Geschlecht abgefragt. Ohne Eingabe dieser Werte lässt sich keine Messung vornehmen. Die Datenbank speichert für jeden dargebotenen Stimulus unter anderem folgende Informationen in einer automatisch generierten Excel-Tabelle:

- Nummer des Probanden
- Nummer der Messung mit dem jeweiligen Probanden
- Abgespielter Schallpegel
- Abgegebenes Urteil (8-Bit-Wert und Wert auf der Heller-Skala)
- Sequenz, die durchlaufen wird
- Dauer des Stimulus
- Länge des Interstimulusintervalls

## 6 Ergebnisse

### 6.1 Statistische Auswertung

Die vorhandenen Daten wurden auf mehrere Faktoren untersucht, welche das Antwortverhalten der Probanden beeinflussen könnten:

1. Probandenspezifische Unterschiede in den Urteilen
2. Einfluss der Urteilmethode
3. Einfluss des Schallpegels
4. Einfluss des Schallpegels des Vorgängerstimulus
5. Einfluss des Vorgängerurteils
6. Einfluss der Position der Urteilmethode im Gesamtablauf

Bei der Durchführung der Versuche entstand eine große Menge an Daten, die statistisch ausgewertet werden mussten. Zuerst wurde - mit einem in *Perl* geschriebenen Skript - eine Roh-tabelle erstellt. Diese vereinigte alle ermittelten Daten von jedem Probanden in einer einzigen Tabelle. Das war vor allem deshalb nötig, weil die Antworten der Probanden je nach Urteilmethode in verschiedenen Tabellen notiert waren (s. Abschnitt 3.8, [Registrierung der abgegebenen Urteile](#), S. 21). Um die statistische Auswertung jedoch nachvollziehbar und automatisiert durchführen zu können, mussten die gesamten Daten erst auf das gleiche Format gebracht werden. Diese Roh-tabelle war der Ausgangspunkt für die gesamte Auswertung. Zur Durchführung der statistischen Tests und zur Erstellung der gesamten Graphiken wurde das Statistikprogramm *R* ([The R Development Core Team, 2006](#)) benutzt. Im Anhang sind die gesamten Rohdaten aufgeführt [B](#) (S. 105 ff.).

### 6.1.1 Verwendete statistische Tests und Verfahren

Die Regressionen der Pegel-Lautheitsfunktionen wurden mit dem nichtlinearen, lokalen „loess“-Regressionsmodell erstellt ([The R Development Core Team, 2006](#), S. 1036 ff.). „loess“ ist die Abkürzung für „locally-weighted regression“.

„Eine loess-Funktion benutzt für jeden Abschnitt der Regressionsfunktion einen gewichteten Abschnitt der Gesamtdatenmenge. Datenpunkte, die einen geringeren Abstand zur Stützstelle aufweisen, werden stärker gewichtet als entferntere Datenpunkte.“ ([Dondrup, 2002](#))

Dadurch wird verhindert, dass Ausreißerwerte übergroßen Einfluss auf den Verlauf der Kurve nehmen. Es können sowohl Regressionsgeraden als auch Polynome höheren Grades berechnet werden.

Die linearen Regressionen der Vorgängerabhängigkeiten wurden mit dem in *R* implementierten „lm“-Verfahren berechnet. „lm“ steht für linear model.

Als globaler, multipler Rangsummentest wurde der H-Test nach Kruskal und Wallis und als Einzeltest der Wilcoxon-Rangsummentest verwendet. Beide Tests sind verteilungsfrei.

„Der Kruskal-Wallis-Test prüft ohne eine Normalverteilung vorauszusetzen (lediglich eine stetige Verteilung wird gefordert), ob die Mittelwerte von  $p$  Meßreihen der Längen  $n_1, \dots, n_p$  als signifikant verschieden angesehen werden können.“ ([Hartung et al., 1998](#))

Anders gesagt vergleicht der H-Test mehrere Stichproben auf Gleichheit bzw. Ungleichheit der zugehörigen Verteilung. Er stellt somit fest, ob ein Unterschied zwischen den Verteilungen besteht oder ob sie statistisch gleich sind. Wenn z.B. nur ein Proband anders skaliert hätte als alle anderen, würde dies vom Kruskal-Wallis-Test erkannt.

Der Wilcoxon-Rangsummentest dagegen vergleicht nicht mehrere, sondern nur zwei vorgegebene Verteilungen auf Un- bzw. Gleichheit. Beide Rangsummentests geben die statistische Signifikanz mit dem p-Wert an. Ist der p-Wert gleich oder unter 0,05 ( $\leq 5\%$ ) so gilt er als signifikant. Man sagt, der p-Wert liegt auf dem 5%-Signifikanzniveau. Ist der p-Wert  $\leq 0,01$  ( $\leq 1\%$ ) spricht man vom 1%-Signifikanzniveau. Das Statistikprogramm gibt aus numerischen Gründen als kleinstmöglichen Wert einen p-Wert von  $p < 2,2 \times 10^{-16}$  an.



### 6.1.2 Umgang mit fehlerhaften Werten

Nach Durchführung der Tests und Durchsicht der erhaltenen Daten fielen einige Werte auf, die fehlerhaft zu sein schienen. Es handelte sich dabei ausschließlich um Werte, die mit den Urteilmethoden A und B erhoben wurden. Deshalb lag die Schlussfolgerung nahe, dass das Skaliertablett einen technischen Defekt haben musste.

Auch bei Nägele (2004) war ein technischer Fehler des Skaliertabletts aufgetreten. Der Computer registrierte bei Druck auf Höhe von 31 hel einen Wert von 5 hel oder 7 hel. Dadurch dass der Fehler so genau spezifiziert werden konnte, war es Nägele möglich die fehlerhaften Werte durch die korrekten Werte zu ersetzen.

Die Fehler, die in den hier vorliegenden Tests auftraten, waren jedoch anderer Art. Es wurden bei Pegeln zwischen 65 dB und 80 dB mehrmals Werte von 0 hel registriert, was der Angabe „nicht gehört“ entspricht. Dies kann nur auf einen technischen Defekt zurückgeführt werden und nicht auf eine Angabe der Probanden. Da alle Probanden auf Normalhörigkeit getestet waren, konnten Stimuli - zumal in diesem Schallpegelbereich - schlichtweg nicht „nicht gehört“ worden sein. Wie bereits in Kapitel 5.6 (Skaliertablett, S. 34) erwähnt, arbeitet das Skaliertablett intern mit einer 8-Bit-Verschlüsselung bzw. 256 möglichen Werten, die registriert werden können. Das Skalierprogramm rechnet die Werte auf die Heller-Skala um. Die Umrechnung ist folgendermaßen definiert: Die internen Werte 0 und 1 werden nicht berücksichtigt, da sie durch technische Gegebenheiten des D/A-Wandlers auch zufällig registriert werden können. Die restlichen 255 möglichen Werte werden durch 50 geteilt ( $255 \div 50 = 5,1$ ). Somit hat 1 hel den Umfang von 5,1 auf der internen Skala. Ein interner Wert von 2 bis 7,1 entspricht damit 0-hel.

Alle gelöschten 0 hel-Werte hatten einen internen Wert von 3. Damit ist erwiesen, dass auch tatsächlich ein Wert von 0 hel registriert wurde und nicht, wie eventuell anzunehmen ein Speicherfehler im Programm vorlag.

Leider konnte der Defekt des Skaliertabletts nicht wie bei Nägele näher verifiziert werden. Weder auf Höhe von 31 hel, noch sonst auf der Skala lies sich eine fehlerhafte Messung registrieren. Ebenso die Annahme, dass der Proband auf den leicht erhöhten Rand des Drucksensorstreifens oder ganz daneben gedrückt haben könnte, ergab keine dergestaltigen Fehlmessungen. Deshalb konnten die fehlerhaften Werte nicht durch einen plausiblen Wert ersetzt werden. Es blieb al-

so nur die Möglichkeit, die sicher als falsch erkannten Werte aus der Auswertung herauszunehmen.

Insgesamt mussten 21 Werte aus dem Rohdatensatz ausgeschlossen werden. Sie wurden als nicht vorhanden gewertet. Das Statistikprogramm *R* kann mit solchen fehlenden Werten umgehen, ohne dass eine Verfälschung der Ergebnisse auftritt.

Eine Konsequenz dieser fehlenden Werte ist, dass sie sowohl als eigentliches Urteil fehlen als auch bei den Untersuchungen zur Vorgängerabhängigkeit als Vorgängerurteil nicht vorhanden sind. Der zugehörige Vorgängerpegel dagegen ist vorhanden, da er ja auch abgespielt wurde.

Bezogen auf die Gesamtzahl von 21125 ermittelten Urteilen (25 Probanden  $\times$  5 Durchläufe  $\times$  169 Stimuli = 21125) entsprechen die fehlenden Werte einem Prozentsatz von nur 0,99 Promille. Alleine deshalb kann eine Verfälschung der Ergebnisse durch das Weglassen der Werte mit sehr hoher Wahrscheinlichkeit ausgeschlossen werden.

Brand et al. (1997) demonstrierten bei einer Simulation am Computer, dass die kategoriale Lautheitsskalierung gegen eine Anzahl von immerhin 5 % Ausreißer robust ist. Die Ursache dieser Ausreißer sahen sie in Aufmerksamkeitslücken der Probanden („gaps in the attention of the subject“). Trotzdem wurde in dieser Arbeit die gesamte Auswertung doppelt durchgerechnet, einmal ohne die fehlerhaften Werte und einmal mit den fehlerhaften Werten. Beide Auswertungen wurden miteinander verglichen. Es konnte jedoch kein relevanter Unterschied gefunden werden. Das Weglassen dieser 21 Werte hat folglich keinen Einfluss auf die Ergebnisse. Die im Folgenden vorgestellten Ergebnisse stammen alleamt aus der Auswertung, die ohne die Fehlerwerte durchgerechnet wurde.

Die Wahl dieses fehlerbereinigten Verfahrens hat mehrere Gründe:

1. Die fehlerbereinigte Variante ist statistisch sauberer, da vor allem Regressionen durch Ausreißerwerte übermäßig beeinflusst werden
2. Da nur bei den Methoden A und B der Fehler auftrat, wären bei Belassen der Fehlerwerte die fehlerfreien Methoden C, D und E mit den fehlerbehafteten Methoden A und B verglichen worden.
3. Bedingt durch den technischen Fehler, wurden besonders die höheren Schallpegel teilweise mit 0 bel beurteilt. Dadurch verschöben sich die Urteils*minima* aufgrund eines einzigen (fehlerhaften) Wertes stark nach unten

## 6.2 Visuelle Betrachtung der Rohdaten

Zur besseren Übersicht werden hier nur drei exemplarische Probanden vorgestellt, an denen man schon durch rein visuelle Betrachtung große Unterschiede im Skalierungsverhalten erkennen kann. Die Daten aller Probanden sind im Anhang aufgeführt [B](#) (S. [105](#) ff.). Die Probanden sind von 50 bis 74 durchnummeriert. Da die Nummerierung zur besseren Nachvollziehbarkeit aus der Patientendatenbank des WHF übernommen wurde, konnten die Probanden nicht von 1 bis 25 beziffert werden.

Die Abbildung [6.1](#), sowie die zugehörigen Abbildungen im Anhang sind wie folgt aufgebaut. Auf der Ordinate sind die möglichen Urteile von 1 hel bis 50 hel aufgetragen, während auf der Abszisse die verwendeten Schallpegel von 30 dB bis 90 dB aufgetragen sind. Jede der senkrecht übereinander stehenden Punktwolken besteht aus 13 Urteilsabgaben, da entsprechend dem Aufbau der Sequenzen (s. Abschnitt [3.9, Aufbau der Stimulussequenzen](#), S. [21](#) ff.) jeder Pegel dreizehnfach vorgegeben und beurteilt wurde. Je geringer die Ausdehnung der Punktwolke in Ordinatendichtung, desto konsistenter hat der Proband den Pegel skaliert. In der ersten Zeile jeder Einzelgraphik steht grau hinterlegt die Urteils-methode. In der zweiten Zeile steht die Nummer des Probanden. Abbildung [6.1](#) zeigt die Ergebnisse der Probanden 57, 59 und 66. Jeder der drei Probanden zeigt besonders deutlich verschiedene Auffälligkeiten, wie sie auch bei anderen Probanden vorkommen können. Vergleicht man die Probanden miteinander, so erkennt man den grundsätzlich etwas unterschiedlichen Gebrauch der Skala. Proband 57 meidet den untersten und den obersten Bereich der Skala. Auffällig ist die Querstruktur zwischen den verschiedenen Schallpegeln. Besonders bei den Urteilmöglichkeiten C und D ist die Querstruktur stark ausgeprägt. Die Extremwerte für die einzelnen Schallpegel liegen weit auseinander. Proband 59 beschränkt sich auf einen engen Bereich der Skala. Mehr als die Hälfte aller Urteile liegen zwischen 15 hel und 35 hel. Urteile kleiner als 10 hel und größer als 42 hel werden praktisch keine abgegeben. Durch diese Beschränkung streuen die Punktwolken automatisch deutlich geringer als die der anderen Probanden. Die Punktwolken stehen sehr flach zueinander, was sich ebenfalls durch die Beschränkung auf einen kleinen Skalenbereich erklären lässt. Proband 66 nutzt fast den gesamten Skalenbereich von ca. 5 hel bis 50 hel aus. Die Streuung der Punktwolken beträgt ca.  $\pm 5$  Skalenteile, was im Normalbereich liegt ([Hellbrück, 1983](#); [Hellbrück und Moser, 1985](#)). Die Steigung der Punktwolken ist

## 6 Ergebnisse

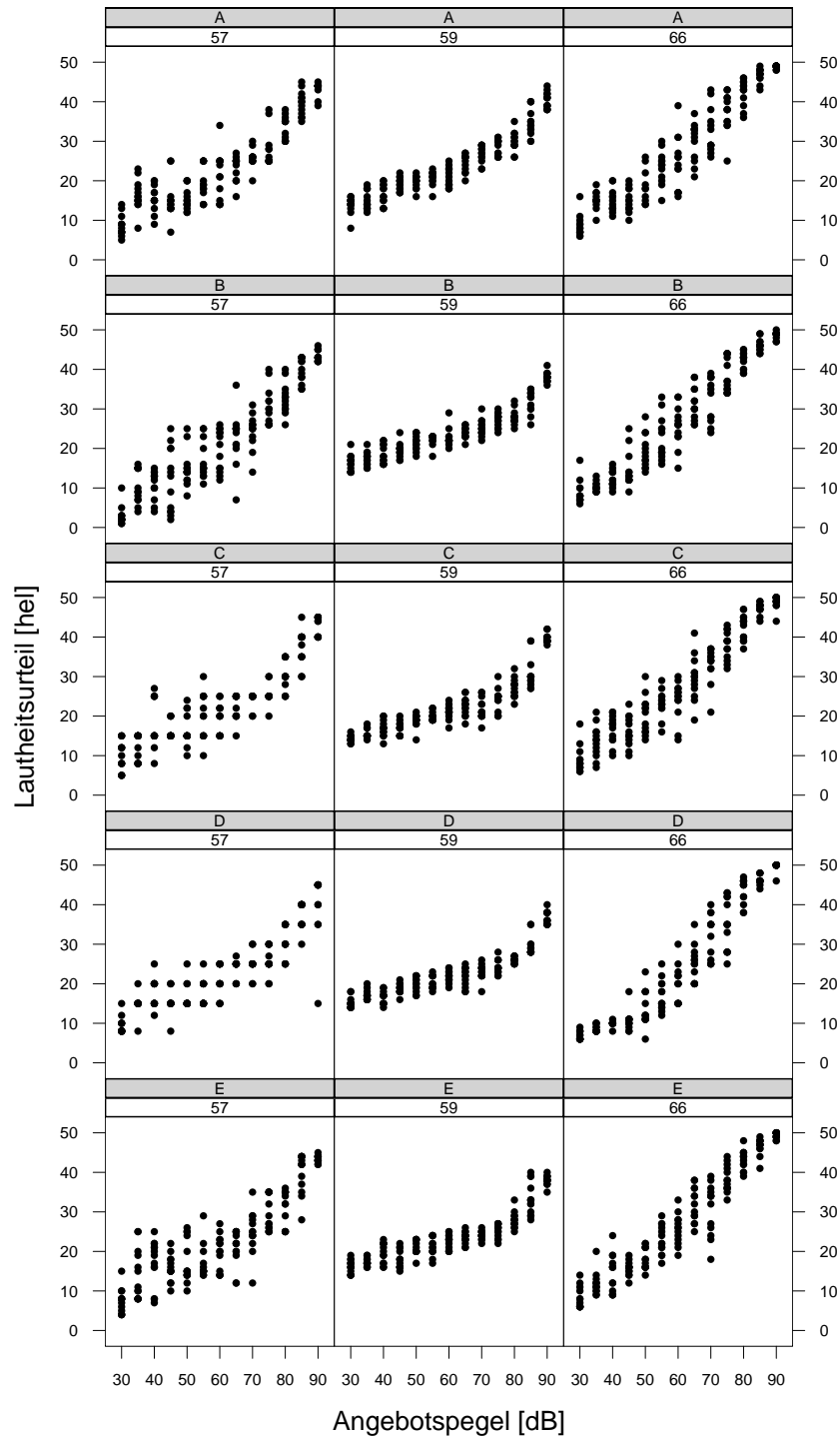


Abbildung 6.1: Rohdaten dreier Probanden für die fünf Urteilsmethoden, Lautheitsurteil aufgetragen über dem Schallpegel

relativ groß. Dieser Proband entspricht am ehesten der Idealvorstellung einer linearen Skalierung, entsprechend einer Geraden durch die Punkte (30 dB|1 hel) und (90 dB|50 hel)

Betrachtet man jeden Probanden einzeln, kann man Unterschiede in der Skalierung mit den verschiedenen Urteilsmethoden erkennen. Bei Proband 57 sieht man eine Zunahme der Querstruktur bei den Urteilsmethoden C und D. Insbesondere bei Urteilsmethode D sieht es so aus, als ob der Proband nur noch in Fünferschritten skaliert und die Bereiche dazwischen völlig ausspart. Bei Proband 59 fallen rein visuell betrachtet keine besonderen Unterschiede zwischen den Urteilsmethoden auf. Auch Proband 66 scheint unabhängig von der Urteilsmethode zu skalieren. Die einzige Auffälligkeit ist die im Vergleich zu den anderen Methoden geringere Streuung im unteren Schallpegelbereich bei Methode D.

## 6.3 Visuelle Betrachtung der Urteilsdichte

In diesem Abschnitt werden die drei gleichen Beispielprobanden betrachtet wie in Kapitel 6.2 ([Visuelle Betrachtung der Rohdaten](#), S. 41). Die Daten werden in Abbildung 6.2 jedoch anders graphisch dargestellt. Auf der Abszisse sind die Lautheitsurteile in „hel“ aufgetragen. Auf der Ordinate wird die Verteilungsdichte der Urteile dargestellt. Die Rohdaten aller Probanden sind im Anhang in Kapitel B (S. 105 ff.) aufgeführt. Genauere Angaben zur Erstellung der entsprechenden Schaubilder folgen in Kapitel 6.3.1 ([Festlegung des Kerndichteschätzers](#), S. 45 ff.).

Beim Vergleich der Probanden fallen die jeweiligen sehr verschiedenen Herangehensweisen der Skalierung auf. Proband 57 zeigt einen zweigipfligen Verlauf der Urteilsdichten. Urteile um 17 hel und 27 hel werden besonders häufig gefällt. Im oberen Bereich der Skala gibt es nochmals einen kleinen Gipfel bei ca. 42 hel. Der Proband nutzt fast die ganze Skala aus, meidet allerdings die Extrembereiche.

Bei Proband 59 erkennt man einen stark zugespitzten Verlauf der Kurve. Die häufigsten Urteile liegen ungefähr zwischen 18 hel und 25 hel. Urteile, die nicht in diesem Bereich liegen, werden weit seltener gefällt; erkennbar an dem sehr steilen Abfall der Kurve links und rechts. Der weitaus kleinere Gipfel bei ca. 38 hel zieht sich konstant durch alle Urteilsmethoden. Die Kurve bedeckt außerdem nur

## 6 Ergebnisse

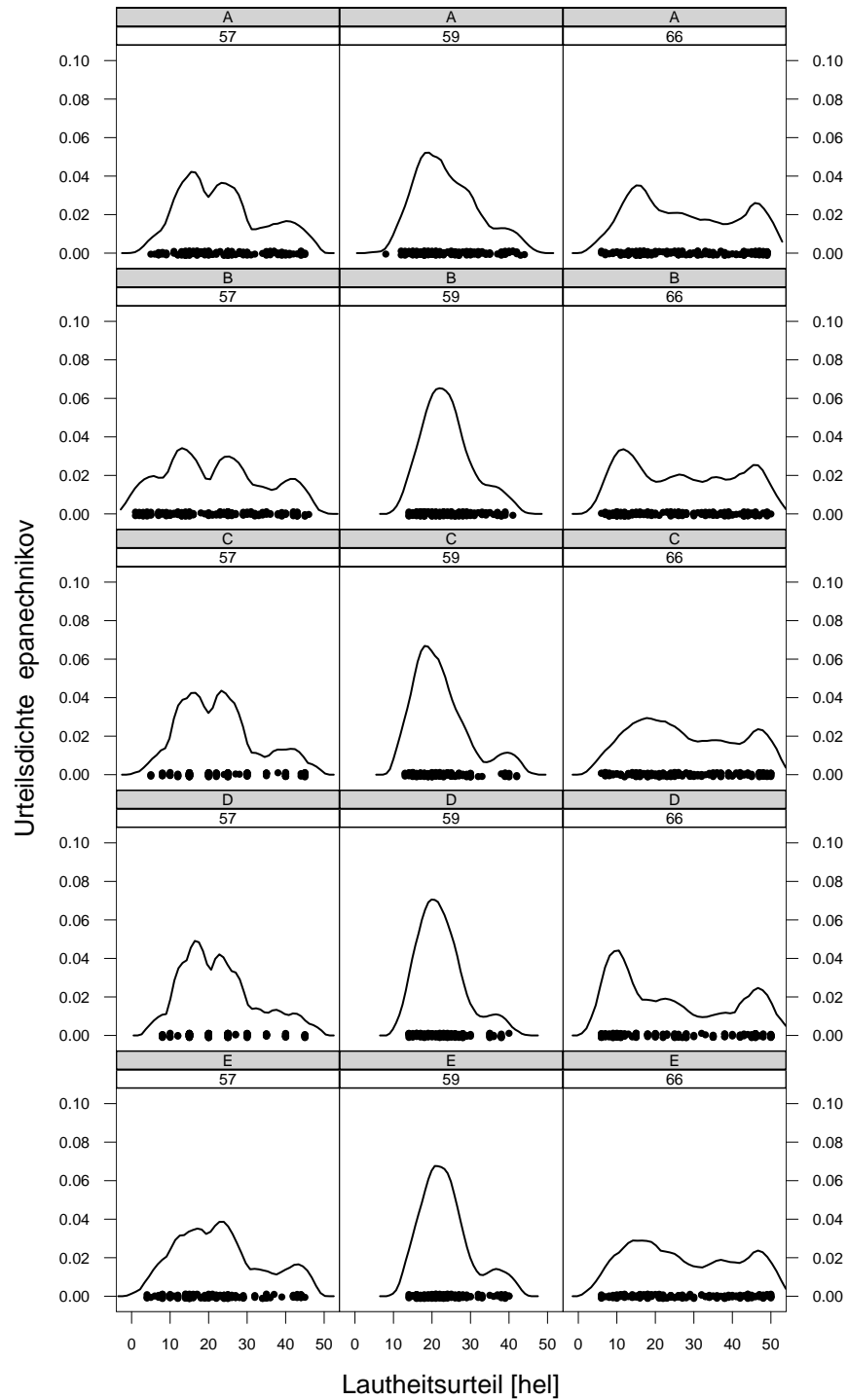


Abbildung 6.2: Urteilsdichte dreier Probanden für die fünf Urteilsmethoden, Dichte aufgetragen über dem Urteil

einen Bereich von ca. 10 hel bis 42 hel. Urteile der untersten und der obersten Kategorie werden quasi völlig weggelassen.

Proband 66 zeigt einen flachen und breiten Verlauf der Kurve. Er nutzt die gesamte Skalenbreite gleichmäßig aus. Die Kategorien „leise“ und „sehr laut“ werden etwas bevorzugt.

Nun zum Vergleich der verschiedenen Urteilsmethoden innerhalb eines Probanden. Bei Proband 57 bestätigen sich die Beobachtungen aus Abschnitt 6.3 ([Visuelle Betrachtung der Urteilsdichte](#), S. 43). Die Querstruktur zeigt sich in dieser Darstellung bei den Urteilsmethoden C und D am stärksten ausgeprägt in den Doppelgipfeln. Die anderen drei Urteilsmethoden sind einander ähnlich.

Bei Proband 59 zeigt nur die Methode A einen etwas abweichenden Verlauf von den anderen Methoden. Hier ist die Zuspitzung der Kurve etwas weniger ausgeprägt, aber immer noch deutlich vorhanden. Sonst sind keine Auffälligkeiten zu sehen.

Die Verläufe der Kurven von Proband 66 sind einander sehr ähnlich. Die Urteilsmethode scheint keinen Einfluss auf sein Skalierungsverhalten zu haben.

#### 6.3.1 Festlegung des Kerndichteschätzers

Im Rahmen einer am selben Institut durchgeführten Dissertation ([Kaulitz, 2008](#)) wurde eine Unregelmäßigkeit in der Darstellung der Dichteverteilungen festgestellt. Um diese verstehen zu können, soll hier kurz auf die Vorgehensweise einer Dichteschätzung eingegangen werden.

Um eine Dichteschätzung vorzunehmen, wird ein einzelner Wert nicht als *ein* konkreter Punkt dargestellt, sondern als eine Wahrscheinlichkeitsverteilung, z.B. als Gauß-Kurve mit Maximum an der Stelle des erhobenen Werts. Diese Kurve wird als Kerndichteschätzer bezeichnet. Für jeden Wert wird ein solcher Kerndichteschätzer erstellt und alle entstanden Kurven zusammenaddiert. So entsteht dann die endgültige Graphik der Verteilungsdichte. Besser verständlich macht dies die Darstellung in [Abb. 6.3](#).

Der Kerndichteschätzer wird durch 2 Parameter definiert. Erstens durch die gewählte Kernfunktion (siehe [Abb. 6.4](#)) und zweitens durch die Bandbreite. Die Bandbreite kann durch ein Streumaß definiert oder konkret festgelegt werden. „Die Wahl der Bandbreite  $b$  spielt [...] eine große Rolle. Bei zu kleiner Bandbreite wird die geschätzte Dichte im Verlauf sehr unruhig und weist viele lokale Maxima auf. Bei großer Bandbreite wird die Dichte zu glatt. [...] Mit wachsendem

$b$  steigt der Bias und sinkt die Varianz des Schätzers.“ (Rinne, 2003). Effekte verschiedener Bandbreiten stellt Abb. 6.5 dar.

Im Rahmen der vorliegenden Dissertation, wie auch der Dissertation von Kaulitz (2008), wurden die Dichteverteilungen mit dem bereits in Abschnitt 6.1 (Statistische Auswertung, S. 37) erwähnten Statistikprogramm *R* (The R Development Core Team, 2006) berechnet. Es zeigten sich allerdings Unstimmigkeiten in der Darstellung der Dichteschätzung zwischen der abgebildeten Kurve und der am unteren Rand im selben Schaubild abgebildeten Punktwolke (vgl. z. B. Abb 6.2). Es stellte sich heraus, dass *R* für jede einzelne Dichteschätzung automatisch den Kerndichteschätzer neu berechnete. Das Berechnungsverfahren ist in *R* implementiert und soll hier nicht näher erörtert werden. Dies erklärte einerseits die Abweichungen in den Graphiken. Außerdem wurde klar, dass die Dichteschätzungen so nicht miteinander vergleichbar waren, da jeder Berechnung ein anderer Kerndichteschätzer zu Grunde lag. Dieser musste also fest vorgegeben werden.

Kaulitz (2008) entschied sich für einen Epanechnikov-Kern mit einer Bandbreite von  $b = 2,5$ . Betrachtete man die von *R* errechneten Kerndichteschätzer, lag deren Bandbreite im Durchschnitt bei  $b = 3,5$ . Kaulitz (2008) berechnete dann alle Dichteschätzungen mit verschiedenen Kernen (unter anderem auch Cosinus-, Dreieck-, Rechteck-Kern) und jeweils Bandbreiten von 0,5 bis 5,0 in 0,5er-Schritten. Für den Epanechnikov-Kern sprach der Gauß-ähnliche Verlauf bei jedoch endlicher Kernfunktion. Da das Spektrum der Urteilsmöglichkeiten begrenzt ist, sollte auch die Kernfunktion nicht unendlich sein. Für die gewählte, etwas kleinere Bandbreite von  $b = 2,5$  sprach eine bessere Detaildarstellung bei immer noch ruhigem Verlauf der Kurven. Auch für die vorliegende Dissertation trafen diese Überlegungen zu. Deshalb wurde der selbe Kerndichteschätzer gewählt. Außerdem bleibt so ein eventuell folgender Vergleich der beiden Arbeiten möglich.

### 6.4 Urteilsunterschiede der Probanden

Eine komprimierten Überblick über die Lage der Urteile eines jeden Probanden gibt Abb. 6.6. Es wird die Verteilung der Urteile in Box-und-Whisker-Plots dargestellt. Jeder Box-Plot ist gleich aufgebaut. Der dicke Mittelstrich gibt den Medianwert der Urteile an. Die rechteckige Box gibt den Wert vom unteren zum



## Konstruktion eines Kerndichteschätzers

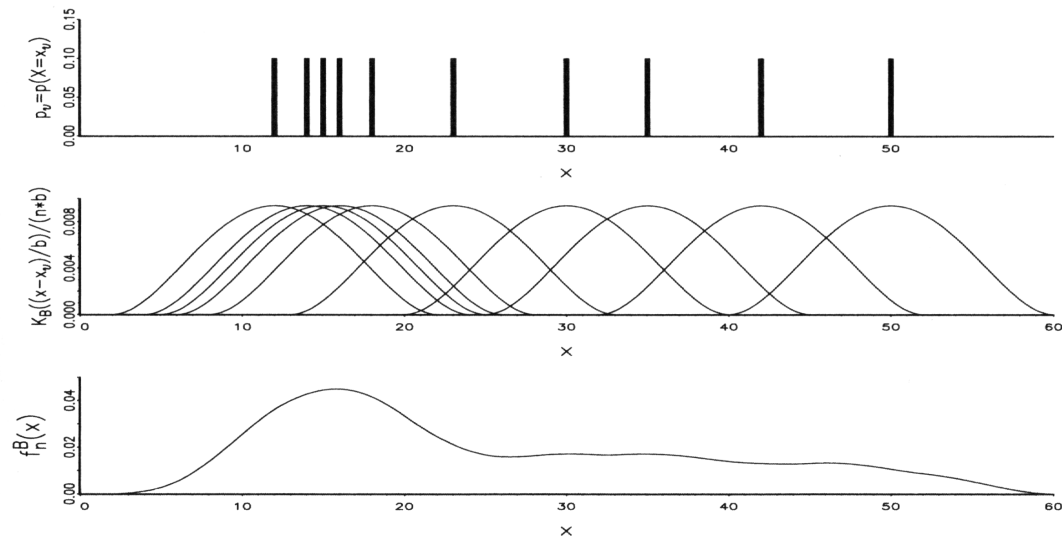


Abbildung 6.3: Konstruktion eines Kerndichteschätzers, Abb. entnommen aus Rinne (2003, S. 485).

## Verlauf einiger Kernfunktionen

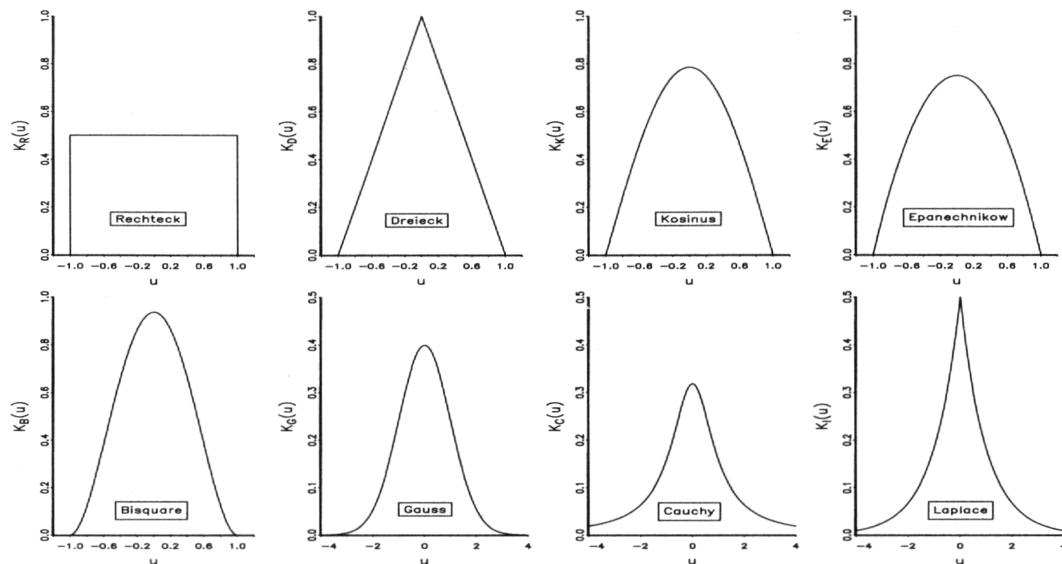


Abbildung 6.4: Verlauf einiger Kernfunktionen, Abb. entnommen aus Rinne (2003, S. 486).

## Bandbreiteneffekt auf die Kerndichteschätzung bei einem Bisquare-Kern

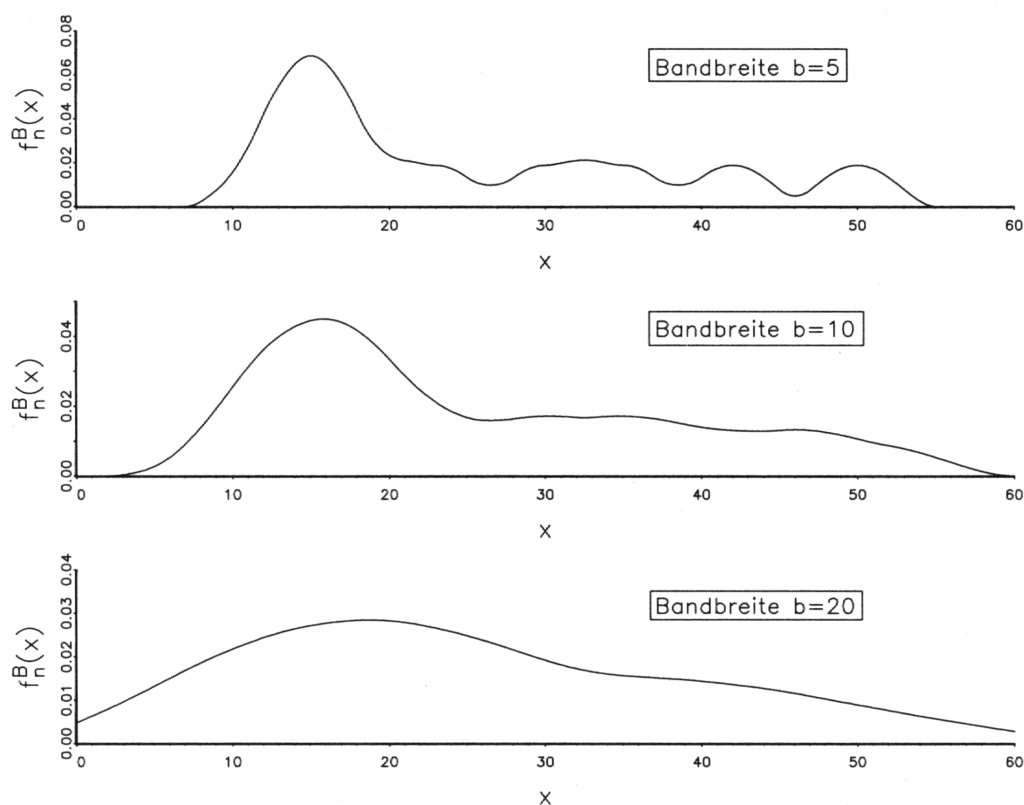


Abbildung 6.5: Bandbreiteneffekt, Abb. entnommen aus Rinne (2003, S. 486).

oberen Quartil an; sprich den Bereich zwischen 25. Perzentile und 75. Perzentile. Die Box enthält also die mittleren 50 % der abgegebenen Urteile. Die gestrichelte senkrechte Linie (Whisker (engl.) Barthaar) erstreckt sich vom Minimalwert zum Maximalwert. Die Weite der Kerbe am Medianwert gibt das 95 %-Konfidenzintervall an. Es wird auch Vertrauensintervall genannt. Es „[...]“ enthält den gesuchten Parameter mit einer Wahrscheinlichkeit von 95 %. Aus dem Konfidenzintervall lassen sich Schlüsse bezüglich der statistischen Signifikanz ziehen.“ (Bender und Lange, 2001). Überschneiden sich die Einkerbungen der Konfidenzintervalle zweier benachbarter Plots nicht, so kann man davon ausgehen, dass die beiden Mediane tatsächlich unterschiedlich sind. Um das durch interindividuelle Unterschiede zustandekommende Urteilsspektrum besser darstellen zu können, wurden die Probanden aufsteigend nach ihrem medianen Urteil angeordnet. Man sieht in Abb. 6.6, dass sich die Urteile der Probanden teilweise stark unterscheiden, sowohl in der Lage des Medians als auch in der

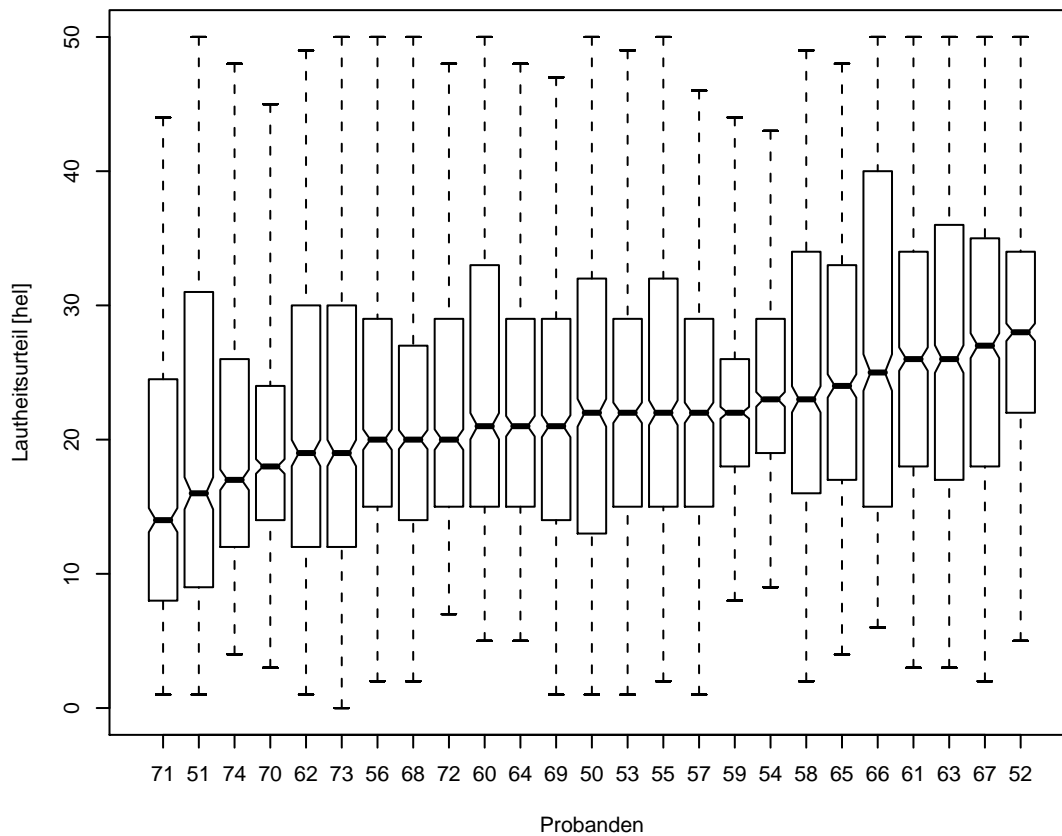


Abbildung 6.6: Lautheitsurteile der 25 Probanden

Ausdehnung der Box und der Extremwerte. Wie aus den oberen Abschnitten zu erwarten war, ist beispielsweise die Box bei Proband 59 äußerst eng und die Extremwerte liegen bei 8 hel und 44 hel, während Proband 66 eine sehr breite Box und Extremwerte von 6 hel und 50 hel hat.

In Tabelle 6.1 werden die der Box-Plot-Darstellung zugrundeliegenden Daten, sowie Mittelwert, Standardabweichung und mittlere absolute Abweichung als Zahlenwerte wiedergegeben. Für die mittlere absolute Abweichung (MAD) gilt: „Die mittlere absolute Abweichung bezieht alle Abweichungen der Merkmalswerte von ihrem arithmetischen Mittel in die Berechnung mit ein.“ (Voß, 2004).

Tabelle 6.1: Statistische Kennwerte der Lautheitsurteile aller Probanden: Mittelwert ( $\bar{x}$ ), Standardabweichung, Minimum, 1. Quartil, Median, 3. Quartil, Maximum und mittlere absolute Abweichung (MAD). Diese Werte werden auch in folgenden Tabellen angegeben.

Proband	$\bar{x}$	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
50	22,98	12,47	1	13,0	22,0	32,0	50	13,34
51	20,15	14,02	1	9,0	16,0	31,0	50	13,34
52	28,97	8,83	5	22,0	28,0	34,0	50	8,90
53	22,79	10,57	1	15,0	22,0	29,0	49	10,38
54	23,68	7,38	9	19,0	23,0	29,0	43	7,41
55	23,78	12,28	2	15,0	22,0	32,0	50	11,86
56	22,01	10,58	2	15,0	20,0	29,0	50	10,38
57	22,68	10,73	1	15,0	22,0	29,0	46	10,38
58	24,81	11,69	2	16,0	23,0	34,0	49	13,34
59	23,26	6,77	8	18,0	22,0	26,0	44	5,93
60	24,22	11,80	5	15,0	21,0	33,0	50	13,34
61	26,42	10,65	3	18,0	26,0	34,0	50	11,86
62	22,08	13,36	1	12,0	19,0	30,0	49	11,86
63	26,75	12,24	3	17,0	26,0	36,0	50	13,34
64	22,59	9,51	5	15,0	21,0	29,0	48	10,38
65	25,28	10,07	4	17,0	24,0	33,0	48	11,86
66	26,93	13,91	6	15,0	25,0	40,0	50	17,79
67	27,27	11,65	2	18,0	27,0	35,0	50	11,86
68	21,51	10,78	2	14,0	20,0	27,0	50	10,38
69	21,53	10,20	1	14,0	21,0	29,0	47	10,38
70	19,29	7,49	3	14,0	18,0	24,0	45	7,41
71	16,30	9,62	1	8,0	14,0	24,5	44	8,90
72	22,55	8,94	7	15,0	20,0	29,0	48	8,90
73	21,86	12,47	0	12,0	19,0	30,0	50	11,86
74	19,62	9,48	4	12,0	17,0	26,0	48	9,64

Tabelle 6.2: Statistische Werte der Lautheitsurteile nach Urteilsmethode geordnet (Erläuterung der Kennwerte in Tab. 6.1)

Urteilsmethode	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
A	24,08	11,26	1	15,0	23,0	32,0	50	11,86
B	24,04	10,87	0	16,0	23,0	31,0	50	11,86
C	22,73	11,11	1	15,0	21,0	30,0	50	10,38
D	22,11	11,59	1	13,0	20,0	30,0	50	11,86
E	22,91	11,07	1	15,0	21,0	30,0	50	10,38

## 6.5 Urteil in Abhängigkeit von der Urteilsmethode

Die Graphik in Abb. 6.7 stellt die Antworten in Abhängigkeit von der Urteilsmethode dar. Jeder Plot beinhaltet alle Antworten aller Probanden mit der jeweiligen Urteilsmethode. Schon visuell kann man die fünf Urteilsmethoden in drei Gruppen unterteilen. Die erste Gruppe bilden die Methoden A und B. Beide sind fast identisch. Die Mediane liegen auf gleicher Höhe bei 23 hel. Auch die Extremwerte sind identisch und liegen an den Enden der Skala (0 hel und 50 hel). Der einzige Unterschied ist die etwas kleinere Box bei Urteilsmethode B. Die Urteilsmethoden C und E gehören zur zweiten Gruppe. Beide sind wiederum annähernd identisch, aber unterscheiden sich deutlich von den Plots der ersten Gruppe. Der Median liegt signifikant niedriger. Auch die 75. Perzentile liegt auf einem niedrigeren Niveau, während die 25. Perzentile bei A, C und E auf gleicher Höhe liegt. Die Urteilsmethode D steht alleine da. Sie hat den niedrigsten Median aller fünf Urteilsmethoden.

Tabelle 6.2 gibt die in Abb. 6.7 aufgetragenen Werte in Zahlenform wieder.

Tabelle 6.3 bestätigt die aufgrund der visuellen Betrachtung naheliegenden Vermutungen. Im statistischen Vergleich mit dem Wilcoxon-Rangsummentest sind die Methoden A und B sowie die Methoden C und E identisch. Alle anderen Gegenüberstellungen zeigen, dass sich die beiden verglichenen Methoden jeweils signifikant voneinander unterscheiden.

## 6.6 Urteil in Abhängigkeit vom Pegel

Abbildung 6.8 zeigt die Lautheitsurteile aufgetragen über dem Schallpegel. Diese Art der Darstellung wird als Pegel-Lautheitsfunktion bezeichnet. Es wurden

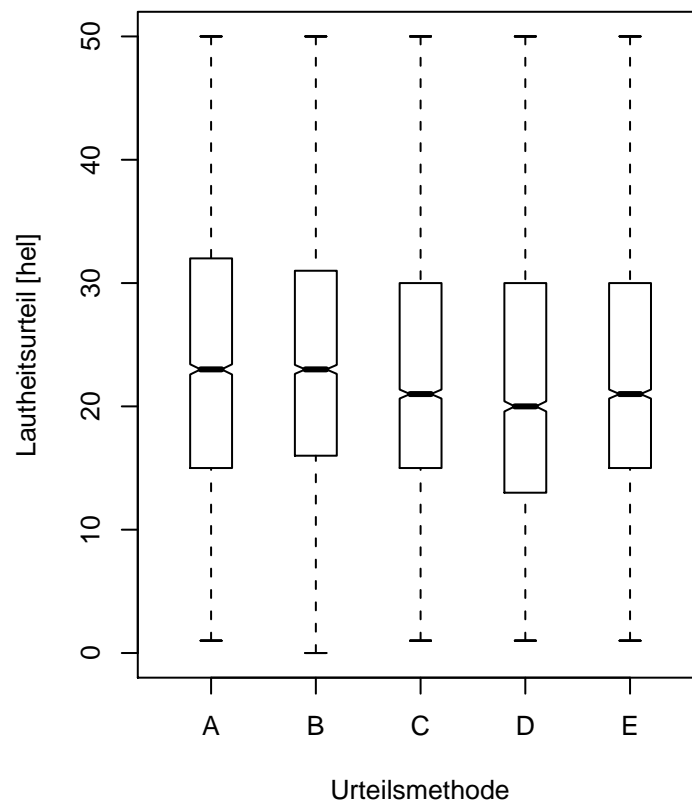


Abbildung 6.7: Lautheitsurteile in Abhängigkeit von der Urteilmethode

Tabelle 6.3: Statistischer Vergleich aller Urteilsmethoden untereinander mit dem Wilcoxon-Rangsummentest

Urteilsmethode			p-Wert
A	—	B	0,835085150927
A	—	C	0,000000012178
A	—	D	$< 1,0 \times 10^{-9}$
A	—	E	0,000000273283
B	—	C	0,000000001214
B	—	D	$< 1,0 \times 10^{-9}$
B	—	E	0,000000031298
C	—	D	0,001755764085
C	—	E	0,581721863378
D	—	E	0,000181687067

alle Probanden und alle Urteilsmethoden zusammengefasst. Die Kurve wurde, wie alle anderen Pegel-Lautheitsfunktionen in dieser Arbeit, mit dem „loess“-Regressionsmodell erstellt (s. Abschnitt 6.1.1, [Verwendete statistische Tests und Verfahren](#), S. 38). Gezeigt wird eine stetige, monoton ansteigende, konkave Kurve, die im höheren Schallpegelbereich einen steileren Verlauf aufweist. Die Lautheit steigt mit zunehmenden Pegel also kontinuierlich, jedoch nicht linear an. Der Kurvenverlauf ist der von [Nägler \(2004\)](#) ermittelten Kurve sehr ähnlich. Die Erkenntnis, dass mit Zunahme des Schallpegels das Ohr empfindlicher gegenüber Schallpegelschwankungen wird, belegten bereits [Zwicker und Feldtkeller \(1967, S. 94ff\)](#).

Die der Pegel-Lautheitsfunktion zu Grunde liegenden Zahlenwerte werden für jeden Pegel einzeln in Tabelle 6.4 aufgelistet. Zur Erläuterung der Werte siehe Tabelle 6.1. Übersichtlicher als in Tab. 6.4 werden die ermittelten Daten in Abb. 6.9 als Boxplot dargestellt. Jede Schallpegelerhöhung um 5 dB bewirkt auch eine Erhöhung der Lautheitsurteile. Nur zwischen den Pegeln 40 dB und 45 dB lässt sich graphisch kein großer Unterschied erkennen. Betrachtet man jedoch die in Tabelle 6.5 aufgeführten p-Werte, so sieht man, dass sich wirklich alle benachbarten Pegeln hochsignifikant voneinander unterscheiden (als statistischer Test diente der ungepaarte Wilcoxon-Test). Tabelle 6.5 gibt an, wie stark sich die Urteile über die benachbarten Pegel jeweils voneinander unterscheiden.

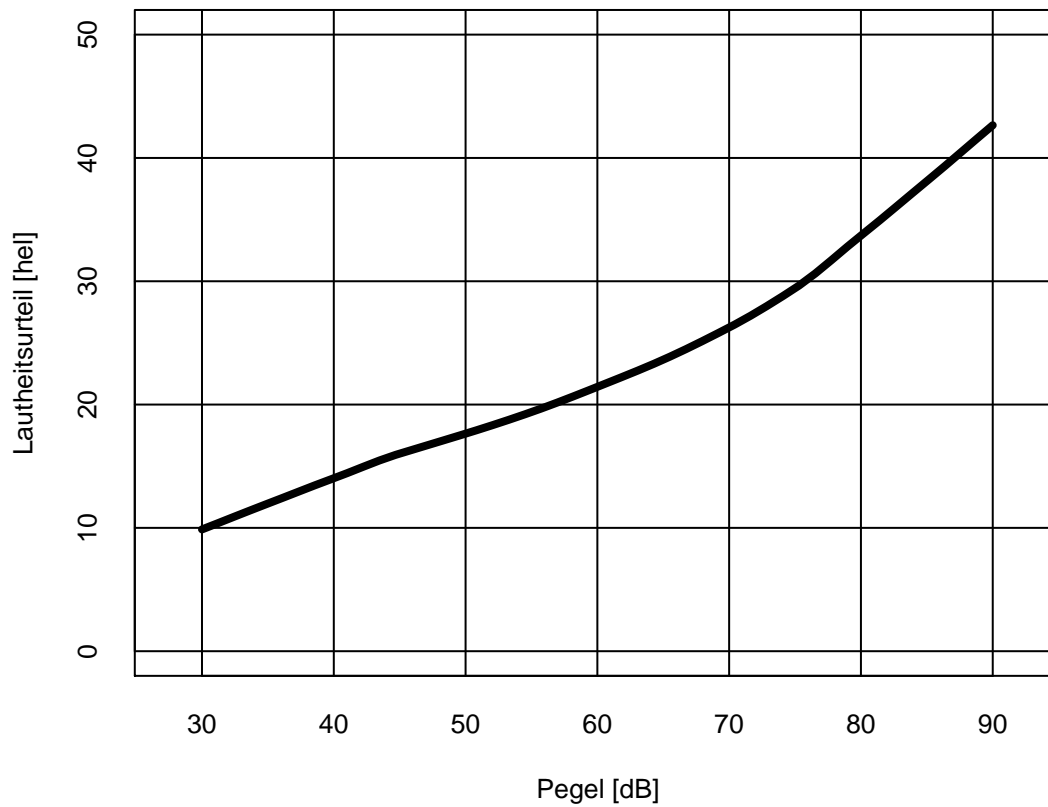


Abbildung 6.8: Pegel-Lautheitsfunktion (nichtlineare „loess“-Regressionskurve), alle Probanden und Urteilsmethoden zusammengefasst



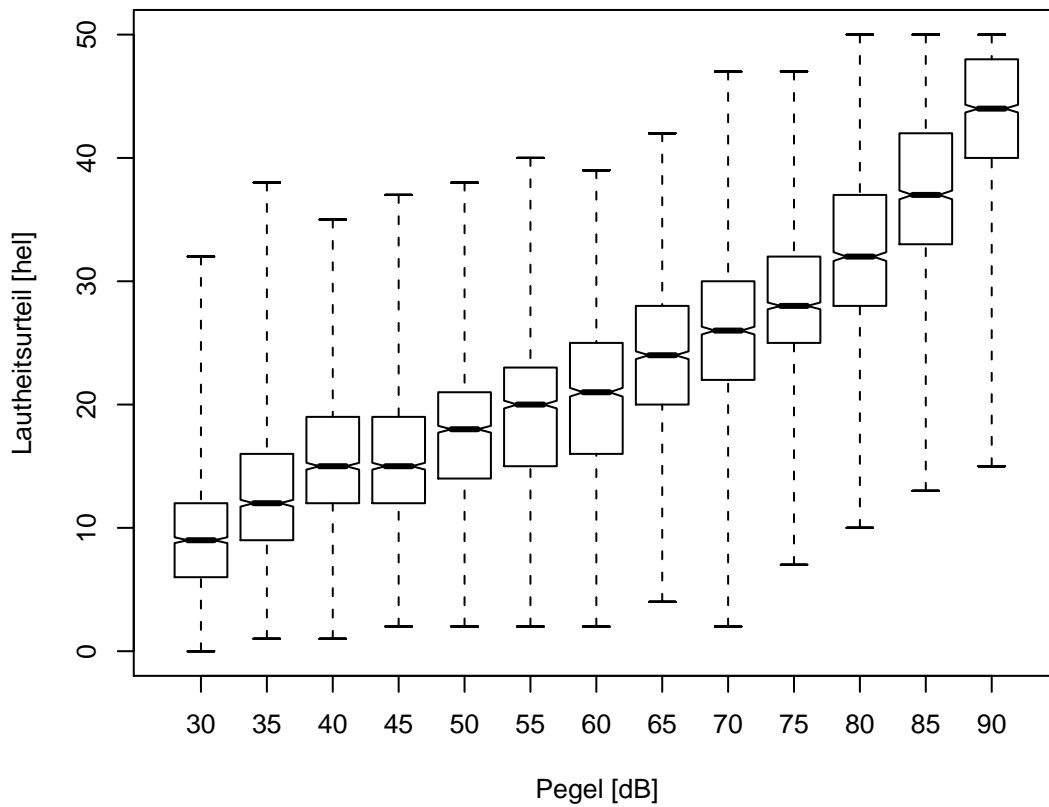


Abbildung 6.9: Lautheitsurteile in Abhängigkeit vom Schallpegel als Boxplots, alle Probanden und Urteilsmethoden zusammengefasst

Tabelle 6.4: Statistische Werte der Lautheitsurteile in Abhängigkeit vom Schallpegel (Erläuterung der Werte in Tab. 6.1)

Stimulus	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
30	9,35	4,82	0	6,0	9,0	12,0	32	4,45
35	12,59	5,28	1	9,0	12,0	16,0	38	5,19
40	15,20	5,61	1	12,0	15,0	19,0	35	5,93
45	15,73	5,45	2	12,0	15,0	19,0	37	5,93
50	17,62	5,38	2	14,0	18,0	21,0	38	4,45
55	19,44	5,66	2	15,0	20,0	23,0	40	5,93
60	20,62	6,18	2	16,0	21,0	25,0	39	5,93
65	23,69	6,18	4	20,0	24,0	28,0	42	5,93
70	25,76	6,16	2	22,0	26,0	30,0	47	5,93
75	28,19	6,39	7	25,0	28,0	32,0	47	5,93
80	32,52	6,37	10	28,0	32,0	37,0	50	7,41
85	37,30	6,45	13	33,0	37,0	42,0	50	7,41
90	43,24	5,22	15	40,0	44,0	48,0	50	5,93

Tabelle 6.5: Statistischer Vergleich der Lautheitsurteile benachbarter Schallpegel: Differenz benachbarter Mittelwerte und Mediane, p-Wert des Wilcoxon-Tests

Pegel	Δ Ø	Δ Med.	p-Wert
30 — 35	3,24	3,0	$< 1,0 \times 10^{-9}$
35 — 40	2,61	3,0	$< 1,0 \times 10^{-9}$
40 — 45	0,53	0,0	0,004853840
45 — 50	1,89	3,0	$< 1,0 \times 10^{-9}$
50 — 55	1,81	2,0	$< 1,0 \times 10^{-9}$
55 — 60	1,18	1,0	0,000000018
60 — 65	3,07	3,0	$< 1,0 \times 10^{-9}$
65 — 70	2,07	2,0	$< 1,0 \times 10^{-9}$
70 — 75	2,43	2,0	$< 1,0 \times 10^{-9}$
75 — 80	4,33	4,0	$< 1,0 \times 10^{-9}$
80 — 85	4,78	5,0	$< 1,0 \times 10^{-9}$
85 — 90	5,94	7,0	$< 1,0 \times 10^{-9}$

## 6.7 Urteil in Abhängigkeit von der Zeitposition der Urteilsmethode im Gesamtablauf

Um Reihenfolge-Effekte zu vermeiden wurden, wie in Kapitel 3.4 ([Reihenfolge der verschiedenen Urteilsmethoden](#), S. 17) beschrieben, die Urteilsmethoden systematisch durcheinandergewürfelt. Ob die Position der Urteilsmethoden im zeitlichen Ablauf des Versuchs überhaupt einen Einfluss hat und ob daraus Effekte auf die Lautheitsskalierung resultieren, soll hier untersucht werden. Beispielsweise könnte es einen Unterschied machen, ob eine Urteilsmethode an erster Stelle oder an letzter Stelle des Versuchs steht. Zuerst wurde nur der Einfluss der Zeitposition betrachtet, ohne zwischen den einzelnen Urteilsmethoden zu unterscheiden. Pro Zeitposition gingen damit alle 25 Probanden in die Auswertung mit ein, wobei jede der fünf Urteilsmethoden fünf mal vertreten war. Weiterhin wurden die Zeitpositionen für jede Urteilsmethode getrennt betrachtet. Aufgrund der in Kapitel 3.4 ([Reihenfolge der verschiedenen Urteilsmethoden](#), S. 17) gezeigten Anordnung der Urteilsmethoden, gehen jetzt pro Zeitposition nur noch fünf Probanden in die statistische Auswertung mit ein. Denn nur fünf Probanden hatten z.B. Urteilsmethode A an erster Stelle, während fünf andere Probanden Methode A an zweiter Stelle hatten, usw. Pro Urteilsmethode gehen also alle 25 Probanden in die Auswertung mit ein. Allerdings stammen die Daten für jede Zeitposition von jeweils verschiedenen Probanden.

Abb. 6.10 zeigt die Abhängigkeit der Lautheitsurteile von der Zeitposition für alle Urteilsmethoden zusammengefasst als Boxplots. Tab. 6.6 gibt die zugrundeliegenden Zahlenwerte wieder.

Schon bei rein visueller Betrachtung fallen folgende Sachverhalte ins Auge: Die Positionen 2, 3 und 4 scheinen identisch zu sein. Bei Zeitposition 1 fallen die Urteile etwas niedriger aus und die Streuung ist größer als bei den folgenden vier Positionen. Bei Zeitposition 5 liegen die Urteile etwas höher, dafür scheint die Streuung am geringsten zu sein.

In Tab. 6.7 werden die verschiedenen Zeitpositionen statistisch miteinander verglichen. Es bestätigen sich die Vermutungen aus der Betrachtung der Boxplotdarstellung. Die Positionen 2, 3 und 4 sind als statistisch gleich anzusehen. Die Positionen 1 und 5 unterscheiden sich von den jeweils vier anderen Positionen hochsignifikant.

In der nach Urteilsmethoden aufgetrennten Darstellung der Zeitpositionseffek-

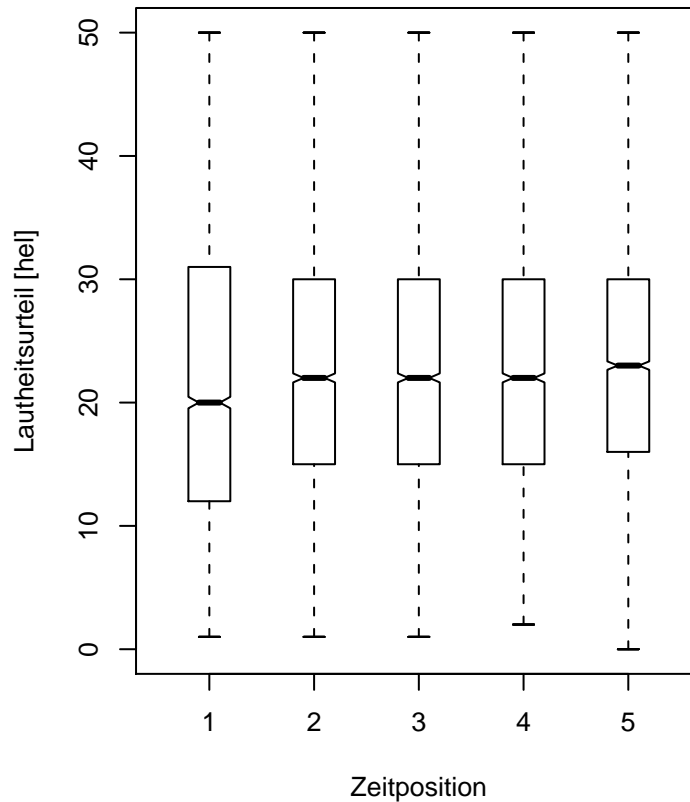


Abbildung 6.10: Lautheitsurteil in Abhängigkeit der Zeitposition als Boxplot, alle Probanden und Urteilsmethoden zusammengefasst

te (Abb. 6.11) fällt auf, dass sich die Abbildungen jeweils erheblich voneinander unterscheiden, ohne ein einheitliches Schema erkennen zu lassen. Eine sinnvolle Interpretation bzw. logische Folgerungen scheinen sich aus dieser Darstellung nicht zu ergeben. Am wahrscheinlichsten liegt der Grund hierfür darin, dass sich für jede Urteilsmethode jede Position aus verschiedenen Probanden zusammensetzt. Dadurch, dass die Probanden individuell sehr verschieden skalieren, könnte in diesem Zusammenhang ein eventuell doch bestehender Einfluss der Positionierung leicht überdeckt werden.

## 6.8 Urteil in Abhängigkeit von Vorgängerpegel und Vorgängerurteil

Nachdem in den vorhergehenden Kapiteln die Abhängigkeit der Urteile von verschiedenen Faktoren präsentiert wurden, soll in diesem Kapitel auf die Effekte

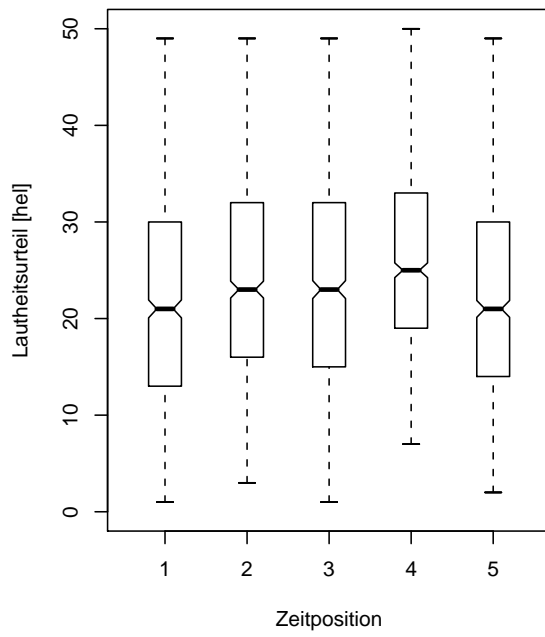
Tabelle 6.6: Statistische Kennwerte zur Abhängigkeit der Urteile von der Zeitposition - alle Probanden und Urteilsmethoden zusammengefasst. (Erläuterung der Kennwerte in Tab. 6.1)

Zeitposition	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
1	22,04	12,50	1	12,0	20,0	31,0	50	13,34
2	23,15	10,96	1	15,0	22,0	30,0	50	10,38
3	23,39	10,94	1	15,0	22,0	30,0	50	10,38
4	23,37	10,92	2	15,0	22,0	30,0	50	10,38
5	23,92	10,54	0	16,0	23,0	30,0	50	10,38

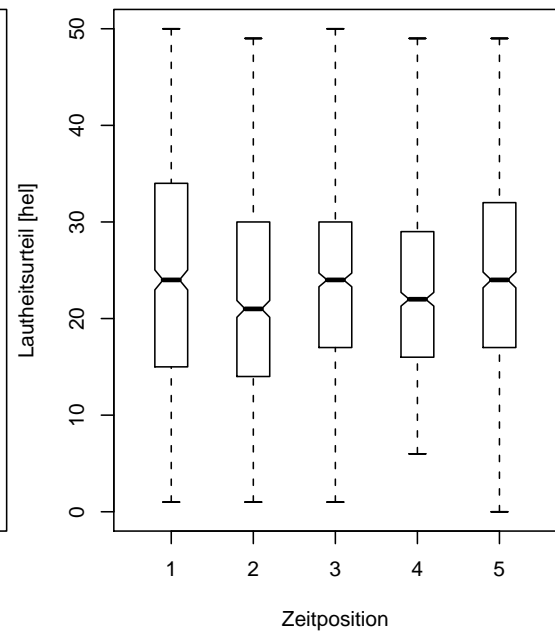
Tabelle 6.7: Statistischer Vergleich der einzelnen Zeitpositionen mit dem Wilcoxon-Rangsummentest - alle Probanden und Urteilsmethoden zusammengefasst.

Zeitposition	p-Wert
1 — 2	0,000000008281
1 — 3	$< 1,0 \times 10^{-9}$
1 — 4	$< 1,0 \times 10^{-9}$
1 — 5	$< 1,0 \times 10^{-9}$
2 — 3	0,278787582395
2 — 4	0,395264343619
2 — 5	0,000459419702
3 — 4	0,817651842952
3 — 5	0,016014739313
4 — 5	0,007029643355

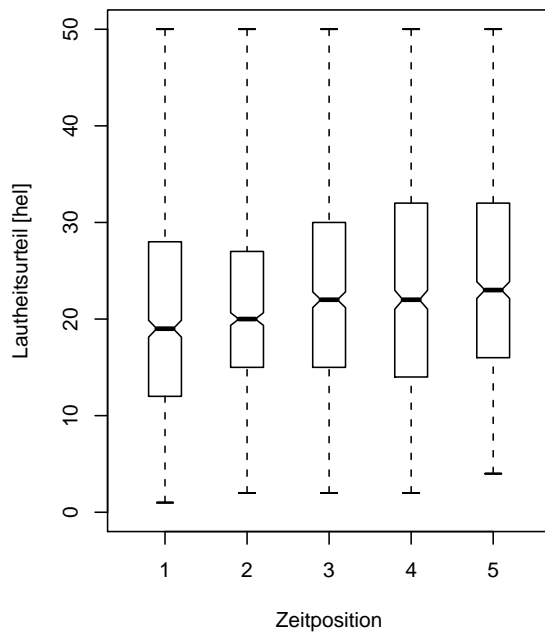
## 6 Ergebnisse



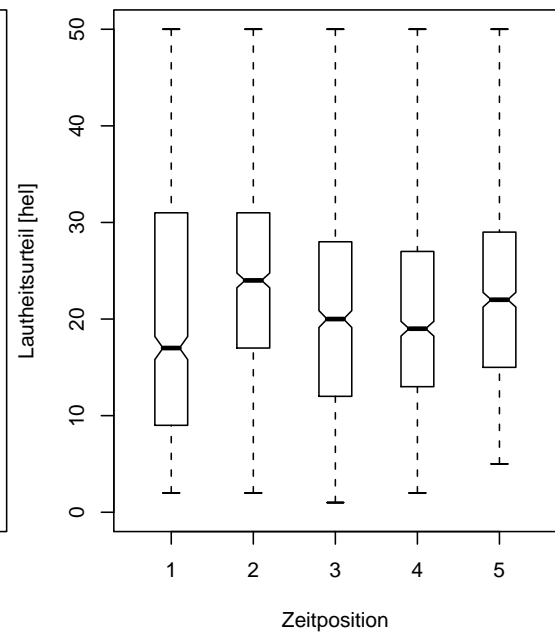
(a)



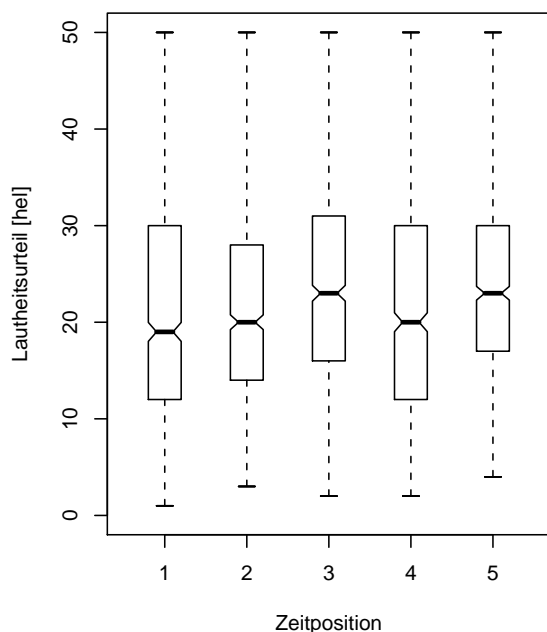
(b)



(c)



(d)



(e)

Abbildung 6.11: Zeitposition nach Urteilsmethoden getrennt; kein systematischer Zusammenhang zwischen den einzelnen Urteilsmethoden erkennbar:

(a) Effekt der Zeitposition für Urteilsmethode A; (b) für Methode B; (c) für Methode C; (d) für Methode D und (e) für Methode E

des *Vorgängerstimulus* auf das Lautheitsurteil eingegangen werden. Die Effekte von Vorgängerurteil und Vorgängerpegel werden dabei getrennt betrachtet. Zur übersichtlichen Veranschaulichung der Ergebnisse werden in Abb. 6.12 wieder die drei aus Kapitel 6.2 ([Visuelle Betrachtung der Rohdaten](#), S. 41) und Kapitel 6.3 ([Visuelle Betrachtung der Urteilsdichte](#), S. 43) bekannten Beispielprobanden 57, 59 und 66 herangezogen.

Die Ergebnisse aller Probanden befinden sich im Anhang (B, S. 105 ff.).

Abbildung 6.12 besteht aus zwei Spalten. Links ist die Abhängigkeit der Lautheitsurteile vom Schallpegel des Vorgängerstimulus, rechts das Lautheitsurteil in Abhängigkeit vom Lautheitsurteil über den Vorgängerstimulus dargestellt. Die Probanden sind von oben nach unten angeordnet.

Auf der Ordinate sind jeweils die Lautheitsurteile von 0 hel bis 50 hel aufgetragen. Die Abszisse ist in der linken Spalte mit den 13 möglichen Vorgängerpegeln in dB und in der rechten Spalte mit den Lautheitsurteilen in hel beschriftet.

Jeder der 13 vorkommenden Schallpegel hatte auch jeden Schallpegel zum Vorgänger. Somit wurde jeder Schallpegel 13 mal bewertet. Die Urteile über

einen Pegel ergeben im Schaubild also eine Punktwolke mit 13 Punkten. Um jeden Punkt im Schaubild eindeutig seinem Pegel zuordnen zu können wurden die Pegel farblich kodiert (s. Farblegende, im Anhang Abb. B.1 in Kapitel B, S. 106).

Für jede Punktwolke wurde außerdem eine lineare Regression berechnet und die sich ergebende Regressionsgerade wurde ebenfalls in das Schaubild eingezeichnet. Das Schaubild beinhaltet also 13 verschiedenfarbige Punktwolken zu jeweils 13 Punkten und die dazugehörige Regressionsgerade.

Jede Gerade kann nach einer einfachen mathematischen Formel exakt definiert werden. Die Formel lautet:

$$f(x) = a * x + c$$

Die Variable  $a$  definiert die Steigung der Geraden und die Variable  $c$  den y-Achsen Schnittpunkt.  $f(x)$  und  $x$  geben die jeweiligen Werte der Funktion an. Dies bedeutet, dass eine Gerade alleine durch ihren Schnittpunkt mit der y-Achse und ihre Steigung definiert ist. Die Steigung kann sowohl negativ als auch positiv sein. In Abb. 6.12 haben die meisten Geraden eine positive Steigung, d.h. sie steigen von links nach rechts an. Eine positive Steigung bedeutet folgendes: Ist der Vorgänger (-pegel oder -urteil) groß, wird ein Stimulus lauter beurteilt als bei einem kleinen Vorgänger. Daraus folgt analog, dass ein Stimulus mit kleinem Vorgänger relativ leiser beurteilt wird.

Für eine negative Steigung (die Gerade fällt von links nach rechts ab) heißt das, dass ein Stimulus mit großem Vorgänger leiser beurteilt wird als derselbe Stimulus mit kleinem Vorgänger.

Der Betrag der Steigung spiegelt das Ausmaß der Abhängigkeitsverhältnisse zwischen Lautheitsurteil und Vorgängerpegel bzw. Vorgängerurteil wider. Je steiler eine Gerade verläuft, desto stärker ist das abgegebene Lautheitsurteil von seinem Vorgänger abhängig. Im Gegenzug besagt eine horizontal verlaufende Gerade, dass keine Abhängigkeit besteht. In der linken Spalte wird die Steigung der Gerade in  $\text{hel} / \text{dB}$  angegeben. In der rechten Spalte in  $\text{hel} / \text{hel}$ . Die Steigung der rechten Gerade ist also dimensionslos, da sich die Einheiten gegeneinander kürzen lassen.

Um die beiden Steigungen miteinander vergleichen zu können, muss die Vorgängerpegelabhängigkeit erst noch mit dem Faktor  $60 \text{ dB} / 50 \text{ hel}$  ( $1,2 \text{ dB/hel}$ ) multipliziert werden. Damit erreicht man eine identische Normierung der beiden Steigungen. Dies wurde in allen in dieser Arbeit vorkommenden Graphiken durchgeführt damit die Geraden auch optisch vergleichbar sind. Im Folgenden



## 6.8 Vorgängerpegel und -urteilsabhängigkeit

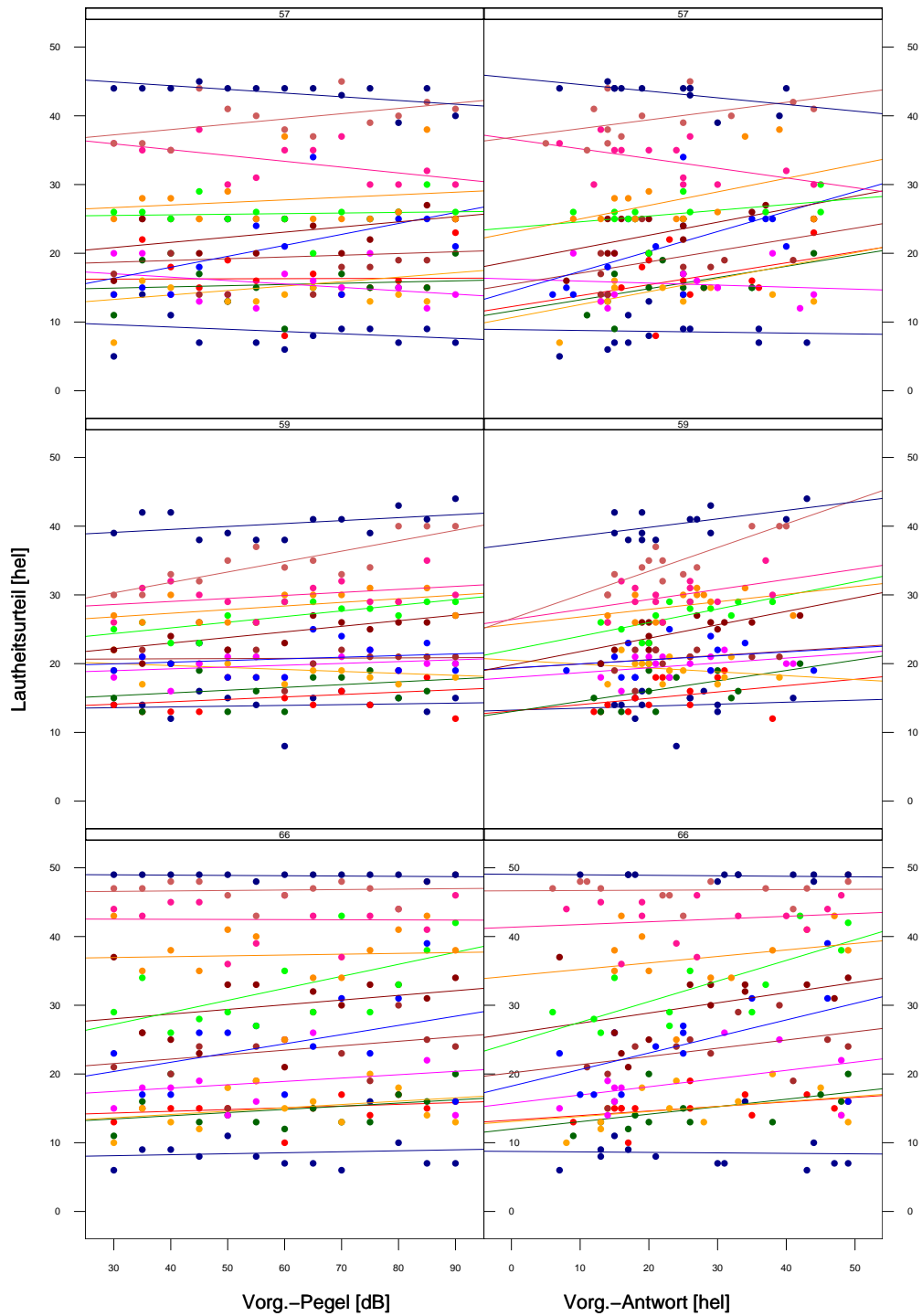


Abbildung 6.12: Lautheitsurteile in Abhängigkeit vom Schallpegel des Vorgängerstimulus (linke Spalte) und vom Lautheitsurteil über den Vorgängerstimulus (rechte Spalte), beispielhaft dargestellt für die Probanden 57, 59 und 66

sollen die erwähnten Eigenschaften in Abb. 6.12 am Beispiel der Geraden des Pegels 90 dB (dunkelblaue Punkte und Regressionsgerade) in der rechten Spalte erläutert werden. Jede der drei Geraden hat eine andere Steigung und somit eine grundlegend andere Aussage.

Bei Proband 57 hat die Gerade eine Steigung von  $-0,096710$ . Dies entspricht einer negativen Vorgängerabhängigkeit. Allerdings ist der p-Wert mit  $p = 0,055694$  gerade eben an der Grenze zum 5%-Signifikanzniveau.

Proband 59 hat eine Steigung von  $0,123871$ . Diese ist mit  $p = 0,050953$ , gerade eben noch signifikant. Es besteht also eine positive Abhängigkeit von Lautheitsurteil und Vorgängerurteil.

Bei Proband 66 verläuft die Gerade mit einer Steigung von  $-0,007247$  ( $p = 0,459813$ ) annähernd horizontal. Das bedeutet, dass keine Vorgängerabhängigkeit besteht.

Der entscheidende Faktor, der eine Aussage über die Vorgängerabhängigkeit macht, ist also die Steigung der Geraden. Mit den Steigungen können nun weiterführende Berechnungen betrieben werden. Pro Proband existieren 130 Steigungen (13 Geraden pro Einzelgraphik  $\times$  5 Urteilsmethoden  $\times$  2 Spalten (Vorgängerurteil und Vorgängerpegel)). Bei 25 Probanden ergibt das insgesamt 3250 Geraden und Steigungen, die je nach Fragestellung in die Auswertung einfließen.

In den folgenden Kapiteln werden die Vorgängerabhängigkeiten - nach Proband, Urteilsmethode und Pegel aufgeschlüsselt - dargestellt.

### 6.8.1 Effekt des Vorgängerpegels und des Vorgängerurteils in Abhängigkeit vom Probanden

In den beiden folgenden Abbildungen werden die Vorgängerabhängigkeiten nach Probanden aufgeschlüsselt dargestellt. Für jeden Probanden werden die Steigungen der Regressionsgeraden aller fünf Urteilsmethoden zusammengefasst und als Box-und-Whisker-Plot dargestellt. Abb. 6.13 zeigt die Urteilsabhängigkeit von *Vorgängerpegel* für jeden Probanden. Die Einkerbung gibt wieder das 95%-Konfidenzintervall an. Für eine übersichtlichere Darstellung wurden die Probanden aufsteigend nach dem jeweiligen Median angeordnet. Die Mediane liegen durchweg über Null. Allerdings muss man beachten, dass bei manchen Probanden das Konfidenzintervall die Null miteinschließt (siehe z.B. Probanden 62 und 63). Deshalb sind die Mediane nicht durchgängig sicher von Null zu unterschei-

den. Für fast alle Probanden lässt sich aber sagen, dass der Median signifikant über Null liegt und somit eine positive Vorgängerpegelabhängigkeit besteht. Im Überblick scheinen alle Probanden auf einem ähnlichem, eher niedrigen Niveau zu liegen. Die Probanden 51 und 61 stechen durch etwas höhere Werte leicht hervor und werden deshalb ganz rechts abgebildet.

Abb. 6.14 schlüsselt die Urteilsabhängigkeit vom Vorgängerurteil nach Probanden auf. Die Probanden sind wiederum nach aufsteigenden Medianen angeordnet. Alle Mediane liegen hier deutlich oberhalb der Nulllinie. Auch die Einkerbungen der Box-Plots beinhalten die Null nicht, so dass alle Mediane auch sicher über Null liegen. Auffallend ist, dass die Werte in dieser Abbildung deutlich höher liegen als bei der Vorgängerpegelabhängigkeit in Abb. 6.13. Dies verdeutlicht eine stärkere Abhängigkeit der Lautheitsurteile vom Vorgängerurteil als vom Vorgängerpegel. Proband 62 scheint sich wie zuvor am wenigsten vom Vorgänger beeinflussen zu lassen. Das Konfidenzintervall grenzt fast an die Nulllinie. Proband 61 dagegen zeigt wieder eine stark ausgeprägte Vorgängerabhängigkeit.

### 6.8.2 Effekt des Vorgängerpegels und des Vorgängerurteils in Abhängigkeit von der Urteilsmethode

In diesem Kapitel werden die Vorgängereffekte bezogen auf die verschiedenen Urteilsmethoden dargestellt. Für jede Urteilsmethode wurden die zugehörigen Geradensteigungen aller Probanden zusammengefasst. Abb. 6.15 zeigt die Urteilsabhängigkeit vom Vorgängerpegel in  $\text{hel/dB}$  zu jeder Urteilsmethode. Wie man sieht, sind die Methoden einander sehr ähnlich. Im Vergleich scheint nur Methode D eine etwas stärker ausgeprägte Vorgängerabhängigkeit aufzuweisen. Alle Methoden zeigen jedoch einen signifikant über Null liegenden Medianwert an. Die den Box-und-Whisker-Plots zugrundeliegenden Zahlenwerte, sowie die Mittelwerte, die Standardabweichungen und die MAD gibt Tabelle 6.8 wieder. In Tab. 6.9 werden die Urteilsmethoden statistisch miteinander verglichen. Es bestätigt sich die aus dem Box-Plot vermutete Tatsache, dass sich einzig und alleine Methode D statistisch signifikant von den anderen Methoden unterscheidet. Allerdings liefert der Vergleich der Methoden C und D nur eine Signifikanz auf dem 10 % Niveau. Der Vergleich der Methoden B und C ergibt ebenfalls einen p-Wert von knapp über 0,05.

In Abb. 6.16 werden die Urteilsabhängigkeiten vom Vorgängerurteil in  $\text{hel/hel}$

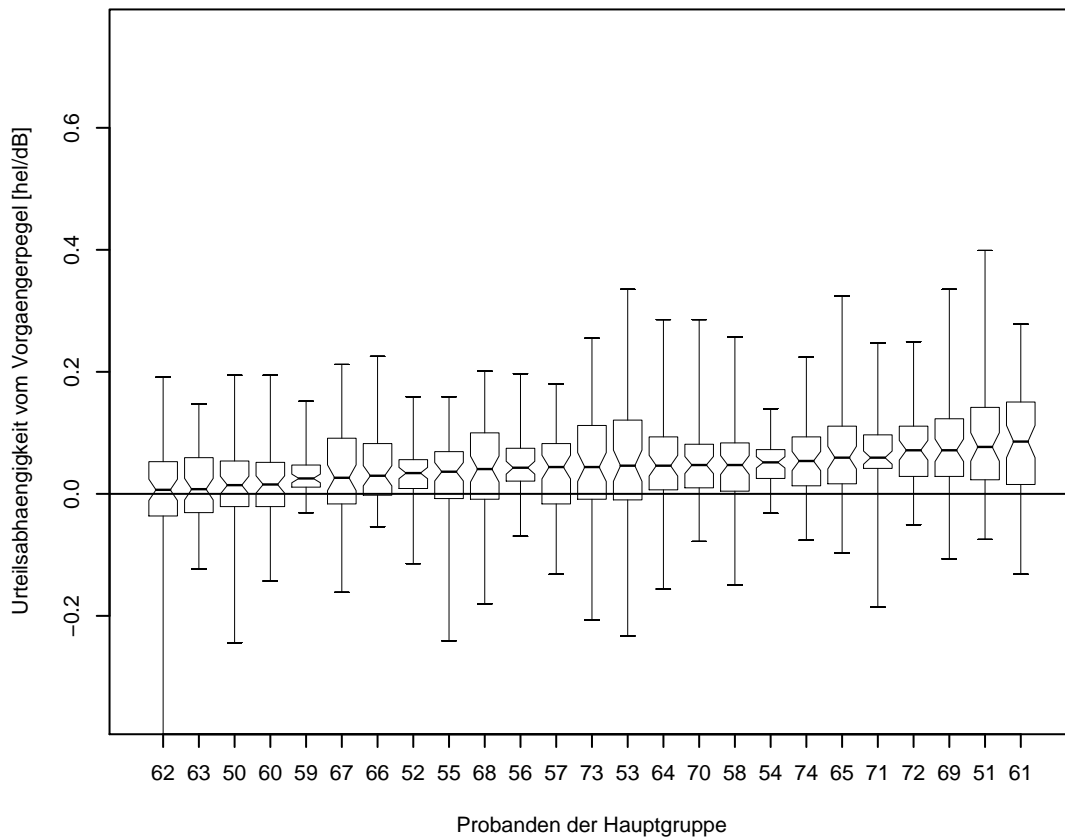


Abbildung 6.13: Lautheitsurteile aller Probanden in Abhängigkeit vom Schallpegel des Vorgängerstimulus, für alle Urteilsmethoden zusammengefasst (nicht normiert)

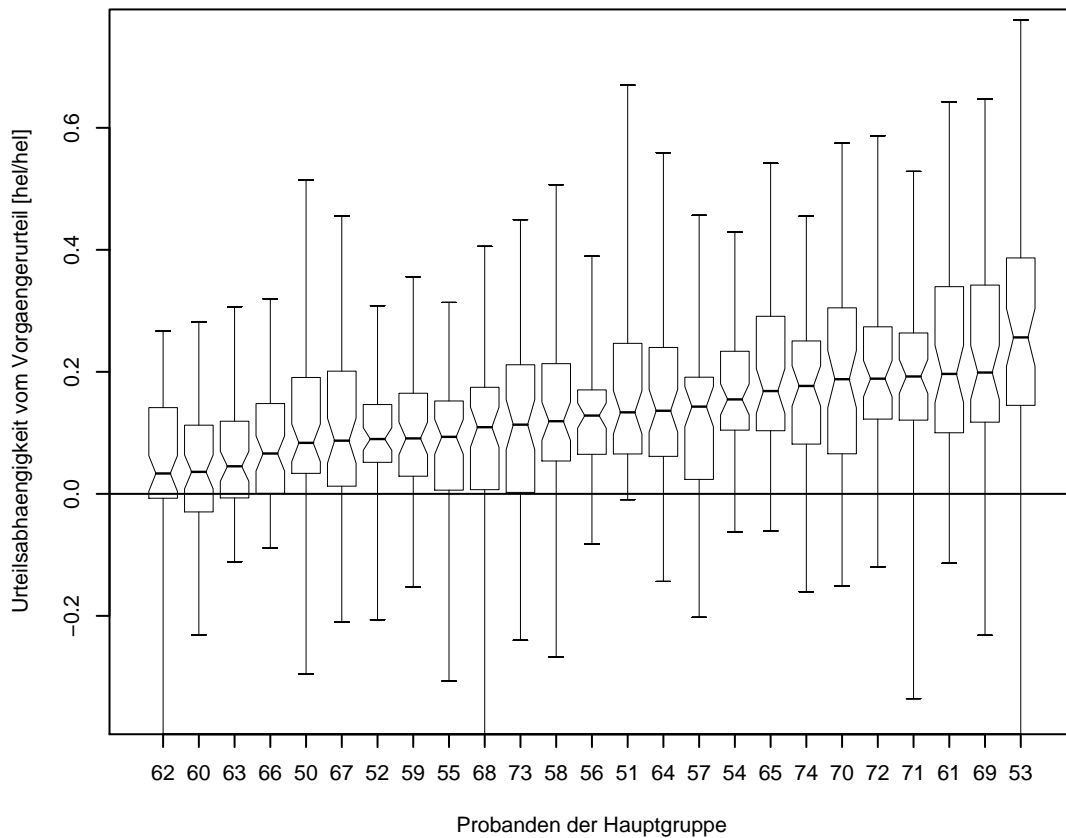


Abbildung 6.14: Lautheitsurteile aller Probanden in Abhängigkeit vom Lautheitsurteil über den Vorgängerstimulus, für alle fünf Urteilsmethoden zusammengefasst

Tabelle 6.8: Statistische Werte der Vorgängerpegelabhängigkeiten je nach Urteilmethode. (zur Erläuterung der Kennwerte s. Tab. 6.1)

Methode	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
A	0,043	0,077	-0,316	0,005	0,041	0,087	0,286	0,059
B	0,038	0,073	-0,177	-0,012	0,036	0,084	0,247	0,072
C	0,049	0,078	-0,277	0,004	0,042	0,096	0,399	0,065
D	0,058	0,079	-0,395	0,014	0,054	0,100	0,335	0,064
E	0,040	0,074	-0,244	0,002	0,038	0,082	0,273	0,057

den Urteilmethoden zugeordnet. Die Mediane und sogar alle ersten Quartile liegen über der Nulllinie. Alleine durch die visuelle Betrachtung lässt sich vermuten, dass die Methoden A, B und E gleich sind, während Methode D sich von den anderen abhebt. Die Methode C lässt sich rein visuell nicht eindeutig zuordnen. Die wichtigsten Zahlenwerte sind in Tab. 6.10 angegeben.

Wie in Abschnitt 6.8.1 ([Vorgängereffekte in Abhängigkeit vom Probanden](#), S. 64) auch, liegen die Werte der Vorgängerurteilsabhängigkeit deutlich höher als die Werte der Vorgängerpegelabhängigkeit. Dies unterstreicht ein zweites Mal den größeren Einfluss des Vorgängerurteils auf das Skalierungsverhalten des Probanden. Den statistischen Vergleich der Urteilmethoden bezogen auf die Vorgängerurteilsabhängigkeit gibt Tabelle 6.11 wieder. Hinsichtlich Methode D ergibt sich dasselbe Ergebnis wie beim Vergleich der Vorgängerpegelabhängigkeit (Tab. 6.9). Sie unterscheidet sich - wieder mit Ausnahme von Methode C - signifikant von allen anderen Urteilmethoden. Der Vergleich der Methoden C und D liegt wiederum ganz knapp über dem 5 % Signifikanzniveau. Anders als bei der Vorgängerpegelabhängigkeit unterscheiden sich auch die Methoden B und C signifikant auf dem 5 % Niveau voneinander. Die durch visuelle Betrachtung gewonnene Vermutung, dass die Methoden A, B und E gleich sind, bestätigt sich.

### 6.8.3 Effekt des Vorgängerpegels und des Vorgängerurteils in Abhängigkeit vom Pegel

Abbildung 6.17 zeigt die Urteilsabhängigkeit vom Vorgängerpegel (in  $\text{hel/dB}$ ) aufgetragen über dem jeweiligen Pegel. Tabelle 6.12 führt die entsprechenden Zahlenwerte auf. Fast alle Mediane liegen signifikant über Null. Für diese Pegel

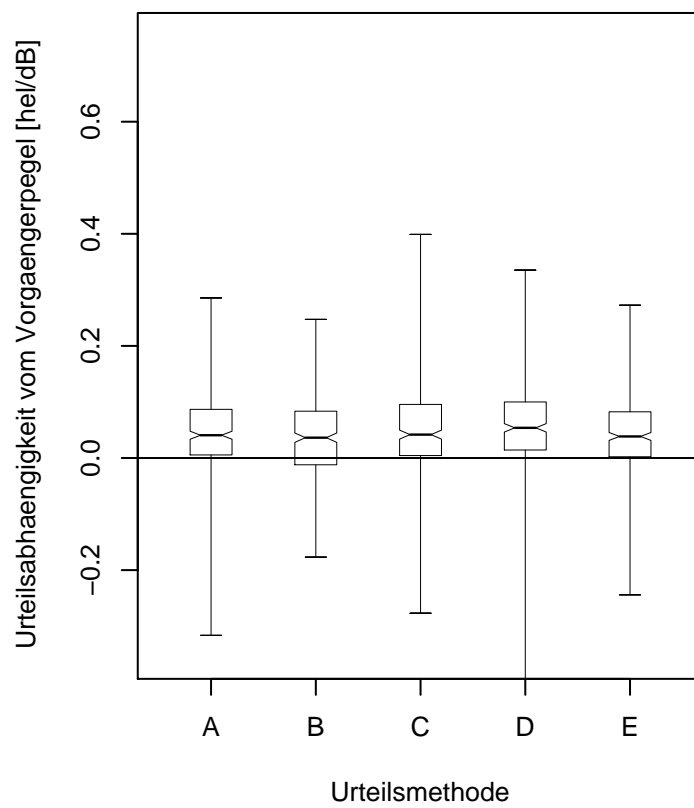


Abbildung 6.15: Lautheitsurteile in Abhängigkeit vom Schallpegel des Vorgängerstimulus, je nach Urteilsmethode

Tabelle 6.9: Vergleich der Vorgängerpegelabhängigkeiten je nach Urteilsmethode

Stimuli	p-Wert
A — B	0,200255165279
A — C	0,658805606338
A — D	0,019056762580
A — E	0,688389873870
B — C	0,075721663616
B — D	0,000414256171
B — E	0,324493464376
C — D	0,079586194250
C — E	0,374466161221
D — E	0,004957341341

Tabelle 6.10: Statistische Werte der Vorgängerurteilsabhängigkeiten je nach Urteilsmethode (zur Erläuterung der Kennwerte s. Tab. 6.1)

Methode	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
A	0,127	0,139	-0,382	0,052	0,120	0,212	0,621	0,115
B	0,115	0,141	-0,362	0,021	0,109	0,198	0,544	0,132
C	0,145	0,149	-0,336	0,039	0,130	0,231	0,670	0,142
D	0,169	0,166	-0,582	0,054	0,157	0,257	0,777	0,152
E	0,127	0,136	-0,497	0,047	0,118	0,215	0,566	0,121



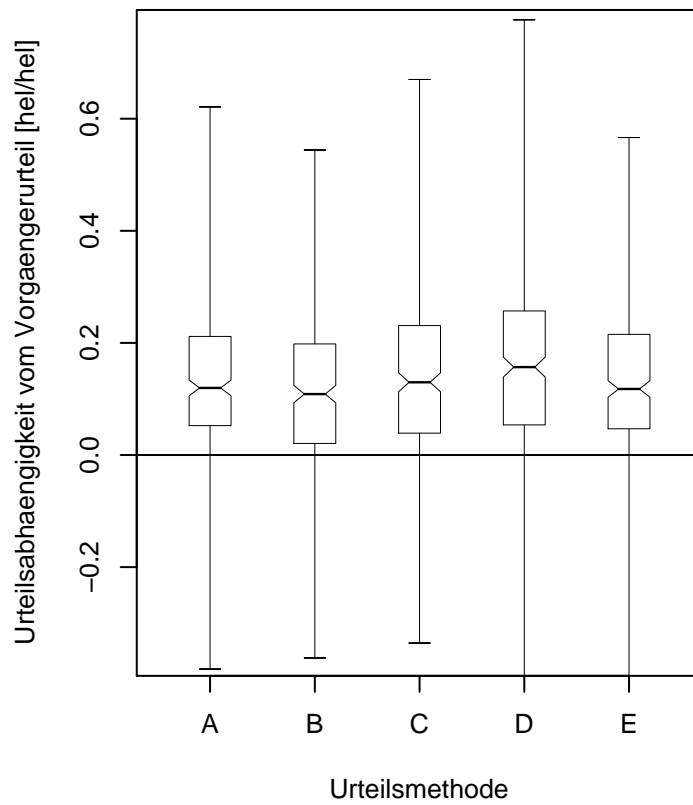


Abbildung 6.16: Lautheitsurteile in Abhängigkeit vom Lautheitsurteil über den Vorgängerstimulus, je nach Urteilsmethode

Tabelle 6.11: Vergleich der Vorgängerurteilsabhängigkeiten je nach Urteilsmethode

Stimuli	p-Wert
A — B	0,225873133624
A — C	0,293039338570
A — D	0,003513575467
A — E	0,800469962525
B — C	0,025722966593
B — D	0,000057375373
B — E	0,130256398797
C — D	0,060181902962
C — E	0,437642682196
D — E	0,007560415564

belegt dies eine positive Urteilsabhängigkeit vom Vorgängerpegel. Auffällig sind jedoch die geringen Werte an den Randbereichen (30 dB und 90 dB) sowie bei 50 dB und 80 dB. Bei diesen vier Schallpegeln ist die Steigung der Regressionsgeraden statistisch nicht von Null verschieden. Dies lässt sich schon rein visuell erahnen, wenn man sich die Übersichten der einzelnen Probanden im Anhang B (S.132 ff.) betrachtet. Graphisch besonders deutlich sind die horizontalen Geraden der entsprechenden Pegel in der Zusammenfassung aller Probanden (s. Abb. 6.20, mittlere Spalte) zu erkennen.

In Abb. 6.18 wird analog zu Abb. 6.17 die Urteilsabhängigkeit vom Vorgängerurteil (dimensionslos in „hel/hel“) über die jeweiligen Pegel aufgetragen. Tabelle 6.13 gibt die Zahlenwerte wieder. Die Position der einzelnen Box-Plots ähnelt stark denen aus Abb. 6.17. So sind sowohl die Mediane als auch die Mittelwerte der randständigen Pegel nahezu Null. Und auch die Werte der Pegel 50 dB und 80 dB liegen deutlich niedriger als die der anderen Pegel.

Im Vergleich zeigen beide Graphiken einen einander sehr ähnlichen Verlauf der Box-Plots. Nur liegen die Werte in der Darstellung der Vorgängerurteilseffektes höher. Dass der Effekt des Vorgängerurteils auf das Lautheitsurteil größer ist als der des Vorgängerpegels hat sich bereits in den vorhergehenden Kapiteln wiederholt gezeigt und bestätigt sich hier nochmals.

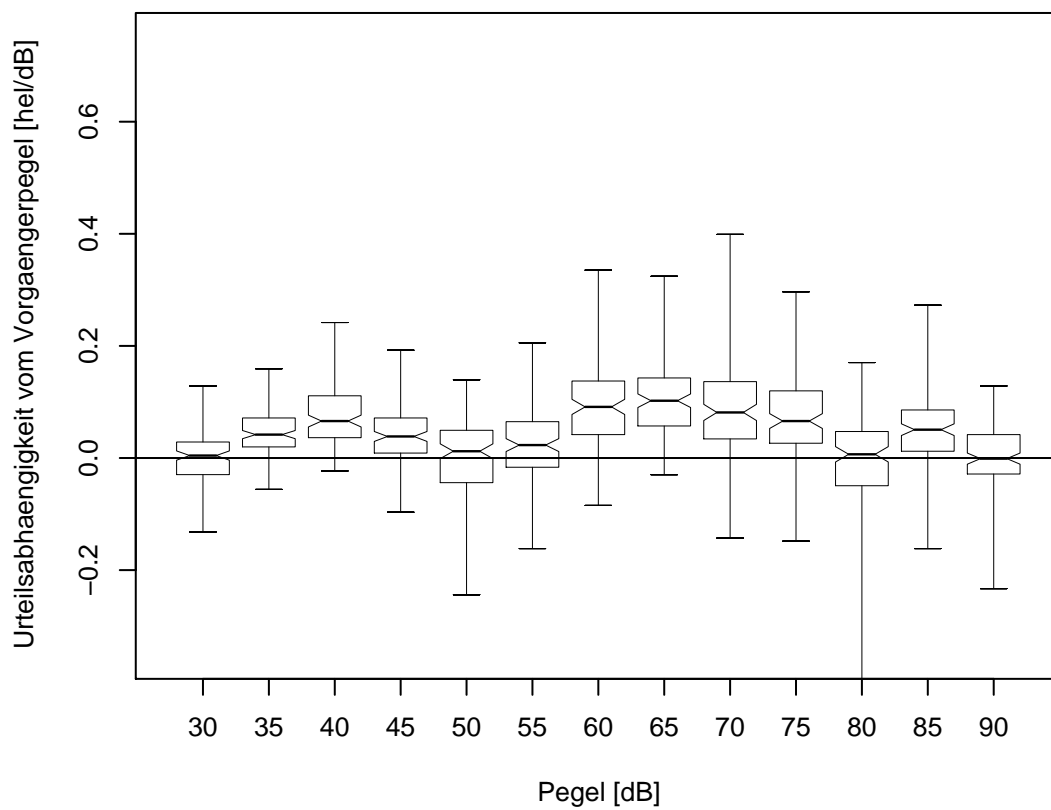


Abbildung 6.17: Lautheitsurteile in Abhängigkeit vom Schallpegel des Vorgängerstimulus, aufgeschlüsselt nach Pegeln

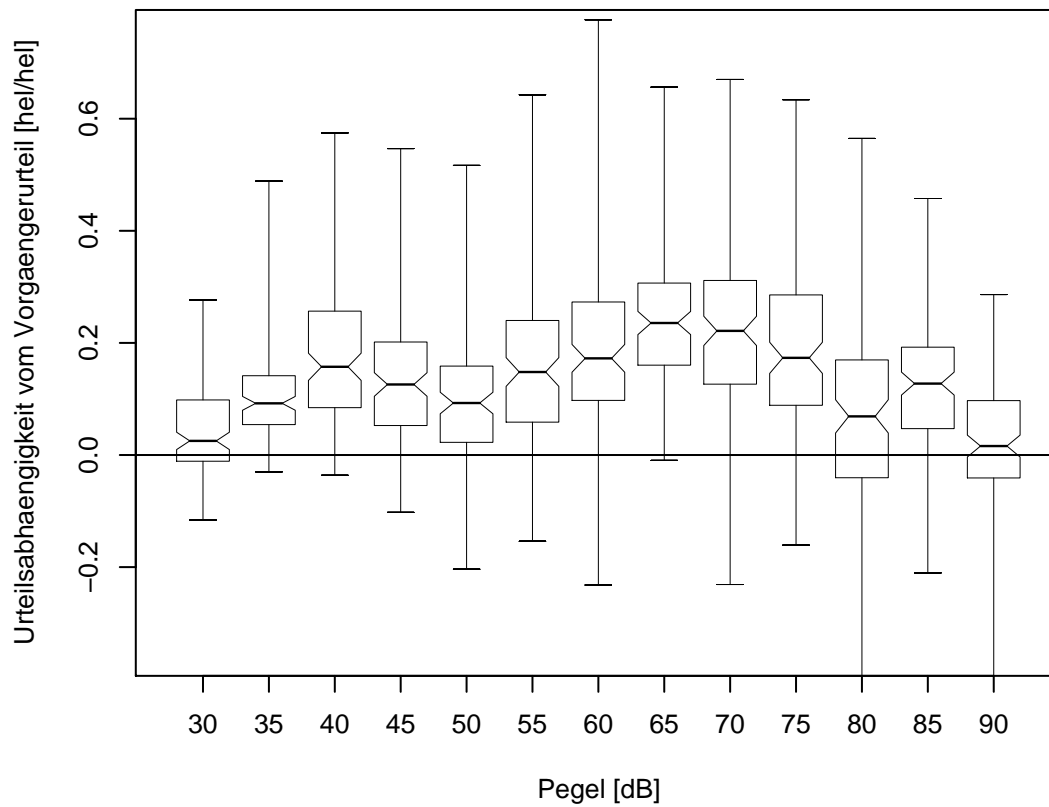


Abbildung 6.18: Lautheitsurteile in Abhängigkeit vom Lautheitsurteil über den Vorgängerstimulus, aufgeschlüsselt nach Pegeln

Tabelle 6.12: Statistische Werte der Vorgängerpegelabhängigkeiten je nach Schallpegel

Pegel	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
30	-0,001	0,042	-0,132	-0,030	0,004	0,029	0,129	0,047
35	0,048	0,043	-0,056	0,020	0,042	0,071	0,159	0,034
40	0,076	0,052	-0,023	0,036	0,066	0,111	0,242	0,049
45	0,042	0,051	-0,097	0,009	0,038	0,071	0,192	0,049
50	0,002	0,067	-0,244	-0,044	0,012	0,049	0,140	0,073
55	0,018	0,066	-0,162	-0,016	0,023	0,065	0,205	0,062
60	0,092	0,072	-0,085	0,042	0,091	0,137	0,335	0,070
65	0,109	0,068	-0,030	0,057	0,102	0,143	0,324	0,065
70	0,090	0,082	-0,143	0,034	0,081	0,136	0,399	0,080
75	0,072	0,081	-0,148	0,026	0,066	0,120	0,297	0,075
80	-0,007	0,095	-0,395	-0,049	0,007	0,047	0,170	0,072
85	0,052	0,065	-0,162	0,012	0,051	0,086	0,273	0,054
90	-0,001	0,058	-0,233	-0,029	-0,001	0,042	0,129	0,051

Tabelle 6.13: Statistische Werte der Vorgängerurteilsabhängigkeiten je nach Schallpegel

Pegel	Ø	Std.	Min.	1. Q.	Med.	3. Q.	Max.	MAD
30	0,043	0,079	-0,116	-0,011	0,025	0,098	0,277	0,072
35	0,112	0,094	-0,030	0,054	0,092	0,141	0,489	0,068
40	0,185	0,129	-0,036	0,084	0,157	0,257	0,575	0,119
45	0,140	0,119	-0,102	0,053	0,126	0,202	0,547	0,111
50	0,096	0,118	-0,204	0,023	0,093	0,159	0,517	0,104
55	0,154	0,146	-0,154	0,059	0,148	0,240	0,643	0,135
60	0,190	0,141	-0,232	0,097	0,172	0,273	0,777	0,129
65	0,241	0,124	-0,010	0,160	0,235	0,307	0,656	0,111
70	0,225	0,157	-0,231	0,126	0,221	0,311	0,670	0,140
75	0,188	0,151	-0,161	0,089	0,173	0,286	0,634	0,155
80	0,059	0,183	-0,582	-0,041	0,069	0,170	0,565	0,159
85	0,123	0,115	-0,210	0,047	0,127	0,192	0,458	0,107
90	0,022	0,127	-0,497	-0,041	0,016	0,097	0,286	0,101

### 6.8.4 Effekt der Medianlage der Urteile

Nachdem sich sowohl für den Vorgängerpegeleffekt als auch für den Vorgängerurteilseffekt eine positive Abhängigkeit zeigte, erschien es sinnvoll, die Korrelation der Vorgängereffekte mit der Medianlage der Urteile eines jeden Probanden zu untersuchen. Weiterhin sollte verifiziert werden, ob zwischen Vorgängerurteilseffekt und Vorgängerpegeleffekt tatsächlich ein positiver Zusammenhang besteht, wie er sich in Kapitel 6.8.1 (Vorgängereffekte in Abhängigkeit vom Probanden, S. 64) und Kapitel 6.8.2 (Vorgängereffekte in Abhängigkeit von der Urteilsmethode, S. 65) vermuten lässt.

Da die in den Versuchen erhobenen Urteile nicht zwingend auf einem Intervallskalenniveau liegen, wurde zur Berechnung der Korrelationen der nicht-parametrische Spearman-Rangkorrelationskoeffizient herangezogen. Im Gegensatz zum häufiger verwendeten Pearson-Korrelationskoeffizienten, der eine lineare Verteilung auf mindestens Intervallskalenniveau voraussetzt, genügt bei der Berechnung nach Spearman eine nichtlineare Verteilung auf Ordinalskalenniveau. Der Vergleich Vorgängerpegeleffekt - Urteilsmedian ergab einen Korrelationskoeffizienten von  $r = -0,279$ , der Vergleich Vorgängerurteilseffekt - Urteilsmedian einen Koeffizienten von  $r = -0,263$ . Beide sind somit negativ, können aber aufgrund der geringen Werte nicht als aussagekräftig angesehen werden. Ein Zusammenhang zwischen Lage der Urteile und Vorgängereffekten scheint somit nicht zu existieren. Dies ließ sich nach dem visuellen Vergleich der einzelnen Probanden jedoch auch nicht vermuten und deckt sich deshalb mit den Erwartungen. Im Gegensatz dazu lässt sich beim Vergleich von Vorgängerurteilseffekt (Abb. 6.14) und Vorgängerpegeleffekt (Abb. 6.13) bereits visuell eine hohe, positive Korrelation vermuten. Der berechnete Korrelationskoeffizient von  $r = 0,875$  bestätigt dies.

## 6.9 Überblick über Vorgängereffekte

Sozusagen als Quintessenz werden in Abb. 6.19 alle im Versuch erhobenen Daten in Hinsicht auf Untersuchung der Vorgängerabhängigkeiten zusammengefasst. Die linke Spalte stellt die Abhängigkeit vom Vorgängerpegel, die rechte Spalte die Abhängigkeit vom Vorgängerurteil dar. In beiden Teilgraphiken wurden alle fünf Urteilsmethoden und alle 25 Probanden zusammengefasst. Deutlich sind die positiven Steigungen der Geraden zu erkennen, wobei die Geraden im rech-

ten Schaubild erkennbar steiler verlaufen. Dies verdeutlicht den stärkeren Effekt des Vorgängerurteils auf die Lautheitsbeurteilung, wie es bereits in Kapitel 6.8 (Vorgängerpegel und -urteilsabhängigkeit, S. 58 ff.) mehrmals aufgezeigt wurde. Außerdem sieht man, dass die Geraden der Pegel 30 dB, 50 dB, 80 dB und 90 dB zumindest auf der linken Seite flacher bzw. annähernd horizontal verlaufen (vgl. Abschnitt 6.8.3 (Vorgängereffekte in Abhängigkeit vom Pegel, S. 68)). Aber auch bei diesen Pegeln ist der Verlauf rechts deutlich steiler.

Eine etwas detaillierte Zusammenfassung der Rohdaten gibt Abb. 6.20. Die Graphik ist gleich aufgebaut wie die Übersichten für jeden einzelnen Probanden im Anhang B (S. 105ff). Die einzelnen Zeilen geben die Urteilmethoden von A nach E aufsteigend an. Für jede Urteilmethode wurden die Urteile aller Probanden zusammengefasst.

Spalte 1 gibt die Pegel-Lautheitsfunktion wieder. Schön zu erkennen ist der bei allen Urteilmethoden sehr ähnliche Verlauf der Kurve. Auch den bereits aus Abb. 6.8 bekannten Knick zwischen 60 dB und 70 dB ist in jeder Zeile gut zu sehen. Spalte 2 zeigt die Abhängigkeit vom Vorgängerpegel. Bei allen fünf Methoden sind die Geraden der Pegel 30 dB, 50 dB, 80 dB und 90 dB wieder annähernd horizontal, während die anderen Geraden eine positive Steigung aufweisen. Spalte 3 zeigt die Abhängigkeit vom Vorgängerurteil. Alle Geraden verlaufen deutlich steiler als in Spalte 2 und bezeugen einen positiven Zusammenhang zwischen Vorgängerurteil und Lautheitsurteil. Auch hier weisen die Geraden der bei Spalte 2 genannten Pegel eine geringere Steigung auf als die übrigen Geraden. Auffällig ist die Konvergenz der Geraden der niedrigeren Pegel für kleine Vorgängerurteile; besonders fällt dies bei den Methoden C, D und E auf. Auf den ersten Blick nicht so deutlich, aber doch erkennbar, zeigt sich dies „fächerförmige“ Verlauf bereits in Abb. 6.19.

Die nach Urteilmethode getrennte Betrachtung in Abb. 6.20 bringt im Vergleich mit der Zusammenfassung in Abb. 6.19 also wenig neue Erkenntnisse. Es scheint so, als ob die verschiedenen Urteilmethoden wenig Einfluss auf Vorgängerabhängigkeiten nehmen.

## 6.10 Verhalten und Aussagen der Probanden

Jeder Proband ging anders mit den Tests um und hatte verschiedene Fragen und unterschiedliche Probleme im Testdurchlauf. So gab es Probanden, die kaum

## 6 Ergebnisse

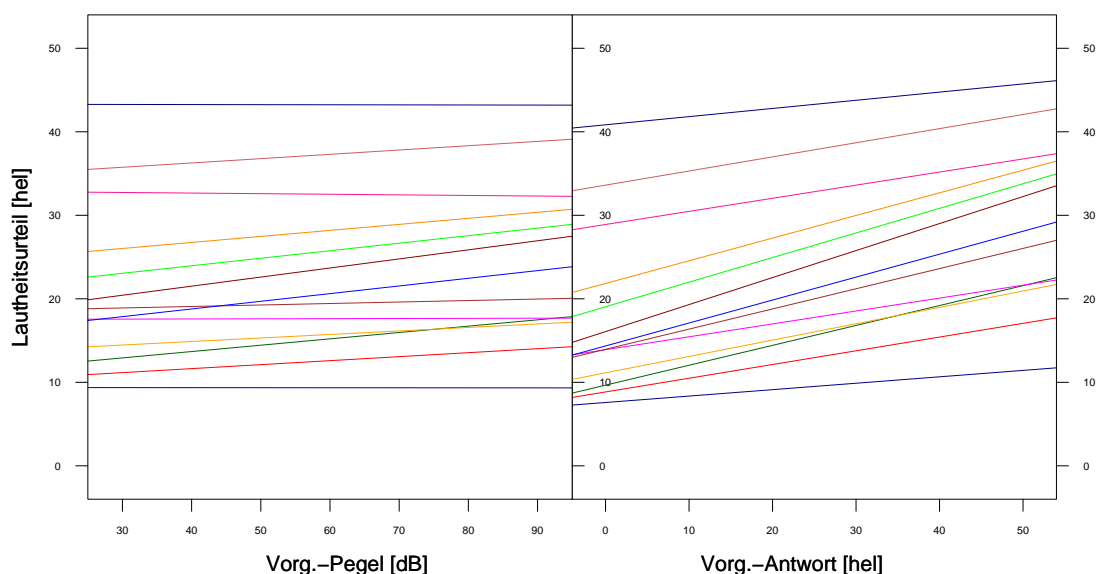


Abbildung 6.19: Abhängigkeit vom Vorgängerpegel (links) und vom Vorgängerurteil (rechts), alle Probanden und alle Urteilsmethoden zusammengefasst. Die Steigungen sind normiert und deshalb miteinander vergleichbar.

nachfragten und auch wenig Schwierigkeiten mit den einzelnen Urteilsmethoden hatten. Andere Probanden dagegen schienen öfters an ihren eigenen Urteilen zu zweifeln und hätten ihr Urteil nach Abgabe am liebsten nochmals korrigiert.

Schon vor Beginn des Messdurchlaufes mit Methode D äußerten viele Probanden Bedenken dahingehend, dass sie ohne eine sichtbare Skala Probleme mit der Beurteilung der Stimuli haben könnten. Speziell diejenigen Probanden, die mit Methode D beginnen mussten, hatten anfangs Schwierigkeiten ein Urteil zu fällen. Mehrere Probanden sagten nach der Durchführung, dass sie sich den Wert 25 hel (also die Mitte der Heller-Skala) als Ankerpunkt gewählt hätten. Von diesem Ankerpunkt ausgehend hätten sie dann ihr Urteil bestimmt. Dieses Vorgehen ist nicht ungewöhnlich. Bereits 1964 veröffentlichte [Helson \(1964\)](#) seine Theorie, dass ein internes Bezugssystem nicht an der absoluten Reizschwelle verankert ist, sondern von einem Mittelpunkt aus in beide Richtungen aufgespannt ist. Er bezeichnete diese Theorie als Adaptationsniveau-Theorie. Der Mittelpunkt entspricht demjenigen Wert, an den der Proband adaptiert ist. Auch [Sarris \(1971\)](#) stimmt hiermit überein.

Einige Probanden bemerkten nach Durchlesen der Probandeninstruktion, dass sie ein absolutes Urteilen, ohne sich vom Vorgängerstimulus beeinflussen zu las-



Alle Probanden Hauptgruppe

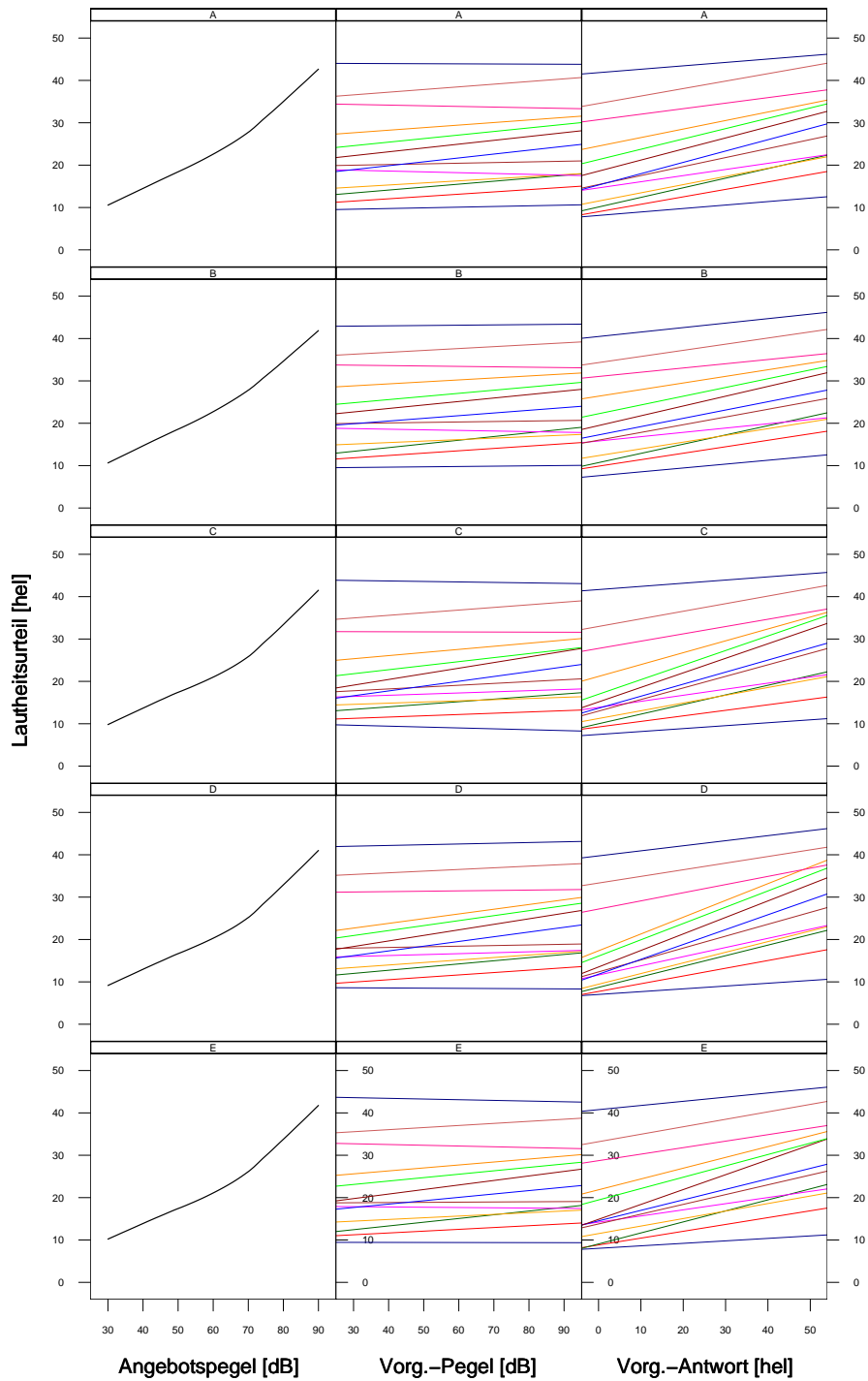


Abbildung 6.20: Pegel-Lautheitsfunktion, Abhängigkeit vom Vorgängerpegel und Abhängigkeit vom Vorgängerurteil für jede Urteils-methode, alle Probanden zusammengefasst.

sen, für sehr schwer bis unmöglich hielten. Obwohl den Probanden nicht explizit erklärt wurde, dass es in dieser Arbeit um mögliche Vorgängerabhängigkeiten geht, äußerten sie also schon den Verdacht, dass sie sich durch den Vorgängerstimulus beeinflussen lassen könnten bzw. würden. Befragt nach ihrer eigenen Vermutung, ob eher eine positive oder negative Vorgängerabhängigkeit bestünde, tippten die meisten, aber nicht alle Probanden - wie Nägle (2004) zu Beginn ihrer Untersuchung auch - auf einen negativen Zusammenhang.

Bei Methode E fiel auf, dass die Probanden relativ häufig erst ein Urteil mit dem Laserpointer anzeigten um dann, nach kurzer Überlegung, noch um einige „hel“ mehr oder weniger nachzujustieren. Dieses Korrigieren des Urteils war aufgrund des Designs bei den Methoden A und B nicht möglich, wurde aber auch bei den Methoden C und D deutlich seltener beobachtet.

Jeder der Probanden verlangte in der Camera silens nach einer nochmaligen Erklärung des Versuchsablaufes und ließ sich den Umgang mit dem Skalieretablett und den anderen Utensilien nochmals zeigen. Die Übungssequenz war allen sehr hilfreich, wobei den meisten Probanden das einmalige Vorspielen der Übungssequenz genügte, um sich in der Skalierung sicher zu fühlen. Nur wenige Probanden wollten die Übungssequenz ein zweites Mal hören.

Viele Probanden empfanden den Test als langwierig und ermüdend. Deshalb konnte jeder Proband zwischen den einzelnen Durchläufen Pausen machen. Die meisten Probanden machten nach dem dritten oder vierten Durchlauf eine kurze Pause, während der sie die Camera silens verließen.

Die Camera silens selbst wurde von den Probanden unterschiedlich aufgenommen. Die meisten waren von der Verkleidung der Wände beeindruckt und schauten sich diese erst einmal genauer an, bevor sie sich dem eigentlichen Versuchsaufbau widmeten. Allerdings gab es auch Probanden, die den Raum als bedrückend und beängstigend empfanden und sich darin nicht wohl fühlten. Viele Probanden empfanden auch die absolute Stille, gepaart mit den Unterbrechungen durch das teilweise als unangenehm empfundene CCITT-Rauschen auf Dauer als belastend.

All diese probandenspezifischen Verschiedenheiten lassen sich durch den Versuchsaufbau nur schwer beeinflussen, geschweige denn können sie in der Auswertung berücksichtigt werden. Trotzdem besteht die Möglichkeit, dass sie das Urteilsverhalten des Probanden beeinflussen.

## 7 Diskussion

### 7.1 Visuelle Betrachtung des Urteilsbereichs und der Urteilsdichte

Die interindividuellen Unterschiede zwischen den Probanden können enorm sein und sich dann auch bereits durch rein visuelle Betrachtung deutlich erkennen lassen. Anhand der drei Beispielprobanden (s. Abschnitt [6.2, Visuelle Betrachtung der Rohdaten](#), S. 41 ff.) wurden verschiedene Skalierungsweisen veranschaulicht. Auch bei den restlichen Probanden (s. Anhang [A](#), S. 99 ff.) zeigen sich diese Abweichungen. Die größten Unterschiede sind zu erkennen in der

- Ausnutzung des Skalenumfangs
- Streuung der Urteile
- Kategorisierung der Urteile bzw. Verteilungsdichte

Proband 57 skaliert z. B. bei den Methoden C und D nur in mehreren Grobkategorien und lässt die dazwischenliegenden Antwortmöglichkeiten fast völlig außer Acht. Erkennbar ist das an der ausgeprägten Querstruktur in [Abb. 6.1](#). Die Urteile beschränken sich quasi auf Fünferschritte. Auch in der Urteilsdichteverteilung ([Abb. 6.2](#)) spiegelt sich diese Grobkategorisierung wieder. Würde dieser Effekt nur bei Methode D auftreten, könnte man vermuten, dass der Proband Probleme gehabt hätte, sich die Skala vorzustellen. Jedoch war bei Methode C die Skala sichtbar und die Kategorisierung hätte dann nicht so stark ausfallen dürfen. Eine andere Vermutung wäre, dass der Proband Schwierigkeiten im Umgang mit der Skala gehabt hätte und sich diese im Laufe der Untersuchung durch Üben gebessert hätten. Eindeutig dagegen spricht jedoch, dass bei diesem Probanden die Methoden C und D am Schluss der Untersuchung durchgeführt wurden. Die Positionierung am Ende könnte allerdings Grund für eine Ermüdung des Probanden oder zunehmendes Desinteresse sein, was in der Folge zu diesem vereinfachten Urteilen geführt haben könnte. Eine plausible Erklärung für die ab-

weichende Skalierung bei den beiden Methoden mit verbaler Urteilsabgabe lässt sich allerdings nicht finden.

Beim intraindividuellen Vergleich zeigt sich im Gegensatz zum interindividuellen Vergleich eine deutliche Konstanz. Die meisten Probanden skalieren mit allen fünf Methoden sehr ähnlich. Es scheint so, als ob sich jeder Proband ein eigenes Urteilsmuster zurecht legt und dieses dann über den ganzen Versuchsablauf beibehält. Es gibt kaum Probanden, die bei den verschiedenen Methoden von ihrem Skalierungsmuster abweichen. Bei Proband 59 (Abb. 6.1) erkennt man sehr gut ein solches konstantes Muster. Es besteht vor allem darin, fast nur den Bereich zwischen 12 *hel* und 40 *hel* zu nutzen. Dies mag an einer ausgeprägten zentralen Tendenz des Probanden liegen. Das Phänomen der zentralen Tendenz ist in der Psychologie schon lange bekannt. Man versteht darunter „jene [...] Versuchspersoneneffekte, bei denen mittlere Bewertungen zur Charakterisierung einzelner Merkmale bevorzugt und extreme Bewertungen vermieden werden“ (Rau und Pauli, 1995, S. 14). Für die Kategorienskalierung von Lautheit bestätigten dies Payne et al. (1985). Bei Proband 59 fällt weiterhin bei allen Methoden ein kleiner Sprung zwischen der Beurteilung von 85 dB und 90 dB auf. Dies könnte daher rühren, „[...] daß bei hohen Pegeln die Unterschiedsschwelle bekanntermaßen kleiner ist als bei niedrigen. Der Unterschied beispielsweise zwischen 80 und 85 dB wird deutlicher wahrgenommen als beispielsweise der zwischen 50 und 55 dB [...]“ (Hellbrück et al., 1995).

Eine weniger konstantes Urteilsverhalten zeigt Proband 57. Bei ihm ist in den Methoden C und D eine Abweichung vom üblichen Muster zu sehen. Betrachtet man die Dichteverteilung in Abb. 6.2, erkennt man jedoch bei allen fünf Methoden die zumindest angedeuteten Doppelgipfel. Prinzipiell erkennt man ein Muster besser in den Dichteverteilungen als in den Punktwolken der Urteile. In den Dichteverteilungen sind teilweise Abweichungen vom Antwortmuster zu erkennen, die man in der Punktwolkendarstellung nicht erkennt.

Es ist selbstverständlich, dass nicht alle Probanden gleich skalieren. Wodurch die Abweichungen bedingt sind, lässt sich jedoch nicht leicht fassen. Die Bedingungen zur Teilnahme an den Untersuchungen wurden so gestellt, dass relevante und objektivierbare Unterschiede zwischen den Probanden möglichst klein gehalten wurden. Dazu gehören die Altersbeschränkung, die nachgewiesene Normalhörigkeit, die standardisierte Probandeninstruktion sowie der standardisierte Versuchsaufbau und -ablauf. Trotzdem scheint es weitere Einflussfaktoren zu geben, die vor allem der Persönlichkeit des Probanden zuzuordnen sind.

## 7.2 Urteilsunterschiede der Probanden

Es soll hier nicht auf jeden Probanden einzeln eingegangen werden, da Daten eines einzelnen Probanden keine Rückschlüsse auf allgemeingültige Erkenntnisse zulassen. In den vorhergehenden Kapiteln wurden bereits typische Auffälligkeiten am Beispiel einzelner Probanden beschrieben und mögliche Gründe dafür angeführt. Dass die Ergebnisse der einzelnen Probanden trotz weitgehender Reduzierung von äußeren Einflüssen stark voneinander abweichen können liegt in der Natur des Menschen. Während in der klinischen Diagnostik und Therapie die individuellen Voraussetzungen jedes Patienten berücksichtigt werden müssen, werden hier die Daten aller Probanden je nach Fragestellung zusammengefasst um einen möglichst allgemeingültigen Konsens herzustellen.

## 7.3 Auswirkungen der Urteilsmethoden auf das Urteil

Zur Rekapitulation werden die einzelnen Charakteristika der verschiedenen Urteilsmethoden in Tabelle 7.1 nochmals übersichtlich aufgezeigt.

Abbildung 6.7 und Tabelle 6.3 zeigen, dass sich die Methoden A und B sowie C und E gleichen. Alle anderen Paarungen der Urteilsmethoden sind signifikant voneinander verschieden. Explizit erwähnt sei, dass sich die Methode D von *allen* anderen Methoden unterscheidet.

Mit A und B wird generell lauter skaliert als mit C und E. Methode D liefert die kleinsten Urteile (s. Mittelwerte in Tab. 6.2).

Für den Vergleich der Methoden A und B, also der beiden Skaliertablettmethoden, bedeutet die statistische Gleichheit, dass die Strecke, die die Hand bzw. der Arm zurücklegen muss, keinen Einfluss auf die Absolutlage der Urteile nimmt.

Erstaunlich ist, dass die Methoden C und E gleiche Ergebnisse liefern, während die Methode D (die C sehr ähnlich ist) zu beiden signifikant verschieden ist.

Da bei den Methoden C und E das Poster als Referenz der Heller-Skala dient, liegt die Vermutung nahe, dass nicht die Urteilsabgabe selbst, sondern die Art und Weise der Präsentation der Skala relevanten Einfluss nimmt.

Obwohl im Versuchsaufbau auf gleiche räumliche Proportionen geachtet wurde (s. Abschnitt 3.6, [Mündliche Abgabe des Urteils](#), S. 19), scheinen diese doch das Urteilsverhalten der Probanden beeinflusst zu haben. Eventuell liegt die Ursache darin, dass die Probanden bei C und E den Kopf heben mussten um das

Poster zu betrachten und dass sie dabei eher auf die unteren Kategorien geschaut haben. Dieses vermehrt auf die unteren Kategorien gelegte Augenmerk könnte der Grund für die niedriger ausfallende Lautheitsskalierung in C und E sein. Diese Annahme bleibt jedoch reine Spekulation und kann hier leider nicht durch Fakten untermauert werden.

Die signifikante Verschiedenheit der Methoden A bzw. B zur Methode C steht im Gegensatz zu [Kießling et al. \(1993\)](#), die keine signifikanten Unterschiede zwischen der Ansage- und der Tablettmethode fanden. Allerdings benutzten sie anderes technisches Equipment. Im einen Durchgang wurde anstatt des Skaliertabletts ein LCD-Touchscreen benutzt und im anderen Durchgang hatte der Proband die Skala auf Papier in der Hand und nicht als Poster vor sich. Außerdem wurde eine 12-Punkte-Skala verwendet und die Urteile anschließend auf die 50-Punkte-Skala nach Heller umgerechnet. Diese verschieden konstruierten Versuchsaufbauten könnte eine Ursache für die gegensätzlichen Ergebnisse sein.

Es wäre aber auch denkbar, dass die gefundenen Unterschiede zwischen den Methoden A/B und C einzig und allein auf der verschiedenen Präsentation der Skala beruhen und von der restlichen technischen Ausstattung unbeeinflusst sind.

Warum aber fällt die Lautheitsskalierung mit Methode D dann noch niedriger aus als mit den anderen Methoden?

Hier wurde die Skala weder als Poster noch auf dem Tablett vorgegeben. Eine mögliche Erklärung bieten wiederum [Kießling et al. \(1993\)](#):

„Grundsätzlich ist jedoch festzustellen, daß alle Versuchspersonen dazu neigen, sich durch tendenziell niedrige Lautheitsurteile Handlungsspielraum im oberen Dynamikbereich zu verschaffen. Infolgedessen sind alle derartigen Verteilungen schief im Sinne einer Bevorzugung der unteren Lautheitskategorien. Diesem Verhalten kann durch die Vorschaltung eines Orientierungsdurchganges lediglich partiell entgegengewirkt werden.“

Denkbar ist, dass die Probanden, da sie nun keine sichtbare Skala vor sich haben noch stärker verunsichert sind und sich deshalb ein noch größeres Sicherheitspolster für den höheren Pegelbereich verschaffen wollen. Und zwar indem sie ihre Lautheitsskalierung nochmals weiter nach unten verschieben. Dass sich praktisch alle Probanden mit Methode D am unsichersten fühlten ging aus den

Tabelle 7.1: Überblick über die Charakteristika der verschiedenen Urteilsmethoden.

Urteilsmethode	Urteilsabgabe	Bewegung des Armes	Visuelle Repräsentation
A	Tablett	maximal	Tablett
B	Tablett	konstant	Tablett
C	mündlich	keine	Poster
D	mündlich	keine	keine
E	Laserpointer	minimal	Poster

Rückmeldungen klar hervor (s. Abschnitt 6.10, Verhalten und Aussagen der Probanden, S. 77).

## 7.4 Auswirkungen des Pegels auf das Urteil

Abbildung 6.8 zeigt die Pegel-Lautheitsfunktion zusammengefasst für alle Probanden und Urteilsmethoden. Die Kurve verläuft monoton konkav ansteigend, wobei sie ab einem Schallpegel von ca. 70 dB einen steileren Anstieg aufweist. Die in der Literatur angegebenen Pegel-Lautheitsfunktionen ergeben ähnliche Verläufe (Kollmeier, 1997, S. 47 ff.). Ebenso ist die Kurve annähernd mit der von Nägle (2004, S. 44) ermittelten Kurve identisch.

Die Urteile können auch als Box-und-Whisker-Plot über jedem Pegel dargestellt werden (s. Abb. 6.9). Für jede Anhebung des Pegels ergibt sich auch eine signifikante Erhöhung des Urteils. Dies beweist der Vergleich der benachbarten Pegel in Tab. 6.5 insbesondere auch für die Pegel 40 dB und 45 dB, deren Beurteilungen in der Boxplotdarstellung fast identisch aussehen. Die Probanden konnten also eine Erhöhung des Schallpegels zuverlässig als Lautstärkezuwachs erkennen und beurteilen. Dies überrascht insofern nicht, als bekannt ist, dass bereits Schallpegeldifferenzen von 1 dB sicher erkannt werden können (vgl. Schmidt und Schaible (2006, S. 310)).

Wie bei Nägle (2004) auch, trat die bei Hellbrück et al. (1995) erwähnte Nicht-Monotonität zwischen 60 dB und 65 dB nicht auf. Bei Hellbrück et al. (1995) zeigte sich bei diesen Pegeln - allerdings nur bei bestimmten Reizabfolgen - ein Abfall der Lautheitsempfindung, dessen Ursache als Sequenzeffekt im Rahmen der speziellen Reizabfolgen gedeutet wurde. Nägle (2004) ging davon aus, diese

Effekte durch die durchdachte Konstruktion der Reizserien eliminiert zu haben, was sich in der vorliegenden, ähnlich aufgebauten Arbeit bestätigt. Speziell die Eigenschaft, dass jeder Pegel jedem anderen vorangestellt ist, verringert das Ausmaß der Effekte der Darbietungsreihenfolge auf die Lautheitsskalierung.

### 7.5 Auswirkungen der Zeitposition der Urteilsmethoden

Betrachtet man die Zeitposition unabhängig von den Urteilsmethoden (Abb. 6.10 und Tab. 6.7) zeigt sich, dass zwischen den Positionen 2, 3 und 4 kein Unterschied besteht. Dies entspricht der Idealvorstellung eines konstanten Skalierungsverhalten seitens der Probanden.

An Zeitposition 1 ergibt sich eine zu allen anderen Positionen signifikant niedrigere Lautheitsskalierung. Dies bestärkt die bereits in Abschnitt 7.3 ([Auswirkungen der Urteilsmethoden auf das Urteil](#), S. 83 ff.) geäußerte Vermutung, dass die Probanden bei größerer Verunsicherung leiser skalieren, um sich für die größeren Pegel ein Sicherheitspolster nach oben zu schaffen. Dafür spricht auch, dass gerade bei Betrachtung der Zeitpositionseffekte für die Urteilsmethode D (Abb. 6.11) ein besonders ausgeprägtes Absinken bei großer Streuung der Lautheitsurteile an Position 1 beobachtet wird. Auch bei den Urteilsmethoden A, C und E fallen die Urteile an Position 1 am niedrigsten aus. Jedoch ist der Unterschied zu den anderen Zeitpositionen nicht so ausgeprägt wie bei Methode D. Hier scheint die Verunsicherung des ersten Durchganges mit der Verunsicherung durch die Urteilsmethode zusammenzufallen und sich zu einer noch tiefer absinkenden Lautheitsskalierung zu addieren.

An Zeitposition 5 erkennt man in Abb. 6.10 ein Ansteigen des Lautheitsempfindens. Allerdings lässt sich dieses Phänomen in den Auswertungen der einzelnen Urteilsmethoden nicht so gut erkennen wie es bei Position 1 der Fall war. Die Ursache für das lautere Einschätzen des CCITT-Rauschens könnte in einer Ermüdung der Probanden am Ende der Untersuchung liegen. Bereits [Schechter und Buchsbaum \(1973\)](#) zeigten, dass sich bei veränderter Aufmerksamkeit im EEG messbare Veränderungen der evozierten Potentiale feststellen lassen. Eine veränderte Aufmerksamkeit auf einen Stimulus führt also zu einer physiologisch messbaren Veränderung der Verarbeitung im zentralen Nervensystem. Mehrere Probanden berichteten, dass sie im Verlauf des Versuchs das Rauschen als zunehmend störend empfunden hätten. Dadurch könnte sich auch unbewusst die



Skalierung in Richtung lauterer Urteile verschoben haben.

Die großen Unterschiede innerhalb und zwischen den Urteilmethoden rühren wahrscheinlich weniger von Effekten der Zeitpositionen. Plausibler ist, dass die ausgeprägten probandenspezifischen Unterschiede den größeren Einfluss haben. Durch die kleine Zahl von nur fünf Probanden pro Zeitposition und den Umstand, dass für jede Zeitposition verschiedene Probanden herangezogen werden, können solche Schwankungen entstehen. Ein Indiz dafür ist auch, dass sich bei der urteilsmethodenunabhängigen Auswertung, bei der in jede Zeitposition alle 25 Probanden eingehen, signifikante und bei weitem nicht so extreme Schwankungen zeigen.

Die Betrachtung für die einzelnen Urteilmethoden bringt also keinen großen Fortschritt, da probandenspezifische Skalierungsdifferenzen die durch die Zeitposition bedingten Effekte überdecken.

### 7.6 Abhängigkeit der Urteile von Vorgängerpegel und Vorgängerurteil

In der Arbeit von [Nägle \(2004\)](#) wurde zunächst von einer negativen Vorgängerabhängigkeit ausgegangen. In der Auswertung ergab sich dann jedoch ein positiver Zusammenhang zwischen dem Vorgängerstimulus und seinem Nachfolger; und zwar sowohl beim Vorgängerpegel als auch beim Vorgängerurteil. Positiv heißt, dass ein mittlerer Stimulus bei lautem Vorgänger lauter und bei leisem Vorgänger leiser beurteilt wird, während bei negativer Abhängigkeit der mittlere Stimulus bei lautem Vorgänger leiser beurteilt wird und umgekehrt. Die positive Vorgängerabhängigkeit war eine unerwartete Erkenntnis; auch wenn [Cross \(1973\)](#) und [Staddon et al. \(1980\)](#) bereits über positive Abhängigkeiten vom Vorgängerpegel berichteten. Deshalb konnte Nägle über mögliche Ursachen nur spekulieren. Eine Möglichkeit sah sie in einem im Laufe des Versuches sich verändernden Bezugssystem des Probanden, im Sinne von „gleitenden Kategorien“. Weiterhin spekulierte sie, dass vielleicht doch eine negative Abhängigkeit bestünde, die aber durch die oben veranschlagte „Armträgheit“ überdeckt würde und deshalb positiv erschiene. Ebendiese Idee wurde in dieser Arbeit aufgenommen und weiterentwickelt. Nägle schlug vor, den Probanden so zu instruieren, dass er nach jedem Urteil den Arm in einer definierten Ausgangslage positionieren sollte. Dieser Vorschlag entspricht in der vorliegenden Arbeit am ehesten

der Methode B. Weiterhin wurden noch andere Urteilsmethoden so konstruiert, dass die Armträgheit schrittweise ausgeschaltet wurde.

### 7.6.1 Probandenabhängigkeit

Dass sich die Probanden in ihrem Urteilsverhalten voneinander unterscheiden, und sich das auch auf die Vorgängerabhängigkeiten auswirkt, war abzusehen. Das Spektrum reicht von Probanden, die sich fast gar nicht vom Vorgängerstimulus beeinflussen lassen bis zu Probanden mit extrem ausgeprägter Vorgängerabhängigkeit (s. Abb. 6.13 und 6.14). Wie in Abschnitt 7.2 (Urteilsunterschiede der Probanden, S. 83) ist es jedoch nicht sinnvoll, die Probanden einzeln zu betrachten, da daraus kein auf die Allgemeinheit übertragbares Schema ableitbar wird. Wodurch die unterschiedlichen Vorgängerabhängigkeiten hervorgerufen werden ist wahrnehmungspsychologisch nicht direkt greifbar. Einige mögliche Einflussaktoren sind in Kapitel 7.1 (Visuelle Betrachtung des Urteilsbereichs und der Urteilsdichte, S. 81) aufgeführt.

### 7.6.2 Abhängigkeit von der Urteilsmethode

Für alle Urteilsmethoden ergaben sich nur positive Vorgängerabhängigkeiten. Sowohl für die Abhängigkeit vom Vorgängerpegel (Abb. 6.15 und Tab. 6.9) wie für die Abhängigkeit vom Vorgängerurteil (Abb. 6.16 und Tab. 6.11) zeigt sich dasselbe Schema. Die Methode D hebt sich von den anderen Methode ab. Bei ihr ist die Vorgängerabhängigkeit stärker ausgeprägt als bei den anderen Methoden. Sie ist im Vergleich der Methoden die einzige, die sich quasi immer signifikant von den anderen Methoden unterscheidet. Nur im Vergleich mit Methode C ergibt sich jeweils kein signifikanter Unterschied. Allerdings ist zu beachten, dass für beide Betrachtungen der Vergleich von C und D jeweils noch auf dem 10 %-Signifikanzniveau liegt (Vorgängerpegel  $p = 0,080$ ; Vorgängerurteil  $p = 0,060$ ). Somit ist es nicht möglich einen Unterschied zwischen diesen beiden Methoden völlig auszuschließen. Besonders, wenn man berücksichtigt, dass ein  $p > 0,05$  die Nullhypothese (in unserem Fall  $C = D$ ) keineswegs bestätigt (Bortz und Lienert, 2003, S. 38).

Nur Methode D scheint eine *systematische* Änderung der Vorgängerabhängigkeiten zu bewirken; allerdings eher zum Schlechten. Denn die Intention, die Vorgängerabhängigkeit zu verringern wird nicht erreicht. Im Gegenteil verstärken sich hier die Abhängigkeiten. Ursachen hierfür lassen sich nur vermuten. Eine

Idee ist, dass sich die Urteilsfindung des Probanden aus zwei Komponenten zusammensetzt. Erstens einer absoluten Komponente, die den Stimulus absolut, d.h. ohne Einfluss vom Vorgängerstimulus und Umgebungsgeräusch beurteilt. Zweitens einer Komponente, die sich vom letzten oder auch von mehreren Vorgängerstimuli beeinflussen lässt und versucht die Urteile in Übereinstimmung mit den vorangegangenen Stimuli zu bringen. Ein solches System beschreibt [Ward \(1987\)](#), indem er das Langzeitgedächtnis mit seiner Hörerfahrung dem Kurzzeitgedächtnis, welches die Kontrasteffekte zu unmittelbar davor Gehörtem verarbeitet, gegenüberstellt. Unter diesen Voraussetzungen wäre denkbar, dass der Proband bei fehlender visueller Repräsentation der Skala stärker verunsichert wird und sich deshalb mehr an den Vorgängern orientiert. Die Urteilsfindung verschiebt sich in Richtung der zweiten Komponente, der Proband hangelt sich sozusagen von einem Stimulus zum nächsten. Es geschieht also genau das, was versucht wurde zu verhindern, nämlich ein Verlassen der absoluten Lautheitsbeurteilung zu Gunsten einer relativen Beurteilung abhängig vom Vorgänger. Weiterhin soll noch auf den Vergleich der Methoden B und C eingegangen werden. Bei der Vorgängerpegelabhängigkeit (Tab. [6.9](#)) liegt der p-Wert nicht signifikant bei  $p = 0,076$ , bei der Vorgängerurteilsabhängigkeit (Tab. [6.11](#)) jedoch signifikant bei  $p = 0,026$ . Eine mögliche Erklärung für die unterschiedlichen Signifikanzen sind die höheren Absolutwerte der Abhängigkeiten vom Vorgängerurteil. Durch die größeren Zahlenwerte kann eventuell eine Anhebung der Signifikanz und dadurch ein Unterschreiten der 5 %-Marke bewirkt werden. Warum sich gerade die Methoden B und C unterscheiden ist unklar. Besonderes Augenmerk verdient der Vergleich der Methoden A und B, der sowohl für Vorgängerpegel- (Tab. [6.9](#)) als auch für Vorgängerurteilsabhängigkeit (Tab. [6.11](#)) nicht signifikant ist. Wäre die Vorgängerabhängigkeit durch die postulierte Armträgheit bedingt, müsste sich alleine beim Vergleich dieser beiden Methoden ein hochsignifikanter Unterschied ergeben. Denn während bei Methode A die Armträgheit und somit die Vorgängerabhängigkeit maximal sein müsste, sollte bei Methode B die Armträgheit keine Rolle spielen, da der zurückgelegte Weg des Armes konstant war.

Als Fazit lässt sich sagen, dass die postulierte Armträgheit wohl nicht existiert bzw. keinen Einfluss auf des Skalierungsverhalten der Probanden nimmt. Zwar wird mit den verschiedenen Methoden in Einzelfällen unterschiedlich skaliert; eine systematische und verallgemeinerbare Beeinflussung ist aber nicht erkennbar.

### 7.6.3 Abhängigkeit vom Pegel

Die Abhängigkeit vom Pegel zeigt für den Vorgängerpegel und das Vorgängerurteil ein symmetrisches Bild. Der Verlauf der einzelnen Box-und-Whisker-Plots ist in den beiden Schaubildern 6.17 und 6.18 gleich, nur dass in der Darstellung der Vorgängerurteilsabhängigkeiten größere Zahlenwerte vorherrschen. Daraus folgt, dass das vom Probanden selbst abgegebene Urteil einen stärkeren Einfluss auf sein folgendes Urteil nimmt als der Vorgängerpegel.

Weiterhin fallen die äußerst niedrigen Vorgängerabhängigkeiten für die Extrempiegel 30 dB und 90 dB, sowie für die Pegel 50 dB und 80 dB auf.

Für die Pegel 30 dB und 90 dB ist dies zu erwarten gewesen und durch den Versuchsaufbau begründbar. Die randständige Position als kleinster bzw. größter vorkommender Pegel bewirkt einen Randeffect, da jeweils kein leiserer bzw. lauterer Vorgänger existiert (Kollmeier, 1997, S. 126). Dadurch ergibt sich die designbedingte Reduzierung der Vorgängerabhängigkeiten.

Die Ergebnisse für die Pegel 50 dB und 80 dB sind nicht so leicht erklärbar. Allerdings sind sie reproduzierbar, da sie in gleicher Form auch bei Nägle (2004) aufgetreten sind. Auch sie fand keine Erklärung für dieses Phänomen. Eine zumindest theoretisch denkbare Ursache könnte in der verwendeten Stimulussequenz stecken. Diese wurde zwar für jeden Durchgang um einige Stimuli durchrotiert, war aber letztendlich immer die gleiche mit jeweils unterschiedlichem Startpunkt. Dass in der Anordnung der Stimuli tatsächlich ein höheres, unentdecktes Geheimnis liegt ist zwar eher unwahrscheinlich, aber dennoch möglich.

Eine zweite, weitaus interessantere Möglichkeit muss jedoch auch darin gesehen werden, dass dieses Phänomen nicht etwa designbedingt entstanden ist, sondern tatsächlich existiert. Denkbar ist, dass die Probanden im Pegelbereich um 50 dB und 80 dB schlicht besser absolut urteilen können und sich weniger vom Vorgängerstimulus beeinflussen zu lassen. Ob dies aus einem in diesem Bereich gefestigteren Bezugssystem resultiert, bleibt offen. Denkbar wäre dies eventuell dadurch, dass es sich ungefähr um die Pegelgrenzen des Hauptsprachbereichs (s. „Sprachbanane“ in Abb. 1.1) handelt.

### 7.6.4 Abhängigkeit von der Medianlage der Urteile

Die Auswertung in Kapitel 6.8.4 (Effekt der Medianlage der Urteile, S. 76) zeigte keinen Zusammenhang zwischen der Lage der Urteile und den Vorgänger-

abhängigkeiten. Die hohe positive Korrelation zwischen Vorgängerurteileffekt und Vorgängerpegel-effekt verwundert indes nicht. Bereits in den zugehörigen Graphiken (Abb. 6.14 und Abb. 6.13) war dieser Effekt anhand des sehr ähnlichen Verlaufs der Box-und-Whisker-Plots deutlich zu sehen. Es ist zu beachten, dass eine nachgewiesene Korrelation nur einen *systematischen*, aber keinen *kausalen* Zusammenhang zwischen den Variablen beweist. Nichtsdestotrotz ist bewiesen, dass ein Proband mit großer Vorgängerpegelabhängigkeit auch eine große Vorgängerurteilsabhängigkeit aufweist. Dies gilt auch umgekehrt für *kleine* Vorgängerabhängigkeiten. Außerdem zeigt sich, dass sowohl für Vorgängerpegel als auch für Vorgängerurteil konstant positive Abhängigkeiten bestehen, und nicht mal eine positive und mal eine negative.

### 7.7 Mögliche unberücksichtigte Einflussfaktoren und Fehler

Es existieren eine Vielzahl von Faktoren, die eine Person während eines Tests beeinflussen können. Um deren Einfluss zu verringern wurde versucht, die äußeren Umstände und die Voraussetzungen der Probanden möglichst konstant zu halten. Außerdem sollten potenzielle Fehlerquellen in Vorversuchen erkannt und beseitigt werden.

Dennoch blieben Einflussfaktoren bestehen, die entweder nicht ausgeschaltet werden konnten, noch nicht näher untersucht oder auch (noch) gar nicht als solche bekannt waren bzw. immer noch sind.

Bereits in Kapitel 6.1.2 ([Umgang mit fehlerhaften Werten](#), S. 39) wurde auf die fehlerhaften Werte eingegangen, die bei den Methoden A und B aufgetreten sind. Durch das weitere Vorgehen in der statistischen Auswertung dürfte eine fehlerhafte Beeinflussung der Ergebnisse nicht aufgetreten sein, kann jedoch nicht vollständig ausgeschlossen werden.

Auch die lange Zeitdauer, die die Versuche beanspruchten, könnte die Lautheitsskalierung der Probanden beeinflussen. Es war klar, dass die Versuche recht umfangreich und langwierig waren und auch, dass das Verfahren so in der Klinik nicht angewandt werden kann. Dennoch wurde der Ablauf bewusst so gewählt um statistisch einwandfreie Daten zu erhalten und um eine Vergleichbarkeit der unterschiedlichen Methoden zu erreichen. Es ist möglich, dass gegen Ende der Tests die Probanden ermüdet waren und ihre Konzentration nachgelassen hat, womit die Skalierung unsauberer geworden wäre (vgl. Abschnitt 7.5

([Auswirkungen der Zeitposition der Urteilsmethoden](#), S. 86)).

Eine große Rolle spielte sicherlich die persönliche Einstellung und Motivation der Probanden gegenüber der Untersuchung. Diese war jedoch nicht objektivierbar und konnte somit nicht in der Auswertung berücksichtigt werden. Einfluss auf den Probanden können auch die Camera silens mit ihrer außergewöhnlichen Wand- und Deckenverkleidung, die Tageszeit und die Stimmung des Probanden selbst genommen haben.

## 8 Ausblick

Diese Arbeit hat gezeigt, dass die Urteilmethode die Lautheitsskalierung im WHF beeinflussen kann. Daneben gibt es noch weitere Faktoren, die Auswirkungen auf das Skalierverhalten der Probanden haben können und die noch weiter untersucht werden müssen.

Dazu gehört einerseits die Dauer des Interstimulusintervalls, auf die bereits in Abschnitt 3.3 ([Voruntersuchungen zum Testablauf](#), S. 17) eingegangen wurde. Eine Arbeit zu dieser Thematik läuft bereits an der HNO-Universitätsklinik Würzburg. Sie ist aber leider noch nicht fertiggestellt.

Auch die aus der vorliegenden Arbeit gewonnenen Erkenntnisse können weitergesponnen werden. Speziell die Auswirkung von verschiedenen Arten der Skalenpräsentation wäre interessant weiterzuverfolgen. So könnte die Heller-Skala zum Beispiel in verschiedenen räumlichen Relationen präsentiert werden, während die Urteilmethode immer dieselbe wäre. Am besten würde sich dazu die mündliche Skalierung eignen.

Da das Spektrum der klinischen Patienten sehr breit ist, wäre auch eine Untersuchung der Lautheitsskalierung in verschiedenen Populationen denkbar. Besonders wichtig wäre es, vergleichbare Untersuchungen mit älteren Probanden, dem Hauptklientel der Audiometrie, sowie mit Trägern von Cochlea-Implantaten durchzuführen.

Auch die individuellen Voraussetzungen der Probanden, wie Hörerfahrung, Intelligenz oder Konzentrationsfähigkeit können die Lautheitsskalierung beeinflussen. Sicher wäre es äußerst schwierig diese individuellen Vorbedingungen wissenschaftlich zu untersuchen, wenngleich die daraus gezogenen Erkenntnisse sehr interessant wären.

Eine weitere Untersuchungen an der HNO-Universitätsklinik Würzburg, die aber noch am Anfang steht, soll den Einfluss des Pegelumfanges auf das Skalierverhalten untersuchen. Die Frage ist, inwiefern bzw. ob sich bei kleinerem Pegelumfang (z.B. 45 dB bis 75 dB) die Lautheitsskalierung der Probanden aufdehnt. Es wäre z.B. denkbar, dass die Probanden dann die Kategorien »sehr leise« und »sehr laut« nicht benutzen oder aber die Urteile weiter auseinander-

ziehen.

Ein möglicher Effekt der Stimulusdauer, wie ihn [Nägle \(2004\)](#) vorschlägt, ist weniger wahrscheinlich. [Zwicker und Fastl \(1990\)](#) zeigten, dass die Stimulusdauer nur bis zu einer Länge von 300 ms eine entscheidende Rolle in der Lautheitsempfindung spielt.



## 9 Zusammenfassung

Das Würzburger Hörfeld (WHF) ist ein computergestütztes Verfahren der kategorialen Lautheitsskalierung, das in der Audiometrie Verbreitung gefunden hat. Die akustischen Reize werden computergesteuert abgegeben und die Lautheitsurteile des Probanden mittels eines Skaliertabletts an den Computer übermittelt und von diesem registriert. Bei aufeinanderfolgender Darbietung der Reize treten Sequenzeffekte auf, so dass das Urteil über den aktuellen Stimulus vom vorangehenden Stimulus beeinflusst wird. Die Korrelation zwischen dem Urteil über den Vorgänger- und den aktuellen Reiz ist positiv, d.h. das aktuelle Urteil wird in Richtung des Vorgängerurteils beeinflusst. Diese Beeinflussung ist sowohl für die Methode der „magnitude estimation“ als auch für die kategoriale Lautheitsskalierung nachgewiesen. Die zugrundeliegenden Mechanismen sind jedoch weitgehend ungeklärt und Thema dieser Arbeit.

Um die jeweiligen Anteile verschiedener möglicher Einflußfaktoren an der Vorgängerabhängigkeit zu untersuchen, wurden fünf Urteilsabgabemethoden so konstruiert, dass diese Faktoren Schritt für Schritt ausgeschaltet wurden. Neben der Individualkomponente der Probanden wurde die physikalische Trägheit der Armbewegung, die Art der Abgabe des Urteils, die visuelle Präsenz einer Skala und eine „innere“ Erinnerungskomponente postuliert.

Es wurden 25 audiometrisch überprüft normalhörende Probanden (13 männlich, 12 weiblich) zwischen 20 und 30 Jahren getestet. Als Stimulus wurde CCITT-Rauschen einer Sekunde Dauer und mit Pegeln von 30 dB bis 90 dB SPL in 5 dB-Schritten verwendet. Jede mögliche Pegel-Vorgängerpegel-Kombination kam genau einmal vor. Dadurch konnte der Einfluß der Urteilsabgabemethode sowohl auf die Urteile direkt als auch auf ihre Vorgängerpegel- und -urteilsabhängigkeiten untersucht werden. Das Interstimulus-Intervall betrug konstant vier Sekunden.

Zum Einsatz kamen fünf Urteilsmethoden: zwei Skaliertablettmethoden, eine davon mit Schwebenlassen der Hand über dem soeben gedrückten Urteil (A), und die andere mit Zurücklegen der Hand auf einen definierten Ruhepunkt zwischen den Urteilsabgaben (B), zwei mit mündlicher Abgabe des Urteils als

Zahlenwert, eine davon mit als Poster aufgestellter (C), und die andere ohne sichtbare Skala (D), und eine Methode, bei der das Urteil mit einem Laserpointer auf dem Poster angezeigt wurde (E).

Die aus allen abgegebenen Urteilen gebildete Pegel-Lautheitsfunktion steigt streng monoton, zu höheren Pegeln hin mit zunehmender Steigung. Bei jeder Pegelerhöhung um 5 dB wurde statistisch signifikant lauter geurteilt. Die Urteile sind positiv mit den Vorgängerpegeln und -urteilen korreliert. Diese Übereinstimmungen mit der Literatur und früheren am CI-Zentrum durchgeführten Arbeiten bezeugt die Reproduzierbarkeit der Ergebnisse und spricht für die Verlässlichkeit der Methode.

Die Resultate zeigen eine starke Individualkomponente. Die Antworten der Probanden wichen in Lage und Streuung teils deutlich voneinander ab. Im Vergleich der Absolutlage der Lautheitsurteile stellten sich Methoden A und B als gleich heraus, ebenso wie C und E. Alle anderen paarweisen Vergleiche der Methoden ergaben statistisch signifikante Unterschiede. Am lautesten wurde mit den Methoden A und B geurteilt, mit Methode D am leisesten, C und E lagen dazwischen. Methoden A und B verwendeten das Skaliertablett, C und E das Poster, und D keine sichtbare Skala. Der Unterschied in der Absolutlage der angegebenen Lautheit kann daher auf die Art der visuellen Repräsentation der Skala zurückgeführt werden.

Die Vorgängerabhängigkeit war sowohl für den Vorgängerpegel als auch für das Vorgängerurteil positiv, wobei die Abhängigkeit vom Vorgängerurteil stärker ausgeprägt war als die vom Vorgängerpegel. Die Vorgängerabhängigkeiten waren außerdem pegelabhängig. In den Randbereichen (30 dB und 90 dB) sowie bei 50 dB und 80 dB zeigten sich nur schwache bis keine Abhängigkeiten; bei 65 dB und 70 dB waren sie am stärksten ausgeprägt. Diese Resultate entsprechen denen aus früheren Arbeiten.

Im Vergleich der Stärke der Abhängigkeit vom *Vorgängerpegel* unterschied sich nur Methode D von den Methoden A, B und E signifikant; alle anderen paarweisen Vergleiche ergaben keinen Unterschied. Bezüglich der Abhängigkeit vom *Vorgängerurteil* unterschied sich Methode D von den vier anderen, diese untereinander jedoch nicht. In beiden Fällen war bei Methode D die Vorgängerabhängigkeit am stärksten.

Aufgrund der Gleichheit der Methoden A und B (Schwebenlassen bzw. Zurücklegen der Hand zwischen den Urteilen) kann die physikalische Armträgheit als Ursache der Vorgängerabhängigkeit ausgeschlossen werden. Die Gleichheit der

Methoden A, B, C und E (Skaliertablett, sichtbare Poster-Skala mit Zahlennennung bzw. Zeigen mit dem Laserpointer) schließt auch die Art der Urteilsabgabe (berührend, benennend bzw. zeigend) als Einflußfaktor aus. Die Beobachtung, dass bei Fehlen einer visuellen Skala und benennender Urteilsabgabe (Methode D) die Vorgängerabhängigkeit am stärksten war, legt nahe, dass eine interne Erinnerungskomponente hauptverantwortlich ist. Die visuelle Präsenz einer Skala hat hingegen „stabilisierende“ Wirkung, d.h. sie verringert die Vorgängerabhängigkeit.

Abschließend lässt sich sagen, dass von den postulierten Einflußfaktoren der Vorgängerabhängigkeit die physikalische Armträgheit und die Art der Urteilsabgabe ausgeschlossen, die „interne“ Erinnerungskomponente und die visuelle Präsenz einer Skala jedoch bestätigt werden konnten. Von den modifizierten Verfahren erwies sich keines als günstiger als das bisher verwendete Standardverfahren.



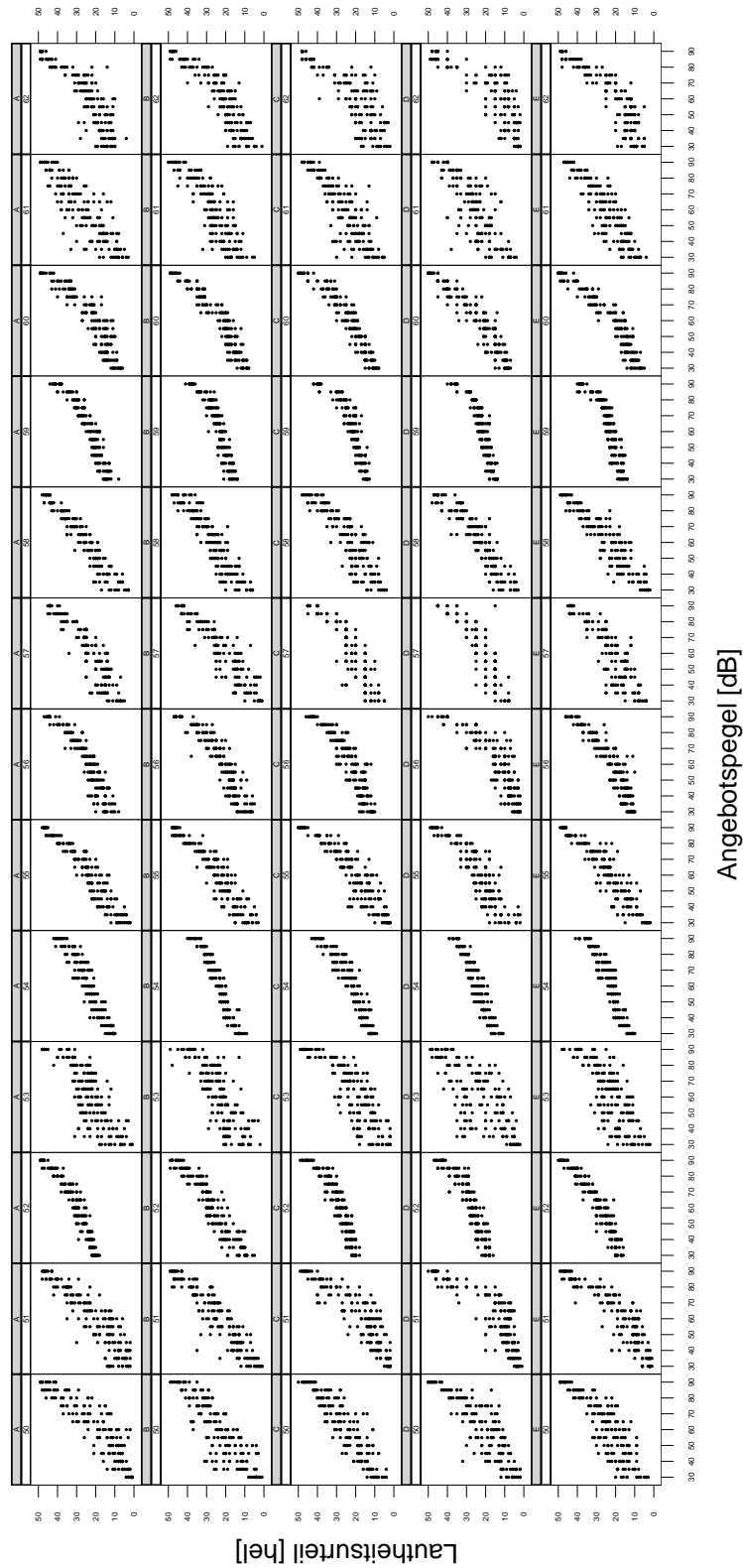
## A Übersichtstafeln Lautheitsskalierung und Urteilsdichte

Die folgenden Schaubilder geben eine Übersicht über die Lautheitsskalierung jedes Probanden. Die Art der Darstellung ist dieselbe wie die für die drei Beispielprobanden aus Kapitel 6 ([Ergebnisse](#), S. 37). Jeweils in der obersten, grau hinterlegten Zeile jeder Einzelgraphik steht die Urteilsmethode, während in der zweiten Zeile die Nummer des Probanden angegeben wird.

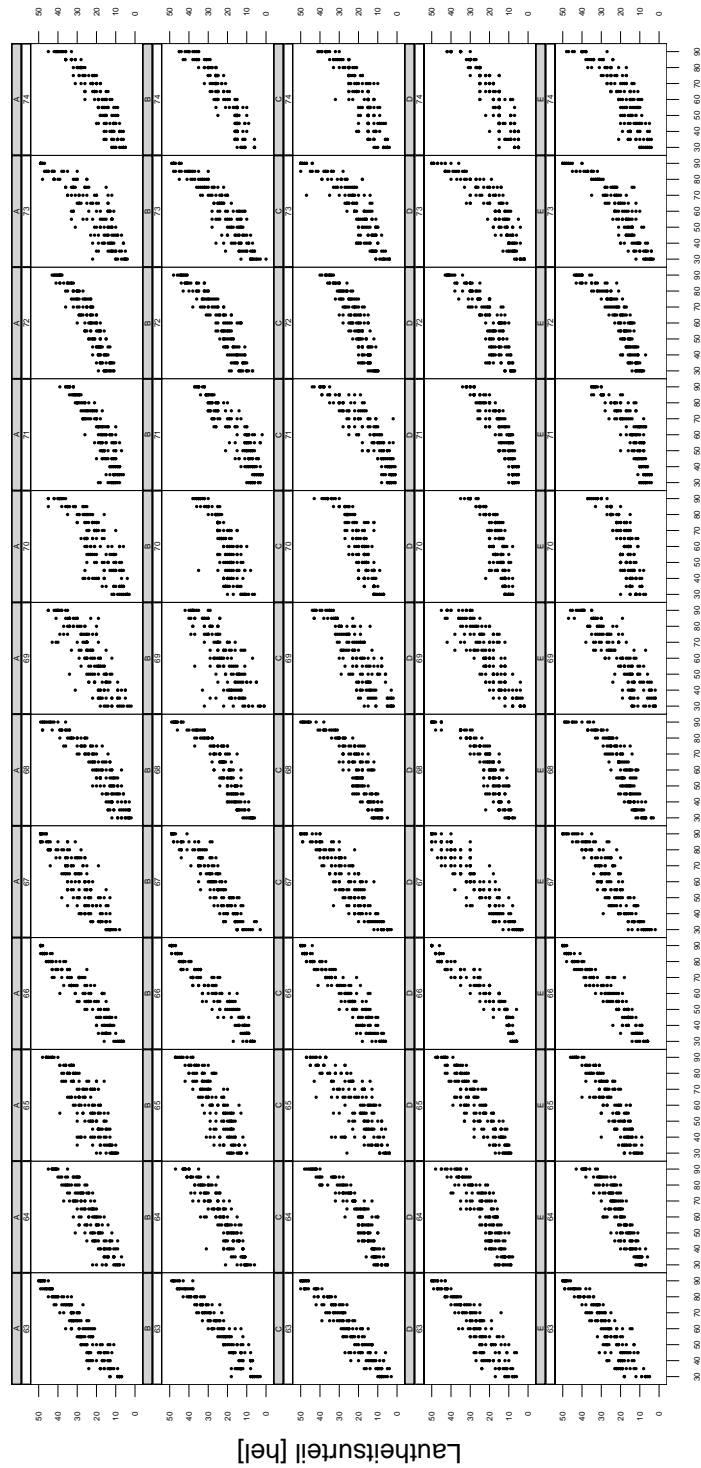
Die beiden ersten Tafeln zeigen die Lautheitsurteile als Punktwolke über dem zugehörigen Pegel. Da jeder Pegel 13 mal bewertet wurde, besteht auch jede zugehörige Punktwolke aus 13 Urteilen. Wenn weniger als 13 einzelne Punkte zu sehen sind, liegt das daran, dass gleiche Urteile übereinander projiziert werden und deshalb als ein Punkt erscheinen. Auf der Abszisse sind die Pegel von 30 dB bis 90 dB aufgetragen und auf der Ordinate die Urteile von 0 hel bis 50 hel.

Die nächsten beiden Tafeln veranschaulichen die Dichteverteilung der Urteile. Hier ist die Abszisse mit dem Lautheitsurteil in hel beschriftet. Auf der Ordinate ist die Urteilsdichte aufgetragen.

# A Übersichtstafeln Lautheitsskalierung und Urteilsdichte

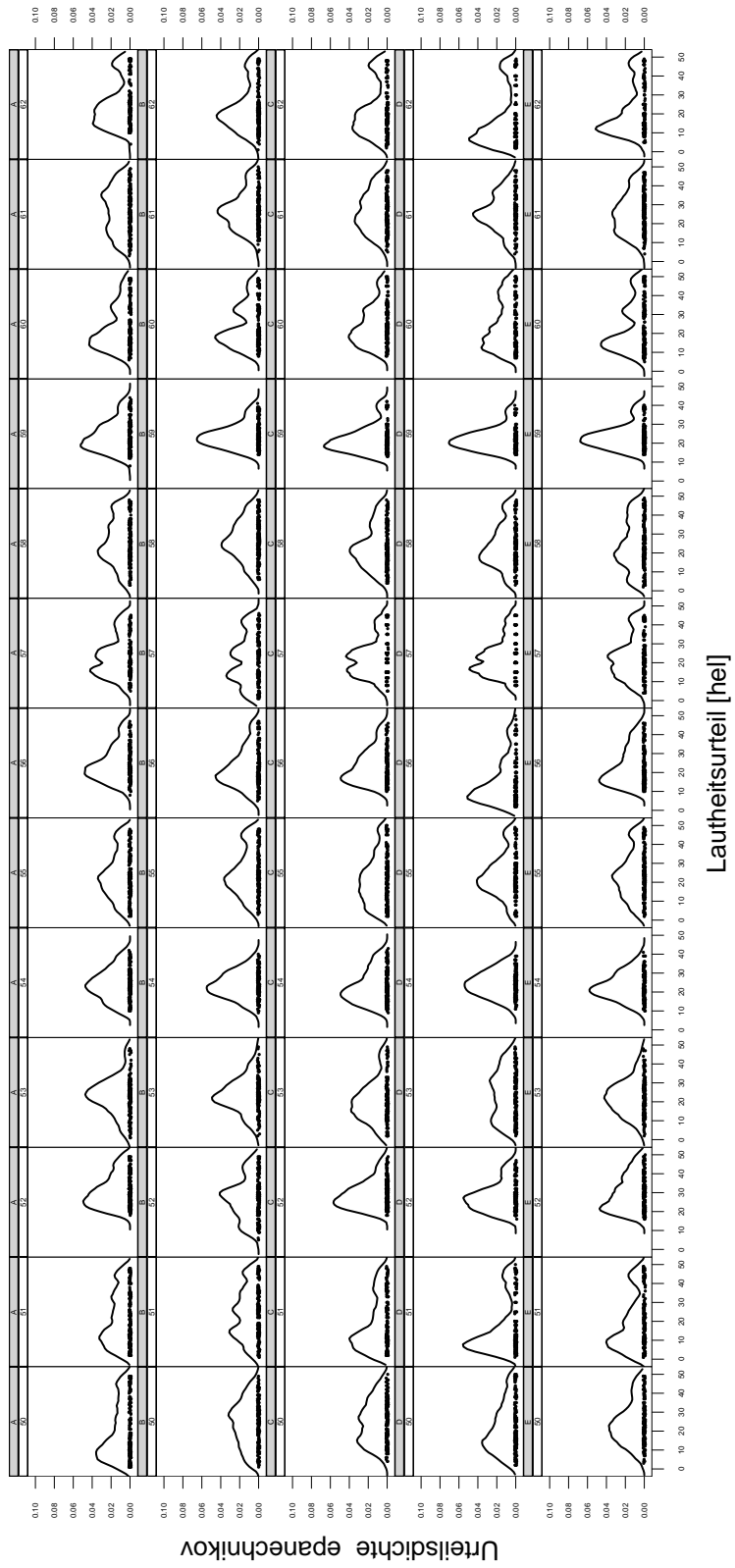


# A Übersichtstafeln Lautheitsskalierung und Urteilsdichte



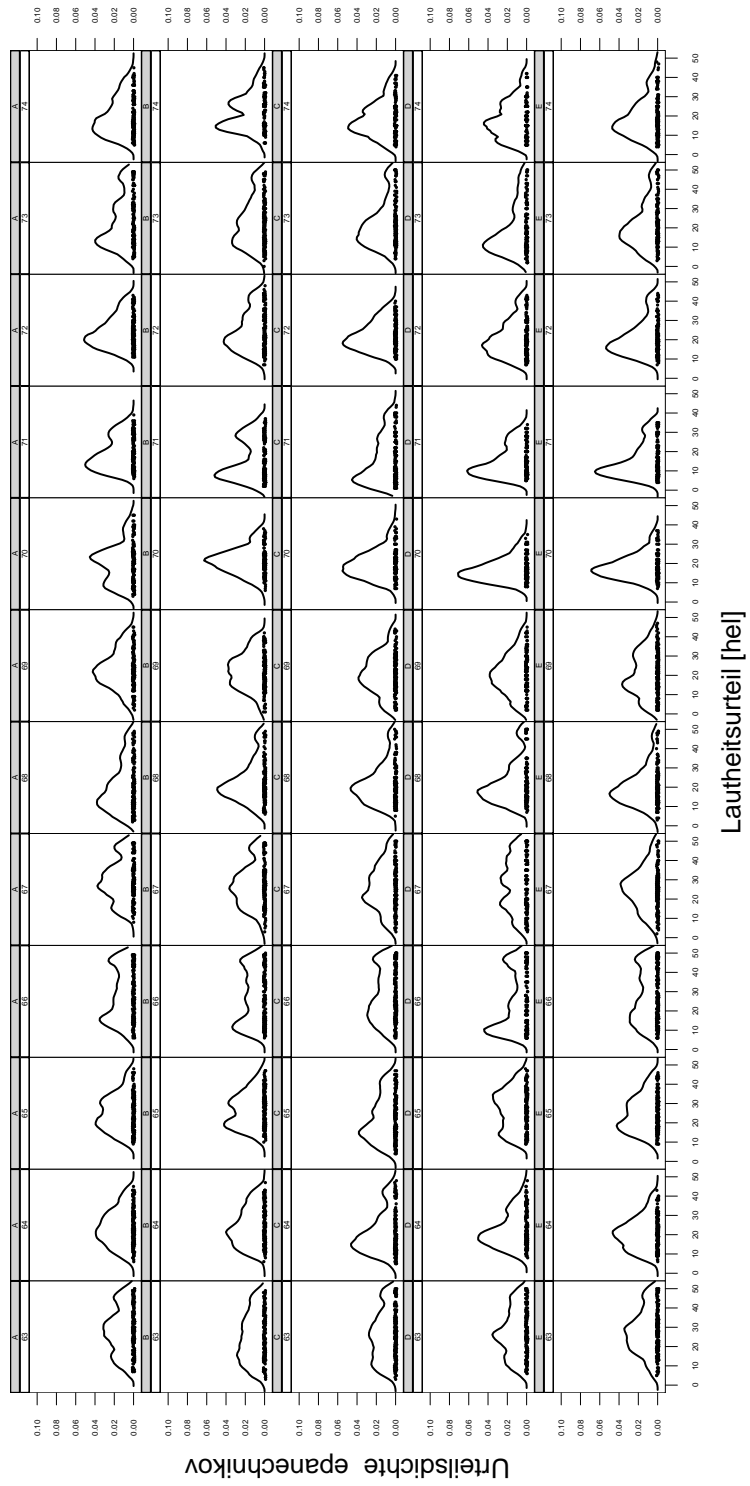
Angebotspegel [dB]

# A Übersichtstafeln Lautheitsskalierung und Urteilsdichte





# A Übersichtstafeln Lautheitsskalierung und Urteilsdichte





## B Skalierungsdaten und Vorgängerabhängigkeiten

Dieses Kapitel gibt im Detail die erhaltenen Daten für jeden einzelnen Patienten wieder. Jede Seite beinhaltet die Daten eines einzelnen Probanden. Dabei sind zeilenweise von oben nach unten die Urteilsverfahren von A bis E angeordnet. Jede Zeile gibt denselben Datensatz dreimal nach verschiedenen statistischen Auswerteverfahren wieder. Gemeinsam ist allen Einzelgraphiken die Farbkodierung der Schallpegel, sowie die auf der Ordinate aufgetragenen Lautheitsurteile in hel. Die Zuordnung der Pegel zu den jeweiligen Farben gibt Abb. B.1 wieder.

In der ersten Spalte sind die Lautheitsurteile über dem zugehörigen Pegel aufgetragen. Die Urteile werden einmal als Punktwolke wiedergegeben. Bei gleichem Urteil überlagern sich die Punkte, so dass es so aussehen kann als ob die Pegel verschieden oft bewertet worden wären. Dies ist aber nicht der Fall gewesen. Außerdem werden in derselben Einzelgraphik die abgegebenen Urteile zu einer Kurve nach dem „loess“-Regressionsmodell zusammengefasst. Dadurch entsteht die Kurve der Pegel-Lautheitsfunktion.

Spalte 2 zeigt die Lautheitsurteile in Abhängigkeit vom Schallpegel des Vorgängerstimulus. Die Vorgängerpegel sind von 30 dB bis 90 dB auf der x-Achse aufgetragen, die Lautheitsurteile auf der y-Achse. Für jeden Vorgängerpegel existieren wieder 13 Urteile, welche als Punkte eingezeichnet sind. Durch jede der 13 Punktwolken wurde eine Regressionsgerade gelegt. Der Betrag der Steigung spiegelt die Größe der Vorgängerabhängigkeit wieder. Je steiler die Gerade, desto größer ist die Abhängigkeit. Das Vorzeichen der Steigung besagt, ob eine negative oder positive Korrelation vorliegt.

In Spalte 3 werden die Lautheitsurteile in Abhängigkeit vom Lautheitsurteil über den Vorgängerstimulus aufgezeigt. Sowohl Abszisse als auch Ordinate tragen hier die Beschriftung in hel. Jedem Angebotspegel sind 13 Lautheitsurteile zugeordnet. Durch jede dieser insgesamt 13 Punktwolken wurde eine Regressionsgerade gelegt. Die Steigung gibt die Abhängigkeit des Lautheitsurteils vom Urteil über den Vorgängerstimulus an. Der Betrag der Steigung steht für das Ausmaß der Vorgängerabhängigkeit, das Vorzeichen für die Richtung der Korrelation.

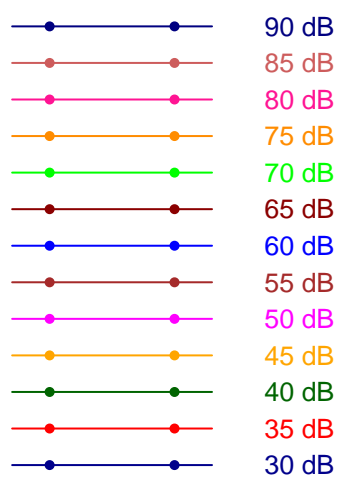
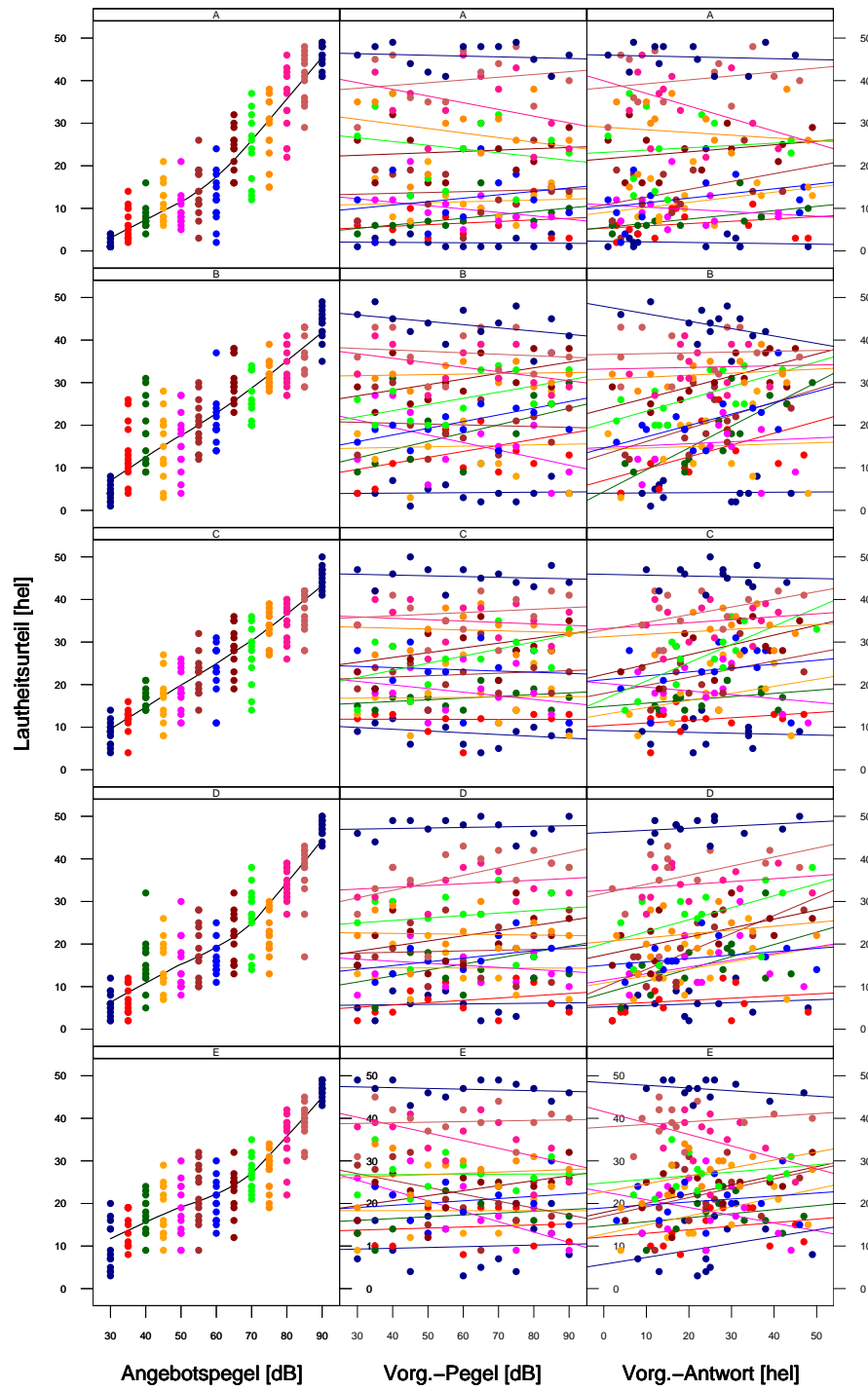


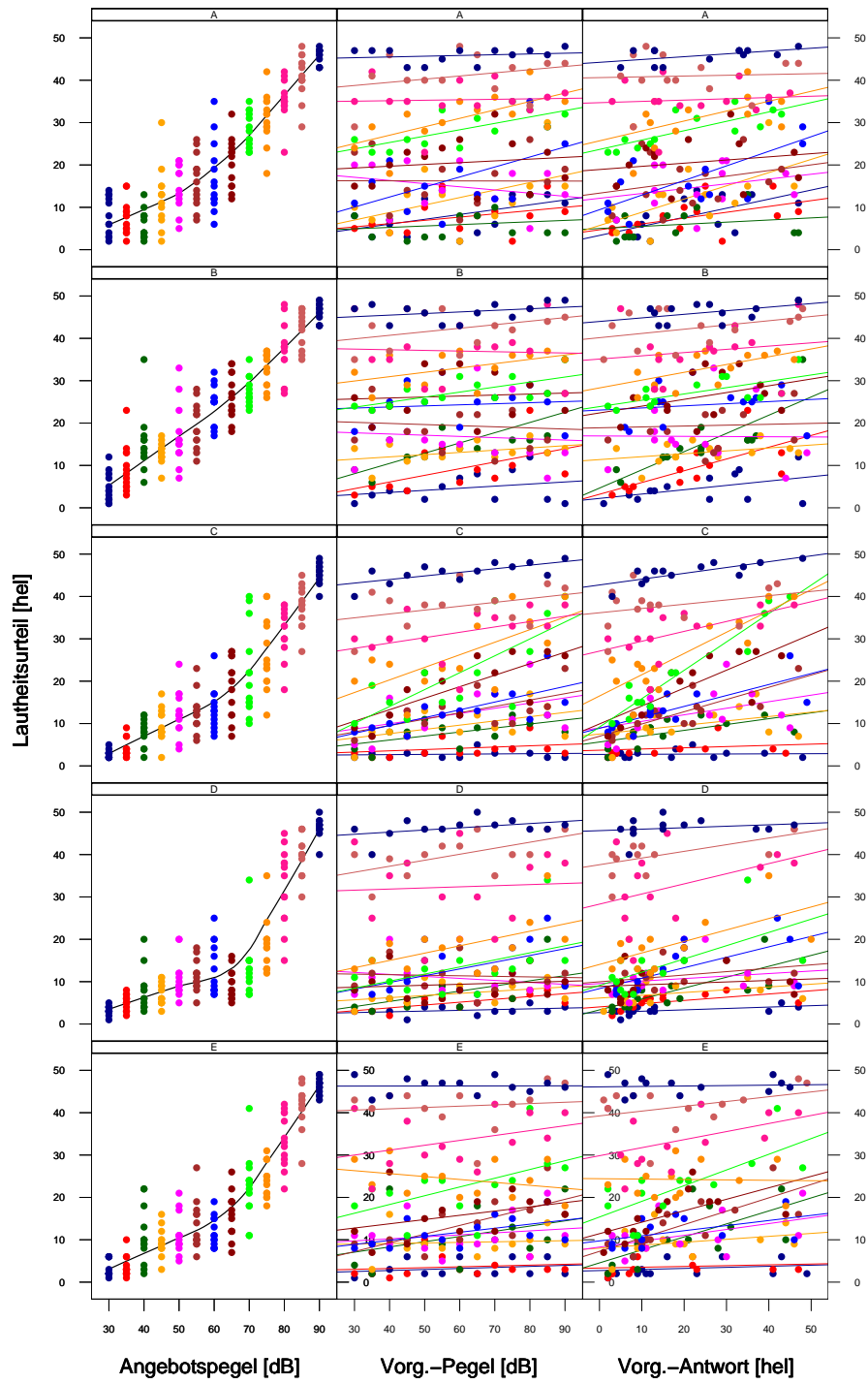
Abbildung B.1: Farblegende für die Skalierungsdaten auf den folgenden Seiten

Proband 50

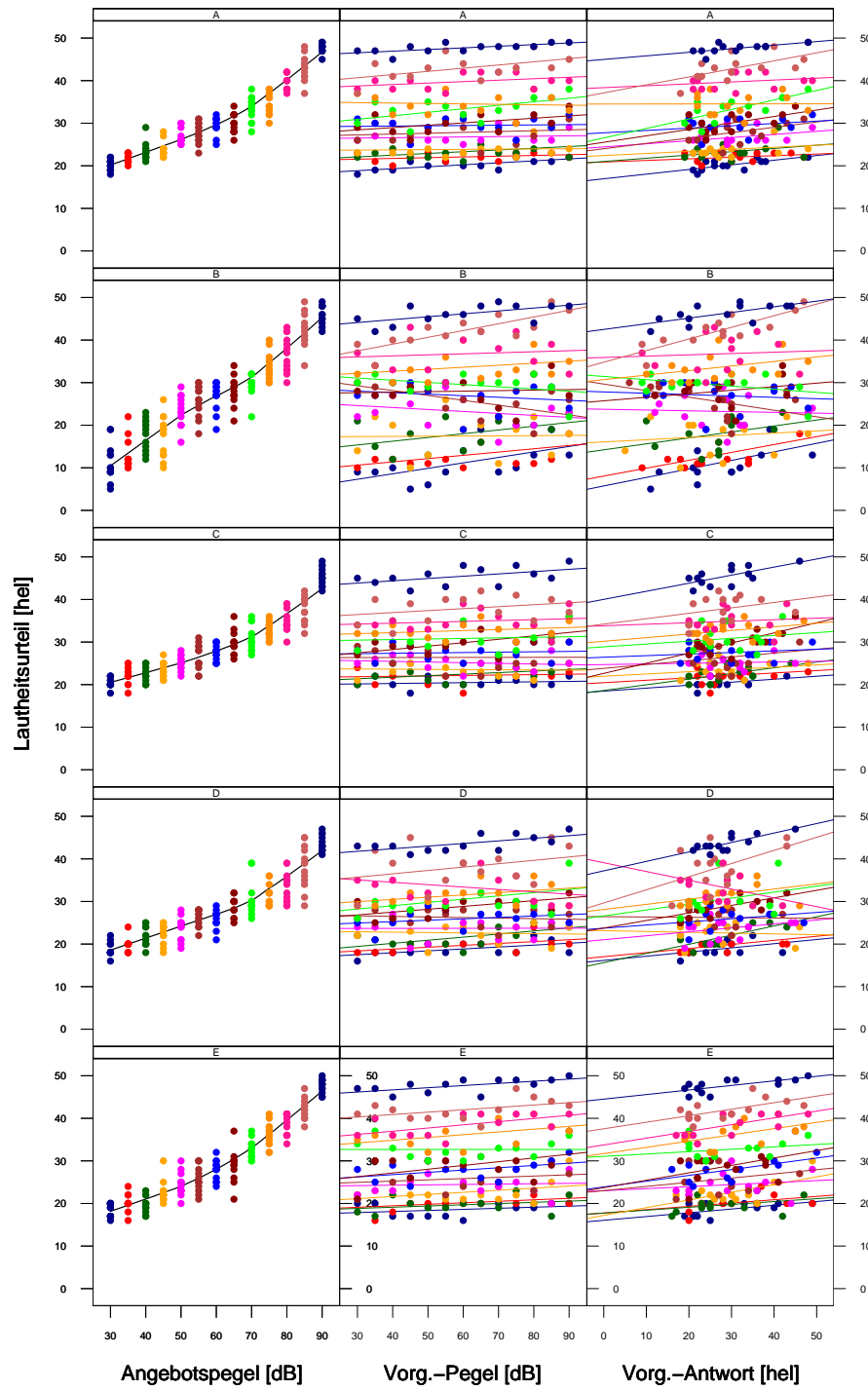


B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 51

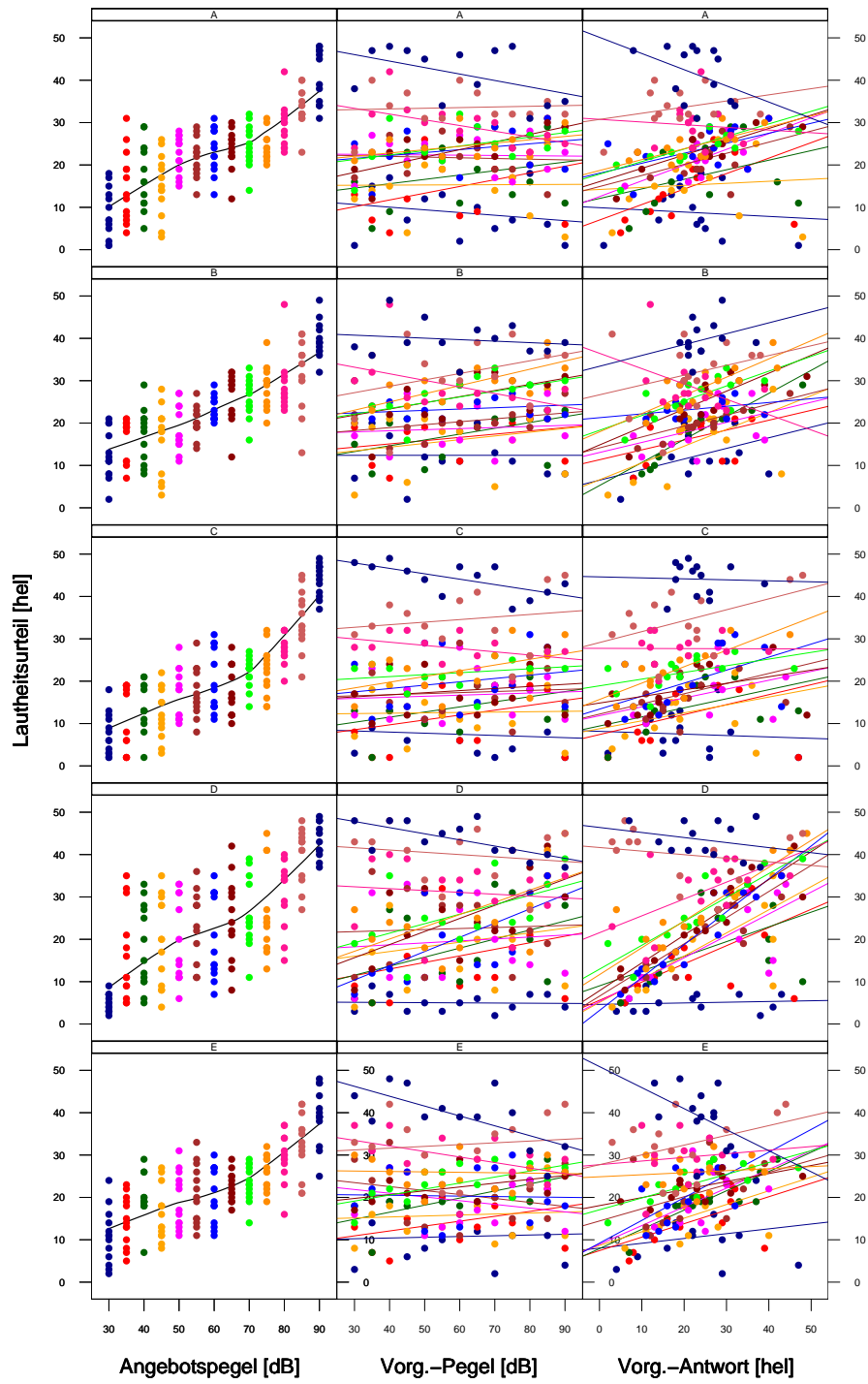


Proband 52



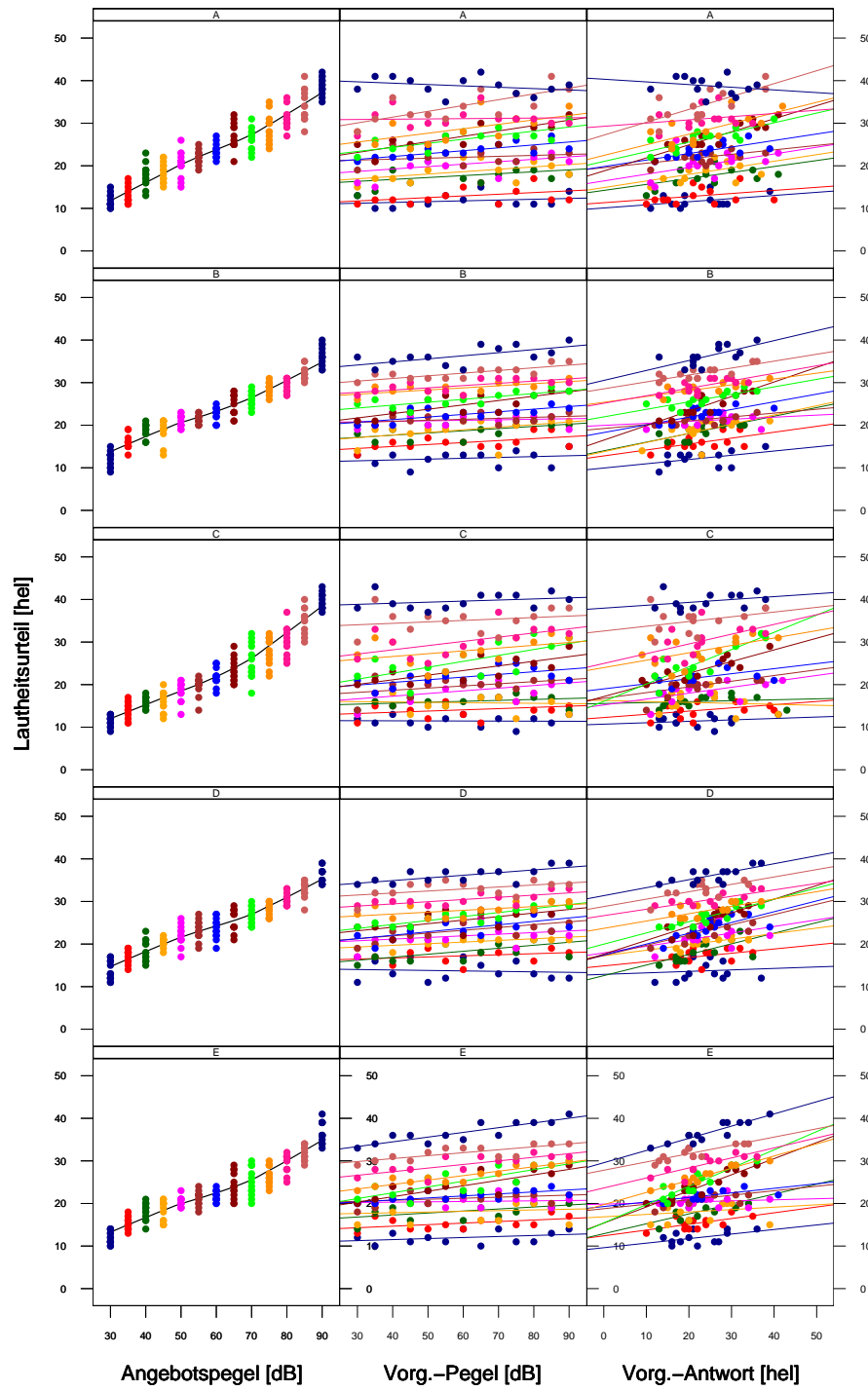
# B Skalierungsdaten und Vorgängerabhängigkeiten

## Proband 53



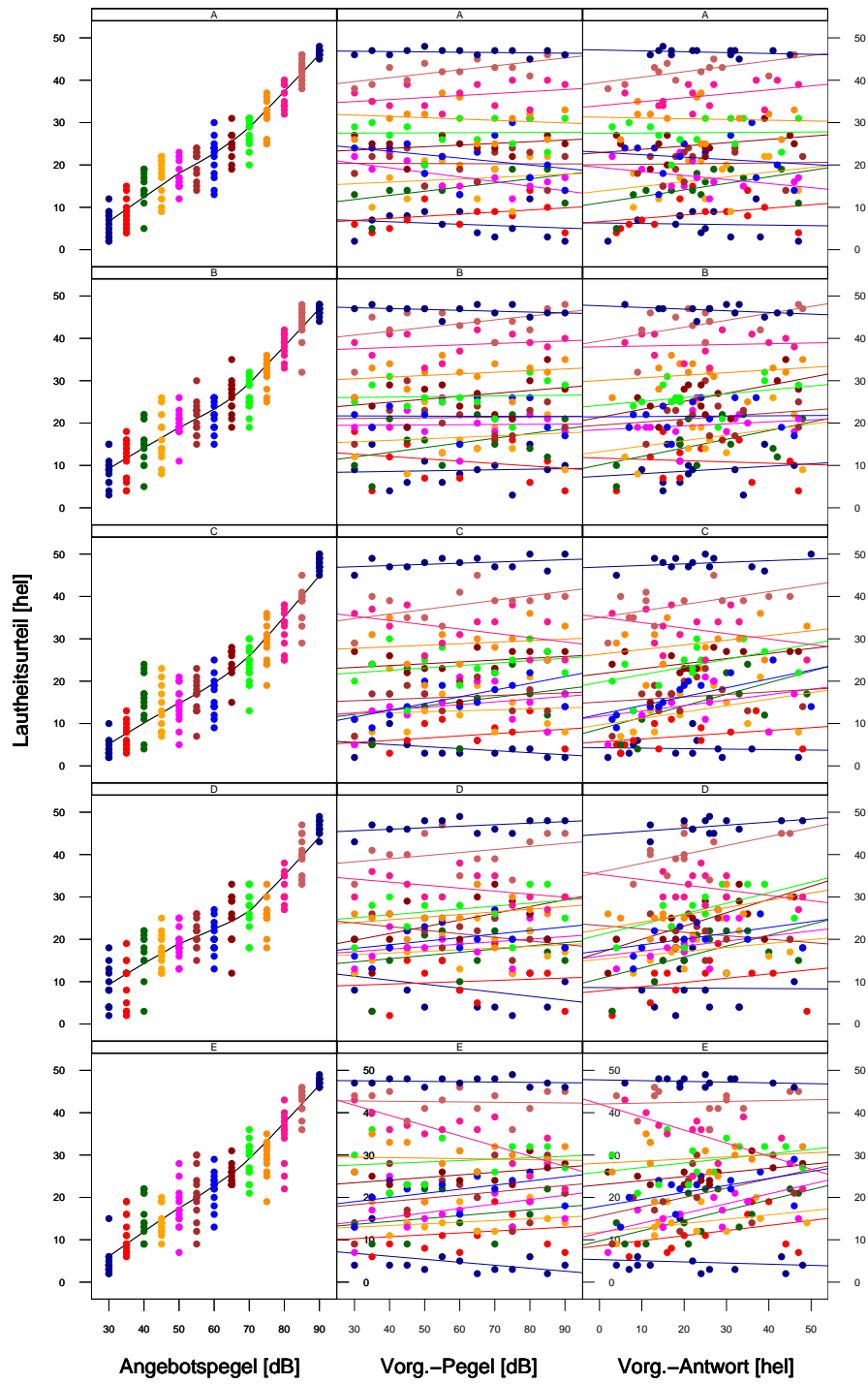


Proband 54

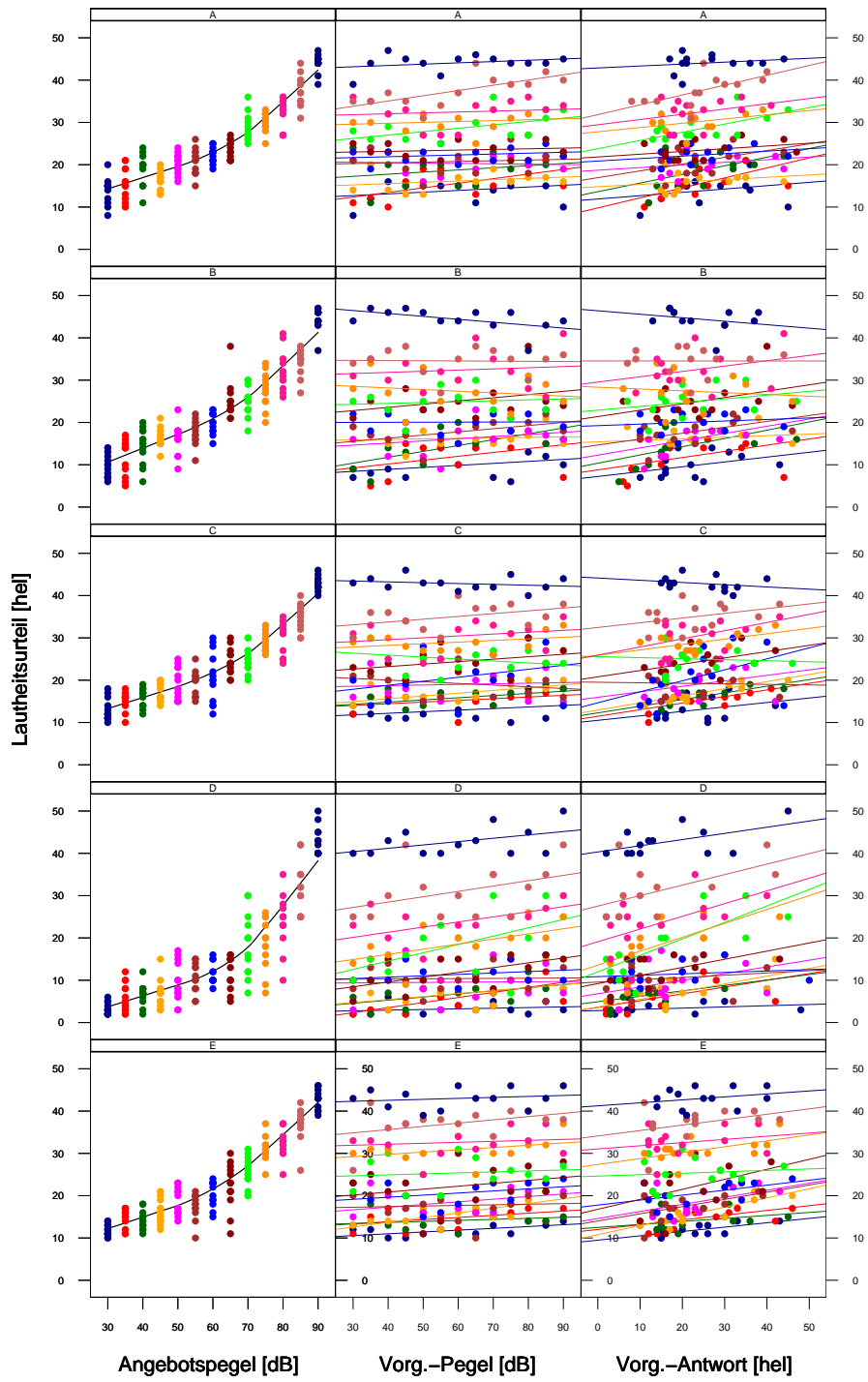


B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 55

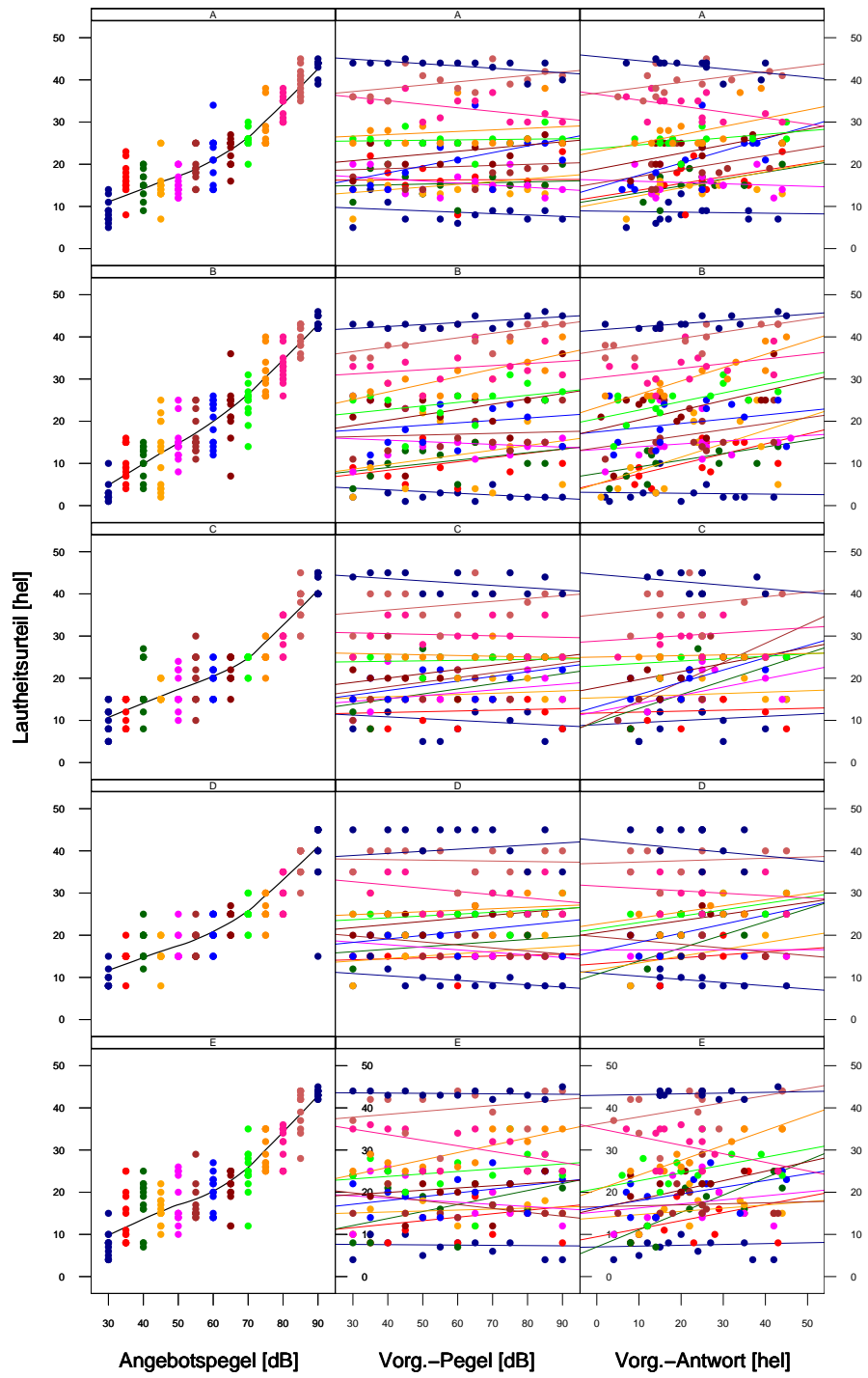


Proband 56

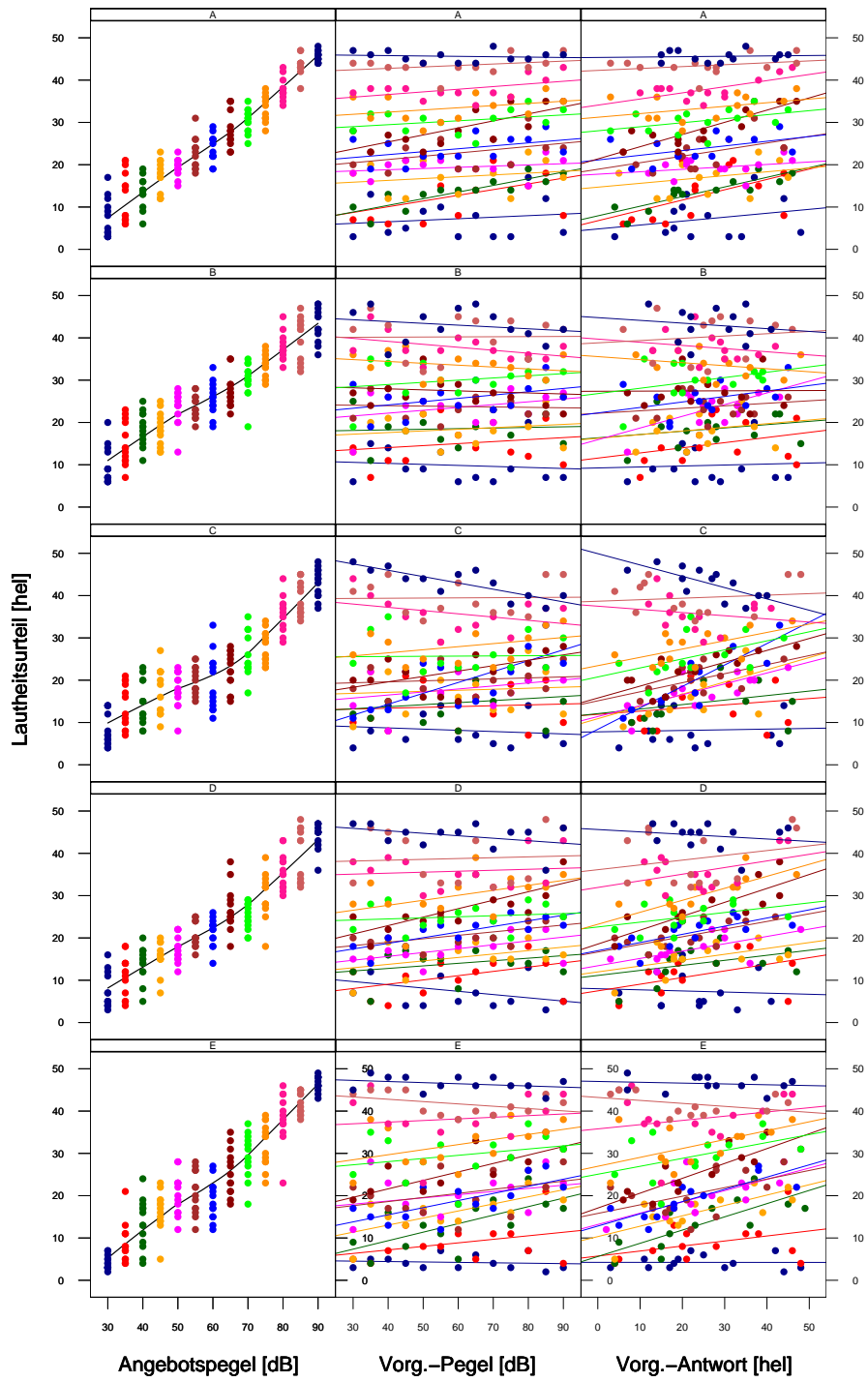


B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 57

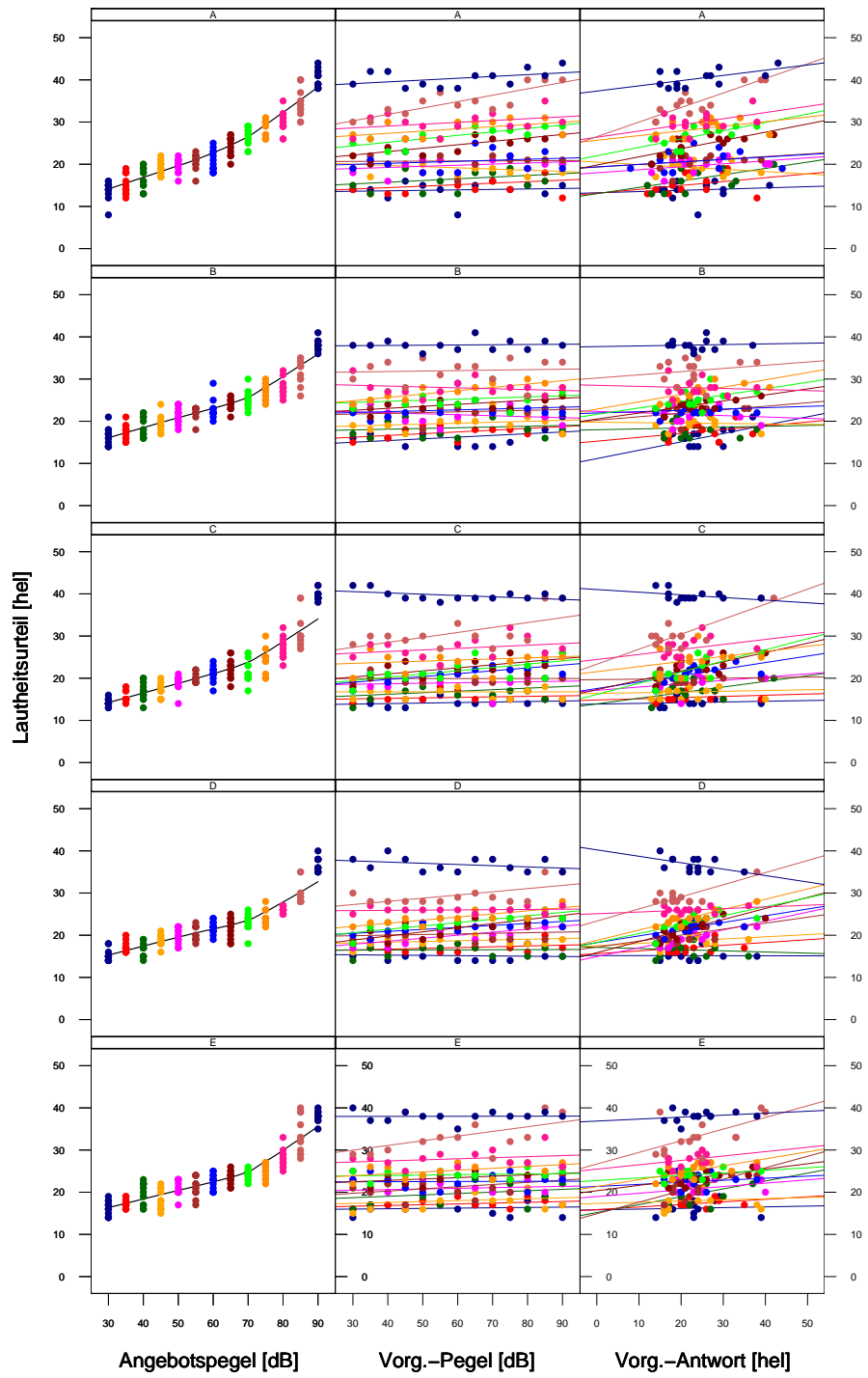


Proband 58

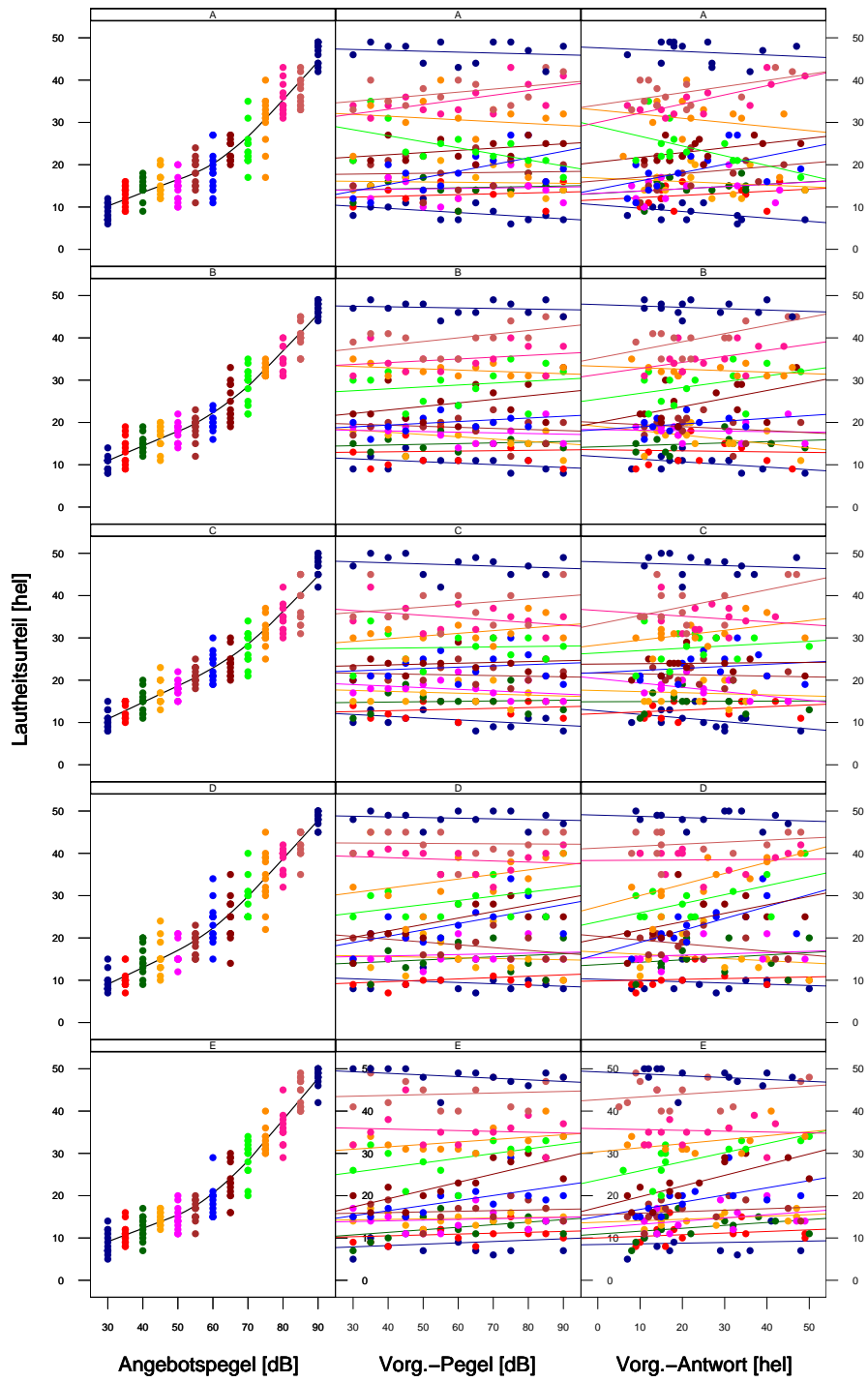


# B Skalierungsdaten und Vorgängerabhängigkeiten

## Proband 59

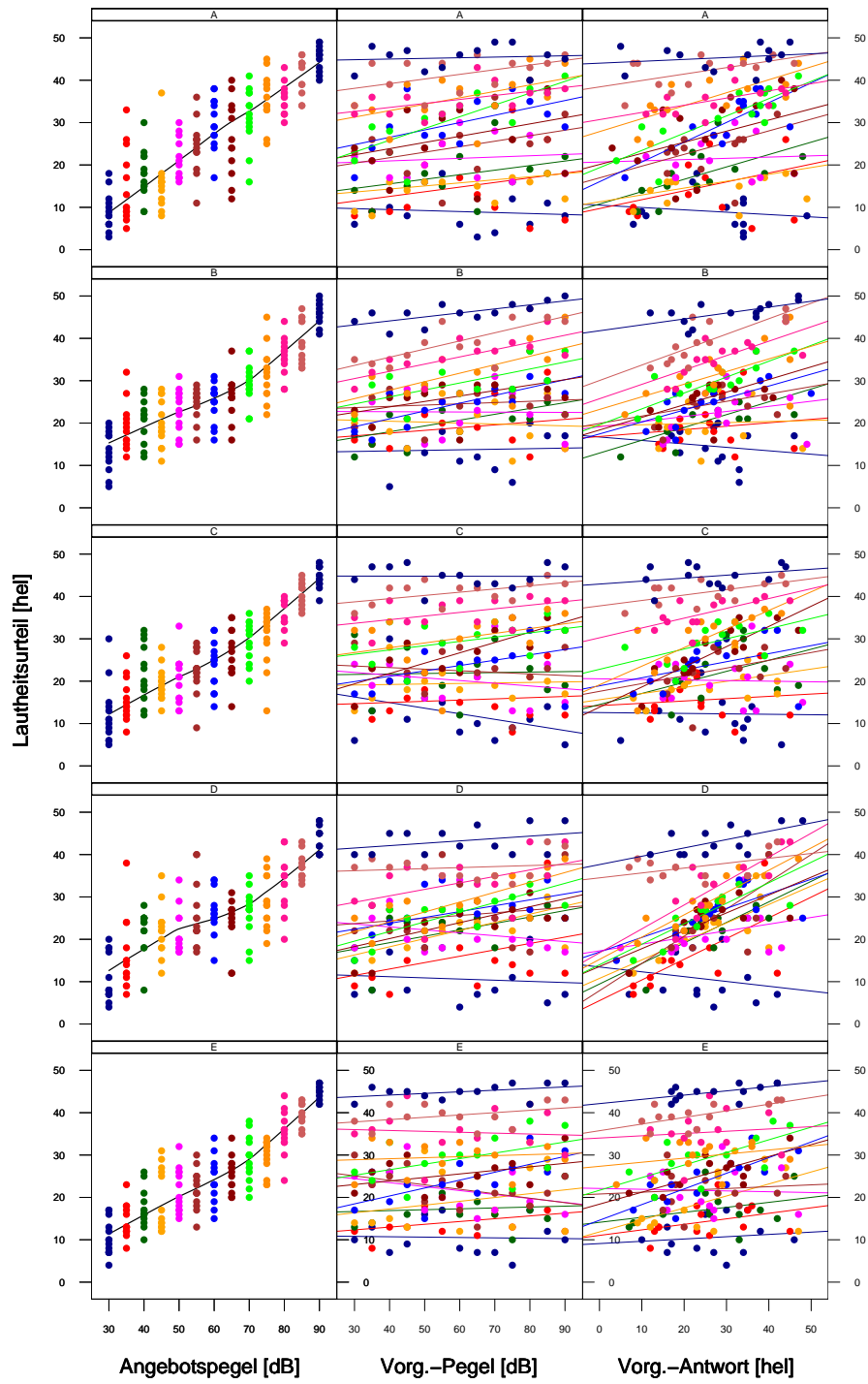


Proband 60



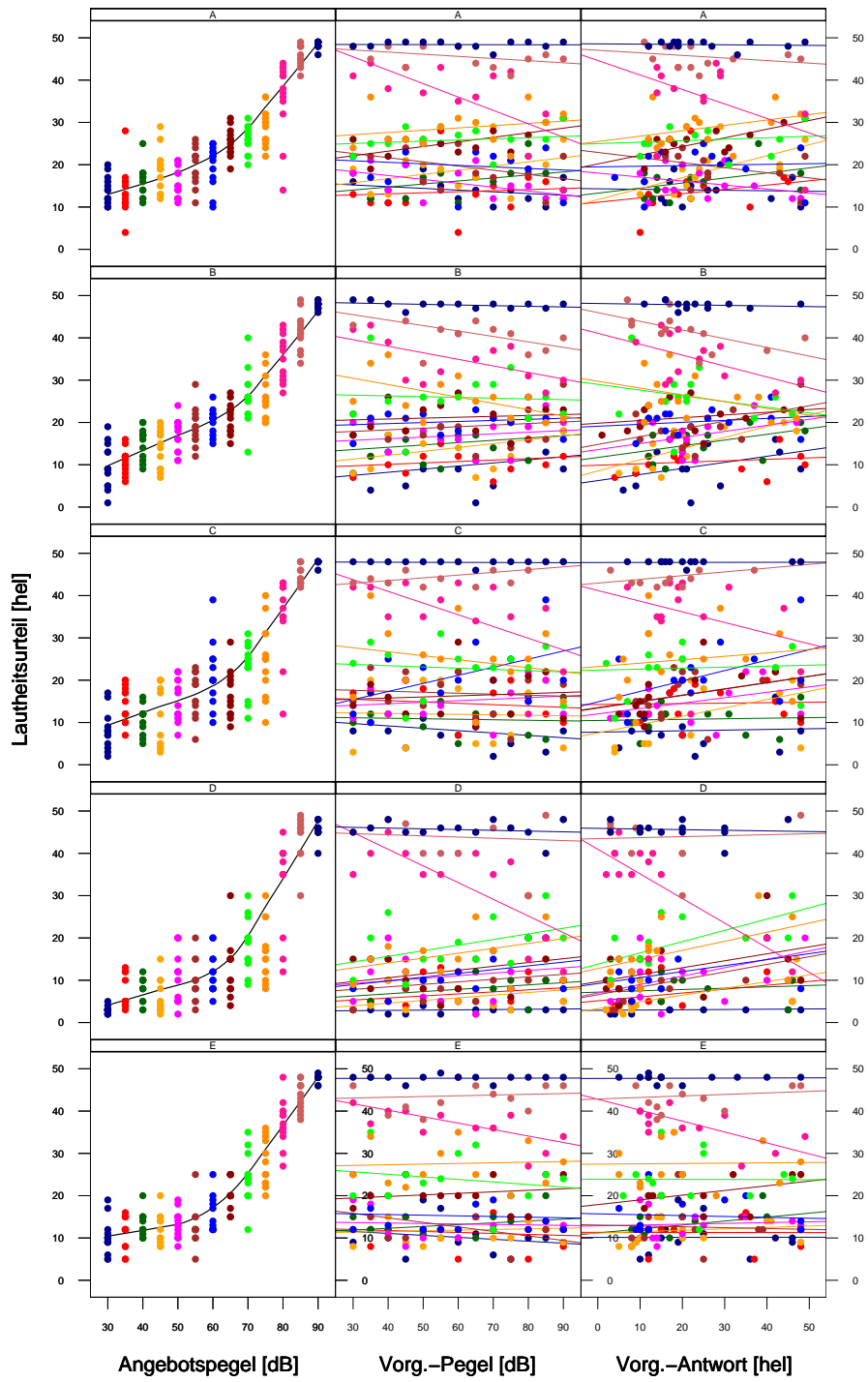
B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 61



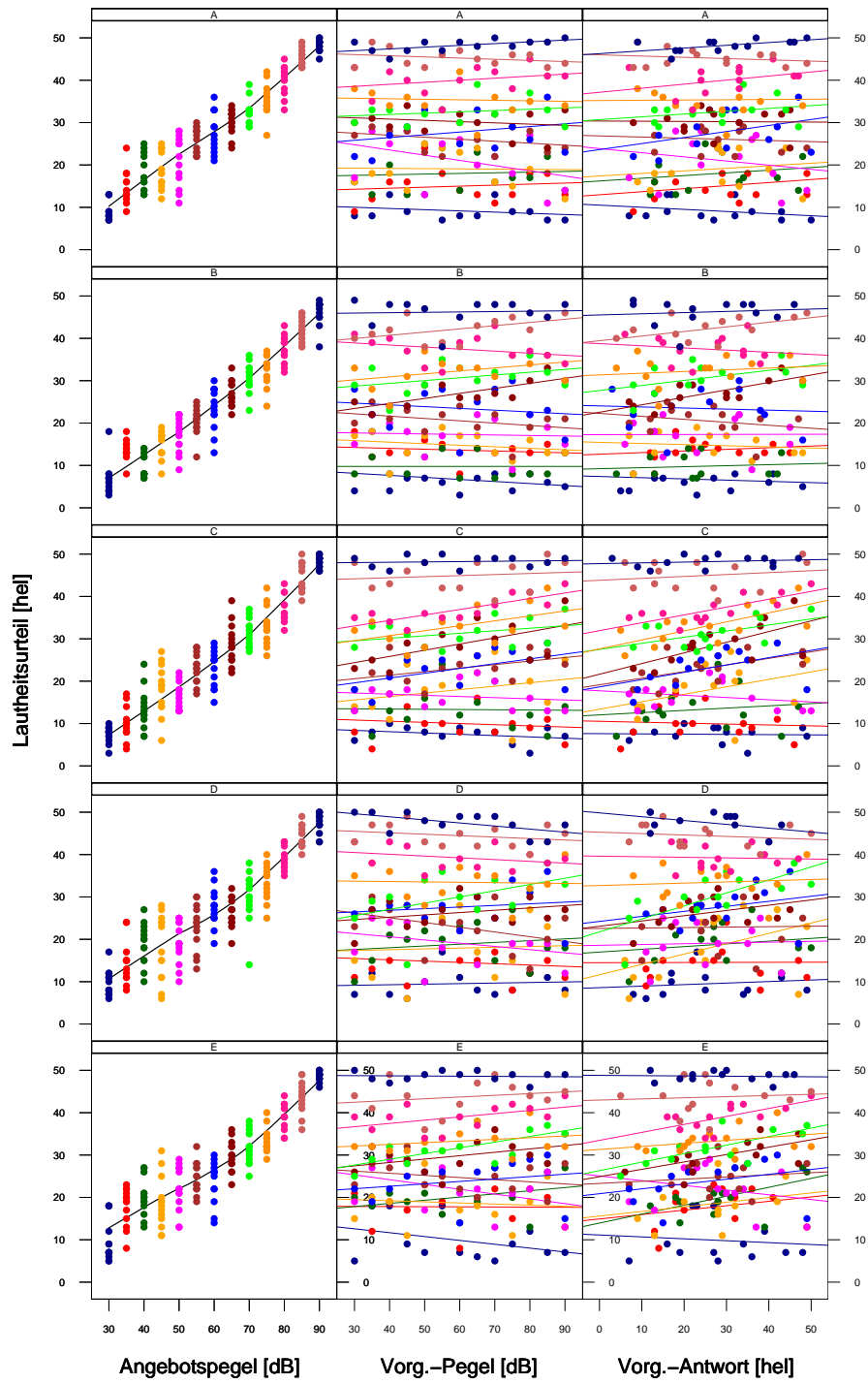


Proband 62



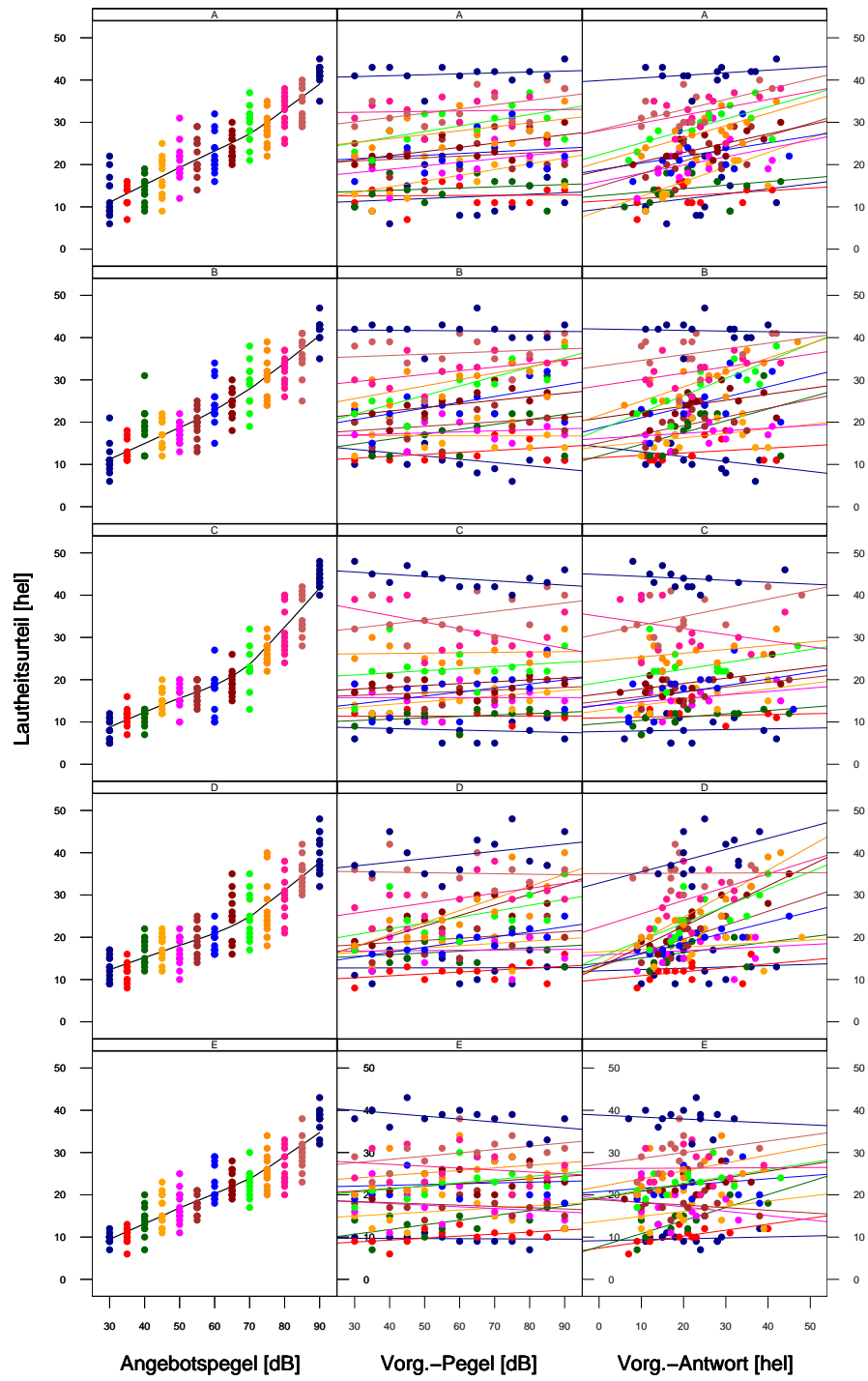
B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 63



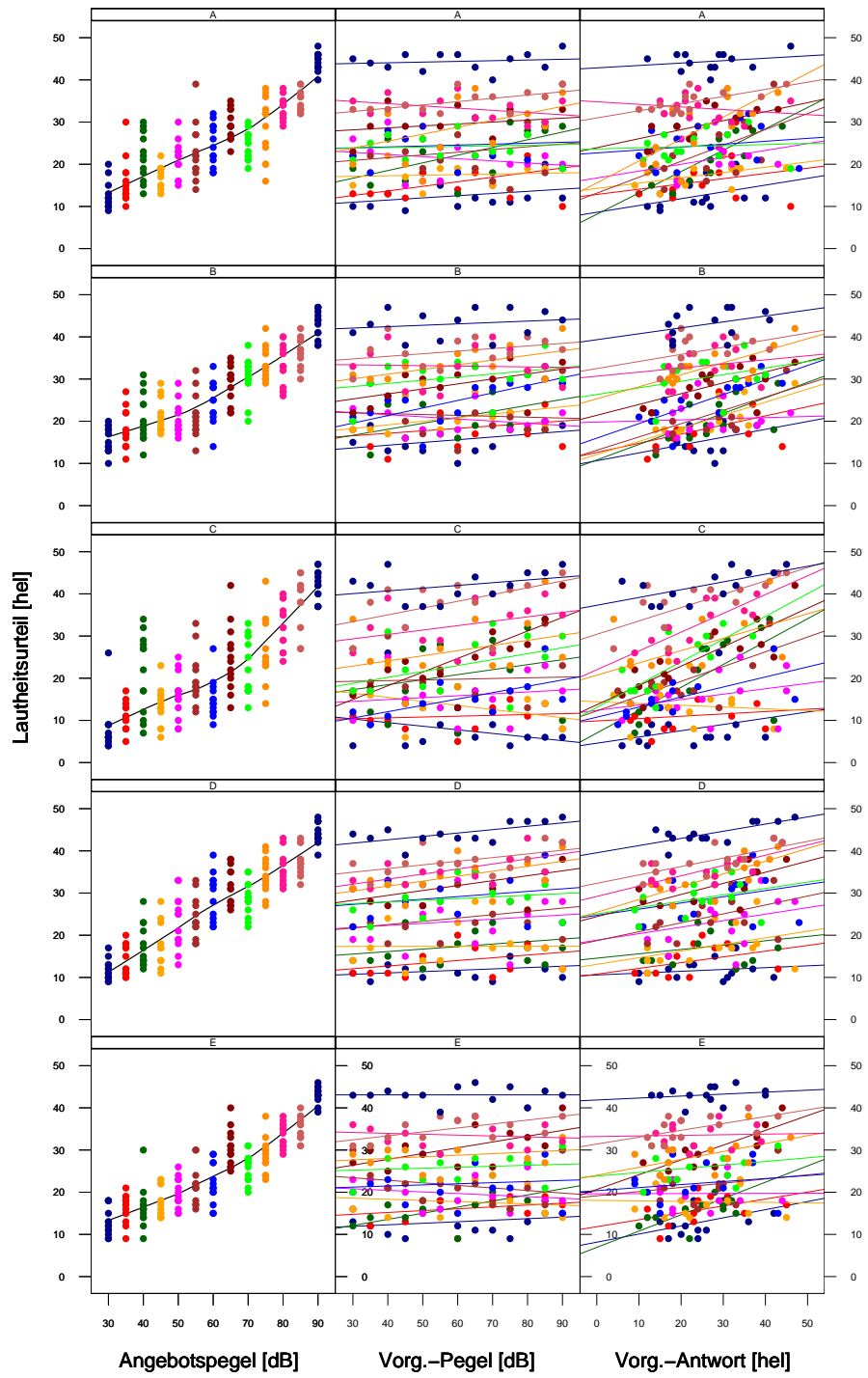
B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 64



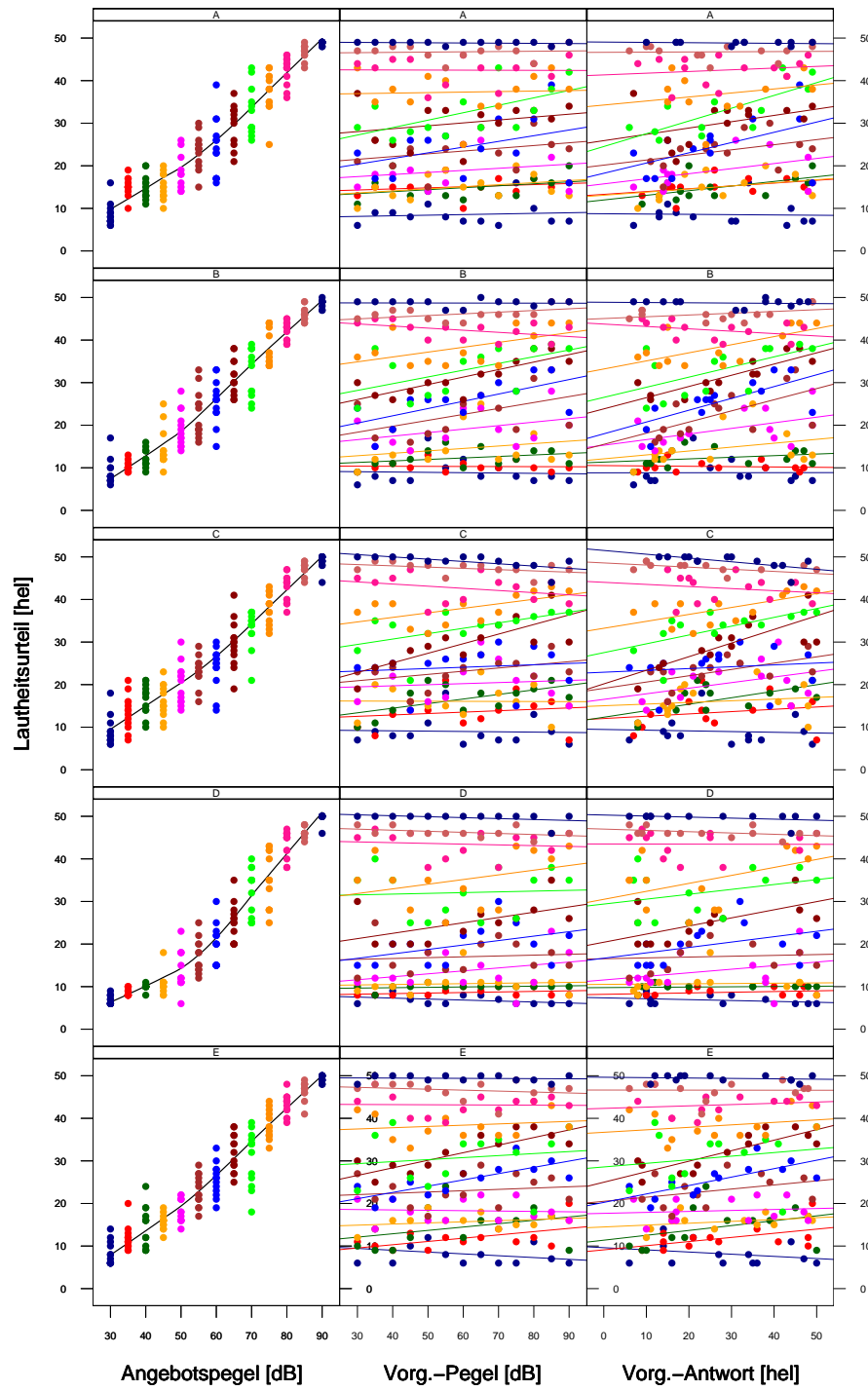
B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 65



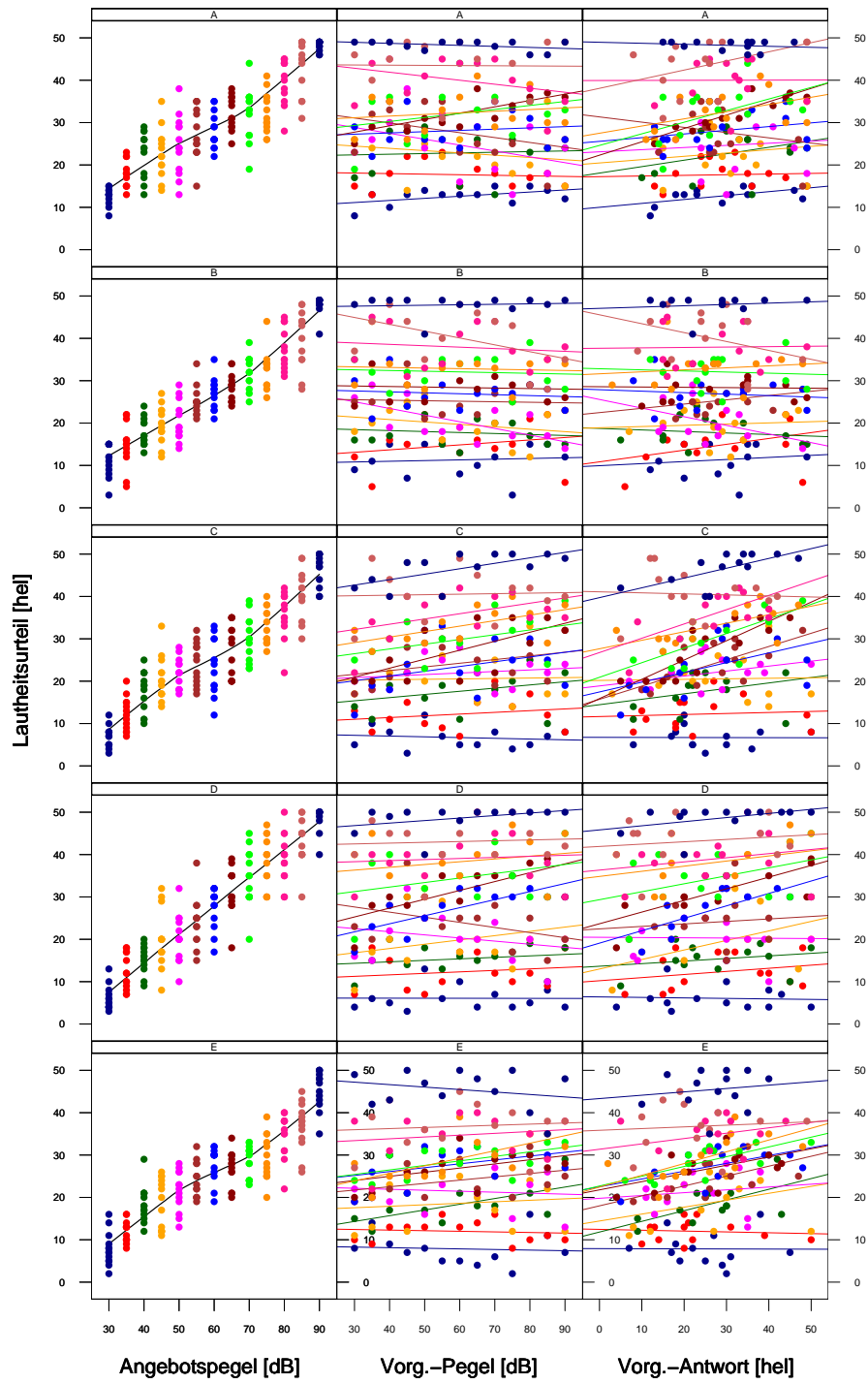
# B Skalierungsdaten und Vorgängerabhängigkeiten

## Proband 66

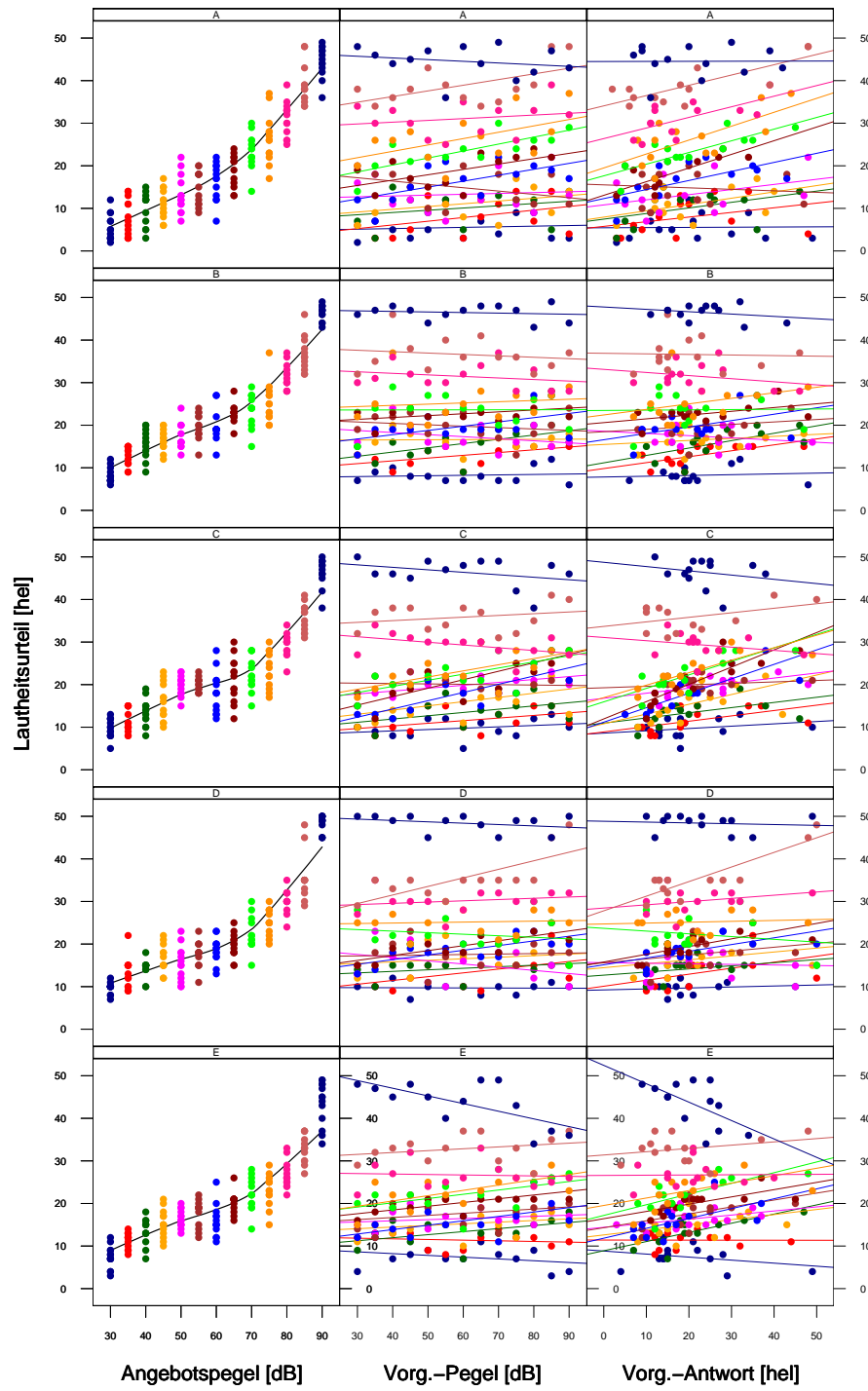


# B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 67

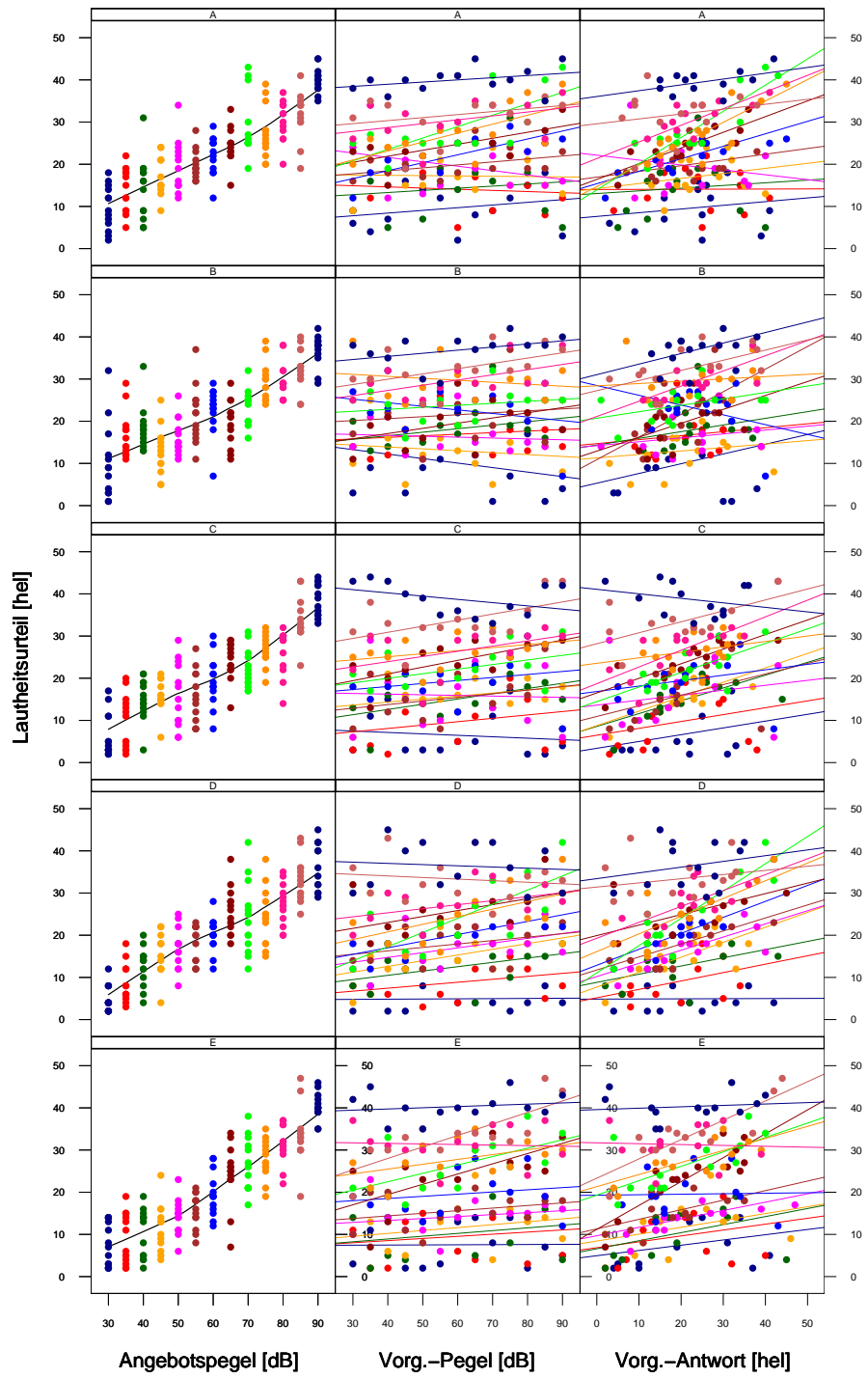


Proband 68



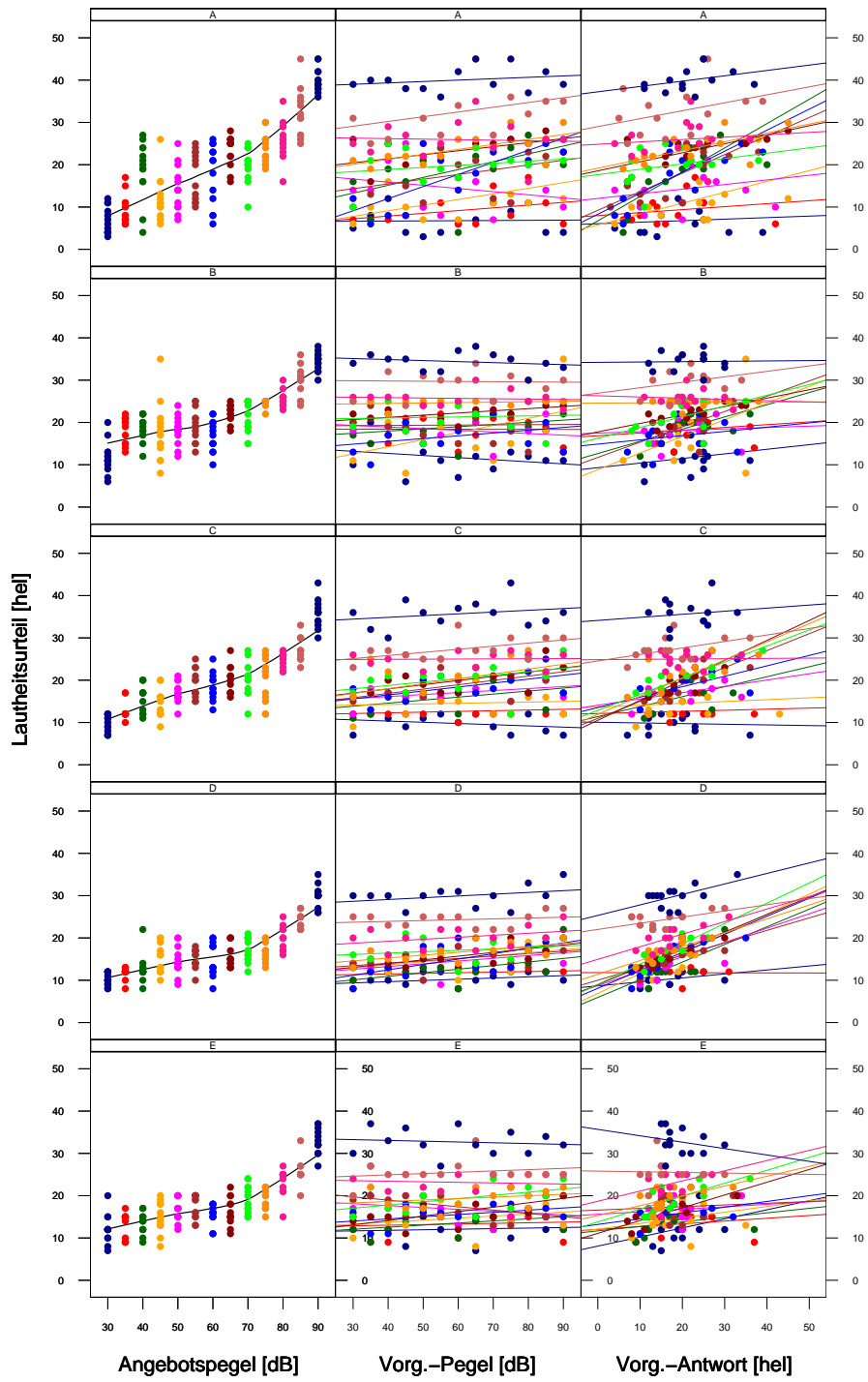
B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 69

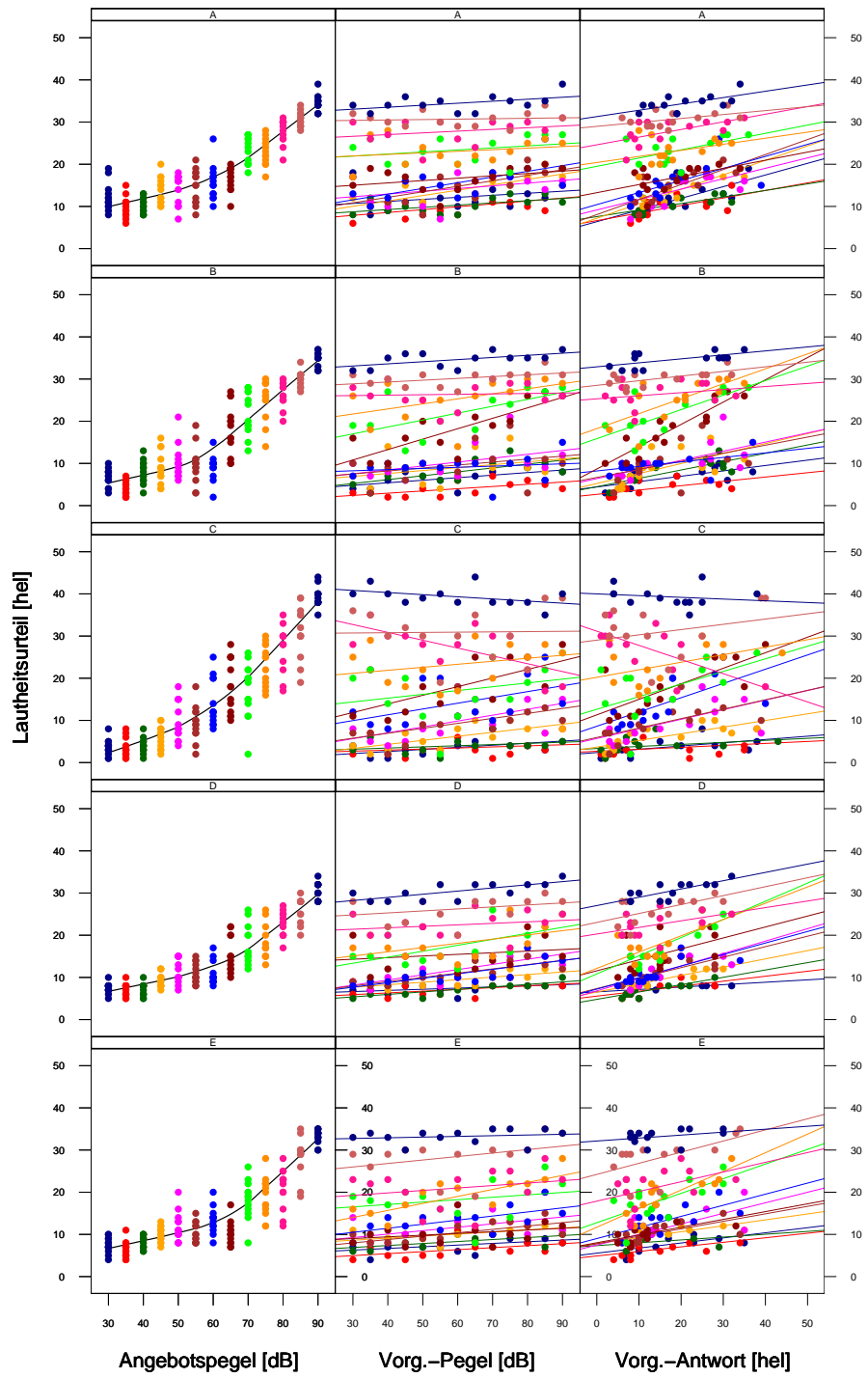




Proband 70

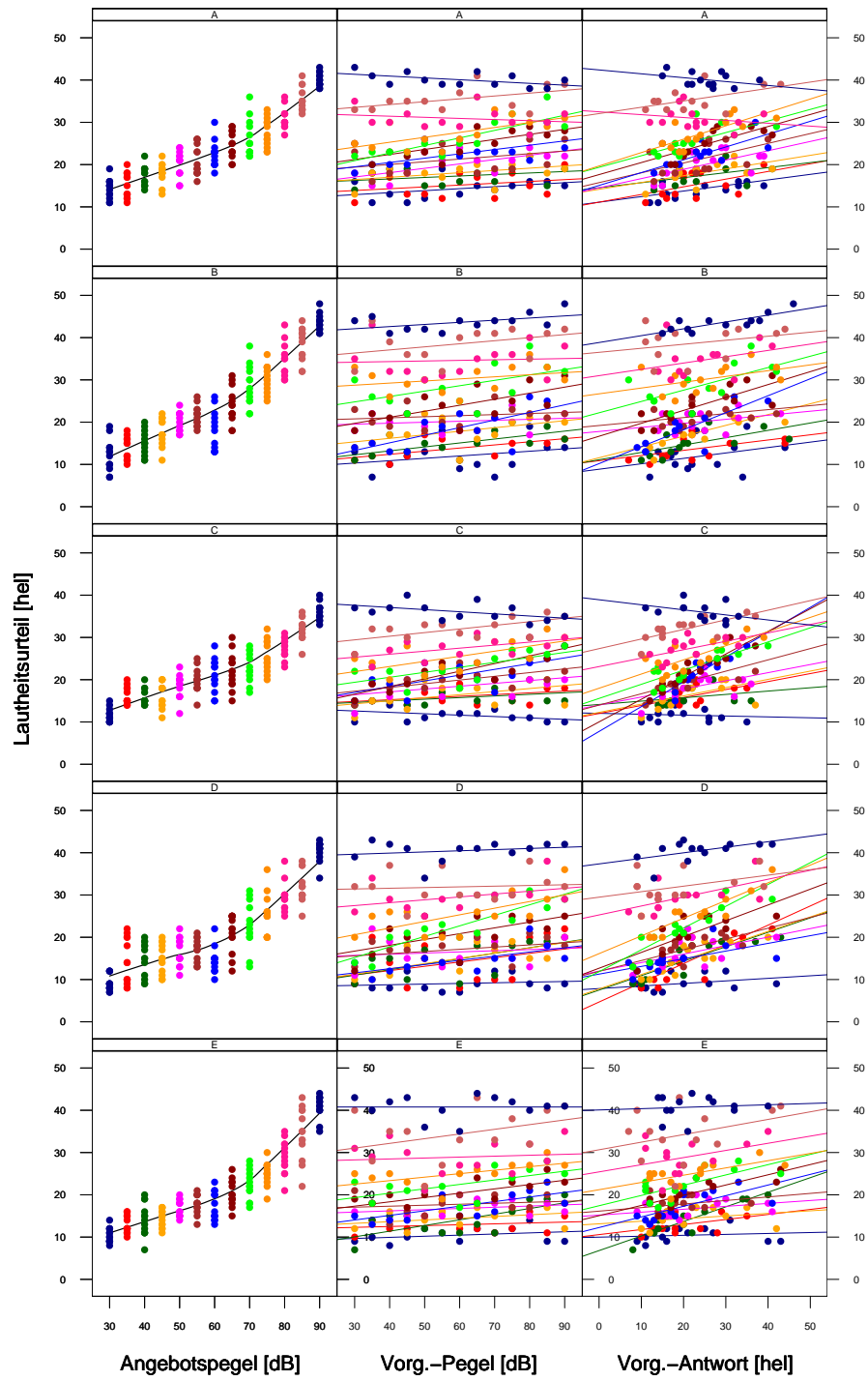


Proband 71



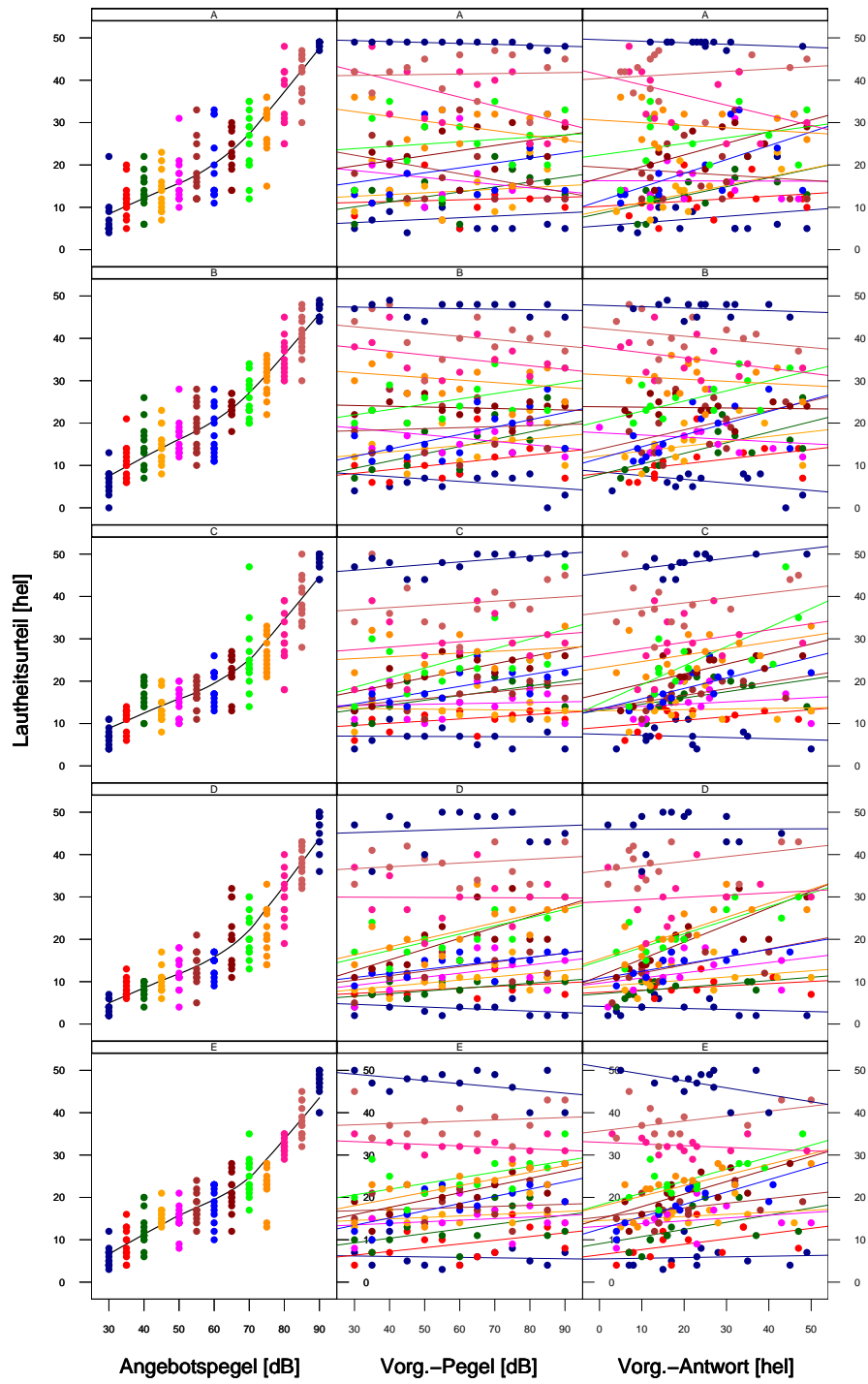
# B Skalierungsdaten und Vorgängerabhängigkeiten

## Proband 72

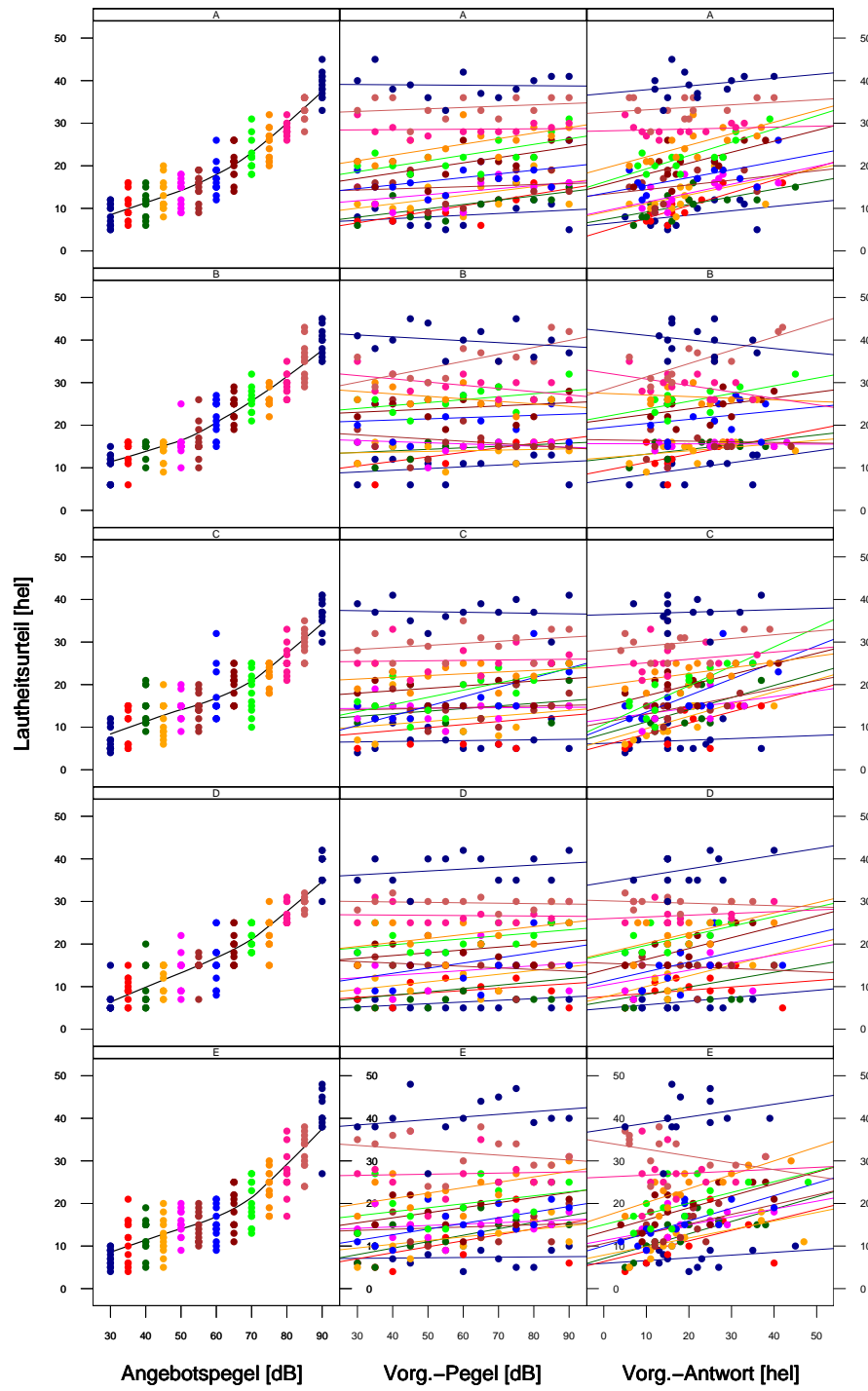


B Skalierungsdaten und Vorgängerabhängigkeiten

Proband 73

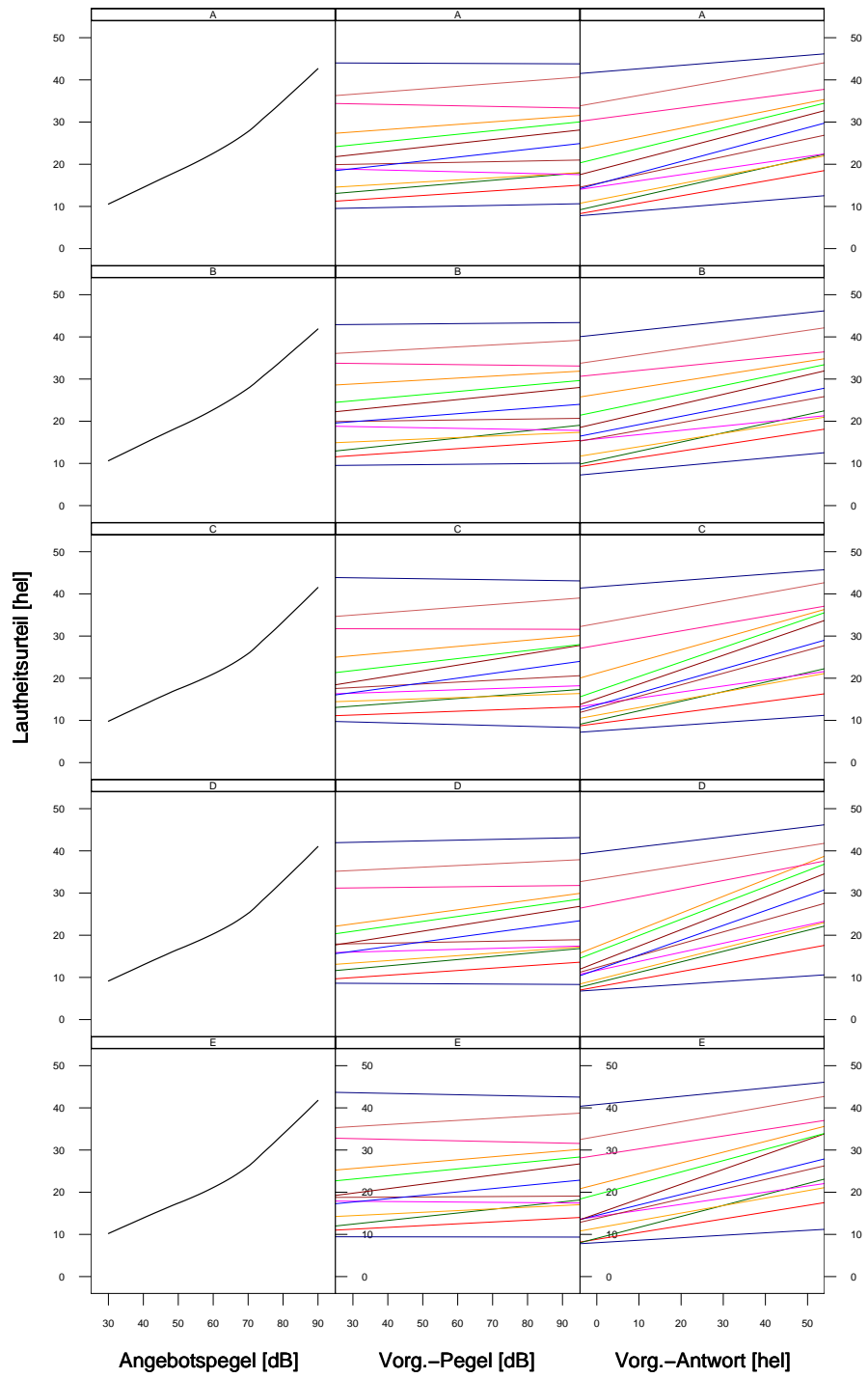


Proband 74



# B Skalierungsdaten und Vorgängerabhängigkeiten

## Alle Probanden Hauptgruppe



## Literaturverzeichnis

- Allen Jont B, Hall J L und Jeng P S. 1990. Loudness growth in 1/2-octave bands (LGOB) - A procedure for the assessment of loudness. *Journal of the Acoustical Society of America*, 88(2):745 – 753.
- Bender R und Lange St. 2001. Was ist ein Konfidenzintervall? *Deutsche Medizinische Wochenschrift*, 126:T41–T41.
- Bortz Jürgen und Lienert Gustav A. *Kurzgefasste Statistik für die klinische Forschung - Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*. Springer Verlag, Berlin Heidelberg New York, 2. Ausgabe, 2003.
- Brand Thomas, Hohmann Volker und Kollmeier Birger. Adaptive categorical loudness scaling. In Schick August und Klatt Maria, Hrsg., *Contributions to Psychological Acoustics. Results of the Seventh Oldenburg Symposium on Psychological Acoustics*, Seiten 603–610, 1997.
- Cox Robyn M, Alexander Genevieve C, Taylor Izel M und Gray Ginger A. 1997. The Contour Test of Loudness Perception. *Ear and Hearing*, 18(5):388–400.
- Crane Hewitt D. *IHC-Tm Connect-Disconnect and Mechanical Interaction among IHCs, OHCs, and TM*, Seiten 125–171. Jerry V Tobias and Earl D Schubert, New York, 1983.
- Cross David V. 1973. Sequential dependencies and regression in psychophysical judgments. *Perception & Psychophysics*, 14(3):547–552.
- DeCarlo L T und Cross D V. 1990. Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, 119:375–396.
- DeCarlo L T und Lawrence T. 1992. Intertrial Interval and Sequential Effects in Magnitude Scaling. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):1080–1088.

- Dondrup Michael. *Dokumentation: Verfahren zur Datennormalisierung mit EM-MA*. Universität Bielefeld, Fakultät für Biologie, Zentrum für Genomforschung, 2002.
- Ellermeier Wolfgang, Hellbrück Jürgen und Nowak Theo. 1985. Direkte Skalierung binauraler Lautheitssummutation: Ein neues psychophysikalisches Modell. *Psychologische Beiträge*, 27:509–519.
- Ellis M R und Wynne M K. 1999. Measurements of loudness growth in 1/2 octave bands for children and adults with normal hearing. *American Journal of Audiology*, 8:40–46.
- Fastl Hugo. 1987. Ein Störgeräusch für die Sprachaudiometrie. *Audiologische Akustik*, 26:2–13.
- Fucci Donald, Harris D und Petrosino L. 1986. Effect of stimulus pulse duration and interstimulus interval on cross-modal matching of auditory and lingual vibrotactile stimuli. *Perceptual and Motor Skills*, 63:1079–1086.
- Fucci Donald, Harris D, Petrosino L und M Banks. 1993. The effect of preference for rock music on magnitude-estimation scaling behavior in young adults. *Perceptual and Motor Skills*, 76:1171–1176.
- Fucci Donald, Petrosino Linda, McColl Doug, Wyatt Denise und Wilcox Corry. 1997. Magnitude estimation scaling of the loudness of a wide range of auditory stimuli. *Perceptual and Motor Skills*, 85:1059–1066.
- Gelfand Stanley A. *Essentials of Audiology*. Thieme, New York, Stuttgart, 2. Ausgabe, 2001.
- Gescheider G A und Hughson B A. 1991. Stimulus context and absolute magnitude estimation: a study of individual differences. *Perception & Psychophysics*, 50:45–57.
- Gocht Ursula Linares. *Würzburger Hörfeld Junior - Ein Verfahren zur Lautheitskalierung bei Kindern*. Dissertation, FU Berlin, 2002.
- Hartung Joachim, Elpelt Bärbel und Klösener Karl-Heinz. *Statistik: Lehr- und Handbuch der angewandten Statistik; mit zahlreichen, vollständig durchgerechneten Beispielen*. R. Oldenbourg Verlag, München, 11. Ausgabe, 1998.



- Haubensak Gert. *Absolutes und vergleichendes Urteil - Eine Einführung in die Theorie psychischer Bezugssysteme*. Springer Verlag, Berlin Heidelberg New York Tokyo, 1985.
- Haubensak Gert. 1992. Sequenzeffekte in absoluten Urteilen. Kritisches zur Methode. *Zeitschrift für experimentelle und angewandte Psychologie*, XXXIX (1):101–113.
- Hellbrück J, Gutwirth I und Ellermeier W. Fechner'sche Skala, Kategorienskala und Verhältnisskala. Experimentelle Beiträge zu einer Psychologisierung der Psychophysik. Technischer Bericht, Psychologisches Institut der Universität Würzburg, Lehrstuhl III, Würzburg, 1980.
- Hellbrück Jürgen. 1983. Geschlechtsspezifische Unterschiede in der Lautstärkeempfindung — Realität oder Artefakt? *Zeitschrift für experimentelle und angewandte Psychologie*, XXX(30):387–399.
- Hellbrück Jürgen. *Hören. Physiologie, Psychologie und Pathologie*. Hogrefe-Verlag, Göttingen, 1993.
- Hellbrück Jürgen und Heller Otto. Direkte Skalierung von Lautstärkedifferenzen. Technischer Bericht, Psychologisches Institut der Universität Würzburg, Lehrstuhl III, Würzburg, 1980.
- Hellbrück Jürgen und Moser Ludwig M. 1985. Hörgeräte-Audiometrie: Ein computerunterstütztes psychologisches Verfahren zur Hörgeräteanpassung. *Psychologische Beiträge*, 27:494–508.
- Hellbrück Jürgen, Oguey Marie-Christin und Seiler Christian. 1984. Sind geschlechtsspezifische Unterschiede in der Lautstärkenempfindung ein Artefakt der Gehörganggröße? *Zeitschrift für experimentelle und angewandte Psychologie*, XXXI(3):439–446.
- Hellbrück Jürgen und Thomamüller Daniela. 1997. Direkte Skalierung der Lautstärke bei normal hörenden Erwachsenen und Kindern mit einem kategorienfreien analogen Skalierungsverfahren. *Audiologische Akustik*, 36:64–78.
- Hellbrück Jürgen, Thomamüller Daniela und Zeitler Alfred. 1995. Hörfeldaudiometrie mit Gruppen normal hörender Personen in natürlichen Schallfeldern. *Audiologische Akustik*, 2:60–81.

- Heller Otto. *Orientierung innerhalb von phänomenalen Steigerungsreihen*, Seiten 107–136. Verlag Hans Huber, Bern Stuttgart Wien, 1980.
- Heller Otto. 1985. Hörfeldaudiometrie mit dem Verfahren der Kategorienunterteilung (KU). *Psychologische Beiträge*, 27:478–493.
- Heller Otto. Oriented category scaling of loudness and speech-audiometric validation. In Schick August, Hellbrück Jürgen und Weber Reinhard, Hrsg., *Contributions to Psychological Acoustics. Results of the Fifth Oldenburg Symposium on Psychological Acoustics*, Seiten 135–159, 1991.
- Hellstrom A. 1979. Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance*, 5(3):460–477.
- Hellstrom A. 2003. Comparison is not just subtraction: effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics*, 65(7):1161–1177.
- Helson H. *Adaptation-level theory. An experimental and systematical approach to behaviour*. Harper und Row, New York, 1964.
- Hohmann Volker und Kollmeier Birger, Hrsg. *Weiterentwicklung und klinischer Einsatz der Hörfeldskalierung*. Nr. 2 in Buchreihe Audiologische Akustik. median-verlag von Killisch-Horn GmbH, Heidelberg, 1995.
- Hosoi Hiroshi, Murata Kiyotaka und Imaizumi Satoshi. 1996. A new method of measurement of most comfortable loudness. *Audiology & Neuro-otology*, 1: 234–242.
- Hubner R und Ellermeier W. 1993. Additivity of loudness across critical bands: a critical test. *Perception & Psychophysics*, 54(2):185–189.
- Jesteadt Walt, Luce R Duncan und Green David M. 1977. Sequential Effects in Judgments of Loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1):92–104.
- Kaulitz Stefan. *Lautheitsbeurteilung unilateraler Cochlea-Implantat-Träger in Abhängigkeit von der Stimulusart und vom vorangehenden Stimulus bei sequenzieller Darbietung der Reize*. Dissertation, HNO-Universitätsklinik Würzburg, 2008.

- Kießling J. 2000. Hörflächenskalierung in der HNO-Heilkunde - Modeerscheinung oder unverzichtbare Methode? *HNO*, 3:167–169.
- Kießling J, Pfreimer C und Schubert M. 1996. Rekrutmentnachweis - Kategorial-Lautheitsskalierung und klassische überschwellige Audiometrie im Vergleich. *Laryngo-Rhino-Otologie*, 75:10–17.
- Kießling J, Schubert M und Wagner I. 1994. Lautheitsskalierung - Ein Verfahren zum quantitativen Rekrutmentnachweis. *HNO*, 42:350–357.
- Kießling Jürgen. 1995. Zum überschwelligen Lautheitsanstieg bei Schallempfindungsschwerhörigen - Konsequenzen für die Hörgeräte-Entwicklung und -Anpassung. *Audiologische Akustik*, 2:82–89.
- Kießling Jürgen, Steffens Thomas und Wagner Ina. 1993. Untersuchung zur praktischen Anwendbarkeit der Lautheitsskalierung. *Audiologische Akustik*, 4: 100–115.
- Kollmeier Birger, Hrsg. *Hörflächenskalierung — Grundlagen und Anwendung der kategorialen Lautheitsskalierung für Hördiagnostik und Hörgeräteversorgung*. Nr. 12 in Buchreihe Audiologische Akustik. median-verlag von Killisch-Horn GmbH, Heidelberg, 1997.
- Körner S. *in Arbeit*. Dissertation, HNO-Universitätsklinik Würzburg.
- Kugler Michael und Moser Ludwig M. 1995. Ein lineares Modell des kindlichen Hörfeldes gemessen mit dem Würzburger Hörfeld. *Audiologische Akustik*, 34: 162–170.
- Launer Stefan, Holube Inga, Hohmann Volker und Kollmeier Birger. 1996. Categorical loudness scaling in hearing impaired listeners - Can loudness growth be predicted from the audiogram? *Audiologische Akustik*, 35:156–163.
- Lauterbach Wolf und Sarris Viktor, Hrsg. *Beiträge zur psychologischen Bezugssystemforschung*. Verlag Hans Huber, Bern Stuttgart Wien, 1980.
- Lehnhardt Ernst und Laszig Roland. *Praxis der Audiometrie*. Thieme, Stuttgart; New York, 8 Ausgabe, 2001.
- Lockhead G R. 1986. Range and sequence effects on judgment. *Perception & Psychophysics*, 40(1):53–62.

- Marks Lawrence E. 1993. Contextual Processing of multidimensional and unidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2):227–249.
- Marks Lawrence E. 1994. “Recalibrating” the Auditory System: the Perception of Loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2):382–396.
- Marshall L und Brand J F. 1980. The relationship between loudness and reaction time in normal hearing listeners. *Acta Oto-Laryngologica*, 90(2):244–249.
- McCull Douglas und Fucci Donald. 1999. Comparison of magnitude estimation and interval scaling of loudness. *Perceptual and Motor Skills*, 88:25–30.
- Meister H, Walger M und Wedel von H. 2001. Hörfeldskalierung mit Kindern. *HNO*, 49:458–464.
- Meister Hartmut und Wedel von Hasso. Untersuchungen zur Lautheitsskalierung in der klinischen Diagnostik und Therapie. In *Fortschritte der Akustik. DAGA '95*, Seiten 199–202. DPG-Verlag, 1995.
- Moser Ludwig M. 1987. Das Würzburger Hörfeld, ein Test für prothetische Audiometrie. *HNO*, 35:318–321.
- Nägler Julia Therese. *Lautheitsbeurteilung in Abhängigkeit von der Stimulusart und vom vorangehenden Stimulus bei sequenzieller Darbietung der Reize*. Dissertation, HNO Würzburg, 2004.
- Nowak Theo. Der Mensch als Schallmessinstrument. Technischer Bericht, Psychologisches Institut der Universität Würzburg, Lehrstuhl III, Würzburg, 1980.
- Pascoe D P. 1978. An approach to hearing aid selection. *Hearing Instruments*, 29:12 – 16.
- Pascoe D P. 1986. Hörgeräte-Auswahlverfahren am Central Institute for the Deaf in Saint Louis. *AUDAK*, 25(3):90 – 106.
- Payne M C, Corso und Gregory M. 1985. Effects of range on category scaling of loudness judgments. *Perceptual and Motor Skills*, 60(2):619–624.
- Rau Harald und Pauli Paul. *Medizinische Psychologie/Medizinische Soziologie systematisch*. UNI-MED Verlag AG, Lorch/Württemberg, 1. Ausgabe, 1995.

- Ricketts Todd A. 1997. Clinical Use Of Loudness Growth Procedures. *The Hearing Journal*, 50(3):10–20.
- Ricketts Todd A und Bentler Ruth A. April 1996. The effect of test signal type and bandwidth on the categorical scaling of loudness. *Journal of the Acoustical Society of America*, 99(4):2281–2287.
- Rinne Horst. *Taschenbuch der Statistik*. Wissenschaftlicher Verlag Harri Deutsch, Frankfurt am Main, 3 Ausgabe, 2003.
- Robinson Ken und Gatehouse Stuart. 1996. Test-Retest Reliability of Loudness Scaling. *Ear and Hearing*, 17(2):120–123.
- Ruscetta Melissa N und Palmer Catherine V. 2000. Effect of Interstimulus Interval on Subjective Categorical Loudness Judgments. *Journal of the American Academy of Audiology*, 11:570–574.
- Sarris Viktor, Hrsg. *Wahrnehmung und Urteil*. Hogrefe, Göttingen, 1971.
- Schechter Gail und Buchsbaum Monte. 1973. The Effects of Attention, Stimulus Intensity and Individual Differences on the Average Evoked Response. *Psychophysiology*, 10(4):392–400.
- Schmidt Robert F und Schaible Hans Georg. *Neuro- und Sinnesphysiologie*. Springer Verlag, Heidelberg, 5 Ausgabe, 2006.
- Schorn K und Baumann U. 1999. Die Verbesserung der Hörgeräteanpassung mit digitalen Geräten und neuen Anpaßmethoden. *Laryngo-Rhino-Otologie*, 78:14–19.
- Staddon J E R, King Michael und Lockhead G R. 1980. On Sequential Effects in Absolute Judgment Experiment. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2):290–301.
- Stevens S S. 1955. The measurement of loudness. *Journal of the Acoustical Society of America*, 27:815–829.
- Stevens S S. 1956. The direct estimation of sensory magnitudes - loudness. *American Journal of Psychology*, 96:1–25.
- Stevens S S. 1957. On the psychophysical law. *Psychol. Rev.*, 64:153–181.

- The R Development Core Team . *R: A Language and Environment for Statistical Computing, Reference Index*. R Foundation for Statistical Computing, Vienna, Austria, 2.3.1 Ausgabe, 2006. ISBN 3-900051-07-0.
- Völter C, Shehata-Dieler W, Moser L, Dieler R und Helms J. 2001. Rekrutmentnachweis bei retrokochleären Raumforderungen mit der Lautheitsskalierung im Würzburger Hörfeld. *Laryngo-Rhino-Otologie*, 80:365–369.
- Voß Werner. *Taschenbuch der Statistik*. Fachbuchverlag Leipzig, 2. Ausgabe, 2004.
- Ward Lawrence M. 1979. Stimulus Information and Sequential Dependencies in Magnitude Estimation and Cross-Modality Matching. *Journal of Experimental Psychology: Human Perception and Performance*, 5(3):444–459.
- Ward Lawrence M. 1987. Remembrance of Sounds Past: Memory and Psychophysical Scaling. *Journal of Experimental Psychology: Human Perception and Performance*, 13(2):216–227.
- Ward Lawrence M und Lockhead G R. 1970. Sequential effect and memory in category judgments. *Journal of Experimental Psychology*, 84(1):27–34.
- Ward Lawrence M und Lockhead G R. 1971. Response system in absolute judgment. *Perception & Psychophysics*, 9(1B).
- Westra GmbH . *Gerätebeschreibung für das Sprachaudiometer CAD 03/1*, a.
- Westra GmbH . *Hörgeräteanpassung, einfach, schnell, erfolgreich.*, b.
- Witte Wilhelm. 1960. Struktur, Dynamik und Genese von Bezugssystemen. *Psychologische Beiträge*, 4 a:218–252.
- Zwicker E und Fastl H. *Psychoacoustics. Facts and Models*. Nr. 22 in Springer Series in Information Sciences. Springer Verlag, 2. Ausgabe, 1990.
- Zwicker E und Feldtkeller R. *Das Ohr als Nachrichtenempfänger*. Nr. XIX in Monographien der elektrischen Nachrichtentechnik. S. Hirzel-Verlag, Stuttgart, 2. Ausgabe, 1967.
- Zwislocki J J. 1983. Absolute and other scales: Question of validity. *Perception & Psychophysics*, 33(6):593–594.

Zwislocki J J und Goodman D A. 1980. Absolute scaling of sensory magnitudes:  
A validation. *Perception & Psychophysics*, 28(1):28–38.





## Danksagung

An erster Stelle möchte ich dem ehemaligen Direktor der Klinik und Poliklinik für Hals-, Nasen- und Ohrenkranke der Universität Würzburg Prof. Dr. Jan Helms, dem amtierenden Direktor Prof. Dr. Rudolf Hagen sowie Herrn Prof. Dr. Joachim Müller für die Überlassung dieser Arbeit herzlich danken.

Herrn Dipl.-Ing. Stefan Brill gilt mein besonderer Dank. Ohne ihn wäre diese Arbeit nicht zustande gekommen. Von ihm stammen die Überlegungen und Ideen, die der hier vorliegenden Arbeit zugrunde liegen. Deren konsequente Umsetzung in einen logischen und einwandfreien Versuchsaufbau wären mir ohne ihn nicht möglich gewesen. Seine Hilfe in Planung und Durchführung der Versuche sowie in der Auswertung und Interpretation der Daten war unersetzlich. Die längeren Nächte im CI-Labor werde ich sicher nicht vergessen.

Weiterhin danke ich Dipl.-Ing. (FH) Alexander Möltner für die technische Betreuung. Ihm verdanke ich auch die Rettung meines Laptops und somit die Ersparnis einigen Ärgers.

Allen weiteren Mitarbeitern der Cochlea-Implantat-Abteilung sowie den Mitarbeiterinnen der Audiometrieabteilung danke ich für die stets freundliche und hilfreiche Unterstützung. Auch all den Anderen, die mich in meiner Arbeit unterstützt haben und hier nicht extra erwähnt wurden, sei hiermit herzlich gedankt.