

Technische Universität Berlin
Fakultät I: Geisteswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation

Erstellung einer Softwarebibliothek für Hörversuche - Programmkonzept und zu implementierende Testverfahren

Magisterarbeit

vorgelegt von: Simon Ciba
geboren am 20. Dezember 1979
in Miltenberg
Matrikel-Nr. 229984

Betreuung: Prof. Dr. Stefan Weinzierl
Dr. Hans-Joachim Maempel

20. November 2008

Die selbständige und eigenhändige Ausfertigung versichert an Eides statt,
Berlin, den 20. November 2008

.....

Simon Ciba

Inhaltsverzeichnis

Inhaltsverzeichnis	5
1 Einleitung	9
2 Gliederung des Entwicklungsprozesses	11
2.1 Implikationen der Software-Ergonomie	11
2.2 Überblick der Entwicklungsschritte und Aufbau dieser Arbeit	12
3 Methodische Aspekte von Hörversuchen	15
3.1 Der experimentelle Hörversuch	15
3.1.1 Unabhängige Variablen	15
3.1.2 Abhängige Variablen	17
3.1.3 Kontrolle von Störeinflüssen	18
3.2 Versuchsdesign	23
3.3 Testverfahren	23
3.4 Familiarisation	27
4 Zu implementierende Testverfahren	31
4.1 Adaptive psychophysische Verfahren	31
4.1.1 Einleitung	31
4.1.2 Relevante Aspekte der psychophysischen Theorie	32
4.1.2.1 Psychometrische Funktion	32
4.1.2.2 Unterschieds- vs. Absolutschwellenmessung	32
4.1.2.3 Antwortparadigmen	33
4.1.2.4 Konventionelle Ratekorrektur	34
4.1.2.5 Logistische Funktion	36
4.1.3 Prinzipielle Funktionsweise und grundlegende Begriffe	38
4.1.4 Annahmen und Klassifizierung	39
4.1.5 Staircase-Verfahren	40
4.1.5.1 Einfache Staircase-Verfahren	40
4.1.5.2 Transformierte Staircase-Verfahren	43
4.1.6 PEST-Verfahren	44

4.1.6.1	Original PEST-Verfahren	45
4.1.6.2	More-Virulent-PEST	48
4.1.7	Best-PEST-Verfahren	49
4.1.7.1	Adaptionsmechanismus	50
4.1.7.2	Terminierung und Ergebnis	54
4.1.7.3	Initialisierung des Verfahrens	56
4.1.7.4	Wahl des Spread-Parameters	58
4.1.7.5	Wahl der Lapsing-Rate	58
4.1.8	Interleaving	59
4.2	Semantisches Differential	61
4.2.1	Einleitung	61
4.2.2	Grundzüge des Verfahrens	61
4.2.3	Operationalisierung des Urteils	63
4.2.4	Wahl der Objekte	63
4.2.5	Wahl der Skalen	64
4.2.5.1	Skalenformat	64
4.2.5.2	Wahl der Adjektivpaare	65
4.3	Repertory-Grid-Technik	67
4.3.1	Einleitung	67
4.3.2	Ablauf des Verfahrens	67
4.3.2.1	Erhebung der Konstrukte durch Triadenvergleiche . .	67
4.3.2.2	Beurteilung der Elemente auf den erhobenen Kon- strukten	69
4.3.3	Auswahl der Elemente	69
4.3.4	Auswahl und Darbietung der Triaden	71
4.3.5	Kritik, Varianten und Modifikationen	71
5	Entwicklung eines Programmkonzepts	73
5.1	Begriffsbestimmung	73
5.2	Nutzungskontext und Anforderungen	73
5.3	Darstellung des entwickelten Programmkonzepts	75
6	Implementierung der Adaptionsmechanismen und Schätzmethode	79
6.1	Die MATLAB-Funktionen im Überblick	79
6.2	Aspekte der praktischen Umsetzung	80
6.3	Beschreibung der Eingabe- und Ausgabeparameter	82
6.3.1	staircaseXd1u.m (X=1, 2, 3)	82

6.3.2	PEST.m und mvPEST.m	83
6.3.3	MLBayes.m	83
6.4	Test durch Simulation	85
7	Entwurf der grafischen Benutzeroberflächen	89
8	Zusammenfassung und Ausblick	91
	Literaturverzeichnis	93
	Abbildungsverzeichnis	99
	Tabellenverzeichnis	101
	Anhang	102
A	Quellcode	103
A.1	MATLAB-Funktionen	103
A.2	Skripte für den Funktionstest durch Simulation	120
B	Benutzerdokumentation der Softwarebibliothek	127

1 Einleitung

Als einer der wesentlichen Aspekte der Audiokommunikation steht die menschliche Wahrnehmung von Schallereignissen im Fokus wissenschaftlicher Fragestellungen. Dabei können Wahrnehmungsmechanismen selbst, die Wirkung von Audioinhalten oder der Einfluss bestimmter Übertragungstechniken Gegenstand des Interesses sein (Maempel, 2007). Der Zugriff auf Wahrnehmungsinhalte erfolgt durch empirische Untersuchungen in Form von Hörversuchen.

Bei deren Durchführung spielen Computer immer häufiger eine unverzichtbare Rolle. Zum einen ermöglichen sie überhaupt erst die Realisierung bestimmter experimenteller Anordnungen, etwa wenn die Aufgaben im Zuge der Interaktion mit der Versuchsperson und der Bedienung der Wiedergabeperipherie zu komplex werden, um noch in Echtzeit durch einen Versuchsleiter „von Hand“ ausgeführt zu werden. Zum anderen vereinheitlichen sie das Vorgehen im Hinblick auf Instruktion, Datenerfassung und Ablaufsteuerung, wodurch der persönliche Einfluss des Versuchsleiters auf die erhobenen Daten geringer wird und die Anzahl potentieller Fehlerquellen reduziert werden kann.

Voraussetzung für die Durchführung von Hörversuchen mit Computern ist eine geeignete Software. Eine Reihe verschiedener Programme ist kommerziell oder frei erhältlich (s. z.B. Bech und Zacharov, 2006, Kap. 9.2.3). Die meisten Programme decken jeweils nur eine kleine Gruppe eng verwandter Testverfahren ab. Eine universitäre Forschungseinrichtung wie das FG Audiokommunikation, welche zahlreiche unterschiedliche Versuchsanordnungen verwendet, ist somit angehalten, auf nahezu ebenso viele verschiedene Programme zurückzugreifen. Für die Praxis in Lehre und Forschung bedeutet das in der Konsequenz ein Auseinandersetzen mit deren jeweiligen Eigenheiten, welche u.a. durch unterschiedliche Benutzer-, Peripherie- und Datenschnittstellen, und nicht zuletzt auch System- und Hardwareanforderungen gekennzeichnet sind.

Eine weitere Schwierigkeit ist die Anpassung von Software an besondere Erfordernisse der Anwendung, wie sie z.B. bei Verwendung spezieller Wiedergabeperipherie entstehen. So sollen am FG Audiokommunikation zukünftig die dort entwickelten Systeme

zur binauralen Echtzeitfaltung und zur Wellenfeldsynthese in Hörversuchen eingesetzt werden. Eine Manipulation der Software dahingehend setzt die Verfügbarkeit des Quellcodes voraus. Zumindest für kommerzielle Programme ist dies in der Regel nicht gegeben.

Diese und andere Probleme haben bereits in der Vergangenheit dazu geführt, dass die Hörversuchssoftware überwiegend eigens angefertigt wird. Dabei handelt es sich meist um Matlab-Skripte, welche für den jeweiligen Einzelfall konzipiert sind. In der Regel wird dabei die Konfiguration des Versuchs auf Skriptebene vorgenommen, auf eine grafische Benutzeroberfläche wird meist verzichtet.

In Anlehnung hieran soll nun ebenfalls unter Matlab eine Software entstehen, welche zumindest einige der genannten Nachteile nicht mehr aufweist. Diese soll einerseits dem Prinzip einer Softwarebibliothek bzw. Toolbox gerecht werden, indem sie eine Ansammlung von Skripten mehrerer Testverfahren enthält und somit offen für Erweiterungen durch neue Testverfahren ist. Andererseits sollen diese Skripte sowohl intern möglichst konsistent aufgebaut, als auch extern in eine sinnvoll übergeordnete Programmstruktur eingebettet sein. Eine grafische Benutzeroberfläche soll dem Entwicklungsobjekt schließlich den Charakter einer in sich geschlossenen und möglichst einfach bedienbaren Software verleihen.

Aufgrund seines Umfangs wird das beschriebene Entwicklungsprojekt in zwei Teilprojekte untergliedert, welche in etwa zeitgleich jeweils im Rahmen einer Masterarbeit durchgeführt werden. Die Entwicklung eines Programmkonzepts und die Aufbereitung der Theorie der zu implementierenden Testverfahren für die praktische Umsetzung stellen die Aufgaben der vorliegenden Arbeit dar. Die Implementierung und die Strukturierung des Quellcodes sind die Inhalte der Arbeit von André Wlodarski (Wlodarski, 2008). Eine strikte Trennung der Arbeiten ist natürlich aufgrund bestehender Abhängigkeiten weder möglich noch wünschenswert. Bestimmte Entwicklungsschritte erfordern eine enge Zusammenarbeit. Die Verteilung der einzelnen Aufgaben und der sich hieraus ergebende inhaltliche Aufbau der vorliegenden Arbeit werden in Abschnitt 2.2 beschrieben. Zuvor erfolgt in Abschnitt 2.1 eine kurze Darlegung softwareergonomischer Implikationen, welche mit dem Ziel einer möglichst hohen Gebrauchstauglichkeit für die Planung und Gliederung des Entwicklungsprozesses Beachtung finden sollen.

2 Gliederung des Entwicklungsprozesses

2.1 Implikationen der Software-Ergonomie¹

Software-Ergonomie ist die Wissenschaft von den Gesetzmäßigkeiten, welche leicht verständlicher und gut zu bedienender Software zugrunde liegen. Sie beschäftigt sich mit der Anpassung von Software an die kognitiven Fähigkeiten des Menschen und dabei insbesondere mit der Darstellung von Information und der Gestaltung von Arbeitsabläufen. Als ein Gütemerkmal für die diesbezügliche Bewertung von Software wird die *Gebrauchstauglichkeit* (engl. *usability*) in der Norm DIN EN ISO 9241-11 definiert:

„Das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen.“

Das Kriterium der Effektivität (Wirksamkeit) beinhaltet, dass die Software den für einen bestimmten Nutzungskontext geforderten Funktionsumfang vollständig und fehlerfrei zu Verfügung stellt. Es ist sowohl für die Entwicklung der internen Softwarestruktur (*Back-End*) als auch der Benutzerschnittstelle (*Front-End*) von grundlegender Bedeutung. Hingegen ist die Effizienz als der bei der Bedienung zu betreibende Aufwand des Benutzers insbesondere von der Gestaltung der Benutzerschnittstelle abhängig. Bestehen hinsichtlich der genannten Kriterien keine Beeinträchtigungen, so kann der Benutzer eine positive Einstellung gegenüber der Nutzung der Software entwickeln und hinsichtlich seiner Erwartungen zufriedengestellt werden.

Ein mögliches Vorgehen für die Entwicklung einer in diesem Sinne gebrauchstauglichen Software ist der *benutzerorientierte Gestaltungsprozess*, wie er in der Norm DIN EN ISO 13407 beschrieben wird. Dieser setzt sich aus vier aufeinanderfolgenden sogenannten Gestaltungsaktivitäten zusammen, welche in der Regel mehrmals zyklisch durchlaufen werden müssen, bis das Entwicklungsobjekt die gestellten Anforderungen erfüllt:

¹Eine ausführlichere Darstellung zum Thema Software-Ergonomie findet sich bereits in der erst vor kurzem am Fachgebiet angefertigten Magisterarbeit von H.-J. Mond (2007).

1. Verstehen und Festlegen des Nutzungskontexts;
2. Festlegen von Benutzeranforderungen und organisatorischen Anforderungen;
3. Entwerfen von Gestaltungslösungen;
4. Beurteilen von Gestaltungslösungen gegenüber Anforderungen.

Für eine detaillierte Beschreibung der einzelnen Schritte sei auf die Norm (ebd., S. 5–10) verwiesen. Das Ausmaß, in welchem die dort dargestellten Gestaltungsaktivitäten anzuwenden sind, hängt vom jeweiligen Kontext ab (ebd., S. 4). Grundsätzliche Merkmale des Vorgehens sind die Erarbeitung eines klaren Verständnisses von Benutzer- und Aufgabenanforderungen, eine aktive Beteiligung des Benutzers und die Iteration von Gestaltungslösungen. Mit dem Aspekt der Iteration ist gemeint, dass bereits vorläufige Gestaltungslösungen Benutzern in Form von Modellen oder Prototypen präsentiert werden und deren Rückmeldungen in schrittweise verbesserte Lösungen einfließen.

Wenngleich im vorliegenden Projekt während der Durchführung der Masterarbeiten eine Evaluierung unter Einbeziehung der zukünftigen Benutzer aufgrund des damit verbundenen Zeitaufwands nicht möglich ist, besitzen dennoch beide Autoren grundlegende Erfahrung mit der computergestützten Durchführung von Hörversuchen. Während der Entwicklungsarbeiten nehmen sie daher jeweils sowohl den Standpunkt des Entwicklers als auch den eines – zumindest potentiellen – Benutzers ein. Nach Abschluss der Masterarbeiten soll der erstellte Prototyp in die Praxis von Lehre und Forschung eingeführt werden. Die hierbei gewonnenen Erfahrungen können für weitere Verbesserungen herangezogen werden.

2.2 Überblick der Entwicklungsschritte und Aufbau dieser Arbeit

Wie aus dem vorherigen Abschnitt hervorgeht, steht in einem benutzer-orientierten Gestaltungsprozess die Erarbeitung eines Verständnisses des Nutzungskontexts zu Beginn der Entwicklungsarbeiten. Im vorliegenden Projekt wird daher als Teil dieser Masterarbeit mit der Aufbereitung allgemeiner methodischer Aspekte begonnen, welche für die Durchführung von Hörversuchen im Kontext wissenschaftlicher Fragestellungen von grundlegender Bedeutung sind (Kap. 3). Anschließend erfolgt eine eingehende Erörterung von Funktionsweisen und Eigenschaften der für die Implementierung vorgesehenen Testverfahren (Kap. 4). Letztere bestehen aus verschiedenen *adaptiven psychophysischen Verfahren* (Abs. 4.1), dem *semantischen Differential*

(Abs. 4.2) und der *Repertory-Grid-Technik* (Abs. 4.3). Auf der Grundlage der auf diese Weise zusammengetragenen Informationen und unter Berücksichtigung von Benutzeranforderungen erfolgt eine zusammenfassende Analyse des Nutzungskontexts und der an die Software gestellten Anforderungen (Abs. 5.2). Davon ausgehend wird ein Konzept im Sinne einer anwenderseitigen Programmstruktur entwickelt (Abs. 5.3). Die Ermittlung technischer Anforderungen und die Implementierung des Back-Ends sind Gegenstand der Arbeit von André Wlodarski (Wlodarski, 2008). Eine Ausnahme hiervon stellt die Implementierung der Adaptionsmechanismen und Schätzmethode der adaptiven psychophysischen Verfahren dar, welche innerhalb dieser Arbeit erfolgt (Kap. 6). Der Entwurf der grafischen Benutzeroberflächen (Kap. 7) wird in gemeinsamer Arbeit durchgeführt. Die anwenderseitige Beschreibung der Softwarebibliothek in Form einer Benutzerdokumentation ist außerdem Teil dieser Arbeit, deren Ergebnis im Anhang B zu finden ist. Hingegen ist die Dokumentation der internen Datenstruktur und des Quellcodes Bestandteil der Arbeit von André Wlodarski.

3 Methodische Aspekte von Hörversuchen

3.1 Der experimentelle Hörversuch

Ein Hörversuch kann allgemein als eine Untersuchung der auditiven Wahrnehmung dargebotener akustischer Reize, oder der daraus resultierenden psychologischen und physiologischen Prozesse betrachtet werden. Im Kontext von perzeptiven Audioevaluationen wird die Bedeutung des Begriffs i.d.R. enger aufgefasst und auf experimentelle Untersuchungen eingegrenzt² (vgl. etwa Maempel, 2007). Bei einer solchen werden systematisch Versuchsbedingungen in Form einer oder mehrerer *unabhängiger Variablen* verändert (s. Abs. 3.1.1) und die Auswirkungen dieser Variationen auf eine oder mehrere *abhängige Variablen* beobachtet (s. Abs. 3.1.2). Das Vorgehen ist zielgerichtet und setzt Vermutungen über bestimmte Wirkungszusammenhänge voraus, welche in Form von wissenschaftlichen Hypothesen formuliert werden. Gleichzeitig werden Störfaktoren durch geeignete Techniken kontrolliert (s. Abs. 3.1.3).

3.1.1 Unabhängige Variablen

Die Variation einer unabhängigen Variable kann entweder durch Selektion oder Manipulation erreicht werden. Im Fall der Selektion kann auf bereits bestehende Voraussetzungen zurückgegriffen werden, wie beispielsweise die Zugehörigkeit eines musikalischen Reizes zu einem bestimmten Genre. Hingegen findet bei der Manipulation ein Eingriff von außen statt, durch welchen bestimmte Merkmale in ihrer Ausprägung verändert werden. Hierzu zählen zum Beispiel die Veränderung der Frequenz eines Tones oder die Verarbeitung eines Signals bei verschiedenen Komprimierungsraten. Außerdem können Bedingungsvariationen nach dem Ort ihrer Entstehung bzw. der relativen Lage bezüglich der Versuchsperson klassifiziert werden. Man spricht in diesem Zusammenhang von einer *Reizvariable*, wenn die zugehörige Quelle der Variation außerhalb der Versuchsperson liegt und somit aus Umständen der Untersuchung

²Eine Ausnahme hiervon stellen beispielsweise Hörversuche dar, welche auf qualitativen Methoden beruhen (vgl. Abs. 3.3).

resultiert. Hingegen spricht man von einer *Organismusvariable*, wenn die Quelle innerhalb der Person liegt, d.h. wenn es sich um ein mit der Person verbundenes Merkmal handelt (vgl. Sarris, 1990, S. 153ff). Im letzteren Fall jedoch können mit dem interessierenden Merkmal eine Reihe weiterer Merkmale konfundiert sein, so dass die beobachteten Veränderungen einer abhängigen Variable nicht mehr ausschließlich durch die Variation der unabhängigen Variable erklärt werden können. Untersuchungen, welche Organismusvariablen als unabhängige Variablen verwenden, werden im strengen Sinne nicht mehr als Experimente bezeichnet (Sarris, 1990, S. 155). Man spricht in diesem Fall von einem *Quasi-Experiment* (Westermann, 2000, S. 23f) oder einer *quasi-experimentellen* Untersuchung (Bortz und Döring, 2006, S. 54ff; vgl. auch Abs. 3.1.3).

Die im Folgenden illustrierten Aspekte stellen eine Auswahl an Gruppen³ von Versuchsbedingungen (ohne Anspruch auf Vollständigkeit) dar, innerhalb welcher mögliche Variationen erfolgen können.

Reizvariablen

Signaleigenschaften des Stimulus: Hierzu zählen die Zugehörigkeit eines Reizes zu einer Signalkategorie (Musik, Sprache, Geräusch) einschließlich weiterer Unterkategorien (Genre, Textgattung, Geräuschtyp, etc.) und seine physikalischen Eigenschaften im Zeit- und Frequenzbereich.

Aufnahmetechnik und Signalverarbeitung: Bei der Signalaufzeichnung können verschiedene Aufnahmeverfahren, bei Abmischung und Mastering unterschiedliche klanggestalterische Mittel und bei der Speicherung verschiedene Audioformate bzw. Kompressionsverfahren eingesetzt werden.

Wiedergabesystem und Raum: Die Wiedergabe des akustischen Signals kann über Kopfhörer oder verschiedene Lautsprecheranordnungen erfolgen. Insbesondere können hierfür folgende Systeme genutzt werden: Monofonie, Zwei- und Mehrkanalstereofonie, Wiedergabe mittels binauraler Echtzeitfaltung und Wellenfeldsynthese. Bei Lautsprecherwiedergabe ist die Akustik des Wiedergaberaums zu berücksichtigen.

Versuchsdurchführung: Es können systematisch unterschiedliche Instruktionen erteilt oder der Versuchsablauf hinsichtlich der Abfolge von Messwiederholungen variiert werden (vgl. Abs. 3.1.3 und 3.4).

³Die vorgenommene Einteilung orientiert sich großteils an der ausführlichen Diskussion *experimenteller Variablen* bei Bech und Zacharov (2006, Kap. 5).

Organismusvariablen

Versuchsperson: Versuchspersonen können nach bestimmten Kriterien klassifiziert werden. Hierzu zählen neben allgemeinen soziodemographischen und soziokulturellen Populationsmerkmalen die physiologische Funktionalität des Gehörs, Hörerfahrung, Diskriminationsfähigkeit, Inter- und Intrasubjektrelabilität sowie Verbalisationsfähigkeit.

Für die Erfassung einer unabhängigen Variable kann grundsätzlich ein beliebiges Skalenniveau herangezogen werden, wobei sie im Fall einer Nominalskalierung auch als *Faktor* bezeichnet wird (vgl. Bortz und Döring, 2006, S. 727). Die für die verschiedenen Versuchsbedingungen definierten Ausprägungen einer unabhängigen Variable nennt man *Stufen*. Bei Variation von mehr als einer unabhängigen Variable, existieren mehrere Kombinationen aus deren Stufen, bei welchen jeweils die Messung von einer oder mehreren abhängigen Variablen erfolgt (s. Abs. 3.1.2). Im Fall nominalskalierter Variablen nennt man diese *Faktorstufenkombinationen*.

Die Wahl geeigneter Stufen kann aufgrund von Vorwissen und theoretischen Überlegungen erfolgen. Für nominalskalierte unabhängige Variablen ist zwischen Faktoren mit *festen* (engl. fixed factor) und mit *zufälligen Effekten* (engl. random factor) zu unterscheiden (Bortz und Döring, 2006, S. 727). Im ersten Fall werden genau die Ausprägungen der unabhängigen Variable untersucht, über welche bezüglich der zugrundeliegenden Hypothese eine Aussage getroffen werden soll. Im letzteren Fall sollen die Aussagen auf eine Population möglicher Stufen verallgemeinerbar sein, aus welcher eine begrenzte Anzahl zufällig ausgewählt wird. Bei metrischen (kardinalen) Variablen ist darauf zu achten, dass die Stufen breit genug gewählt werden, damit so genügend Unterschiedlichkeit in den Daten erzeugt wird, um einen zu erwartenden Effekt erfassen zu können (Laback, 2006). Soll ein funktionaler Zusammenhang zwischen unabhängiger und abhängiger Variable erfasst werden, so sind die Stufen eng genug zu wählen, um eine gewünschte Genauigkeit (Auflösung) zu erzielen. Außerdem ist eine genügend große Anzahl an Stufen zu verwenden, um einen repräsentativen Ausschnitt der Abszissenachse abzubilden.

3.1.2 Abhängige Variablen

Die Auswirkungen der systematischen Bedingungsvariation(en) werden mittels Erhebung der abhängigen Variable(n) erfasst. Damit verschiedene Ausprägungen dieser sowohl unter verschiedenen Versuchsbedingungen als auch zwischen verschiedenen

Versuchsdurchführungen verglichen werden können, werden sie in Form quantitativer Daten erhoben. Diesen Vorgang bezeichnet man auch als *Messen*. Er setzt eine *Operationalisierung* der zu messenden Variable voraus, welche die Festlegung von (Mess-) Operationen und damit verbunden des Skalenniveaus beinhaltet (Bortz und Döring, 2006, S. 736). Als Operationen bezeichnet man dabei Handlungen oder Aufgaben, die während des Messvorgangs zur Aufnahme der Daten durchgeführt werden, wie z.B. die Diskrimination zweier Stimuli und die anschließende Betätigung eines Knopfes durch die Versuchsperson oder das Zählen positiver Antworten durch den Versuchsleiter oder einen Computer. Die Gruppierung von Versuchsperson, Operation und eingesetzter Apparatur (Fragebogen, Knopf, Ratingskala, Benutzerschnittstelle, Computer, etc.) bezeichnet man in der empirischen Forschung auch als *Messinstrument*. In der Regel wird ein bestehendes Datenerhebungsverfahren (*Testverfahren*) gewählt, durch welches u.a. die Art der Operationalisierung bereits festgelegt ist (s. Abs. 3.3).

3.1.3 Kontrolle von Störeinflüssen

Die Behandlung (*Kontrolle*) von Störeinflüssen ist ein wesentliches Merkmal experimenteller Untersuchungen. Als *Störvariablen*⁴ werden Quellen der Variation bezeichnet, welche in einer Untersuchung zusätzlich zu den unabhängigen Variablen auf die abhängige(n) Variable(n) einwirken (vgl. Sarris, 1990, S. 163; Westermann, 2000, S. 303). Diese können entweder auf Umstände der Untersuchung oder Eigenschaften der Versuchspersonen zurückzuführen sein (Reiz- vs. Organismusvariablen, s. Abs. 3.1.1). Zu den untersuchungsbedingten Störvariablen zählen z.B. Unterschiede in der Kalibrierung des Wiedergabesystems, in den Instruktionen der Versuchspersonen oder in den Positionen von Reizdarbietungen innerhalb des Versuchsablaufs im Fall von Wiederholungsmessungen (s.u.). Personengebundene Störvariablen können z.B. Intelligenz, Alter, Hörfähigkeit, Studienfach oder die bisher in Hörversuchen gesammelte Erfahrung sein.

Die Wirkung einer Störvariable auf die gemessenen Werte der abhängigen Variable(n)

⁴Die von den Autoren Bortz und Döring (2006, S. 3) in ihrem Lehrbuch aufgestellte Taxonomie, welche an mehreren Stellen in der Literatur übernommen wird, soll hier nicht verwendet werden. Dort wird eine Störvariable als eine *Moderatorvariable* bezeichnet, welche „nicht beachtet oder schlicht übersehen wird.“ Diese Systematisierung wird von den Autoren nicht weiter begründet und an späteren Stellen im Buch wieder fallen gelassen (vgl. Bortz und Döring, 2006, S. 542ff, S. 536, S. 742). Darstellungen des ursprünglich von Saunders (1956) geprägten Moderatorvariablenbegriffs finden sich z.B. bei Bortz (2005, S.222), Schmitt (1990, S.29f), Shadish et al. (2002, S.159) und Westermann (2000, S.431ff).

kann auf unterschiedliche Weise in Erscheinung treten (vgl. Sarris, 1990, S. 167ff). Zum einen kann es zu einer systematischen Verschiebung der Daten kommen, welche auch als *Datenverzerrung* oder *Bias* bezeichnet wird. Zum anderen kann die Variabilität der Daten erhöht werden, was in einer Zunahme der *Fehlervarianz*⁵ resultiert.

Methoden zur Kontrolle von Störvariablen werden als *Kontrolltechniken* bezeichnet und lassen sich grundsätzlich in *experimentelle* und *statistische* Techniken einteilen (vgl. Sarris, 1990, Kap. 8; s. Abb. 3.1 auf S. 20). Die experimentellen Kontrolltechniken werden bereits vor Durchführung der Datenerhebung im Zuge der Planung und Vorbereitung von Untersuchungen angewendet, während Techniken der statistischen Kontrolle nach deren Abschluss eingesetzt werden.

Instrumentelle Kontrolle Zur ersten Gruppe zählen die instrumentellen Techniken, welche der Kontrolle untersuchungsbedingter Störeinflüsse dienen. Hierzu zählen *Abschirmung*, *Eliminierung* und *Konstanthalten*. Abschirmung hat die Reduktion der Wirkintensität einer Störquelle zum Ziel. Beispielsweise können in einem Hörversuch Störgeräusche durch Verwendung eines geschlossenen Kopfhörers gedämpft werden. Eine extreme Form der Abschirmung ist die Eliminierung, durch welche die Wirkung einer Störvariablen völlig ausgeschaltet wird. In Bezug auf externe Störgeräusche könnte dies z.B. durch Durchführung des Hörversuchs in einer Schallkabine erreicht werden. Lassen sich Störeinflüsse weder reduzieren noch ausschalten, so können sie jedoch in vielen Fällen *konstant gehalten* werden und tragen somit nicht zur Variabilität der Daten bei. Hierbei kann die Verwendung apparativer Techniken, wie etwa einer computergestützten Kommunikation zwischen Versuchsperson und Versuchsumgebung, eine wesentliche Rolle spielen, da hierdurch Versuchsabläufe vereinheitlicht werden (vgl. Bech und Zacharov, 2006, S. 315).

Versuchsplanerische Kontrolle Neben den instrumentellen existieren außerdem eine Reihe von Techniken, welche im Zuge der Versuchsplanung angewendet werden können. Der überwiegende Teil dieser bezieht sich auf die Behandlung von Organismusvariablen.

Die *Randomisierung* ist eine Technik, bei der die Versuchspersonen den Versuchsbedingungen zufällig zugeordnet werden. Bei einer ausreichend großen Anzahl an Versuchspersonen erfolgt hierdurch eine gleichmäßige Verteilung der Ausprägungen personengebinder Störvariablen. Gegenüber einer systematischen Verteilung, wie sie et-

⁵Die Begriffe *Bias* und *Fehlervarianz* sind statistische Termini und sollen hier nicht weiter erläutert werden. Hierfür siehe u.a. Bortz (2005, S. 87, S. 96 u. S. 252ff) oder Sarris (1990, S. 167ff).

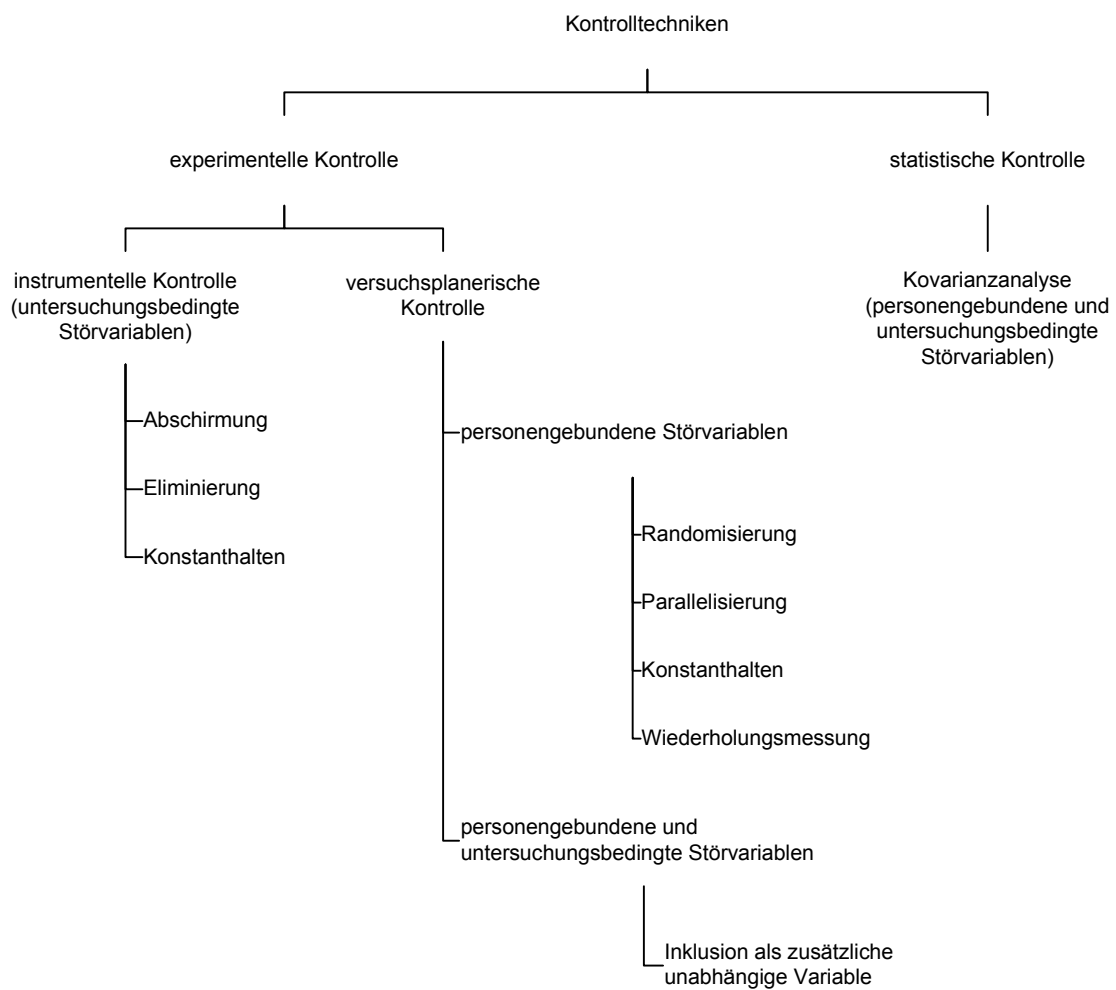


Abbildung 3.1: Übersicht über Techniken zur Kontrolle von Störvariablen in experimentellen/quasiexperimentellen Untersuchungen

wa bei Untersuchung natürlich vorhandener Personengruppen erfolgt, resultiert hierdurch ein reduzierter Bias auf Kosten einer erhöhten Variabilität der Daten innerhalb der Versuchsbedingungen (Westermann, 2000, S. 317f). Eine wesentliche Stärke dieser Technik ist die Tatsache, dass durch sie sämtliche, auch unbekannte, Organismusvariablen in ihrer systematischen Wirkung vermindert werden.

Kann aufgrund der Fragestellung (wenn z.B. Organismusvariablen als unabhängige Variablen untersucht werden, vgl. Abs. 3.1.1) oder aus untersuchungspraktischen Gründen (wenn z.B. unter den verschiedenen Versuchsbedingungen jeweils mit natürlichen Personengruppen gearbeitet wird) keine Anwendung der Randomisierung erfolgen, so sprechen viele Autoren nicht mehr von einer experimentellen sondern einer quasiexperimentellen Untersuchung (vgl. Bortz und Döring, 2006, S. 525; Westermann, 2000, S. 270). In diesem Fall, oder wenn nur eine geringe Anzahl an Versuchspersonen untersucht wird, kann anstelle der Randomisierung die Technik der *Parallelisierung* angewendet werden.

Unter Parallelisierung versteht man die Zusammenstellung von möglichst vergleichbaren Personengruppen unter den jeweiligen Versuchsbedingungen, indem man durch einen Selektionsvorgang für annähernd gleiche Verteilungen hinsichtlich der Organismusvariablen sorgt (Bortz und Döring, 2006, S. 737). Für eine Anwendung dieser Technik ist jedoch die Kenntnis der zu kontrollierenden Organismusvariablen notwendig. Bei kleineren Stichproben wird die Angleichung der Verteilungen nicht mehr hinsichtlich statistischer Kennwerte vorgenommen, sondern indem aus den verschiedenen Gruppen einander Personen mit in Bezug auf die Organismusvariable(n) ähnlichen Ausprägungen zugeordnet werden. Derartige Techniken werden auch als *Blockbildung* (Sarris, 1990, S. 189) oder als *Matching* (Bortz und Döring, 2006, S. 527) bezeichnet, die einander ähnlichen Personen sogenannten *Blöcken* oder *Matched-Samples* zugeordnet.

Eine weitere Kontrolltechnik stellt das *Konstanthalten von Organismusvariablen*, sowohl über die verschiedenen Versuchspersonen als auch über die Versuchsbedingungen, dar. Beispielsweise könnten in einem Hörversuch ausschließlich Versuchspersonen mit einem sehr guten Audiogramm eingesetzt werden, um den Einfluss der Variable Hörfähigkeit zu eliminieren. Der durch diese Maßnahme erreichten Reduktion der Fehlervarianz steht allerdings die Gefahr der Einbuße von externer Validität⁶ der

⁶Auf die klassischen Testgütekriterien wird innerhalb dieser Arbeit nicht eingegangen. Die Kenntnis dieser, insbesondere der Unterschiede zwischen interner und externer Validität, wird für das Verständnis des Textes vorausgesetzt. Diese Inhalte können u.a. bei Bortz und Döring (2006), Westermann (2000), Sarris (1990) oder Shadish et al. (2002) nachgelesen werden.

Untersuchung gegenüber (Bortz und Döring, 2006, S. 526).

Eine völlige Ausschaltung von Unterschieden hinsichtlich der Organismusvariablen zwischen den Versuchsbedingungen kann mit der Technik der *Wiederholungsmessung* (auch *Messwiederholung*) erreicht werden. Dieselben Versuchspersonen werden dabei unter mehreren Versuchsbedingungen untersucht. Wie die Randomisierung betrifft auch diese Maßnahme sämtliche, auch nichtidentifizierte personengebundene Störvariablen. Diese Technik zur Reduktion der Fehlervarianz wird insbesondere für die Überprüfung von Veränderungshypothesen⁷ herangezogen (Bortz und Döring, 2006, S. 548ff). Darüber hinaus wird sie aus ökonomischen Gründen eingesetzt, um die Anzahl benötigter Versuchspersonen zu verringern (Sarris, 1990, S. 189). Ein grundsätzliches Problem bei der Anwendung der Technik ist die Möglichkeit des Auftretens von *Sequenzeffekten*⁸ (Bortz und Döring, 2006, S. 549f). Deren Kontrolle kann ebenfalls versuchsplanerisch erfolgen (s.u.).

Die letzte in diesem Zusammenhang darzustellende versuchsplanerische Technik ist die *Inklusion* einer oder mehrerer Störvariablen als weitere Faktoren in das bestehende Versuchsdesign. Bei den in diesem Zusammenhang als *Kontrollfaktoren* (Bortz und Döring, 2006, S. 536) bezeichneten Variablen kann es sich sowohl um untersuchungsbedingte als auch personenbezogene Variablen handeln. Durch diese Maßnahme können die durch die betreffenden Störvariablen verursachten Varianzanteile separat aufgeklärt werden. Nachteilig hierbei ist jedoch, dass sich durch die zusätzlich entstandenen Faktorstufenkombinationen der zu betreibende Untersuchungsaufwand im Hinblick auf die Anzahl der Versuchspersonen erhöht.

Mittels der zuletzt beschriebenen Technik lassen sich beispielsweise Sequenzeffekte kontrollieren (Bortz und Döring, 2006, S. 550). Hierbei erfolgt die Einführung eines sogenannten *Abfolgefaktors*, dessen Stufen unterschiedliche Abfolgen (Sequenzen) der einzelnen Messungen unter den verschiedenen Versuchsbedingungen repräsentieren. Dieser kann sowohl als Faktor mit festen oder mit zufälligen Effekten konzipiert werden (vgl. Abs. 3.1.1). Entsprechend werden aus der Menge aller möglichen Abfolgen (vollständige Variation) entweder systematisch oder zufällig einige (oder alle) ausgewählt.

⁷Auf die verschiedenen Typen wissenschaftlicher Hypothesen soll in dieser Arbeit nicht eingegangen werden. Siehe hierzu die bereits in Anmerkung 6 aufgeführte Literatur zur Experimentalmethodik.

⁸Auch: Reihenfolgen-, Positions- oder Transfereffekt; mögliche Ursachen können Ermüdung, Lernen, Motivationsverlust, Entwicklung von Erwartungshaltungen, etc. sein.

Statistische Kontrolle Neben den versuchsplanerischen Techniken soll hier noch die *Kovarianzanalyse* als statistische Kontrolltechnik genannt werden (vgl. Bortz und Döring, 2006, S. 527). Gegenüber der bereits beschriebenen Inklusion der Variablen hat diese den Vorteil, dass für die betreffende Variable keine Stufen gebildet werden müssen und das Verfahren somit im Fall kardinalskalierteter Störvariablen genauer ist (ebd., S. 544f). Zudem muss das Versuchsdesign nicht um weitere Faktorstufenkombinationen erweitert werden. Voraussetzung ist u.a., dass die Störvariable während der Untersuchung miterhoben wird. Man spricht dabei in diesem Zusammenhang auch von einer *Kontrollvariablen*. Mittels Regressionsrechnung werden die Werte der abhängigen Variable(n) um den Einfluss dieser bereinigt. Eine detaillierte Darstellung dieser Technik findet man u.a. bei Bortz (2005, Kap. 10).

3.2 Versuchsdesign

Als *Versuchsdesign* oder *Versuchsplan* bezeichnet man ein standardisiertes Strukturschema, „das dem Aufbau, der Kontrolle und der methodologischen Bewertung einer empirischen Untersuchung von unabhängigen (UV) und abhängigen (AV) Variablen sachlogisch zugrunde liegt.“ (Sarris, 1992, S. 4) Auf die Darstellung der verschiedenen existierenden Designtypen soll in dieser Arbeit verzichtet und auf die entsprechende Literatur⁹ verwiesen werden. Für die Konzeption der zu erstellenden Software scheint eine Berücksichtigung dieser und ihrer Unterschiede im Einzelnen nicht notwendig zu sein, da sich die geforderte Funktionalität auf den Vorgang der Datenerhebung beschränkt und somit mehr am verwendeten Testverfahren (s. Abs. 3.3) ausgerichtet ist (vgl. Abs. 5.2). Dennoch können versuchsplanerische Elemente, wie etwa die Durchführung von Wiederholungsmessungen und die Einführung eines Abfolgefaktors (beide s. Abs. 3.1.3), von Bedeutung sein.

3.3 Testverfahren

Für die Durchführung von Hörversuchen stehen der auditiv-perzeptiven Forschung mehrere Testverfahren zu Verfügung, welche zu einem großen Teil der Psychophysik und der Psychologie entstammen. Darüber hinaus existieren Verfahren, welche speziell zum Zweck der Evaluation von Audiokompressionsverfahren (Audiocodex)

⁹U.a. Kirk (1982); Sarris (1992, Kap. 11–15); Shadish et al. (2002, Kap. 4–8); Westermann (2000, Kap. 12.2).

entwickelt wurden. Abbildung 3.2 auf Seite 26 gibt einen Überblick über die im Rahmen des Literaturstudiums zu dieser Magisterarbeit gefundenen Testverfahren. Die Darstellung erhebt weder Anspruch auf Vollständigkeit, noch ist die vorgenommene Einteilung die einzig mögliche. Anhand der Art des zu erfassenden Merkmals (z.B. Hörbarkeit von Reizen bzw. Reizunterschieden, Stärke von Empfindungen) und des Formats der zu erhebenden Daten (quantitativ vs. qualitativ, Skalenniveau) lassen sich drei Kategorien von Testverfahren bilden, welche im Folgenden nur kurz angerissen werden sollen.

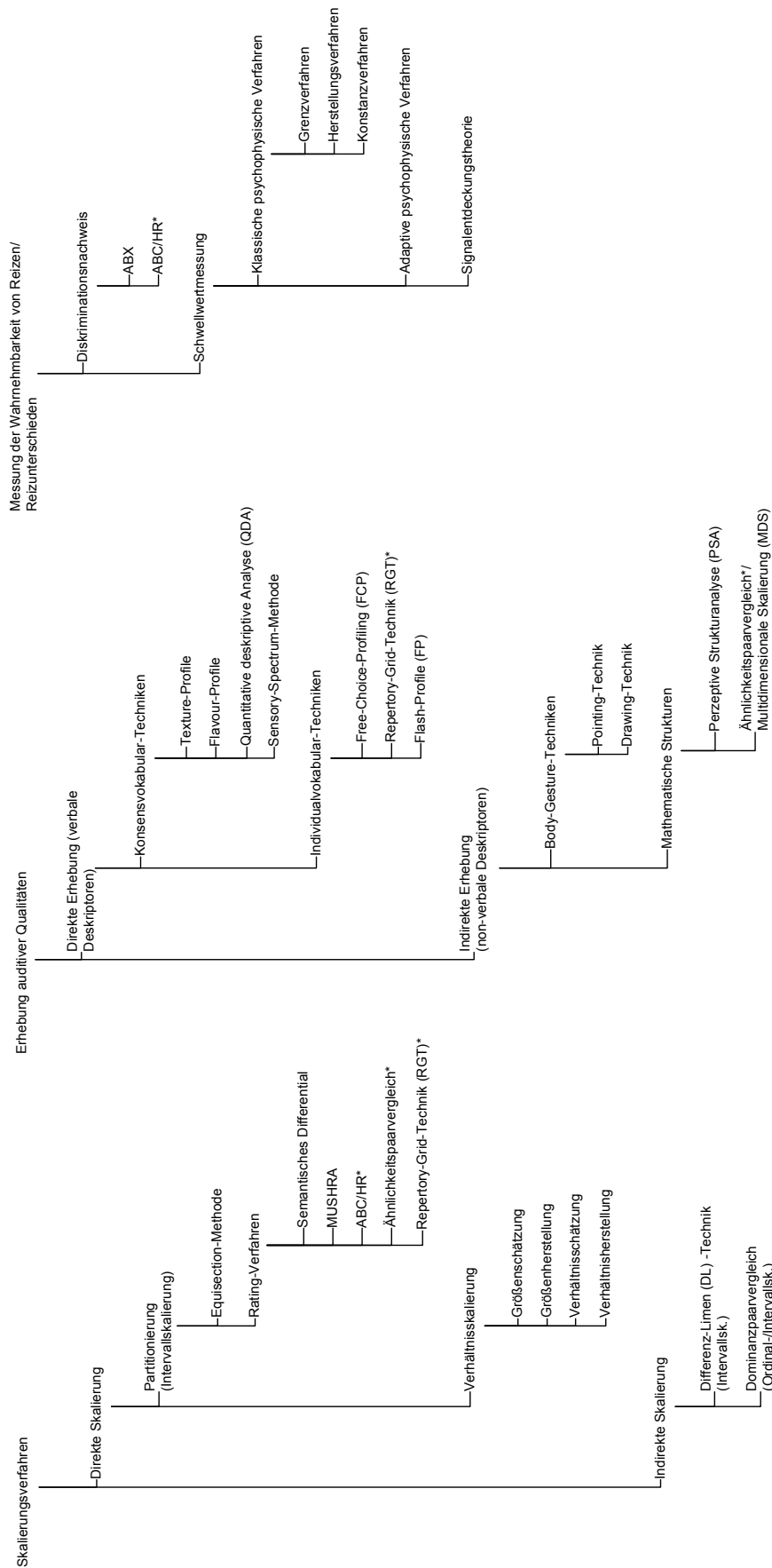
Die sogenannten *Skalierungsverfahren* dienen der quantitativen Erhebung der Stärke mehr oder weniger komplexer (vgl. u.) Empfindungen, welche durch akustische, bzw. physikalische Reize im Allgemeinen, hervorgerufen werden. Sie können wiederum in Verfahren der *direkten* und der *indirekten* Skalierung unterteilt werden, je nachdem ob die durch die Versuchsperson auszuführende Operation eine unmittelbare Schätzung von Merkmalsausprägungen (bzw. von Differenzen oder Verhältnissen unterschiedlicher Merkmalsausprägungen) auf Intervall- oder Verhältnisskalenniveau darstellt, oder ob eine Diskriminations- oder Paarvergleichsleistung zu erbringen ist, deren zugehörige Antwort zunächst auf Nominal- bzw. Ordinalskalenniveau erfasst und durch eine nachgeschaltete Statistik unter bestimmten Annahmen in ein intervallskaliertes Urteil überführt wird. Darstellungen der Verfahren im Einzelnen findet man bei Bech und Zacharov (2006, Kap. 4.2), Bortz und Döring (2006, Kap. 4.2) und Gescheider (1997, Kap. 9–11). Eine Beschreibung des Semantischen Differentials als Vertreter der Rating-Verfahren findet man außerdem in Abschnitt 4.2 dieser Arbeit.

Bei Anwendungen von Skalierungsverfahren in Hörversuchen können den Versuchspersonenurteilen jeweils einzelne auditive Qualitäten als Urteilskriterien zugrunde gelegt werden (*perzeptive Messung*), oder komplexere Urteile durch Integration mehrerer perzeptiver Qualitäten in Interaktion mit kognitiven und emotionalen Einflussgrößen (Stimmung, Kontext, Erfahrungshintergrund, Erwartung, etc.) gefällt werden (*affektive Messung* bzw. *affektive Qualität*). Die Kenntnis der Qualitäten und geeigneter Adjektive zu ihrer sprachlichen Beschreibung ist Voraussetzung für eine präzise Aufgabenstellung und somit für valide Urteile (Bech und Zacharov, 2006, S. 40). Für ihre Erhebung stehen verschiedene qualitative und halbqualitative Verfahren zu Verfügung, wovon einige bei Bech und Zacharov (2006, Kap. 4.1.1) dargestellt werden. Neben Verfahren der *direkten Erhebung* verbaler Deskriptoren zählen dazu Verfahren indirekter Erhebung, in welchen die gesuchten Qualitäten auf non-verbale Deskriptoren wie Elemente des Körperausdrucks (*Body-Gesture-Techniken*) oder mathematische Strukturen abgebildet werden. Verbale Deskriptoren können entweder als Ergebnis

eines Gruppenprozesses (*Konsensvokabular-Techniken*) oder in Form von Einzelerhebungen (*Individualvokabular-Techniken*) gefunden werden. In Abschnitt 4.3 dieser Magisterarbeit wird die Repertory-Grid-Technik (RGT) als eine Individualvokabular-Technik beschrieben.

Die dritte Kategorie wird durch Verfahren gebildet, welche die Messung der Wahrnehmbarkeit von Reizen bzw. Reizunterschieden zum Ziel haben. Hierzu zählen sowohl Verfahren aus der Codeevaluation wie der *ABX*- und der *ABC/HR-Test*¹⁰, welche in erster Linie einen statistischen Nachweis der Unterscheidbarkeit akustischer Reize zum Ziel haben, ohne dass dabei zwangsläufig eine metrische und eindimensionale Merkmalsvariation erfolgen muss. Im Gegensatz hierzu unterscheiden sich die Reize bei Verfahren zur Messung psychophysischer Schwellen grundsätzlich lediglich in einem, mindestens intervallskalierten, physikalischen Merkmal. Eine grundlegende Darstellung der verschiedenen Schwellwertverfahren findet man z.B. bei Gescheider (1997, Kap. 1–8). In Abschnitt 4.1 dieser Arbeit werden einige ausgewählte der adaptiven psychophysischen Verfahren eingehender beschrieben.

¹⁰Auch als *Triple-Stimulus-Hidden-Reference-Test* bezeichnet.



* Das Verfahren lässt sich nicht eindeutig einordnen und wird daher in mehr als einer Kategorie aufgeführt

Abbildung 3.2: Übersicht über Testverfahren der auditiv-perzeptiven Forschung

3.4 Familiarisation

Unter dem englischen Begriff *Familiarisation* versteht man im Kontext der Durchführung von Hörversuchen das vorherige Vertrautmachen der Versuchspersonen mit der zu absolvierenden Aufgabe und den Versuchsbedingungen. Dieser Prozess dient sowohl der Motivation der Versuchspersonen als auch der Reduktion von Fehlereinflüssen. Bech und Zacharov (2006, Kap. 9.2.2) empfehlen die folgenden Schritte:

Allgemeine Einführung Zunächst erhalten die Versuchspersonen in einer allgemeinen Einführung einen umfassenden Überblick sowohl über den Ablauf und sonstige Formalitäten, als auch über Hintergrund und Zielstellung der vorliegenden Untersuchung. Diese Erläuterungen können den Versuchspersonen als Gruppe dargeboten werden und sollten einerseits ausführlich genug sein, um diese zu einer engagierten Mitarbeit zu motivieren, andererseits so neutral wie möglich gehalten werden, um ungewünschte Beeinflussungen hinsichtlich des späteren Antwortverhaltens zu vermeiden.

Schriftliche Instruktion In einem zweiten Schritt erhält jede Versuchsperson eine ausführliche schriftliche Instruktion, in welcher die von ihr während des Hörversuchs durchzuführenden Aufgaben detailliert beschrieben werden. Hierzu zählen Hinweise zum Prinzip des angewandten Testverfahrens, zum Verständnis auditiver Qualitäten oder komplexer Merkmale im Fall von affektiven Messungen, und zur Verwendung von Skalen und sonstiger apparativer Elemente.

Mündliche Instruktion Der dritte Schritt besteht aus einer Phase der mündlichen Instruktion, in welcher der Versuchsleiter den Inhalt der schriftlichen Hinweise nochmals gemeinsam mit den Versuchspersonen rekapituliert. Es soll hierdurch sichergestellt werden, dass diese vollends verstanden und sämtliche individuell auftretenden Fragen vor der Versuchsdurchführung geklärt werden können.

Nachdem die Versuchspersonen ihre Aufgaben im Rahmen des Hörversuchs verstanden haben, sollen sie in den letzten beiden Schritten die Möglichkeit erhalten, sich auf praktischer Ebene mit diesen vertraut zu machen.

Gewöhnung an die Benutzerschnittstelle In der sogenannten *User-Interface-Familiarisation* erfolgt die Eingewöhnung in die Bedienung der Benutzerschnittstelle,

d.h. der eingesetzten Apparatur, welche den Informationsaustausch zwischen Versuchsperson und Versuchsumgebung ermöglicht. Dies kann etwa dadurch geschehen, dass die Versuchspersonen diese vor Versuchsdurchführung probeweise verwenden. Dabei auftretende Probleme sollen durch Rückfragen an den Versuchsleiter geklärt werden können. Wird für die Durchführung des Versuchs eine Hörversuchssoftware eingesetzt, so kann die Benutzerschnittstelle aus Geräten der Ein- und Ausgabeperipherie wie z.B. Maus, Tastatur und Monitor und einer grafischen Benutzeroberfläche bestehen.

Gewöhnung an den Reiz Bei der Verwendung von Ratingskalen kann es zu sogenannten *Ceiling-* oder *Floor-Effekten* kommen, wenn bei extremen Merkmalsausprägungen der entsprechende Randbereich der Skala bereits an mäßigere Ausprägungen vergeben wurde. Der entgegengesetzte Effekt besteht darin, dass die Versuchsperson in der Erwartung extremer Merkmalsausprägungen einen großen Bereich des Skalenrandes reserviert und es zu einer Verdichtung der Urteile im mittleren Bereich der Skala kommt (*Tendenz zur Mitte*). Beide Effekte führen zu einer Verzerrung der Urteile, welche reduziert werden kann, wenn die Versuchsperson die Möglichkeit erhält, vor Versuchsbeginn Stimuli in verschiedenen, insbesondere in extremen, Merkmalsausprägungen kennenzulernen (vgl. Bortz und Döring, 2006, S. 182ff). Ein derartiges Vorgehen erfolgt im letzten Schritt, der sogenannten *Sample-Familiarisation*.

Neben der Gewöhnung an den Variationsbereich (*Range*) der verschiedenen Merkmalsausprägungen kann außerdem je nach Hörerfahrung der Versuchsperson eine Anpassung des Wahrnehmungsapparates an das variierende Merkmal selbst, d.h. an die betreffenden physikalischen Reizeigenschaften bzw. die durch diese angeregten perzeptiven Qualitäten notwendig sein, damit diese während der Versuchsdurchführung zuverlässiger erkannt werden. Aus diesem Grund wird z.B. am FG Audiokommunikation der TU Berlin Sample-Familiarisation vor der Durchführung von Schwellwertmessungen betrieben.

Für die Realisierung der Sample-Familiarisation können verschiedene Vorgehensweisen angewendet werden (Bech und Zacharov, 2006, S. 314). Eine Möglichkeit besteht darin, der Versuchsperson eine Reihe akustischer Reize zu präsentieren, welche hinsichtlich der Ausprägung des interessierenden Merkmals den für den Hörversuch relevanten Variationsbereich abdecken. Die Darbietung kann entweder durch den Versuchsleiter oder durch die Versuchsperson selbst unter Verwendung einer entsprechenden Benutzerschnittstelle (vgl. Abb. 3.3) gesteuert werden. Als eine weitere Möglichkeit kann die Sample-Familiarisation mit der User-Interface-Familiarisation

kombiniert werden, indem bereits das eigentliche Testverfahren für eine geeignete Auswahl an Stimuli durchgeführt wird, ohne die dabei aufgenommenen Antworten der Versuchsperson für die spätere Auswertung der Untersuchungsergebnisse zu berücksichtigen. Dies kann in Form eines als solches ausgewiesenen Trainingsexperiments oder durch verdeckte Integration einiger Dummy-Trials zu Beginn des Hörversuchs erfolgen. Die zuerst genannte Variante besitzt gegenüber der letzteren den Vorteil, dass die Versuchsperson dabei auftretenden Schwierigkeiten mit Rückfragen an den Versuchsleiter begegnen kann.

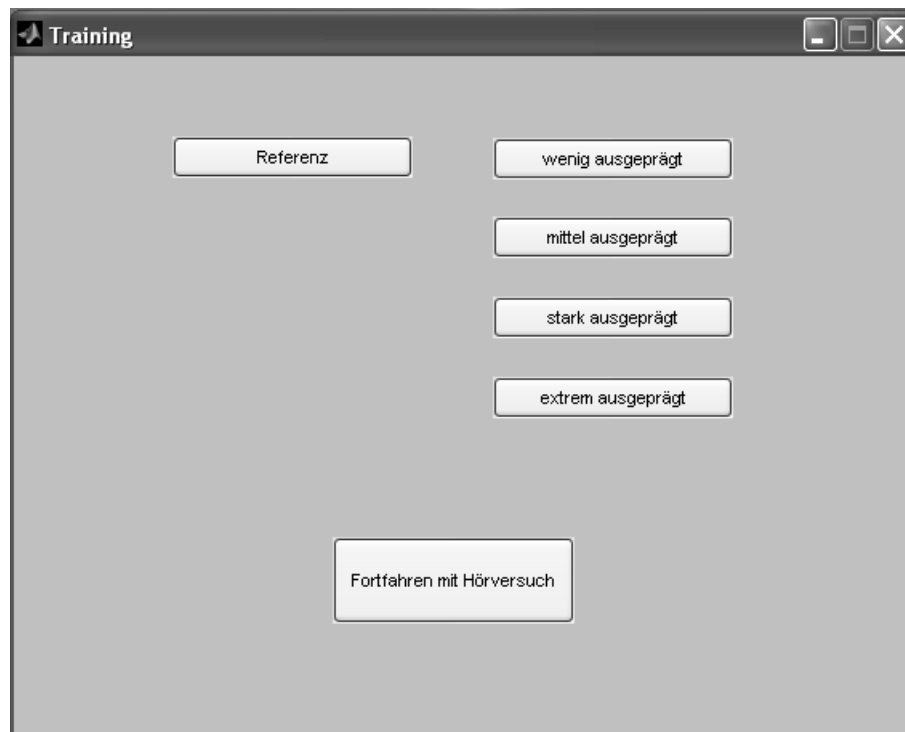


Abbildung 3.3: Beispiel einer grafischen Benutzeroberfläche zur Sample-Familiarisation unter Matlab vor Durchführung eines adaptiven Verfahrens zur Ermittlung einer Unterschiedsschwelle. Das betreffende physikalische Merkmal kann bei der Wiedergabe des Reizes, zusätzlich zur Darbietung des Referenzwertes, in vier verschiedenen Ausprägungen variiert werden. Der zugehörige Untersuchungsabschnitt wird als „Training“ bezeichnet. (Quelle: F.Schultz, FG Audiokommunikation, TU Berlin)

4 Zu implementierende Testverfahren

4.1 Adaptive psychophysische Verfahren

4.1.1 Einleitung

In psychophysischen Experimenten ist es häufig das Ziel, den Zusammenhang zwischen der Ausprägung eines physikalischen Reizmerkmals oder der Differenz zweier solcher Ausprägungen, und der Wahrscheinlichkeit für eine positive Erkennungsleistung einer Person zu untersuchen. Dies geschieht durch Aufnahme einer *psychometrischen Funktion* (s. Abs. 4.1.2.1), wofür unter anderem Verfahren der klassischen Psychophysik, wie z.B. das Konstanzverfahren, oder auch neuere, basierend auf der Signalentdeckungstheorie, zu Verfügung stehen (Gescheider, 1997). Allgemein erfolgen bei diesen Verfahren wiederholt Reizdarbietungen für verschiedene Ausprägungen eines eindimensionalen Reizmerkmals innerhalb eines relevanten Wertebereichs. Daraufhin wird die Antwort der Versuchsperson in einem binären Format aufgenommen, deren konkrete operationale Definition sich nach dem verwendeten Antwortparadigma (s. Abs. 4.1.2.3) ausrichtet. Für den Fall, dass nur ein bestimmter Punkt auf der psychometrischen Funktion als sogenannte *Schwelle* von Interesse ist, stellt die Gruppe der *adaptiven Verfahren* eine effiziente Möglichkeit der Erhebung dar, bei der ein Großteil der Reizdarbietungen in der relativen Umgebung des zu ermittelnden Schwellwertes erfolgt und sich somit die im Laufe des Versuchs gesammelte Information auf diesen Bereich konzentriert.

In den folgenden Abschnitten wird zunächst auf einige grundlegende Aspekte der psychophysischen Theorie eingegangen, deren Kenntnis für das Verständnis der nachfolgenden Abschnitte vorausgesetzt wird (Abs. 4.1.2). Daran schließt sich eine Erläuterung der prinzipiellen Funktionsweise und der zugehörigen Terminologie (Abs. 4.1.3), sowie der zugrundeliegenden Annahmen und einer diesbezüglichen Klassifizierung adaptiver Verfahren (Abs. 4.1.4) an. Davon ausgehend erfolgt eine Darstellung der drei für die Implementierung innerhalb der Software vorgesehenen Verfahren bzw. Verfahrenstypen – *Staircase* (Abs. 4.1.5), *PEST* (Abs. 4.1.6) und *Best-PEST* (Abs. 4.1.7).

Abschließend wird die Technik des *Interleaving* beschrieben, welche u.a. zur Reduzierung von Sequenzeffekten eingesetzt werden kann (Abs. 4.1.8).

4.1.2 Relevante Aspekte der psychophysischen Theorie

4.1.2.1 Psychometrische Funktion

Bezeichnet man die Höhe der Ausprägung eines Reizes entlang einer bestimmten Dimension als *Reizstärke*, so ist die psychometrische Funktion definiert als eine Funktion, welche verschiedenen Werten der Reizstärke die relative Anzahl positiver Antworten in einem *Detektions-* oder *Diskriminationsexperiment* (s. Abs. 4.1.2.2) zuordnet (Gescheider, 1997, S. 46, S. 402). Was dabei konkret unter einer positiven Antwort verstanden wird, hängt vom verwendeten *Antwortparadigma* (s. Abs. 4.1.2.3) ab. Nach ihrer Definition ist die psychometrische Funktion folglich eine empirische Größe, deren Werte nicht nur von den in der Regel eigentlich interessierenden sensorischen, sondern auch methodischen und psychologischen Einflüssen abhängen. Ein methodischer Einfluss ist das konzeptionelle Raten beim *Forced-Choice*-Paradigma (s. Abs. 4.1.2.3). Psychologische Einflüsse sind u.a. Reaktionsneigung und Konzentration der Versuchsperson. In der Psychophysik werden Modelle entwickelt, mit deren Hilfe nichtsensorische Einflüsse abgeschätzt werden können und somit die näherungsweise Angabe einer korrigierten psychometrischen Funktion möglich ist (s. Abs. 4.1.2.4). Bestimmte psychometrische Verfahren, wie die Maximum-Likelihood-Verfahren, erfordern zudem die Zugrundelegung einer bestimmten Form der psychometrischen Funktion, welche durch eine mathematische Modellfunktion angenähert wird (s. Abs. 4.1.2.5).

4.1.2.2 Unterschieds- vs. Absolutschwellenmessung

Verfahren zur Aufnahme psychometrischer Funktionen lassen sich sowohl in Detektions- als auch in Diskriminationsexperimenten einsetzen. In Detektionsexperimenten ist es die Aufgabe einer Versuchsperson einen Reiz bestimmter Reizstärke zu erkennen, während sie in Diskriminationsexperimenten zwei Reize verschiedener Reizstärken unterscheiden soll. In beiden Typen von Experimenten wird in der Regel die Stärke eines Reizes innerhalb eines relevanten Bereichs auf der Abszisse der psychometrischen Funktion variiert. Formal betrachtet, wird dieser Reiz weiterhin in beiden Fällen mit einem *Standardreiz* fester Reizstärke verglichen, welcher im Fall des Detektionsexperiments der *Nullreiz*, also ein Reiz minimaler Reizstärke,

ist. Detektionsexperimente können daher (formal) auch als Sonderfall von Diskriminationsexperimenten betrachtet werden (Treutwein, 1995). In beiden Fällen kann die durch die Versuchsperson erbrachte Leistung, mittels einer durch das Antwortparadigma festgelegten Operation, im binären Format (positive/negative Antwort) quantifiziert werden.

4.1.2.3 Antwortparadigmen

Der Begriff des Antwortparadigmas wird in Anlehnung an die von Marvit et al. (2003) aufgestellte Taxonomie verwendet. Im Zusammenhang psychophysischer Messungen ist hiermit eine bestimmte Systematik gemeint, welche dem Vorgang der Aufnahme der Versuchspersonenantwort zugrunde liegt. Neben dem Modus der Reizdarbietung zählt hierzu die von der Versuchsperson zu absolvierende Aufgabe. Im folgenden Abschnitt werden zwei häufig verwendete und elementare Antwortparadigmen beschrieben.

ja/nein-Paradigma Beim *ja/nein*-Paradigma soll die Versuchsperson in jedem Trial angeben, ob sie einen im Folgenden als *Testreiz* bezeichneten Reiz im Fall eines Detektionsexperiments, oder einen Unterschied zwischen diesem und einem Standardreiz im Fall eines Diskriminationsexperiments, erkennen kann (*ja-Antwort*) oder nicht (*nein-Antwort*). Die Stärke des Testreizes wird im Verlauf des Experiments variiert, die des Standardreizes bleibt unverändert.

Um auf diese Weise jeweils zu einer Entscheidung zu gelangen, muss sich die Versuchsperson an einem internen Kriterium orientieren, welches ihrer Willkür unterliegt. Es gibt keine richtigen oder falschen Antworten und somit für den Versuchsleiter keine Möglichkeit abzuschätzen, zu welchem Grad Antworten auf reiner Wahrnehmungsleistung beruhen, bzw. wie stark die Reaktionsneigung der Versuchsperson ein Ergebnis beeinflusst. Dieser Nachteil wird als *Kriterienproblem* bezeichnet (Maempel, 2007).

Ein Lösungsansatz hierfür stellt die Hinzunahme von Trials dar, welche Darbietungen deutlich unterschwelliger Reize bzw. Reizunterschiede enthalten. Das Verhältnis der Anzahl dieser sogenannten *blinden* Trials zur Gesamtanzahl ist üblicher aber nicht zwingender Weise 0.5 und wird der Versuchsperson zuvor mitgeteilt. Beide Sorten von Trials sollen sich in ihrer Abfolge zufällig abwechseln. Innerhalb der klassischen Psychophysik werden die blinden Trials auch als *Catch-Trials* bezeichnet und verwendet, um den Anteil fehlerhafter ja-Antworten, die sogenannte *False-Alarm-Rate* zu schätzen (Gescheider, 1997, S. 98f). Bei Kenntnis letzterer kann basierend auf dem

Schwellenverständnis der klassischen Psychophysik eine rechnerische Korrektur der psychometrischen Funktion erfolgen (s. Abs. 4.1.2.4). Auch im Rahmen der moderneren Signalentdeckungstheorie werden blinde Trials verwendet. Ein weiterentwickeltes Wahrnehmungsmodell bietet hierbei die Möglichkeit sensorische und nichtsensorische Einflüsse getrennt zu erheben (Gescheider, 1997, S. 105ff). Alternativ zur Verwendung des ja/nein-Paradigmas kann das Kriterienproblem durch Einsatz des Forced-Choice-Paradigmas umgangen werden, welches im folgenden Absatz beschrieben wird.

Forced-Choice- oder nAFC-Paradigma Bei Verwendung des Forced-Choice-Paradigmas werden der Versuchsperson in jedem Trial in aufeinanderfolgenden Intervallen mehrere Reize dargeboten, von denen einer als *Zielreiz* (engl. *target stimulus*) bezeichnet und dessen Reizstärke während des Experiments variiert wird. Die übrigen sind entweder Nullreize für Detektionsexperimente oder Standardreize einer konstanten Reizstärke für Diskriminationsexperimente. Die Wahrscheinlichkeit für das Enthalten des Zielreizes soll dabei für alle Intervalle gleich groß sein. Bei einer Darbietung von n Intervallen in einem Trial, spricht man auch von einem *n-Alternative-Forced-Choice* (kurz: *nAFC*)-Verfahren. Anschließend muss die Versuchsperson jeweils angeben, an welcher Position in der Darbietungsreihenfolge sich der Testreiz befand. Erfolgt keine sensorische Detektion bzw. Diskrimination des Zielreizes, muss die Versuchsperson willkürlich eines der n Darbietungsintervalle auswählen. Dieser Vorgang wird auch als *Raten* bezeichnet. Die Trefferwahrscheinlichkeit beträgt hierbei $p=1/n$.

Der Vorteil dieser Art der Erhebung gegenüber der ja/nein-Methode liegt in der objektiven Auswertbarkeit des Versuchspersonenurteils, da hierbei jede Antwort eindeutig als richtig oder falsch eingestuft werden kann. Unter Zugrundelegung der klassischen psychophysischen Theorie (vgl. etwa Gescheider, 1997, Kap. 4) wird hierdurch das Kriterienproblem gelöst: Das Urteil der Versuchsperson kommt entweder durch rein sensorische Leistung oder durch Raten zustande, dessen Einfluss basierend auf der High-Threshold-Annahme herausgerechnet werden kann (s. Abs. 4.1.2.4). Tendenzen zu positiven oder negativen Antworten können ausgeschlossen werden.

4.1.2.4 Konventionelle Ratekorrektur

Die näherungsweise Bereinigung der Werte einer aufgenommenen psychometrischen Funktion von nichtsensorischen Einflüssen kann durch Anwendung einer Korrekturformel erfolgen, welche auch als *Ratekorrekturformel* oder *Abott'sche Formel* bezeich-

net wird. Deren Herleitung basiert auf dem Schwellenverständnis der klassischen Psychophysik, welches auf der sogenannten *High-Threshold-Annahme* beruht. Letztere besagt, dass sich ein Beobachter bei Darbietung eines Reizes (bzw. eines Reizunterschieds) stets in einem von zwei Zuständen befindet: entweder er kann den Reiz sensorisch detektieren (bzw. diskriminieren) oder nicht¹¹.

Für die Entwicklung¹² der Formel betrachte man den Fall einer positiven Antwort einer Versuchsperson auf einen dargebotenen Reiz der Reizstärke x . Diese ist eine ja-Antwort im Fall der ja/nein-Methode, bzw. die Bezeichnung des korrekten Darbietungsintervalls im Fall der nAFC-Methode. Sie kann zum einen dadurch zustande kommen, dass der Reiz tatsächlich sensorisch detektiert (bzw. diskriminiert) wird. In diesem Fall darf es jedoch nicht zu einer (z.B. konzentrationsbedingten) Fehlleistung (*Lapse*) kommen. Eine positive Antwort kann zum anderen auch dann zustande kommen, wenn der Reiz nicht detektiert wird. Bei der nAFC-Methode geschieht dies durch Raten. Bei der ja/nein-Methode wird durch die Reaktionsneigung der Versuchsperson eine falsch-positive Antwort (*False-Alarm*) produziert.

Die Wahrscheinlichkeit $\psi(x)$ einer positiven Antwort – entsprechend dem Wert der psychometrischen Funktion an der Stelle x – kann folglich ausgedrückt werden durch:

$$\psi(x) = \psi^*(x) \cdot [1 - \lambda] + [1 - \psi^*(x)] \cdot \gamma , \quad (4.1)$$

mit

- $\psi(x)^*$: Wahrscheinlichkeit einer rein sensorischen Detektion bzw. Diskrimination,
- λ : Wahrscheinlichkeit einer negativen Antwort unter der Bedingung, dass der Reiz detektiert bzw. diskriminiert wurde (Lapsing-Rate),
- γ : Wahrscheinlichkeit einer positiven Antwort unter der Bedingung, dass der Reiz nicht detektiert bzw. diskriminiert wurde (auch: False-Alarm-Rate oder Guessing-Rate). Für die nAFC-Methode gilt: $\gamma = \frac{1}{n}$.

Gleichung (4.1) lässt sich in eine handlichere Form bringen:

$$\psi(x) = \gamma + [1 - \gamma - \lambda] \cdot \psi^*(x) . \quad (4.2)$$

¹¹Für weitergehende Erläuterungen, Kritik an diesem Konzept und die Darstellung eines weiterentwickelten Wahrnehmungsmodells basierend auf der Signalentdeckungstheorie, siehe z.B. Gescheider (1997). Möglichkeiten der Kombination letzterer mit adaptiven Verfahren werden bei Klein (2001) diskutiert.

¹²Eine Herleitung findet man auch bei Gescheider (1997, S. 99f), eine Diskussion bei Klein (2001) oder Treutwein (1995).

$\psi^*(x)$ stellt einen rein sensorischen Term dar, welcher dem um nichtsensorische Einflüsse korrigierten Wert der psychometrischen Funktion entspricht. Ein Umstellen von Gleichung (4.2) nach diesem führt daher auf die Korrekturformel:

$$\psi^*(x) = \frac{\psi(x) - \gamma}{1 - \gamma - \lambda} . \quad (4.3)$$

4.1.2.5 Logistische Funktion

Für die Anwendung bestimmter psychophysischer Testverfahren ist die Annahme einer konkreten Form der psychometrischen Funktion erforderlich. Für die mathematische Beschreibung dieser werden meist sigmoidförmige Modellfunktionen herangezogen, wie z.B. die *kumulierte Normalverteilung*, die *Weibull-Funktion* oder die *logistische Funktion* (Treutwein, 1995). Letztere wird von den Autoren des Best-PEST-Verfahrens (s. Abs. 4.1.7) verwendet und deshalb an dieser Stelle erläutert.

Die logistische Funktion $l(x; \theta, \beta)$ kann wie folgt beschrieben werden:

$$l(x; \theta, \beta) = \frac{1}{1 + e^{\frac{-(x-\theta)}{\beta}}} , \quad (4.4)$$

mit

- x : Abszissenwert; Definitionsbereich: $x \in] - \infty, +\infty[$,
- θ : Lageparameter: Lage des Wende- und Symmetriepunktes auf der Abszissenachse,
- β : *Spread-* (Streuungs-) *Parameter*¹³; alternativ wird in der Literatur auch $-\beta'(x - \theta)$ als Argument der e-Funktion verwendet. Dabei ist β' der sogenannte *Slope-* (Steigungs-) *Parameter*¹⁴.

Der zugehörige Funktionsgraph ist symmetrisch bezüglich seines Wendepunktes in θ mit $l(\theta) = 0.5$ und besitzt die Asymptoten $\lim_{x \rightarrow -\infty} l = 0$ und $\lim_{x \rightarrow \infty} l = 1$ (vgl. Abb. 4.1). Für $\beta = \sigma/1.7$ und $\theta = \mu$ stellt die logistische Funktion eine gute Approximation an die Kumulation einer Normalverteilung $\mathcal{N}(x; \mu, \sigma)$ dar (Treutwein, 1995).

Interpretiert man $l(x)$ als den Wert der Wahrscheinlichkeit $\psi^*(x)$ für die rein sensorische Detektion bzw. Diskrimination eines Reizes der Reizstärke x (vgl. Abs. 4.1.2.4),

¹³Nicht zu verwechseln mit der Streuung bzw. Standardabweichung der zugehörigen Verteilungsdichte, welche durch $\beta\pi/\sqrt{3} \approx 1.81 \cdot \beta$ gegeben ist (Hastings und Peacock, 1975, S. 80).

¹⁴In der Literatur wird die Bezeichnung Slope-Parameter mitunter auch für β verwendet (vgl. etwa Madigan und Williams, 1987). Daher sollte man für eine richtige Deutung von Aussagen zur Sicherheit das Argument der e-Funktion betrachten.

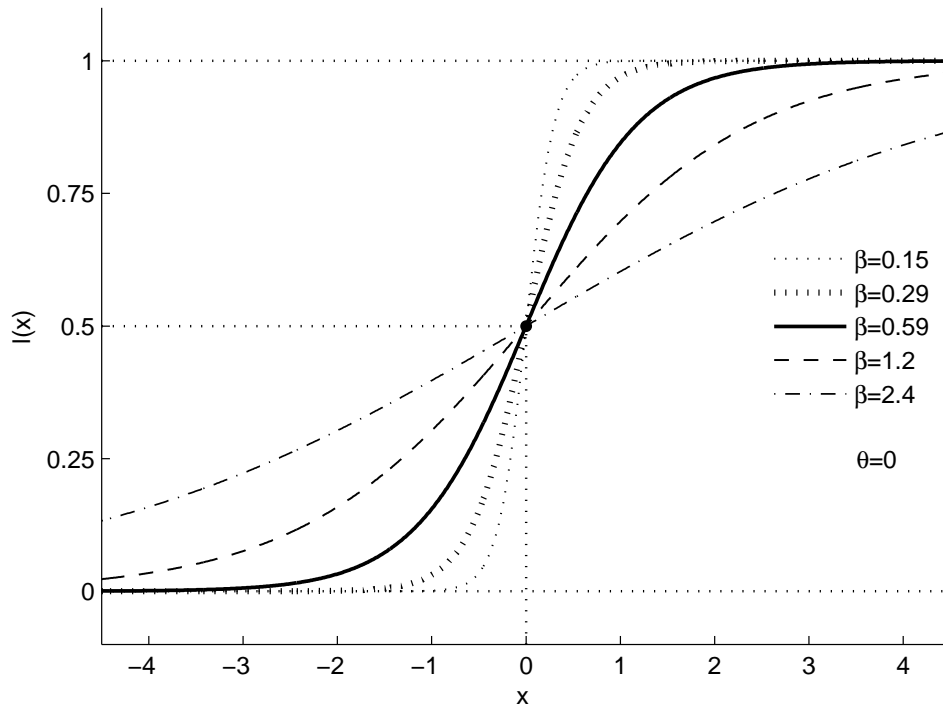


Abbildung 4.1: Verläufe der logistischen Funktion für verschiedene Spread-Parameterwerte β bei festem Lageparameterwert $\theta = 0$ (modifiziert nach Treutwein, 1995, fig. 2(d₂)).

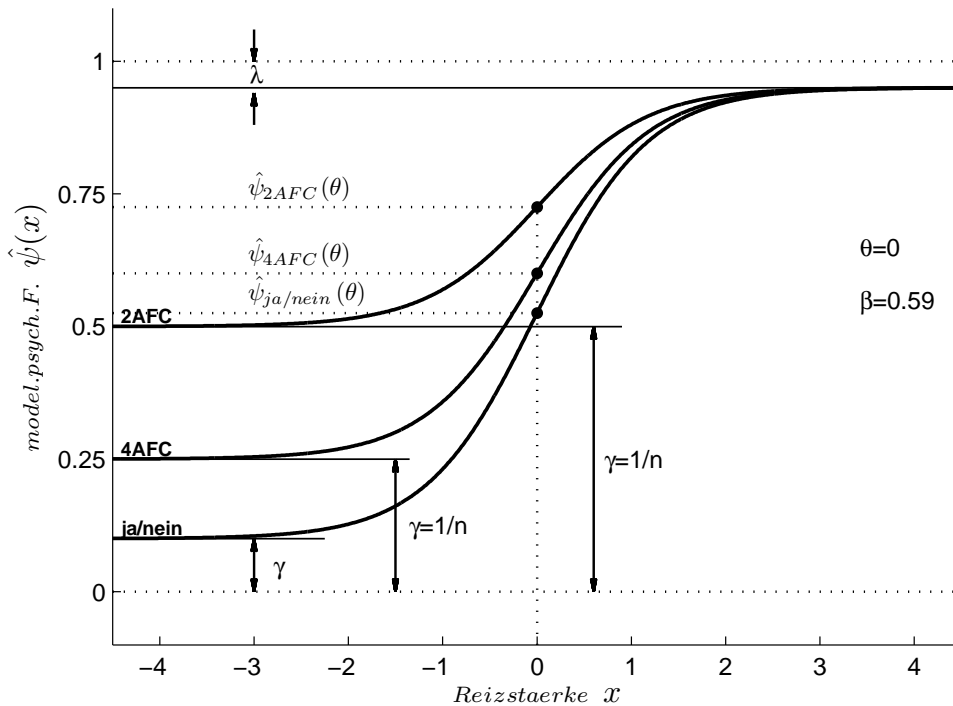


Abbildung 4.2: Verläufe der mittels logistischer Funktion modellierten psychometrischen Funktionen für verschiedene Antwortparadigmen (modifiziert nach Zuberbühler, 2002, fig. 1).

so gibt der Parameter θ die Lage der 50%igen sensorischen Detektions- bzw. Diskriminationsschwelle an.

Die modellierte psychometrische Funktion $\hat{\psi}$ ergibt sich durch Einsetzen von Gleichung (4.4) in Gleichung (4.2) zu:

$$\hat{\psi}(x) = \hat{\psi}(x; \theta, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) \cdot \frac{1}{1 + e^{\frac{-(x-\theta)}{\beta}}}. \quad (4.5)$$

Abbildung 4.2 auf Seite 37 zeigt die Verläufe verschiedener Funktionsgrafen für unterschiedliche Antwortparadigmen. Wie man erkennen kann bleibt dabei die Lage des Wendepunktes unverändert, während sowohl der Ordinatenwert $\hat{\psi}(\theta)$ als auch die Steigung (*Slope*) S im Wendepunkt vom verwendeten Antwortparadigma abhängen. Es lässt sich letztere wie folgt berechnen:

$$S = \left. \frac{d\hat{\psi}(x)}{dx} \right|_{x=\theta} = \frac{(1 - \gamma - \lambda)}{\beta} \cdot \frac{e^{\frac{-(x-\theta)}{\beta}}}{\left(1 + e^{\frac{-(x-\theta)}{\beta}}\right)^2} \bigg|_{x=\theta},$$

$$S = \frac{(1 - \gamma - \lambda)}{4\beta}. \quad (4.6)$$

4.1.3 Prinzipielle Funktionsweise und grundlegende Begriffe

Ein Durchlauf eines adaptiven Verfahrens (auch: *Run*) lässt sich prinzipiell wie folgt beschreiben. Ausgehend von einem Startwert (engl. *initial value*) wird die Reizstärke (engl. *stimulus intensity*) x nach jedem Trial und der zugehörigen, mittels eines bestimmten Paradigmas (Abs. 4.1.2.3) erhobenen Antwort der Versuchsperson, systematisch innerhalb eines relevanten (*Reizstärken-*) *Bereichs* (auch: *Range*), $[x_{min}, x_{max}]$, variiert. Der zugrundeliegende Mechanismus, auch als *Adaptions-* oder *Platzierungsmechanismus* (engl. *adaption mechanism* oder *placing mechanism*) bezeichnet, berücksichtigt dabei bereits erhaltene Antworten, um die darauffolgende Darbietung entlang der Ausprägung der Reizstärke in größerer Nähe zur gesuchten Schwelle (engl. *threshold*) $x = \vartheta$ zu platzieren. Die dabei vorgenommene Veränderung der Reizstärke kann sowohl positiv als auch negativ sein. Ihr Betrag wird als *Schrittweite* (engl. *step size*) bezeichnet. Ziel des Verfahrens ist eine fortschreitende Annäherung des Reizstärkenverlaufs (auch: *Track*) an die gesuchte Schwelle, wobei dieser Vorgang häufig auch als *Konvergenz* bezeichnet wird. Ändert der Track seine Richtung im Sinne eines Wechsels von einer Zunahme der Reizstärke zu einer Abnahme oder umgekehrt, so

liegt ein *Umkehrpunkt* (auch: *Reversal*) vor. Die Charakterisierung der zu bestimmenen Schwelle ϑ erfolgt durch die ihr zugehörige Wahrscheinlichkeit für das Auftreten einer positiven Antwort, d.h. durch den Wert der psychometrischen Funktion $\psi(x)$ an der Stelle $x = \vartheta$. Dieser Wert wird auch als *Konvergenzniveau* (engl. *target probability*) P_t bezeichnet.

Das Verfahren wird in Abhängigkeit eines bestimmten (*Abbruch-*) Kriteriums (engl. *stopping rule* oder *termination criterion*) beendet. Anschließend erfolgt die Schätzung des Schwellwertes anhand einer bestimmten (*Schätz-*) Methode (engl. *estimation method*), die in einigen Fällen eine statistische Auswertung des Reizstärkenverlaufs und der gegebenen Antworten beinhaltet, durch welche die Genauigkeit der Schwellenschätzung $\hat{\vartheta}$ erhöht werden soll.

4.1.4 Annahmen und Klassifizierung

Bezüglich der Durchführung adaptiver Verfahren existieren eine Reihe verschiedener Annahmen (vgl. Harvey, 1986; Levitt, 1971; Treutwein, 1995), welche im Folgenden dargestellt werden sollen.

- (1) *Die psychometrische Funktion ist streng monoton.* Dies ist eine notwendige Voraussetzung für die Definition der zu messenden Schwelle, da nur dann eine eindeutige Zuordnung zwischen Abszissen- und Ordinatenwerten besteht.
- (2) *Die erhaltenen Antworten sind sowohl voneinander als auch von den Reizstärken der vorhergehenden Trials statistisch unabhängig.* Nur bei Gültigkeit dieser Annahme kann die in einem Trial auf eine erfolgte Reizdarbietung gegebene Antwort als Ergebnis eines Zufallsexperiments aufgefasst werden, dessen Auftretenswahrscheinlichkeit allein vom aktuellen Wert der psychometrischen Funktion der Versuchsperson abhängt.
- (3) *Die psychometrische Funktion ist in Form und Lage zeitinvariant.* Ist diese Annahme erfüllt, so stellt (bei gleichzeitiger Gültigkeit von Annahme 2) der Verlauf des adaptiven Verfahrens einen *stationären* stochastischen Prozess dar. Die Antworten auf n Darbietungen bei einer festen Reizstärke x folgen dann einer Binomialverteilung $\mathcal{B}(n, \psi(x))$.
- (4) *Die Form der psychometrischen Funktion ist bekannt und kann durch ein Modell der parametrischen Statistik (vgl. Abs. 4.1.2.5) beschrieben werden.* Für

die meisten Verfahren welche dieser Annahme bedürfen, wird in diesem Zusammenhang die Kenntnis eines Formparameterwertes vorausgesetzt. Einige wenige lassen zu, dass dieser unbekannt ist und zusätzlich zum Schwellwert gemessen wird.

Annahmen 1 und 2 sind grundlegend und daher für adaptive Verfahren im Allgemeinen immer zu treffen. Die Wichtigkeit der Erfüllung von Annahme 3 hängt vom einzelnen Verfahren ab. Sie ist grundsätzlich von umso größerer Bedeutung, je stärker vorhergehende Reizdarbietungen und Antworten für die Platzierungen der Reizstärken und/oder die abschließende Schätzung der Schwelle berücksichtigt werden. Anhand von Annahme 4 werden adaptive Verfahren in zwei verschiedene Klassen eingeteilt. Die sogenannten *parametrischen* Verfahren setzen deren Gültigkeit im Gegensatz zu den *nichtparametrischen* voraus.

Innerhalb der beiden genannten Klassen existieren jeweils zahlreiche verschiedene Verfahren bzw. Verfahrenstypen, welche sich u.a. in Adaptionmechanismus und Schätzmethode unterscheiden. Für einen umfassenden Überblick und eine vergleichende Darstellung wird z.B. auf Treutwein (1995) verwiesen. In den folgenden Abschnitten werden die nichtparametrischen Verfahrenstypen *Staircase* und *PEST* behandelt. Als Vertreter der parametrischen Verfahren wird das Maximum-Likelihood-Verfahren *Best-PEST* dargestellt. Dabei liegt der Schwerpunkt auf den Funktionsweisen der Verfahren. Deren Robustheit gegenüber Verletzungen der genannten Annahmen soll im Kontext dieser Arbeit nicht diskutiert werden, da dies den Rahmen sprengen würde.

4.1.5 Staircase-Verfahren

Staircase-Verfahren zählen zu den nichtparametrischen Verfahren (s. vorhergehender Abs.) und zeichnen sich durch einen einfachen Platzierungsmechanismus aus, welchen man in diesem Zusammenhang als *Adaptionsregel* bezeichnet.

4.1.5.1 Einfache Staircase-Verfahren

Sogenannte *einfache* Staircase-Verfahren machen die Wahl der Reizstärke nur von einer vorhergehenden Antwort abhängig. In jedem Trial wird die Reizstärke nach einer positiven Antwort um den Betrag einer konstanten (vgl. weiter unten) Schrittweite erniedrigt bzw. in Folge einer negativen Antwort um denselben Wert erhöht. Der zugehörige Track pendelt sich um die gesuchte Schwelle bei einem Konvergenzniveau

von $P_t = 50\%$ ein (Levitt, 1971). Bezüglich der Wahl des Anfangswertes werden von verschiedenen Autoren unterschiedliche Empfehlungen gegeben. Im Hinblick auf eine möglichst geringe Anzahl benötigter Trials wird ein Startwert empfohlen, welcher möglichst nahe am wahren Schwellwert liegt (Levitt, 1971; Cornsweet, 1962). Andererseits kann es für die Familiarisation (vgl. Abs. 3.4) der Versuchsperson förderlich sein, wenn mit einem überschwelligen Reizstärkewert begonnen wird, wie etwa bei Leek (2001) beschrieben. Abbildung 4.3 gibt exemplarisch einen möglichen Reizstärkenverlauf wieder, wobei letztere Variante der Initialisierung verwendet wird.

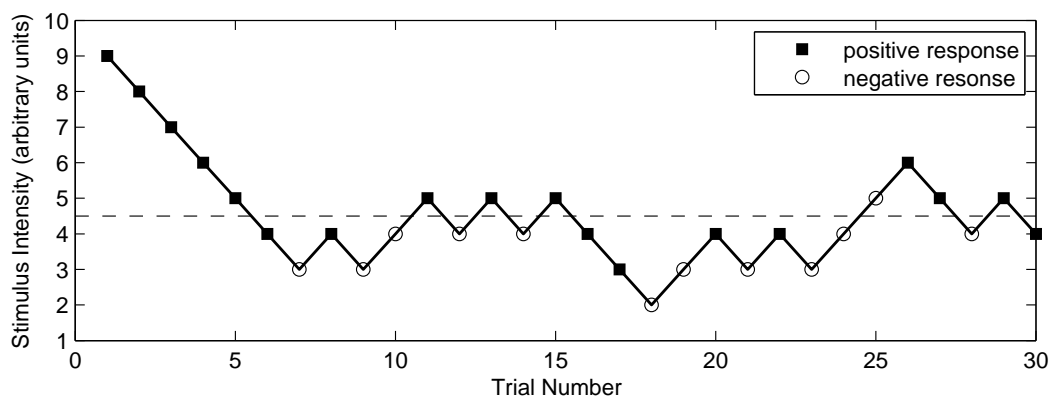


Abbildung 4.3: Ausschnitt eines möglichen Reizstärkenverlaufs für ein einfaches Staircase-Verfahren mit konstanter Schrittweite. Das Verfahren beginnt im gegebenen Beispiel bei einem überschwelligen Wert der Reizstärke. Die zu messende Schwelle liegt bei $\vartheta_{0.5} = 4.5$, wobei die Einheit der Reizstärke beliebig gewählt wurde. Der abgebildete Verlauf enthält 16 Umkehrpunkte.

Als Abbruchkriterium für ein Staircase-Verfahren kann die Anzahl der Umkehrpunkte im bisherigen Reizstärkenverlauf herangezogen werden. Für letztere werden Werte von mindestens 6 oder 8 empfohlen (Wetherill und Levitt, 1965, zit. nach Levitt, 1971). Grundsätzlich hängt die Wahl eines im Hinblick auf die Genauigkeit der Schätzung genügend großen Wertes jedoch von der Steilheit bzw. Streuung der psychometrischen Funktion ab und sollte daher in Vorversuchen ermittelt werden (vgl. Cornsweet, 1962). Die Schätzung des Schwellwertes erfolgt üblicherweise durch arithmetische Mittelung der Reizstärkenwerte bei den Umkehrpunkten (Levitt, 1971; Treutwein, 1995). Hierbei wird stets eine gerade Anzahl aufeinanderfolgender Werte berücksichtigt, so dass benachbarte lokale Maxima (auch: *Gipfel*, engl. *peak*) und Minima (auch: *Tal*, engl. *valley*) jeweils paarweise in die Berechnung eingehen. Die Mittelwerte der einzelnen Paare werden in der Literatur auch als *Mid-Run-Estimates* bezeichnet (ebd.). Zu beachten ist, dass für die beschriebene Schätzmethode, zumindest über den in die Mittelung eingehenden Verfahrensabschnitt hinweg, die Gültigkeit von Annahme 3 (s. Abs. 4.1.4) vorausgesetzt wird.

In der klassischen Verwendung von Staircase-Verfahren wird die Größe der Schrittweite als konstant gewählt (vgl. Cornsweet, 1962). Hierbei ist allerdings problematisch, dass die Wahl eines hohen Wertes zu einer geringen Genauigkeit führt und bei Wahl eines kleinen Wertes für die Konvergenz des Tracks eine große Anzahl an Trials benötigt wird. Ein im Hinblick auf die Effizienz¹⁵ des Verfahrens optimaler Wert liegt zwischen der 0.5 und 2.4fachen Standardabweichung der psychometrischen Funktion, unter der Annahme dass letztere durch eine kumulierte Normalverteilung beschrieben wird (Dixon und Mood, 1948, zit. nach Treutwein, 1995). Da die Standardabweichung der psychometrischen Funktion in vielen Fällen vor einer Untersuchung nicht bekannt ist, ist die Angabe eines optimalen Wertes häufig nicht möglich. Als Versuch einer Lösung dieses Problems und zur generellen Steigerung der Effizienz kann die Schrittweite im Laufe des Verfahrens reduziert werden¹⁶. Als ein für die Praxis geeignetes Vorgehen beschreibt beispielsweise Levitt (1971) die Vornahme von Halbierungen nach dem „1., 3., 7., 15., etc.“ Umkehrpunkt. Andere Autoren bezeichnen eine einmalige Halbierung unter bestimmten Voraussetzungen bereits als ausreichend (vgl. etwa Wetherill, 1963, zit. nach Levitt, 1971). Zu beachten ist jedoch, dass bei derartigen Vorgehensweisen die Reduktion der Schrittweite nach einem unflexiblen Kriterium erfolgt, welches eine evtl. Verschiebung der wahren Schwelle während des Versuchs nicht berücksichtigt. Sind derartige Veränderungen in der Lage der psychometrischen Funktion zu erwarten (vgl. wiederum Annahme 3, Abs 4.1.4), empfiehlt sich die Verwendung feiner ausgearbeiteter Regelwerke, die in einem solchen Fall auch mit einer Erhöhung der Schrittweite reagieren (Levitt, 1971). Ein derartiger Adaptionsmechanismus findet sich beim PEST-Verfahren (s. Abs. 4.1.6).

¹⁵Effizienz wird im Kontext adaptiver Verfahren aufgefasst als eine inverse Funktion des Produkts aus der Varianz der Messergebnisse und der durchschnittlichen Anzahl der dafür benötigten Darbietungen (Taylor und Creelman, 1967). Eine vergleichende Diskussion der dargestellten adaptiven Verfahren hinsichtlich ihrer Effizienz ist aufgrund der schweren Vergleichbarkeit hierzu veröffentlichter Untersuchungsergebnisse ein umfangreiches Unterfangen und kann deshalb im Rahmen dieser Magisterarbeit nicht geleistet werden.

¹⁶Mit jeder Reduzierung der Schrittweite im Laufe des Verfahrens erhöht sich dessen Genauigkeit. Für eine Mittelwertbildung der Reizstärken sollten dann sinnvoller Weise nur spätere Umkehrpunkte nach dem Erreichen einer minimalen Schrittweite berücksichtigt werden. Entsprechend konzipiert ist z.B. die Matlab-Implementierung PsyLab der FH Oldenburg (Hansen, 2008). Dort wird die Anfangsphase eines Runs, innerhalb welcher noch Schrittweithalbierungen erfolgen, als *Familiarisation-Phase* bezeichnet.

4.1.5.2 Transformierte Staircase-Verfahren

Transformierte Staircase-Verfahren berücksichtigen in ihren Adaptionsregeln Antworten auf mehr als nur eine vorhergehende Darbietung. Es werden Sequenzen von Antworten festgelegt, welche jeweils entweder eine Verringerung (*Down*-Sequenz) oder eine Erhöhung (*Up*-Sequenz) der Reizstärke zur Folge haben. Durch die Wahl unterschiedlicher Kombinationen derartiger Sequenzen können verschiedene Konvergenzniveaus erzeugt werden. Eine Zusammenstellung mehrerer Adaptionsregeln und ihrer zugehörigen Konvergenzniveaus findet man bei Levitt (1971, Tabelle I). Aus der Verwendung des nAFC-Paradigmas ergibt sich insbesondere die Forderung nach Konvergenzniveaus von größer als 50% um dem Einfluss des konzeptionellen Ratens entgegenzuwirken (vgl. nächster Absatz). Für die zu erstellende Software wurden die transformierten Staircase-Verfahren 2-Down/1-Up und 3-Down/1-Up ausgewählt. Es werden dafür zwei bzw. drei positive Antworten für eine Reizstärkenreduktion und eine negative Antwort für eine Erhöhung benötigt. Die zugehörigen Reizstärkenverläufe konvergieren bei $P_t = 70.7\%$ bzw. $P_t = 79.4\%$. Abbildung 4.4 zeigt einen möglichen Verlauf unter Zugrundelegung der 3-Down/1-Up-Regel.

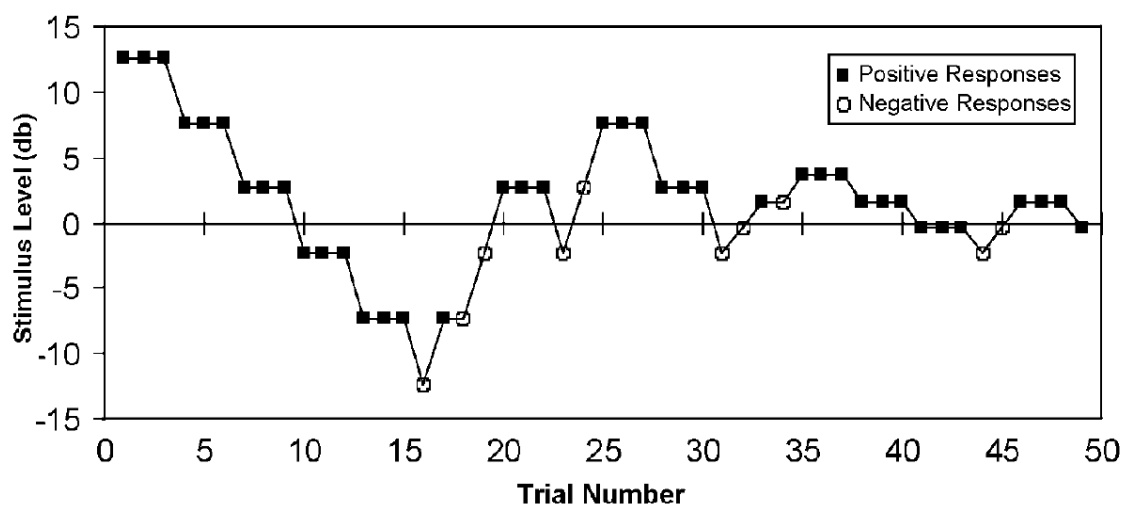


Abbildung 4.4: Ausschnitt eines möglichen Reizstärkenverlaufs für ein 3-Down/1-Up-Staircase-Verfahren mit einer Halbierung der Schrittweite nach dem 5. Umkehrpunkt und einer Schwelle von $\vartheta_{0.794} = 0dB$. [Quelle: Leek (2001, fig. 4B)]

Die Notwendigkeit der Wahl unterschiedlicher Konvergenzniveaus resultiert vor allem aus der Verwendung verschiedener Antwortparadigmen, da letztere der psychometrischen Funktion $\psi(x)$ jeweils unterschiedlich hohe nichtsensorische Anteile beisteuern. In der Regel wird eine dem zu messenden Schwellwert ϑ zugeordnete reine Detektions- bzw. Diskriminationswahrscheinlichkeit $\psi^*(\vartheta)$ von möglichst nahe bei 50% angestrebt.

Unter der High-Threshold-Annahme kann der Zusammenhang zwischen $\psi^*(\vartheta)$ und $P_t = \psi(\vartheta)$ durch Gleichung (4.3) auf Seite 36 beschrieben werden. Tabelle 4.1 gibt einen Überblick über die für die zur Implementierung vorgesehenen Adaptionsregeln, ihre zugehörigen Konvergenzniveaus und deren ratekorrigierte Entsprechungen bei Verwendung des nAFC-Paradigmas ($n=2,3,4$).

Adaptionsregel	Konvergenzniveau	»ratekorrigierte Konvergenzniveaus«		
		2-AFC	3-AFC	4-AFC
1-Down/1-Up	0.5	0	0.25	0.333
2-Down/1-Up	0.707	0.414	0.561	0.609
3-Down/1-Up	0.794	0.588	0.691	0.725

Tabelle 4.1: Übersicht der für die Implementierung ausgewählten Adaptionsregeln, ihre zugehörigen Konvergenzniveaus und ihre ratekorrigierten Detektions- bzw. Diskriminationswahrscheinlichkeiten für das nAFC-Paradigma in Abhängigkeit von der Anzahl n der Darbietungsintervalle ($n = 2, 3, 4$). Die Ratekorrektur erfolgte nach Gleichung (4.3) auf Seite 36 für $\gamma = 1/n$ und $\lambda = 0$.

4.1.6 PEST-Verfahren

Um die Effizienz von Schwellwertmessungen im Vergleich zur Staircase-Prozedur zu erhöhen, wurde *Parameter-Estimation-by-Sequential-Testing (PEST)* entwickelt. Grundlage des Verfahrens ist ein statistischer Signifikanztest, der nach jedem Trial, unter Berücksichtigung aller bei einer bestimmten Reizstärke gegebenen Antworten, eine Entscheidung über eine evtl. Erhöhung oder Verminderung der Reizstärke fällt. Eine Reihe heuristisch gefundener Regeln sorgt für eine Anpassung der Schrittweise und somit insgesamt für eine raschere Konvergenz des Verfahrens. Die Prozedur vermag auf Verschiebungen der Schwelle mittels Erhöhung der Schrittweite zu reagieren. Wie auch das Staircase-Verfahren zählt PEST zur Klasse der nichtparametrischen Verfahren und setzt daher kein spezielles Wissen des Experimentators über die Form der psychometrischen Funktion voraus.

Im Folgenden wird zunächst das Verfahren in der ursprünglichen Fassung nach Taylor und Creelman (1967) erläutert (Abs. 4.1.6.1). Daran schließt sich die Darstellung des *More-Virulent-PEST*, einer modifizierten Version nach Findlay (1978), an (Abs. 4.1.6.2).

4.1.6.1 Original PEST-Verfahren

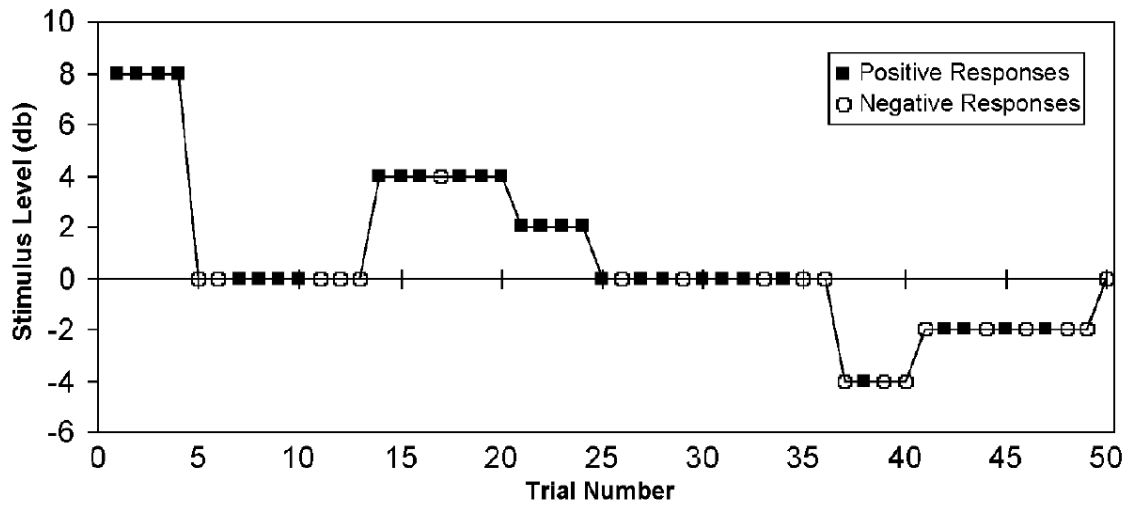


Abbildung 4.5: Ausschnitt eines nach dem PEST-Verfahren aufgenommenen Reizstärkenverlaufs für ein Konvergenzniveau von $P_t = 0.5$ und eine Schwelle von $\vartheta_{0.5} = 0dB$. [Quelle: Leek (2001, fig. 2)]

Signifikanztest Die Entscheidung darüber, ob im Verlaufe des adaptiven Verfahrens nach einer Darbietung die Stärke des darauf folgenden Reizes verändert, d.h. erniedrigt oder erhöht wird, ergibt sich als Ergebnis einer leicht abgewandelten¹⁷ Version des *Sequential-Likelihood-Ratio-Tests (SLRT)* nach Wald (1947). Dieser soll überprüfen, ob die augenblickliche Auftretenswahrscheinlichkeit positiver Antworten größer bzw. kleiner als, oder gleich des geforderten Konvergenzniveaus P_t ist. Hierfür erfolgen mehrere Darbietungen bei einer gegebenen Reizstärke, wobei positive Antworten in $N(C)$ und Antworten insgesamt (d.h. positive und negative) in T gezählt werden. Nach einer Veränderung der Reizstärke beginnt der Zählvorgang von neuem. In jedem Trial werden nach gegebener Antwort für $N(C)$ Ober- und Untergrenzen – $N_{b,u}(C)$ und $N_{b,l}(C)$ – definiert. Liegt $N(C)$ innerhalb dieser, so wird im nächsten Trial eine weitere Darbietung bei derselben Reizstärke erfolgen. Liegt es oberhalb oder auf der oberen Grenze, so wird die derzeitige Auftretenswahrscheinlichkeit für eine positive Antwort im Vergleich zu P_t als zu hoch eingeschätzt und daher die Reizstärke für die folgende Darbietung verringert. Umgekehrt führt das Erreichen bzw. Unterschreiten der unteren Grenze zu einer Erhöhung der Reizstärke. Die Grenzen ergeben sich aus dem erwarteten Wert $E[N(C)]$ der Auftretenshäufigkeit einer positiven Antwort im

¹⁷Im Original wird das *Deviation-Limit* (s. weiter unten im Text) analytisch in Abhängigkeit von α - und β -Fehlerwahrscheinlichkeit ausgedrückt. Bei Taylor und Creelman (1967) hingegen wird ein heuristisch gefundener Wert verwendet.

Zustand der vollständigen Annäherung des Tracks an die Schwelle und einer Konstante W , welche als *Deviation-Limit* bezeichnet wird.

Die Entscheidung des Tests lässt sich wie folgt darstellen:

$$N(C) \geq N_{b,u} = E[N(C)] + W \Rightarrow \text{Reduktion der Reizstärke,} \quad (4.7)$$

$$N(C) \leq N_{b,l} = E[N(C)] - W \Rightarrow \text{Erhöhung der Reizstärke,} \quad (4.8)$$

$$E[N(C)] - W < N(C) < E[N(C)] + W \Rightarrow \text{Beibehaltung der Reizstärke,} \quad (4.9)$$

$$\text{mit } E[N(C)] = P_t \cdot T. \quad (4.10)$$

$E[N(C)]$ wird dabei als Erwartungswert der Binomialverteilung $\mathcal{B}(T, \psi(x))$ berechnet (vgl. etwa Lehmann, 2002, Gl. (3.1a) auf S. 54). Voraussetzung für die Anwendbarkeit des Signifikanztests ist daher ein während eines Testzyklus bei gegebener Reizstärke x konstanter Wert $\psi(x)$ der psychometrischen Funktion und daher zumindest abschnittsweise die Erfüllung von Annahme 3 (vgl. Abs. 4.1.4).

Regeln zur Variation der Schrittweite Liefert oben beschriebener Signifikanztest eine Entscheidung für eine Erhöhung oder Verminderung der Reizstärke, so ist zusätzlich festzulegen, um welchen Betrag der Schrittweite diese Änderung vorgenommen werden soll. Hierfür dienen die folgenden vier Regeln, welche im Einzelnen bei Taylor und Creelman (1967) heuristisch begründet werden.

1. Bei jeder Richtungsumkehr des Tracks wird die Schrittweite halbiert.
2. Im zweiten Schritt in eine gegebene Richtung bleibt die Schrittweite unverändert.
3. Der vierte und die darauf folgenden Schritte in eine gegebene Richtung sind jeweils doppelt so groß wie der ihnen vorher gehende.
4. Die Größe des dritten Schrittes in eine gegebene Richtung hängt von den Schritten ab, welche zur letzten Richtungsumkehr führten: Wenn der Schritt, welcher dieser vorausging, aus einer Verdopplung der Schrittweite hervorgegangen war, dann wird die Schrittweite nicht verdoppelt (sondern beibehalten). Falls dieser allerdings nicht das Ergebnis einer Verdopplung war, so wird die Schrittweite verdoppelt.

Abbruchkriterium und Schätzmethode Ein Durchlauf des PEST-Verfahrens ist beendet, wenn obige Regeln einen vorher festgelegten minimalen Wert¹⁸ Δx_{stop} der Schrittweite aufrufen. Der aus diesem Schritt folgende Reizstärkewert wird nicht mehr für eine Darbietung verwendet, sondern stellt die Schätzung des Schwellwertes dar.

Wahl der Anfangswerte für Reizstärke und Schrittweite Die Wahl der Anfangswerte für Reizstärke und Schrittweite unterliegt keinen Einschränkungen durch das Verfahren. Beide Werte haben nach Taylor und Creelman (1967) keine wesentlichen Auswirkungen auf das Ergebnis. Für die Reizstärke wird im Hinblick auf die Familiarisation der Versuchsperson an das variierende Merkmal (vgl. Abs. 3.4) die Wahl eines deutlich überschwelligen Reizes empfohlen. Bezüglich der Schrittweite warnen die Autoren vor zu großen Werten, da diese zur Verunsicherung von Versuchspersonen führen könnten und legen die Einführung einer oberen Grenze nahe, welche sowohl am Anfang als auch während des Verfahrens zu keinem Zeitpunkt überschritten wird.

Wahl des Deviation-Limits W Die Wahl des Deviation-Limits W beeinflusst sowohl die Geschwindigkeit als auch die Genauigkeit des Signifikanztests in gegenläufiger Weise. Höhere Werte erzeugen ein größeres Intervall $]N(C)_{l,b}, N(C)_{u,b}[$, weshalb der Test restriktiver wird, da die Wahrscheinlichkeit für die Lage des empirischen Wertes $N(C)$ innerhalb dieses Intervalls, und somit für eine Beibehaltung der Reizstärke, zunimmt. Hierdurch steigt die Anzahl der für einen Reizstärkenwechsel notwendigen Trials und, bedingt durch die größere Stichprobe, gleichzeitig die Sicherheit der Entscheidungen. Aufgrund ihrer Erfahrungen aus Simulationen empfehlen Taylor und Creelman (1967) die Wahl eines Wertes von $W = 1.0$ für $P_t = 0.75$ bei Verwendung des 2AFC-Paradigmas. Für Werte von P_t näher bei 0.5 wird eine Wahl von $W = 1.5$ oder 2 empfohlen. Die Notwendigkeit der Wahl höherer Werte für W lässt sich in diesem Fall dadurch begründen, dass die Antwortsequenzen bei einer gegebenen Reizstärke x binomialverteilt sind (s.o.) und die Varianz dieser unimodalen Verteilung ihr Maximum bei $\psi(x) = 0.5$ besitzt.

Anzahl effektiver Umkehrungen Zur handlichen Beschreibung der Schrittweite führt Findlay (1978) den Laufindex n mit der Bezeichnung *Anzahl effektiver Umkehrungen* (engl. *number of effective reversals*) ein. Zu Beginn des Verfahrens besitzt dieser den Wert 0. Jedes mal, wenn die Regeln eine Verdopplung der Schrittweite verlangen, wird n um 1 erniedrigt und bei jeder Halbierung um 1 erhöht. Dadurch

¹⁸Von den Autoren im englischen Text auch als *stopping step size* bezeichnet.

lässt sich die Schrittweite Δx schreiben als:

$$\Delta x(n) = \frac{\Delta x_1}{2^n}, \quad (4.11)$$

mit

- Δx_1 : Startwert der Schrittweite,
- n : Anzahl effektiver Umkehrungen.

Der maximale Wert für n ist gegeben durch:

$$n_{max} = \log_2\left(\frac{\Delta x_1}{\Delta x_{stop}}\right). \quad (4.12)$$

Das Erreichen dieses Wertes ist dem bereits formulierten Abbruchkriterium gleichwertig.

4.1.6.2 More-Virulent-PEST

Um die Effizienz des PEST-Verfahrens weiter zu erhöhen, wurden durch Findlay (1978) zwei Modifikationen des Adaptionsmechanismus vorgenommen.

Modifikation 1 Die erste Modifikation berücksichtigt den Fall, dass der beschriebene Signifikanztest¹⁹ bei einer gegebenen Reizstärke eine große Anzahl T an Trials benötigt, um zu einer Entscheidung für einen Reizstärkenwechsel zu gelangen. Hierbei ist die Wahrscheinlichkeit, dass sich die aktuelle Wahrscheinlichkeit einer positiven Antwort nahe des Konvergenzniveaus P_t befindet, größer als bei einem schnellen Wechsel. Findlay erweitert das Verfahren dahingehend, dass in einem solchen Fall eine zusätzliche Verminderung der Schrittweite erfolgt, um zu einer rascheren Konvergenz des Tracks zu gelangen. Sobald die Gesamtzahl T aller Darbietungen bei einer Reizstärke einen vorgegebenen Wert M , oder jedes ganzzahlige Vielfache davon erreicht, wird die Schrittweite jeweils halbiert, bzw. die Anzahl effektiver Umkehrungen n um 1 erhöht.

¹⁹Findlay (1978) schließt die Fälle der Gleichheit ohne Begründung von den eine Reizstärkenänderung bedingenden Vergleichsrelationen in Ungleichungen (4.7) und (4.8) aus und ordnet sie stattdessen der Entscheidung für die Beibehaltung der Reizstärke (4.9) zu. In der Publikation des PEST-Verfahrens von Taylor und Creelman (1967) wird dies allerdings explizit umgekehrt gehandhabt, wie auch in der originalen Darstellung des SLRT bei Wald (1947). Bezüglich der Implementierung des Verfahrens in der zu erstellenden Software wurde zugunsten der letzteren ursprünglichen Variante entschieden.

Formal betrachtet erfolgt daher nach jeder Reizstärkenänderung eine zusätzliche Erhöhung von n um den Wert

$$\Delta n = \text{integerpart} \left(\frac{T}{M} \right) . \quad (4.13)$$

Modifikation 2 Wie bereits in Abschnitt 4.1.6.1 dargestellt, werden Genauigkeit und Konvergenzgeschwindigkeit des Verfahrens von der Wahl des Deviation-Limits W gegenläufig beeinflusst. Findlays Ansatz ist es, diesen Parameter während des Verfahrens nicht mehr konstant zu halten, sondern schrittweise zu erhöhen. Zu Beginn sorgen geringe Werte für schnelle Reizstärkenwechsel, gegen Ende hohe Werte für eine hohe Genauigkeit. Für die konkrete Umsetzung wird folgender funktionaler Zusammenhang vorgeschlagen:

$$W = \begin{cases} 0.5 & \text{für } n \leq 1, \\ \frac{n}{2} & \text{für } n > 1. \end{cases} \quad (4.14)$$

Wahl des Parameters M Hinsichtlich der Wahl des Parameters M sind Findlays Publikation keine eindeutigen Empfehlungen zu entnehmen. Zwar wurden Computersimulationen für verschiedene Werte $M = 4, 8$ und 12 durchgeführt. Die Angabe von Optima für die gegenüber dem Originalverfahren erreichte Effizienzsteigerung erfolgt jedoch nur für die erste Modifikation für sich allein genommen. Dies sind $M = 4$ für $W = 1$, bzw. $M = 8$ für $W = 2$. Weiterhin wird angegeben, dass durch die Kombination beider Modifikationen die in diesem Zusammenhang größte Effizienzsteigerung erreicht wird, eine Diskussion verschiedener Werte von M bleibt allerdings aus. Findlays Daten berücksichtigen nur den Fall $P_t = 0.5$.

4.1.7 Best-PEST-Verfahren

Die Effizienz adaptiver Verfahren soll mit Hilfe von Methoden der parametrischen Statistik gegenüber dem PEST-Verfahren nochmals gesteigert werden. Hierfür sind im Vorfeld zusätzliche Annahmen über die Form der psychometrischen Funktion zu treffen (s. Abs. 4.1.4). Einen derartigen Ansatz verfolgen die *Maximum-Likelihood-Verfahren*, bei welchen sowohl für die Platzierung der Reizstärken als auch für die abschließende Schätzung der Schwelle sämtliche im bisherigen Verlauf des Verfahrens gesammelten Daten für eine Parameterschätzung nach der gleichnamigen statistischen Methode verwendet werden (Gescheider, 1997). Innerhalb dieser Klasse wurden seit

den 1960er Jahren zahlreiche Verfahrensweisen veröffentlicht, welche sich hauptsächlich im Hinblick auf das verwendete parametrische Modell, die Art der Initialisierung und die Terminierung des Verfahrens unterscheiden. Einige wenige Verfahren bieten zusätzlich zur Schwellenschätzung über bestimmte Formparameter die Schätzung der Steigung im Wendepunkt der psychometrischen Funktion (Slope), die meisten setzen letztere als bekannt voraus. Eine Übersicht verschiedener veröffentlichter Maximum-Likelihood-Verfahren findet man etwa bei Treutwein (1995). In den folgenden Abschnitten wird die Funktionsweise des Best-PEST-Verfahrens nach Pentland (1980) und Lieberman und Pentland (1982) beschrieben, welches häufig zitiert wird (vgl. etwa Gescheider, 1997) und im Vorfeld der Magisterarbeit für die Implementierung innerhalb der Softwarebibliothek ausgewählt wurde. An verschiedenen Stellen werden außerdem Möglichkeiten für Erweiterungen und Modifikationen aufgezeigt, welche auf Ideen von Autoren anderer Verfahren zurückgehen.

4.1.7.1 Adaptionmechanismus

Platzierung der Reizstärken Die Schätzung des Reizstärkewertes für die jeweils nächste Darbietung erfolgt in jedem Trial durch Anwendung der Maximum-Likelihood-Methode²⁰. Diese besagt, dass ein Satz von Populationsparametern π_1, \dots, π_m so geschätzt wird, dass die Wahrscheinlichkeit für das Auftreten der empirischen Daten maximiert wird. Die Population besteht im vorliegenden Fall aus der Menge aller möglichen Paare (x, r) von Reizstärkewerten x und zugehörigen Antworten r einer Person. Als zu schätzende Populationsparameter kommen grundsätzlich sämtliche Parameter des für die psychometrische Funktion angenommenen Modells in Betracht. Ist x_i die im i . Trial dargebotene Reizstärke und r_i die daraufhin abgegebene Versuchspersonenantwort, so stellen die während des Verfahrens bis zum n . Trial erhaltenen Paare $(x_1, r_1), \dots, (x_n, r_n)$ zu diesem Zeitpunkt die gezogene Stichprobe, bzw. den empirischen Datensatz dar. Während die Auswahl der Reizstärkenwerte x_i jeweils durch den Adaptionmechanismus erfolgt, stellen die aufgenommenen Antworten r_i Realisierungen einer diskreten Zufallsvariablen R mit einer von x abhängigen Wahrscheinlichkeitsfunktion $PF(r, x)$ dar. R kann dabei zwei mögliche Werte annehmen, welche durch eine beliebige binäre Kodierung ausgedrückt werden können, wie etwa

$$r = \begin{cases} +1 : & \text{positive Antwort,} \\ -1 : & \text{negative Antwort,} \end{cases} \quad (4.15)$$

²⁰Die Grundlagen der Maximum-Likelihood-Schätzung werden beispielsweise bei Lehmann (2002) beschrieben. Das Lehrbuch diente als Quelle für die nachfolgende Darstellung.

welche im Folgenden verwendet werden soll. Mit Hilfe der psychometrischen Funktion ψ kann die Wahrscheinlichkeitsfunktion PF wie folgt angegeben werden:

$$PF(r, x) = \begin{cases} \psi(x) & \text{für } r = +1, \\ 1 - \psi(x) & \text{für } r = -1. \end{cases} \quad (4.16)$$

Um durch Ziehung einer Stichprobe die Werte bestimmter Parameter einer Population schätzen zu können, ist grundsätzlich vorauszusetzen, dass sich die Wahrscheinlichkeitsfunktion bzw. das zu deren Beschreibung verwendete parametrische Modell, zumindest während des Vorgangs der Datenerhebung nicht ändert. Für die Anwendung von Parameterschätzungen auf adaptive Verfahren ist daher zu fordern, dass die psychometrische Funktion in Lage und Form während eines Runs unverändert bleibt (vgl. Annahme 3, 4.1.4). Bei Gültigkeit dieser Annahme kann die Likelihood-Funktion \mathcal{L}_n im n . Trial als Wahrscheinlichkeit des bis dorthin erhobenen Datensatzes $(x_1, r_1), \dots, (x_n, r_n)$ in Abhängigkeit der Populationsparameter π_1, \dots, π_m definiert werden:

$$\mathcal{L}_n(\pi_1, \dots, \pi_m) = P\left[(x_1, r_1), \dots, (x_n, r_n); \pi_1, \dots, \pi_m\right]. \quad (4.17)$$

Weiterhin ist die Maximum-Likelihood-Schätzung definiert durch die Wahl derjenigen Werte der Parameter π_1, \dots, π_m , für welche $\mathcal{L}_n(\pi_1, \dots, \pi_m)$ maximiert wird.

Unter der zusätzlichen Annahme der Unabhängigkeit der Trials (vgl. Annahme 2, Abs. 4.1.4) lässt sich \mathcal{L}_n durch das Produkt der einzelnen Wahrscheinlichkeiten für das Auftreten der Antworten r_i bei den einzelnen Darbietungen x_i berechnen:

$$\mathcal{L}_n(\pi_1, \dots, \pi_m) = \prod_{i=1}^n P(x_i, r_i; \pi_1, \dots, \pi_m). \quad (4.18)$$

Die Einzelwahrscheinlichkeiten $P(x_i, r_i; \pi_1, \dots, \pi_m)$ entsprechen dabei jeweils dem Wert der nach Gleichung (4.16) definierten Wahrscheinlichkeitsfunktion PF für $x = x_i$ und $r = r_i$, wenn die psychometrische Funktion $\psi(x)$ durch eine Modellfunktion $\hat{\psi}(x; \pi_1, \dots, \pi_m)$ beschrieben wird. Für letztere kann grundsätzlich jede beliebige Funktion herangezogen werden, sofern sie eine hinreichend genaue Approximation darstellt (vgl. Annahme 4, Abs. 4.1.4). Die Autoren des Best-PEST-Verfahrens beschreiben in ihren Publikationen (Pentland, 1980; Lieberman und Pentland, 1982) die Verwendung der logistischen Funktion, welche bei Einsatz des nAFC-Paradigmas mit einer konventionellen Ratekorrektur kombiniert wird. Diese Modellfunktion $\hat{\psi}(x; \theta, \beta, \gamma, \lambda)$ und deren Parameter wurden bereits in Abschnitt 4.1.2.5 dargestellt. Dabei gibt θ die

Lage des Wendepunkts an und entspricht gleichzeitig der zu ermittelnden Schwelle ϑ . Der Wert dieses Parameters ist unbekannt und soll im Laufe des Verfahrens geschätzt werden. Alle anderen Parameter werden als bekannt angenommen und sind vor der Durchführung des Verfahrens festzulegen (vgl. Abs. 4.1.7.4 und Abs. 4.1.7.5)²¹. Es kann nun Gleichung (4.18) für das Best-PEST-Verfahren präzisiert werden:

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n P(x_i, r_i; \theta) , \quad (4.19)$$

mit

$$P(x_i, r_i; \theta) = \begin{cases} \hat{\psi}(x_i; \theta, \beta, \gamma, \lambda) & \text{für } r_i = +1, \\ 1 - \hat{\psi}(x_i; \theta, \beta, \gamma, \lambda) & \text{für } r_i = -1. \end{cases} \quad (4.20)$$

Die Maximum-Likelihood-Schätzung $\hat{\theta}_{ML,n}$ nach dem n . Trial und damit der Wert der Reizstärke für die $(n+1)$ -te Darbietung x_{n+1} ergibt sich schließlich durch Ermittlung des Maximums der Likelihood-Funktion:

$$x_{n+1} = \hat{\theta}_{ML,n} \in \{\theta \in \Theta \mid \mathcal{L}_n(\theta) = \max\} , \quad (4.21)$$

wobei Θ die Menge aller möglichen Werte des Parameters θ repräsentiert.

Die Bedingungen für die Existenz einer eindeutigen Lösung dieser Gleichung werden in den Publikationen des Verfahrens nicht thematisiert. Bei Lieberman und Pentland (1982) wird lediglich darauf hingewiesen, dass Likelihood-Funktionen mit einem Plateau im Maximum auftreten können. Deren Zustandekommen wird nicht erklärt. Neben einer funktionstheoretischen Begründung kommen hierfür auch praktische Ursachen wie die Diskretisierung der Reizstärke (s.u.) oder die Genauigkeit der numerischen Verarbeitung in Betracht. Wenn ein derartiger Funktionsverlauf vorliegt, wählen die Autoren den Mittelwert aus den Randpunkten des Plateaus.

Zur Berechnung der Likelihood-Funktion Aufgrund der grundsätzlich endlichen Rechen- und Speicherkapazität von Computern, muss die Berechnung der Likelihood-Funktion für eine begrenzte Auswahl diskreter möglicher Werte des Populationsparameters θ , $\Theta = \{\theta_1, \theta_2, \dots, \theta_t\}$ erfolgen. Durch die Berücksichtigung bereits vorhandener

²¹In der bisherigen Darstellung der Maximum-Likelihood-Methode wurde die Anzahl der zu schätzenden Populationsparameter nicht eingeschränkt. Theoretisch ist es also denkbar, mehrere oder sogar alle Parameter der psychometrischen Funktion innerhalb eines Verfahrens zu schätzen (vgl. etwa Hall, 1981; Treutwein, 1995; Treutwein und Strasburger, 1999). Derartige Ansätze sollen jedoch im Rahmen dieser Arbeit nicht weiter diskutiert werden.

Informationen kann in der Regel für die Lage der Schwelle im Vorfeld einer Untersuchung ein endliches Intervall angegeben werden. Zudem wird in der Praxis nur eine begrenzte Genauigkeit für die Parameterschätzung gefordert, weshalb eine endliche Auflösung der Parameterwerte ausreichend ist. Andererseits ist in der Praxis auch die Stärke möglicher Reizdarbietungen aus verschiedenen Gründen auf ein endliches Intervall $[x_{min}, x_{max}]$ begrenzt und die einzelnen Reizstärken sind nur mit einer endlichen Genauigkeit reproduzierbar, so dass insgesamt auch nur eine endliche Anzahl an Stimuli mit den Reizstärkenwerten $\{x_s \in [x_{min}, x_{max}] | s = 1, 2, \dots, S\}$ generiert werden kann. Zugunsten einer einfachen Struktur des Algorithmus werden in der Implementierung des Best-PEST nach Lieberman und Pentland (1982) für die Berechnung der Likelihood-Funktion die selben Werte des Definitionsbereichs verwendet, welche auch für Reizdarbietungen zu Verfügung stehen²², d.h. es gilt:

$$\Theta = \{\theta_1, \theta_2, \dots, \theta_T\} = \{x_s \in [x_{min}, x_{max}] | s = 1, 2, \dots, S\} . \quad (4.22)$$

Wie aus Gleichung (4.19) auf Seite 52 hervorgeht, muss die Likelihood-Funktion nicht in jedem Trial vollständig neu berechnet werden, sondern kann aus Multiplikation der Einzelwahrscheinlichkeit $P(x_n, r_n; \theta)$ mit dem Funktionswert $\mathcal{L}_{n-1}(\theta)$ des vorherigen Trials, hervorgehen:

$$\mathcal{L}_n(\theta) = \mathcal{L}_{n-1}(\theta) \cdot P(x_n, r_n; \theta) . \quad (4.23)$$

Abbildung 4.6 zeigt zur Veranschaulichung exemplarisch die Entwicklung der Likelihood-Funktion $\mathcal{L}_n(\theta)$ in den ersten Trials eines Runs und den zugehörigen Verlauf der Reizstärke x_n .

Da die einzelnen Faktoren in Gleichung (4.23) Werte zwischen 0 und 1 besitzen, nimmt die Likelihood-Funktion im Verlauf des Verfahrens Werte nahe bei Null an, weswegen es nach einer großen Anzahl an Trials zu Ungenauigkeiten der Zahlendarstellung (sogenannten *arithmetischen Underflows*) kommen kann, wodurch geringe Unterschiede in den Funktionswerten evtl. nicht mehr aufgelöst werden²³. Lieberman und Pentland (1982) führen deshalb eine Logarithmustransformation der Ordinatenwerte durch, welche die Lage des Maximums aufgrund deren strenger Monotonie nicht verändert. Durch diese Maßnahme konnte außerdem zum damaligen Zeitpunkt Rechenzeit eingespart werden, da Multiplikationen durch Additionen ersetzt werden

²²Nach Treutwein (1995) gilt dies auch für alle vergleichbaren Verfahren, welche dort genannt werden.

²³Die Höhe der Wahrscheinlichkeit für das Auftreten derartig geringer Zahlenwerte – insbesondere unter den heutigen technischen Gegebenheiten, konnte im Rahmen dieser Arbeit nicht eingeschätzt werden.

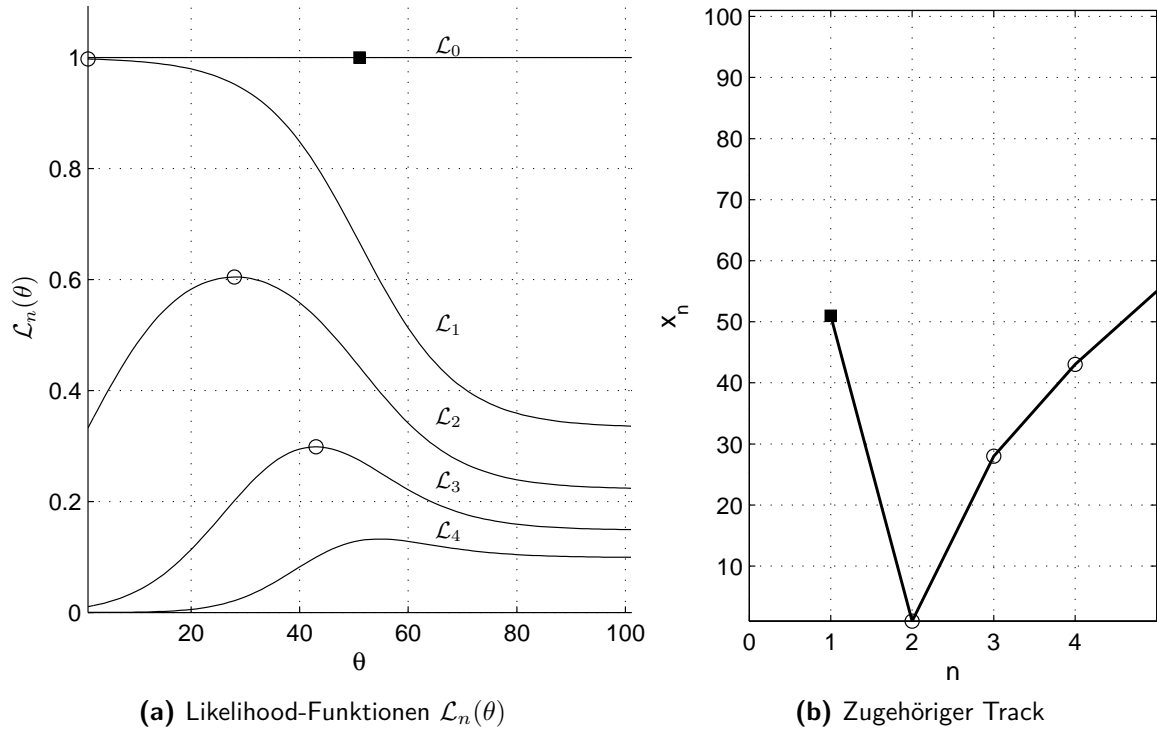


Abbildung 4.6: Exemplarische Darstellung der Likelihood-Funktionen $\mathcal{L}_n(\theta)$ und des zugehörigen Reizstärkenverlaufs für die ersten 4 Trials eines Runs. Verfahrensparameter: 3AFC-Paradigma, $\lambda = 0$, $\beta = 9$, $\Theta = \{1, 2, \dots, 101\}$. Die erste Reizdarbietung x_1 erfolgt in der Mitte des Range, die übrigen x_2, \dots, x_4 jeweils für den Abszissenwert θ im Maximum der Funktion. Die zugehörigen Versuchspersonenantworten sind wie folgt gekennzeichnet: ■: positive Antwort, ○: negative Antwort.

konnten. Zur Vermeidung von Underflows kann anstelle der Logarithmierung auch eine Normierung der Likelihood-Funktion nach jedem Multiplikationsschritt bzw. Trial erfolgen. Eine bestimmte Art der Normierung ist ohnehin dann notwendig, wenn das Verfahren per vorgegebenem Konfidenzintervall beendet werden soll (s. Gl. (4.25) auf S. 55).

4.1.7.2 Terminierung und Ergebnis

Das Verfahren wird standardmäßig nach einer vorgegebenen Anzahl an Trials beendet. Der nach der letzten Reizdarbietung ermittelte Wert der Maximum-Likelihood-Schätzung stellt das Ergebnis des Verfahrens dar.

Als eine alternative Möglichkeit der Terminierung wird in der Literatur die Berücksichtigung eines dynamischen Kriteriums beschrieben. Hierbei wird das Verfahren solange ausgeführt, bis die Breite des Konfidenzintervalls CI_n der aktuellen Schätzung

$\hat{\vartheta}_{ml,n}$, für ein vorgegebenes Konfidenzniveau γ (auch als *Konfidenzkoeffizient* bezeichnet) einen festgelegten Wert CI_{max} erreicht bzw. unterschreitet. Für die Berechnung des Konfidenzintervalls gibt verschiedene Ansätze.

Die Autoren des Best-PEST deuten ein Vorgehen an, bei welchem die Werte der Likelihood-Funktion innerhalb des Bereiches $\pm CI_{max}/2$ um die aktuelle Schätzung $\hat{\vartheta}_{ml,n}$ herum aufsummiert werden und diese Summe mit dem Konfidenzkoeffizienten verglichen wird. Formal kann ein derartiger Ansatz durch folgende Ungleichung ausgedrückt werden:

$$\int_{\hat{\vartheta}_{ml,n} - \frac{CI_{max}}{2}}^{\hat{\vartheta}_{ml,n} + \frac{CI_{max}}{2}} \mathcal{L}_{n,N}(\theta) d\theta \geq \gamma . \quad (4.24)$$

mit

γ : Konfidenzkoeffizient; $0 \leq \gamma \leq 1$,
 $\mathcal{L}_{n,N}(\theta)$: normierte Likelihood-Funktion nach dem n.Trial, welche wie folgt definiert ist:

$$\mathcal{L}_{n,N}(\theta) = \frac{\mathcal{L}_n(\theta)}{\int_{-\infty}^{\infty} \mathcal{L}_n(\theta) d\theta} . \quad (4.25)$$

Treutwein (1995) beschreibt die Berechnung eines sogenannten *symmetrischen zweiseitigen bayesschen Wahrscheinlichkeitsintervalls* (vgl. etwa Martz und Waller, 1982, S. 208). Es ist dann $CI_n = \theta_u - \theta_l$, wobei die obere und die untere Grenze θ_u bzw. θ_l , durch folgende Gleichungen definiert werden:

$$\int_{-\infty}^{\theta_l} \mathcal{L}_{n,N}(\theta) d\theta = \frac{1-\gamma}{2} \quad \text{und} \quad \int_{\theta_u}^{\infty} \mathcal{L}_{n,N}(\theta) d\theta = \frac{1-\gamma}{2} . \quad (4.26)$$

Die Berechnung der auf die Fläche unter dem Funktionsgrafen normierten Likelihood-Funktion²⁴ $\mathcal{L}_{n,N}(\theta)$ kann für den Fall diskreter Parameterwerte $\theta \in \{\theta_1, \theta_2, \dots, \theta_T\}$ erfolgen, indem das Integral in Gleichung (4.25) durch eine Summe angenähert wird:

$$\mathcal{L}_{n,N}(\theta_i) = \frac{\mathcal{L}_n(\theta_i)}{\sum_{j=1}^T \mathcal{L}_n(\theta_j)} . \quad (4.27)$$

²⁴Im Kontext der bayesschen Theorie wird anstelle der nach Gleichung (4.25) normierten Likelihood-Funktion von einer *A-posteriori-Wahrscheinlichkeitsdichte* gesprochen.

Voraussetzung für die Gültigkeit dieser Approximation ist, dass die Likelihood-Funktion an den Rändern des betrachteten Bereiches möglicher Parameterwerte verschwindet. Hieraus ergibt sich die Konsequenz, dass dieser Bereich genügend groß gewählt werden sollte, so dass die Wahrscheinlichkeit für die Lage des wahren Parameterwertes (d.h. der wahren Schwelle) an dessen Rändern vernachlässigt werden kann.

Bei der Verwendung der bisher aufgeführten Ansätze ist zu beachten, dass die Dauer eines Verfahrensdurchlaufs dabei nicht nur von γ und CI_{max} , sondern auch wesentlich von der Wahl des Spread-Parameters β abhängt (Madigan und Williams, 1987; Treutwein, 1995). Da der wahre Wert von letzterem in der Forschungspraxis selten bekannt ist, kann ein Konfidenzintervall auf die beschriebene Weise in der Regel nicht als absolute, sondern nur als relative Größe bestimmt werden. Für einen Überblick weiterer Ansätze und Möglichkeiten sei auf die zusammenfassende Darstellung bei Treutwein (1995) verwiesen.

Ob durch die Anwendung derartiger dynamischer Abbruchkriterien gegenüber der Terminierung anhand einer festen Anzahl an Trials, im Hinblick auf Bias und Fehlervarianz eine Verbesserung der Güte der Schwellenschätzung erzielt werden kann, ist umstritten (vgl. Treutwein, 1997; Anderson, 2003).

4.1.7.3 Initialisierung des Verfahrens

Im Fall des Best-PEST-Verfahrens erfüllt der Vorgang der Initialisierung zwei verschiedene Funktionen. Zum einen beinhaltet er die Festlegung des Anfangswertes der Reizstärke. Zum anderen kann aufgrund von bereits vorhandener Information über die Lage der Schwelle eine Vorformung (auch: *A-priori-Formung*) der Likelihood-Funktion vorgenommen werden, durch welche in den ersten Trials Sprünge der Reizstärke in uninteressante Regionen vermieden werden (Pokorny, 1998; Treutwein, 1995; vgl. auch Abb. 4.6 auf S. 54). Es ist hierbei allerdings dafür Sorge zu tragen, dass die A-priori-Formung der Likelihood-Funktion nicht derart stark ausgeprägt ist, dass hieraus eine relevante Verschiebung des Schätzwertes (Bias) resultiert (vgl. Martz und Waller, 1982, S. 184).

Nach Lieberman und Pentland (1982) erfolgt die Initialisierung durch Definition impliziter Trials, welche keine tatsächlichen Reizdarbietungen enthalten, sondern deren Antworten bereits vor Versuchsdurchführung anhand von Vorwissen festgelegt werden. Für das ja/nein-Paradigma sind dabei die Ränder x_{min} und x_{max} des Reizstärken-

bereichs so zu wählen, dass sie im Fall einer Darbietung mit hoher Wahrscheinlichkeit nicht detektiert bzw. detektiert werden und entsprechend eine negative bzw. positive Antwort der Versuchsperson evozieren. Unter dieser Annahme geben die Autoren zu Beginn des Verfahrens die zwei folgenden impliziten Trials vor:

$$(x_1, r_1) = (x_{max}, +1) \text{ und } (x_2, r_2) = (x_{min}, -1) . \quad (4.28)$$

Anhand von Gleichung (4.21) auf Seite 52 wird hieraus der Wert für die erste tatsächliche Reizdarbietung (in diesem Fall x_3) berechnet. Für das 2AFC-Paradigma berücksichtigen die Autoren ausschließlich die Information bezüglich des unteren Randes x_{min} , nach welcher eine positive Antwort mit der Ratewahrscheinlichkeit von $1/n = 0.5$ auftritt. Das Auslassen des oberen Randes x_{max} wird nicht begründet. Es werden die folgenden zwei impliziten Trials definiert:

$$(x_1, r_1) = (x_{min}, -1) \text{ und } (x_2, r_2) = (x_{min}, +1) . \quad (4.29)$$

Ein verallgemeinerte Festlegung für das nAFC-Paradigma wird nicht ausdrücklich beschrieben. Es kann daher nur induktiv geschlossen werden, dass allgemein n implizite Trials am unteren Rand x_{min} definiert werden, wobei genau einem eine positive Antwort zugeordnet wird. Grundsätzlich wird das Vorgehen der Initialisierung mittels impliziter Trials von den Autoren nicht weiter begründet. Es wird insbesondere nicht erklärt, nach welchem Kriterium die Anzahl der impliziten Trials im Hinblick auf den Grad der Vorformung der Likelihood-Funktion in geeigneter Weise bestimmt werden kann.

Eine weitere Möglichkeit der Initialisierung besteht in der expliziten Vorgabe einer bestimmten A-priori-Likelihood-Funktion $\mathcal{L}_0(\theta)$. So verwenden z.B. Watson und Pelli (1983) im Kontext des dem Best-PEST ähnlichen *QUEST* eine Normalverteilungsdichte, deren Lage- und Formparameter μ und σ aufgrund von bereits vorhandenem Wissen festgelegt werden²⁵. Diese beeinflusst die Platzierungen der Reizstärkenwerte während des Verfahrens, wird jedoch von den Autoren unmittelbar von der Berechnung der abschließenden Schwellenschätzung ausgeschlossen, indem die Likelihood-Funktion durch diese dividiert wird.²⁶

In der Literatur werden darüber hinaus die Durchführung des Best-PEST und die-

²⁵Watson und Pelli (1983) verwenden die Terminologie der bayesschen Theorie und sprechen daher anstelle einer A-priori-Likelihood von einer *A-priori-Wahrscheinlichkeitsdichte*.

²⁶Hiermit sei nichts darüber ausgesagt, inwieweit durch diese Maßnahme die Ausprägung eines Bias vermindert werden kann.

sem ähnlicher Verfahren unter Verzicht auf eine Vorformung der Likelihood-Funktion beschrieben (vgl. Treutwein, 1995). In diesem Fall wird letztere zu Beginn des Verfahrens mit einer über den gesamten Reizstärken- bzw. Parameterbereich konstanten Wahrscheinlichkeit (vgl. Abb. 4.6 auf S. 54) belegt. Die Wahl des Anfangswertes der Reizstärke ist damit grundsätzlich beliebig, fällt jedoch durch den Best-PEST-Algorithmus auf die Mitte des Intervalls (Verhalten bei Vorliegen eines Plateaus, s. S. 52).

4.1.7.4 Wahl des Spread-Parameters

Die Wahl eines geeigneten Wertes für den Spread-Parameter β ist anhand von Vorwissen über die Form der psychometrischen Funktion festzulegen, welche häufig auch anhand der Steigung (Slope) S im Wendepunkt beschrieben wird (Treutwein, 1995). Der Zusammenhang zwischen S und β ist durch Gleichung (4.6) auf Seite 38 gegeben. Zwar beschreiben Lieberman und Pentland (1982) das Verfahren als robust gegenüber Fehleinschätzungen von β , empfehlen aber eher die Wahl eines zu großen Wertes, falls der wahre unbekannt ist oder negative Antworten der Versuchspersonen trotz überschwelliger Reizstärke (*Lapsing*, s. Abs. 4.1.2.4) zu erwarten sind. Beide Aussagen werden durch Simulationsergebnisse von Madigan und Williams (1987) unterstützt.

4.1.7.5 Wahl der Lapsing-Rate

Die Festlegung der Lapsing-Rate λ (vgl. Abs. 4.1.2.5) ist – ähnlich wie die der False-Alarm-Rate γ im Fall des ja/nein-Paradigmas – mit der Schwierigkeit verbunden, dass deren Höhe von der Versuchsperson im Einzelnen abhängt und daher in der Regel nicht im Vornherein angegeben werden kann. In der Publikation des Best-PEST-Verfahrens wird dieser Parameter von den Autoren im Modell der psychometrischen Funktion nicht berücksichtigt und somit implizit auf den Wert 0 gesetzt. Andere Autoren empfehlen die Wahl eines pauschalen Wertes von 1% (vgl. Hall, 1981; Harvey, 1986; Klein, 2001). In der Literatur wird außerdem als ein möglicher Lösungsansatz ein Vorgehen beschrieben, bei welchem durch die Beimengung deutlich überschwelliger Reize in Form sogenannter Catch-Trials²⁷, λ als der Anteil der negativen Antworten in diesen Trials geschätzt und für die abschließende Berechnung der Schwellenschätzung berücksichtigt werden kann (Treutwein, 1995). Darüberhinaus ist es grundsätzlich auch denkbar, die Lapsing-Rate als einen zusätzlichen unbekannten

²⁷Zum Einsatz unterschwelliger Reize in Catch-Trials vgl. Abs. 4.1.2.3.

Parameter mittels der Maximum-Likelihood-Methode zu schätzen (vgl. Anm. 21 auf S. 52 oder auch Treutwein und Strasburger, 1999).

4.1.8 Interleaving

Als *Interleaving* wird die gleichzeitige Durchführung mehrerer adaptiver Verfahren auf ineinander verschachtelte Weise verstanden. Dabei folgt der Verlauf jedes einzelnen Tracks ausschließlich seinem eigenen Adaptionsmechanismus. Die Festlegung des aktuellen Trials erfolgt durch Auswahl des aktuellen Tracks²⁸ aus der Menge der noch nicht beendeten Tracks. Dies kann grundsätzlich je nach Fragestellung auf unterschiedliche Weise geschehen. Eine Möglichkeit, von welcher in der Literatur an mehreren Stellen berichtet wird (u.a. bei Cornsweet, 1962; Leek, 2001; Levitt, 1971; Treutwein, 1995), ist die zufällige Auswahl des Tracks entsprechend einem Ziehen mit Zurücklegen²⁹ (engl. *randomly interleaving*). Sie wird vor allem eingesetzt um innerhalb eines Verfahrens die Unabhängigkeit der Trials voneinander zu erhöhen, welche eine grundlegende Annahme bei der Durchführung adaptiver Verfahren ist (vgl. Abs. 4.1.4). Durch die Zufallsabhängigkeit der Trial-Abfolgen wird es der Versuchsperson erschwert, den Verlauf eines Tracks zu antizipieren und ihr Antwortverhalten danach auszurichten. Dies ist bei Anwendung des ja/nein-Paradigmas aufgrund des vorliegenden Kriterienproblems von Bedeutung (vgl. Abs. 4.1.2.3). Außerdem werden sensitivitätsbedingte Sequenzeffekte wie der Hysterese-Effekt³⁰ reduziert, weshalb diese Variante auch (sinnvoll) für das nAFC-Paradigma verwendet werden kann. Darüberhinaus werden auch Untersuchungen beschrieben, in welchen die Auswahl des aktuellen Tracks ohne Zurücklegen (etwa bei Leek und Watson, 1984) oder auch nach bestimmten Regeln erfolgt (etwa bei Levitt, 1968).

Beim Interleaving können sich die einzelnen Tracks u.U. hinsichtlich bestimmter physikalischer Parameter oder ihrer Adaptionsmechanismen unterscheiden. So verwenden beispielsweise Leek und Watson (1984) in einer Untersuchung zum Einfluss von Lern-

²⁸Im Kontext des Interleavings wird in der Literatur anstelle eines *Runs* (vgl. Abs. 4.1.3) auch von einem *Track* gesprochen, auch wenn dabei nicht explizit die Reizstärkenwerte gemeint sind, sondern allgemein der Durchlauf eines Verfahrens einem bestimmten Adaptionsmechanismus folgend (vgl. etwa Hansen, 2008; Leek, 2001).

²⁹Bei einem Ziehen ohne Zurücklegen würde ein gerade verwendeter Track erst dann wieder zur Auswahl stehen, wenn alle übrigen für jeweils eine Reizdarbietung herangezogen wurden.

³⁰Als Hysterese-Effekt bezeichnet man in diesem Zusammenhang das Phänomen, dass sich die Detektions- bzw. Diskriminationswahrscheinlichkeiten eines Reizes unterscheiden, je nach dem ob sich der Track dem Konvergenzniveau von überschwelligen oder unterschwelligen Werten her nähert (Levitt, 1968).

prozessen auf die Detektion von Tönen innerhalb eines festen Tonmusters zehn Tracks von jeweils unterschiedlicher Frequenz des Zielreizes.

4.2 Semantisches Differential

4.2.1 Einleitung

Das semantische Differential ist ein in den 1950er Jahren von Osgood und Mitarbeitern entwickeltes Verfahren zur Messung von Bedeutungen beliebiger Objekte. Es handelt sich dabei weniger um einen vollständig ausgearbeiteten Test mit fest vorgeschriebenen Items, als vielmehr um eine generelle Technik, deren konkrete Ausgestaltung an die für die jeweilige Anwendung interessierende Fragestellung anzupassen ist (Osgood et al., 1967). Dementsprechend vielfältig sind die Einsatzmöglichkeiten. Neben zahlreichen psychologischen Anwendungen (vgl. Bortz und Döring, 2006, S. 187) bezeichnet Merten (1983, S. 253) die Gebiete Marktforschung, Medienanalyse und Sozialforschung. Ein Beispiel für eine Anwendung in der perzeptiven Audioevaluati-on stellt eine Untersuchung von Maempel (2001) zum Einfluss klanggestalterischer Mittel auf das Hörerurteil dar.

4.2.2 Grundzüge des Verfahrens

Das Verfahren kann sowohl für die Messung konnotativer als auch denotativer Bedeutungen herangezogen werden (Bortz und Döring, 2006, S. 185f). Auf einem Satz von 10–30 Ratingskalen (Maempel, 2007) werden die Ausprägungen perzeptiver oder affektiver Qualitäten als Werte ebenso vieler abhängiger Variablen abgefragt. Dabei wird jedes Urteilsobjekt auf sämtlichen Skalen bewertet. Während die Reihenfolge der Skalen für jedes Objekt unverändert bleibt³¹, sollte die der Objekte zur Kontrolle von Sequenzeffekten zwischen den Versuchspersonen variieren (vgl. Abs. 3.1.3). Die Mittelwerte über die intervallskalierten Urteile der Versuchspersonen zu einem bestimmten Objekt, können mittels eines sogenannten *Polaritätsprofils* grafisch veranschaulicht werden (Hofstätter, 1980). Abbildung 4.8 zeigt exemplarisch ein solches für die Eingangs genannte Untersuchung.

³¹Osgood et al. (1967) beschreiben ursprünglich noch eine alternative Möglichkeit des Verfahrensablaufs: die sogenannte *Form I* besteht aus einem rotierenden Vorgehen, bei welchem die Urteilsaufgaben derart angeordnet werden, dass eine maximale Anzahl an verschiedenen Objekten und Skalen zwischen den Wiederholungen jeweils der Objekte und der Skalen dargeboten wird. Dieses Vorgehen benötigt jedoch einen wesentlich höheren Zeitaufwand, da die Exposition der Objekte jeweils mehrmals (für jede Skala einmal) erfolgt.

4. Zu implementierende Testverfahren

Klassifikation	Kurzbezeichnung	Positiver Skalenpol	Negativer Skalenpol
EPA-Indikatoren	Gefallen	gefällt mir	gefällt mir nicht
	Abwechslung	abwechslungsreich	eintönig
	Härte	hart	weich
musikalisch-strukturelle Merkmale	Melodie	melodisch	unmelodisch
	Rhythmus	rhythmisch	unrhythmisch
	Tempo	schnell	langsam
klangbeschreibende Merkmale	Lautstärke	laut	leise
	Helligkeit	hell	dumpf
	Baß	lauter Baß	leiser Baß
	Deutlichkeit	deutlicher Gesang	undeutlicher Gesang
emotionsbeschreibende Merkmale	Laune	macht gute Laune	macht schlechte Laune
	Tanzen	gut zum Tanzen	nicht gut zum Tanzen
	Träumen	gut zum Träumen	nicht gut zum Träumen

Abbildung 4.7: Tabellarische Übersicht und Einteilung in Klassen der von Maempel in einer Untersuchung zum Einfluss klanggestalterischer Maßnahmen auf das Hörerurteil verwendeten Skalenpolbeschriftungen. [Quelle: Maempel (2001, Tab. 1)]

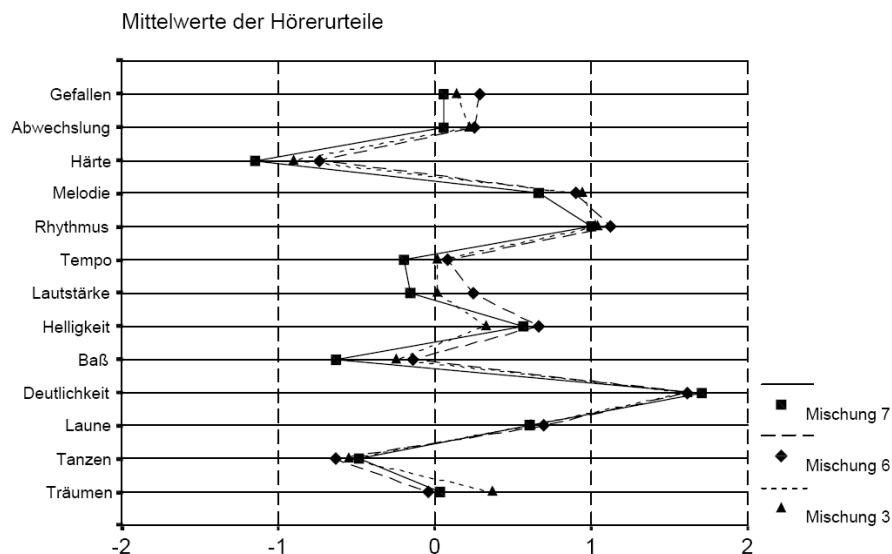


Abbildung 4.8: Polaritätsprofil eines Popmusiktitels aus der genannten Untersuchung von Maempel. Die drei verschiedenen Linienzüge stellen die Profilverläufe verschiedener Abmischungen dar. [Quelle: Maempel (2001, Abb. 1)]

4.2.3 Operationalisierung des Urteils

Als ein sogenanntes *Rating-Verfahren* zählt das semantische Differential zu den Techniken der *direkten Skalierung* (vgl. Abb. 3.2 auf S. 26). Letztere zeichnen sich dadurch aus, dass Versuchspersonen unmittelbar über die Stärke von Sinneseindrücken (perzeptive Qualitäten) oder der durch diese hervorgerufenen psychologischen Prozesse (affektive Qualitäten) Auskunft geben, indem sie direkt kardinale Urteile über diese fällen. Im Fall des Rating-Verfahrens, welches auch als *kategoriales Rating* oder *kategoriale Skalierung*³² bezeichnet wird (Gescheider, 1997, S. 214ff; Bech und Zacharov, 2006, S. 72), sollen intervallskalierte Urteile erzeugt werden. Dabei ist eine Skala mit markierten Abschnitten (*Stufen*) des zu erfassenden Merkmalskontinuums gegeben, welche von der Versuchsperson als äquidistant aufgefasst werden sollen. Jedem Urteilsobjekt ordnet die Versuchsperson entsprechend der Merkmalsausprägung eine Stufe zu. Alternativ zur Vorgabe diskreter, auszuwählender Stufen kann die Versuchsperson selbst eine Aufteilung des Kontinuums vornehmen, indem sie die Möglichkeit erhält, jedem Objekt eine beliebige Position auf der Skala zuzuweisen. Diese Variante des sogenannten *grafischen Ratings* bietet besonders gute Voraussetzungen für intervallskalierte Urteile (Bortz und Döring, 2006, S. 179f). Die Auswertung der sich individuell ergebenden Skalenpositionen kann insbesondere durch computergestützte Datenerhebung vereinfacht werden. Hierbei werden üblicherweise Schieberegler (auch: *Slider*) verwendet. Darüber hinaus kommen Mischformen der beiden Skalenvarianten zum Einsatz, sogenannte *kontinuierliche Skalen*, welche zwar markierte Stufen besitzen, jedoch eine Auswahl jedes beliebigen Wertes dazwischen zulassen (vgl. Bech und Zacharov, 2006, S. 71ff; Kap. 10). Die beim semantischen Differential abgegebenen Urteile sind für sich genommen absolut, d.h. die Versuchsperson bezieht sich dabei auf eine interne Referenz. Für die Vermittlung der Bedeutung einer Skala, bzw. des zu beurteilenden Merkmals (Qualität), werden beim semantischen Differential die Skalenpole beschriftet. Von großer Bedeutung ist dabei die Wahl geeigneter Adjektivpaare (vgl. Abs. 4.2.5.2).

4.2.4 Wahl der Objekte

Die Wahl der Objekte richtet sich nach der jeweiligen interessierenden Forschungsfrage und unterliegt grundsätzlich keinen Einschränkungen. Ursprüngliche Anwendungen

³²Der Zusatz *kategorial* ist in diesem Fall nicht auf das Skalenniveau zu beziehen, sondern resultiert aus der Einteilung einer Skala in gleichabständige Stufen, welche in diesem Zusammenhang missverständlicher Weise als „*Kategorien*“ bezeichnet werden.

des semantischen Differentials untersuchen die Bedeutungen von Begriffen, welche sowohl Personen, soziale Rollen, Gruppen oder auch Gegenstände beschreiben können. Darüber hinaus stehen auch nonverbale Objekte, wie künstlerische Gemälde, TAT- oder Rorschach-Figuren³³, oder auch Unterwasser-Schallsignale im Mittelpunkt von Untersuchungen (Osgood et al., 1967). Für die Anwendung in der Audioevaluation interessieren in erster Linie akustische Reize, evtl. in Kombination mit visuellen Reizen. Dies können Ausschnitte von Musikstücken, Sprachsignalen oder Geräuschen sein. Die dabei systematisch variierenden Merkmale entsprechen den unabhängigen Variablen (s. Abs. 3.1.1) der Untersuchung. Die Anzahl der Objekte ist somit durch die Wahl des Versuchsdesigns (s. Abs. 3.2) festgelegt. In der eingangs genannten Untersuchung verwendet Maempel (2001) beispielsweise drei Popmusikstücke in jeweils drei verschiedenen Abmischungen.

4.2.5 Wahl der Skalen

4.2.5.1 Skalenformat

Osgood et al. (1967) verwenden siebenstufige bipolare Skalen, da diese laut ihren Untersuchungen gegenüber fünf- und neunstufigen Skalen zu einer gleichmäßigeren Verwendung der einzelnen Skalenstufen durch die Versuchspersonen führen. Unter bestimmten Bedingungen sind diesen aber auch andere Stufenanzahlen vorzuziehen. Maßgeblich für die Größenordnung sind u.a. die Schwierigkeit der Urteilsaufgabe und die Differenzierungskapazität der Versuchspersonen (Bortz und Döring, 2006, S. 180). Die Wahl einer ungeraden vs. geraden Stufenanzahl wird in der Literatur im Fall bipolarer Skalen (vgl. Abs. 4.2.5.2) vor dem Hintergrund des *Ambivalenz-Indifferenz-Problems* diskutiert (Bortz und Döring, 2006, S. 180). Unter diesem versteht man die Schwierigkeit der Interpretation einer neutralen mittleren Antwortkategorie, da diese gleichermaßen auf eine mögliche Unentschlossenheit (Ambivalenz) des Urteilers sowie auf dessen Gleichgültigkeit gegenüber dem Urteilsobjekt (Indifferenz) zurückgeführt werden kann. Andererseits birgt die Verwendung von Skalen mit geraden Stufenanzahlen den Nachteil, dass diese auch dann ein tendenzielles Urteil in Richtung eines der beiden Skalenpole erzwingen, wenn tatsächlich eine neutrale Merkmalsausprägung vorliegt.

³³TAT- und Rorschach-Figuren sind doppeldeutige Bilder bzw. abstrakte grafische Figuren, welche als Interpretationsvorlagen für sogenannte *projektive Tests* dienen (vgl. Bortz und Döring, 2006, S. 191). Diese sollen in diesem Zusammenhang nicht näher erläutert werden.

Bei der Konstruktion des semantischen Differentials sollten die Skalen, welche gleiche Dimensionen des semantischen Raums repräsentieren (s. Abs. 4.2.5.2), jeweils in ihren Polaritäten abwechseln, um der Ausbildung einer Präferenz der Versuchsperson für einen der beiden Skalenpole bei der Positionierung ihrer Urteile entgegenzuwirken. Ebenfalls sollte die Abfolge der Skalen bezüglich der Reihenfolge der zugrundeliegenden Dimensionen rotierend gewählt werden (Osgood et al., 1967, S. 82).

4.2.5.2 Wahl der Adjektivpaare

Bezüglich der inhaltlichen Auswahl der Skalen, welche sich durch die Wahl der zugehörigen Adjektivpaare auszeichnet, geben Osgood et al. (1967, S. 78ff) Kriterien vor, die im Folgenden kurz dargestellt werden sollen.

Faktorielle Zusammensetzung Für jeden Faktor³⁴ wird die Wahl von drei Skalen empfohlen, welche jeweils möglichst hoch auf diesen laden und ihn möglichst reliabel erfassen. Sprachvergleichende Untersuchungen der Autoren ergeben meistens ein dreidimensionales System, bestehend aus den Faktoren *Evaluation*, *Potency* und *Activity* (*EPA-Struktur*). Diese sind jedoch nicht erschöpfend, so dass noch weitere bedeutungsunterscheidende Dimensionen vorliegen können (Osgood et al., 1967, S. 79).

Kontextspezifische Relevanz Hiermit ist die Forderung nach einem inhaltlichen Bezug der Adjektive zu den Urteilsobjekten, bzw. einer natürlich gegebenen Bedeutung innerhalb des Verwendungskontexts, gemeint. Die Einhaltung dieses Kriteriums hängt in hohem Maße von der jeweiligen Fragestellung ab. Häufig sind konnotative Bedeutungen von Interesse, so dass gerade Adjektivpaare verwendet werden sollen, welche dieses Kriterium nicht erfüllen. Ferner weisen die Autoren darauf hin, dass nicht selten gerade für besonders relevante Skalen die faktorielle Zusammensetzung unbekannt ist. Zwar können diese verwendet werden wenn es die Fragestellung erfordert. Jedoch sollen dann zusätzlich Skalen von hoher Faktorladung als Referenzskalen (s.o.) eingeschlossen werden.

Semantische Stabilität Die aufgefasste Bedeutung einer Skala soll über alle Versuchspersonen und für sämtliche Objekte konstant sein.

³⁴Als *Faktor* wird in diesem Zusammenhang eine Dimension des semantischen Raumes bezeichnet. Der Begriff geht aus dem Kontext der Faktorenanalyse hervor, deren Kenntnis an dieser Stelle vorausgesetzt wird.

Linearität der Skala und ihr Verlauf durch den Nullpunkt Diese Forderung gilt ausschließlich für bipolare Skalen, deren Verwendung für das Verfahren ursprünglich vorausgesetzt wurde. Sie verlangt, dass die beiden polaren Adjektive einer Skala durch Punkte des semantischen Raumes repräsentiert werden, welche im selben Abstand zum Ursprung, jedoch in entgegengesetzter Richtung liegen (Malmstrom und French, 1963). Nach Mann et al. (1979, zit. nach Bortz und Döring, 2006, S. 187) führt jedoch die Anwendung unipolarer Skalen zu nur unbedeutend abweichenden Untersuchungsergebnissen.

Eine Auflistung geläufiger Adjektivpaare (in Finnisch, mit Übersetzung ins Englische) für die Anwendung in der Audioevaluation findet man z.B. bei Bech und Zacharov (2006, Appendix B). Diese beinhalten u.a. die Beschreibung auditiver Qualitäten für die Beurteilung von Sprache und Räumlichkeit. In Abbildung 4.7 auf Seite 62 werden die von Maempel (2001) in seiner Untersuchung verwendeten Beschriftungen der Skalenpole dargestellt.

4.3 Repertory-Grid-Technik

4.3.1 Einleitung

Die Repertory-Grid-Technik (RGT) ist ein auf der Basis seiner *Konstrukttheorie* von Kelly (1955) entwickeltes Verfahren zur Erhebung von Dimensionen (*Konstrukten*), mit welchen eine Person subjektiv ihre Umwelt wahrnimmt und strukturiert. Nach Kelly offenbaren sich diese Konstrukte als Urteilkriterien von Versuchspersonen beim Vergleich verschiedener Objekte. Die RGT ist ein halbqualitatives Verfahren, da neben den Konstrukten noch quantitative Daten erhoben werden, welche die spätere Anwendung datenreduzierender statistischer Verfahren, wie z.B. Cluster- und Faktorenanalysen³⁵, ermöglichen. Sie zählt ferner zu den Individualvokabular-Techniken, da die Konstrukte dem Wortschatz jeweils einer einzelnen Person entspringen. Das Verfahren wurde von Berg (2002) und Berg und Rumsey (1999) für den Einsatz in der Audioevaluation überarbeitet (Bech und Zacharov, 2006, S. 56). An der Universität Luleå (Schweden) wurde in Kooperation mit dem *Interactive-Institute* bereits eine Software zur computergestützten Datenerhebung (einschließlich clusteranalytischer Auswertung) bis in das Stadium eines Prototyps entwickelt (Berg, 2005).

4.3.2 Ablauf des Verfahrens

4.3.2.1 Erhebung der Konstrukte durch Triadenvergleiche

Aus einer Menge von Objekten, welche Kelly auch als *Elemente* bezeichnet, werden Gruppen von jeweils drei Stück ausgewählt und sogenannten *Triaden* zugeordnet (vgl. Abs. 4.3.4). Nach der Präsentation einer Triade soll die Versuchsperson beschreiben, *auf welche bedeutende Weise sich zwei der drei Elemente ähneln und sich dadurch von dem dritten unterscheiden*. Anschließend wird die Versuchsperson gefragt, *inwiefern sich das dritte Element von den ersten beiden unterscheidet*³⁶. Auf diese Weise sollen verbale Deskriptoren für *Initial- und Kontrastpol* eines Konstrukts gefunden werden, dessen bipolare Natur von Kelly angenommen wird, jedoch nicht zwangsläufig erreicht werden muss (vgl. Abs. 4.3.5). Diese Prozedur wird für mehrere, in der Regel 10–20 verschiedene Triaden von Elementen wiederholt.

³⁵Auf Verfahren der statistischen Auswertung soll im Rahmen dieser Arbeit nicht näher eingegangen werden.

³⁶Die kursiv gesetzten Passagen sind an dieser Stelle bewusst so gewählt, dass sie eine möglichst direkte Übersetzung des englischen Originaltexts (Kelly, 1955, S. 222) darstellen.

Für die Aufnahme der Versuchspersonenurteile unterscheidet Kelly grundsätzlich ein individuelles, durch einen Interviewer geführtes, von einem vereinheitlichten, weitgehend automatisierten Vorgehen. Im ersten Fall werden die Antworten der Versuchsperson in einem Gespräch mit dem Versuchsleiter gewonnen und durch diesen notiert. Im letzteren trägt die Versuchsperson die Pole eines Konstrukts auf einem Fragebogen in zwei leere Felder ein, welche sich jeweils in einer Zeile zusammen mit einer Kodierung der betreffenden Triade befinden. Letztere Variante wird von Kelly wegen der Möglichkeit einer gleichzeitigen Anwendung auf mehrere Versuchspersonen, als *Gruppenform* bezeichnet (Kelly, 1955, S. 227).

Ein durch den Versuchsleiter geführtes Vorgehen bietet die Möglichkeit, den Erhebungsvorgang direkt zu steuern und somit, falls nötig, korrigierend einwirken zu können. Bezüglich der Klassifikation unzulässiger Konstrukte wird bei Kelly (1955, S. 222f) die folgende Einteilung vorgenommen³⁷: (1) *situationsbezogene*, (2) *übermäßig leicht zu durchdringende*, (3) *übermäßig schwer zu durchdringende*, (4) *oberflächliche*, (5) *vage* Konstrukte und (6) Konstrukte, die ein *direktes Produkt der Rollenbezeichnung* sind³⁸. Als zu ergreifende Maßnahmen werden von Kelly Formulierungen empfohlen, in welchen die Versuchsperson zur Reflexion der gegebenen Antworten und dem wiederholten Betrachten der dargebotenen Triade angehalten wird. Eine vergleichbare Möglichkeit der Korrektur für die Gruppenform wird nicht diskutiert. Im Zuge eines automatisierten Vorgehens kann diese jedoch nur a posteriori erfolgen. Die computergestützte Durchführung des Verfahrens ist grundsätzlich für beide der beschriebenen Varianten denkbar. Notizblock des Interviewers bzw. Fragebogen werden durch eine grafische Benutzeroberfläche ersetzt. In der zu entwickelnden Software soll daher die Möglichkeit einer Bearbeitung der Konstruktpole vorgesehen werden, bevor diese im nächsten Verfahrensschritt (s. folgender Abs.) für die Beurteilung der Elemente herangezogen werden.

³⁷Die Begriffe wurden vom Autor in Bemühen um eine möglichst originalgetreue Wiedergabe ihrer Bedeutungen übersetzt. Dennoch sollen zur Vermeidung von Fehlinterpretationen an dieser Stelle die englischen Formulierungen genannt werden: (1) *situational constructs*, (2) *excessively permeable c.*, (3) *excessively impermeable c.*, (4) *superficial c.*, (5) *vague c.*, (6) *c. which are a direct product of the role title*.

³⁸Diese Kategorie bezieht sich ausschließlich auf den ursprünglichen Anwendungsbereich der RGT, welcher die Verwendung von *Rollen* vorsieht (vgl. Abs. 4.3.3).

4.3.2.2 Beurteilung der Elemente auf den erhobenen Konstrukten

In einem zweiten Schritt soll die Versuchsperson die verwendeten Elemente³⁹ mit Hilfe der zuvor erhobenen Konstrukte bewerten. Hierzu kommt eines der drei folgenden Urteilsverfahren zum Einsatz (Bech und Zacharov, 2006, S. 55): Die Zuordnung des Elements zu einer von zwei Kategorien, welche den beiden Konstruktpolen entsprechen (*Dichotomisierung*), die Bildung von Rangordnungen (*Ranking*) oder die Bewertung auf Intervallskalenniveau (*Rating*). Im Kontext der zu erstellenden Hörversuchssoftware soll ausschließlich letztere verwendet werden.

Eine von Kelly traditionell verwendete Form der Datenaufbereitung ist die Erstellung eines sogenannten *Grid*. Hierfür wird eine zweidimensionale Gitternetzstruktur (engl. *grid*) erstellt, deren eine Dimension den erhobenen Konstrukten und die andere den bewerteten Elementen entspricht. Die Einträge in den so definierten Zellen bestehen jeweils aus einem numerischen Wert für das Urteil der Versuchsperson über das zugehörige Element auf dem jeweiligen Konstrukt. Abbildung 4.9 zeigt exemplarisch das Grid einer Versuchsperson als Ergebnis eines Hörversuchs.

4.3.3 Auswahl der Elemente

Ursprüngliche Einsatzgebiete der RGT sind im Bereich der Persönlichkeitspsychologie und klinischen Diagnostik angesiedelt. Zu untersuchende Konstruktsysteme beziehen sich dabei bevorzugt auf Personen und ihre Einstellungen im Kontext sozialer Beziehungen. Als Elemente werden daher von Kelly (1955, S. 220ff) sogenannte *Rollen* verwendet, welchen konkrete Namen aus dem Bekanntenkreis einer Person zugeordnet werden.

In der Audioevaluation steht das Konstruktsystem einer Person in Bezug auf seine auditorische oder auch audiovisuelle Erlebenswelt im Mittelpunkt des Interesses. Deshalb werden hierbei akustische, ggf. in Kombination mit visuellen Reizen, verwendet. In ihrer Untersuchung zur Identifikation von Kriterien der auditiven Wahrnehmung von Räumlichkeit, verwenden Berg und Rumsey (1999) als Stimuli Sprachsignale und Musikstücke unterschiedlicher Genres, welche sich innerhalb der Triaden systematisch hinsichtlich der verwendeten Aufnahme- und Signalverarbeitungsverfahren unterscheiden.

³⁹ Aus Gründen der begrenzten Versuchszeit beschränkt man sich in der Praxis u.U. auf eine Auswahl von Elementen (vgl. Berg und Rumsey, 1999).

<i>Inside head</i>	4	3	3	5	5	5	3	5	1	<i>In front of head</i>
<i>Live</i>	1	2	3	1	5	2	2	5	4	<i>Metallic</i>
<i>Sounds from one point</i>	5	5	5	5	2	2	5	4	5	<i>Sounds bigger</i>
<i>Natural</i>	1	3	4	1	5	2	2	4	5	<i>Artificial</i>
<i>Unpleasant</i>	5	3	3	5	2	3	5	3	1	<i>Pleasant</i>
<i>Like listening at home to the hi-fi system</i>	4	2	2	4	2	2	4	2	3	<i>Like listening in the room/live</i>
<i>Enhanced mid frequencies</i>	5	5	5	5	5	5	5	5	5	<i>Wide frequency response</i>
<i>Larger</i>	1	1	1	1	4	2	1	3	2	<i>Smaller</i>
<i>Realistic</i>	1	3	5	1	4	2	1	3	5	<i>Less realistic</i>
										P6 STR Pop
										P1 5CH Speech (2nd)
										P3 5CH Outdoor environment
										P2 5CH Saxophone
										P1 5CH Speech (1st)
										P4 5CH Symph orch (2nd)
										P6 4CH Pop
										P5 5CH Big band
										P4 5CH Symph orch (1st)

Abbildung 4.9: Grid einer Versuchsperson als Ergebnis eines Hörversuchs im Rahmen einer von Berg und Rumsey durchgeführten Untersuchung zur Identifikation von Kriterien zur auditiven Wahrnehmung und Beurteilung von Räumlichkeit. Es wurden 9 Elemente auf 9 erhobenen Konstrukten unter Verwendung fünfstufiger Ratingskalen bewertet. Die Elemente stellen Sprachausschnitte- oder Musikstücke unterschiedlicher Genres (*Programme* P1–P6) dar und liegen jeweils in drei verschiedenen Varianten bezüglich der verwendeten Aufnahme- und Signalverarbeitungstechniken vor, von denen nicht alle für die Bewertung herangezogen wurden (STR: Stereo, 4CH/5CH: 4-/5-Kanalverfahren, für weitere Informationen s. Publikation). [Quelle: Berg und Rumsey (1999, fig. 16a)]

Die Beschreibung der RGT nach Kelly geht von einer Anzahl von 15–25 Urteilsobjekten (Elementen) aus. Als Voraussetzung für eine nachfolgende faktoren- oder clusteranalytische Auswertung empfiehlt Danielsson (1991, zit. nach Berg und Rumsey, 1999) eine minimale Anzahl von 6–7.

4.3.4 Auswahl und Darbietung der Triaden

Für die Auswahl und Zusammenstellung der Triaden aus den Elementen existieren nach Kelly eine Reihe von Möglichkeiten, für deren Diskussion an dieser Stelle allerdings auf die entsprechende Literatur⁴⁰ verwiesen wird, da deren Unterschiedlichkeiten vor allem durch die Verwendung von Rollen als Elemente begründet sind. Für die Implementierung des Verfahrens im Rahmen der Magisterarbeiten ist die Möglichkeit einer zufälligen Zusammenstellung in Anlehnung an Kellys *Minimum-Context-Card-Form* vorgesehen, welche eine in der psychodiagnostischen Praxis gebräuchliche Variante darstellt (Scheer, 1992). Außerdem soll die Möglichkeit bestehen, bestimmte ausgewählte Konstellationen von Elementen vorzugeben, wie auch geschehen in der Untersuchung von Berg und Rumsey (1999).

Bei Anwendung des Verfahrens in der Audioevaluation ist das Augenmerk besonders auf die Präsentation der Elemente einer Triade zu richten. Für länger andauernde Stimuli (z.B. Ausschnitte von Musikstücken) ist hierbei die Begrenztheit der Gedächtnisleistung der Versuchsperson zu berücksichtigen, welche einen Vergleich der Stimuli erschweren kann. Zur Kompensation dieses Einflusses können innerhalb einer Triade ein oder mehrere Stimuli mehrmals dargeboten werden. Beispielsweise wiederholen Berg und Rumsey den ersten Stimulus nach Darbietung des dritten, so dass von den drei verschiedenen Stimuli immer zwei paarweise verglichen werden können. Alternativ hierzu ist ein interaktives Vorgehen denkbar, bei welchem die Versuchsperson selbst die Wiedergabe der akustischen Reize beliebig oft und in beliebiger Reihenfolge starten und jeweils nach einer beliebigen Zeit wieder beenden kann. Für die zu erstellende Software wurde letztere Variante gewählt.

4.3.5 Kritik, Varianten und Modifikationen

Die auf die beschriebene Weise erhobenen Konstrukte weisen nach Kelly eine bipolare Struktur auf (Kelly, 1955, S. 106). Diese Annahme wird in der Literatur jedoch in Frage gestellt (s. z.B. Bech und Zacharov, 2006, S. 56). Es wird außerdem angegeben,

⁴⁰Siehe hierfür Kelly (1955, S. 222ff) und Fransella und Bannister (1977, S. 14ff).

dass durch eine explizite Frage nach dem Gegenteil des Initialpols ein eindeutigerer Kontrast erzielt werden kann (Epting et al., 1971, zit. nach Fransella und Bannister, 1977, S. 15). Darüber hinaus gibt es neben der Verwendung von Triaden grundsätzlich andere Möglichkeiten für die Gruppierung der Urteilsobjekte, z.B. den Einsatz von Dyaden oder Vierergruppen mit gepaarten Stimuli. Eine Darstellung dieser und weiterer Varianten findet man bei Fransella und Bannister (1977, S.16–19).

Ist aufgrund der Fragestellung eine spezielle inhaltliche Fokussierung gefordert, so kann die Anzahl aller möglichen durch zusätzliche Instruktion der Versuchsperson auf eine geringere Anzahl relevanter Konstrukte eingeschränkt werden. So geben Berg und Rumsey (1999) in der genannten Untersuchung einer Auswahl an Versuchspersonen die zusätzliche Anweisung, nur auf die „dreidimensionale Natur“ der Stimuli zu achten. Sollen innerhalb eines gegebenen inhaltlichen Kontexts möglichst viele Konstrukte ermittelt werden, kann der Erhebungsvorgang für dieselbe Triade mehrmals durchgeführt werden, wie ebenfalls dort geschehen.

5 Entwicklung eines Programmkonzepts

5.1 Begriffsbestimmung

Die Entwicklung eines Programmkonzepts beinhaltet im gegebenen Kontext vor allem den Entwurf einer anwenderbezogenen Programmstruktur. Aspekte der internen Quellcode- und Datenstruktur werden dabei ausgeklammert, da deren Berücksichtigung Teil der Arbeit von André Wlodarski ist. Das in dieser Arbeit zu entwickelnde Konzept soll skizzieren, welche Funktionen dem Benutzer zu Verfügung stehen und wie die Interaktion zwischen Benutzer und Software strukturiert wird, d.h. welche Funktionen dem Benutzer auf welche Weise, an welcher Stelle der Benutzeroberfläche und zu welchem Zeitpunkt zugänglich sind. Es ist das Ziel, den Vorgang der Interaktion nach plausiblen, dem Nutzungskontext entlehnten Kriterien zu gliedern. Der Vorgang der Entwicklung des Programmkonzepts war zu keinem Zeitpunkt während der Entstehung der Software völlig abgeschlossen. Vielmehr wurden nach der Erkundung des Nutzungskontexts und der Herausarbeitung der Anforderungen (s. Abs. 5.2) durch den Autor einige grundlegende Merkmale festgelegt, welche im weiteren Verlauf der Entwicklungsarbeiten zunehmend präzisiert werden konnten (s. Abs. 5.3).

5.2 Nutzungskontext und Anforderungen

Die primäre Aufgabe der zu entwickelnden Software besteht in der Automatisierung bestimmter Prozesse, welche Bestandteile eines Testverfahrens (s. Abs. 3.3 und 4) sind. Hierfür sind eine Reihe verschiedener Funktionen notwendig. Dazu zählt die Steuerung des Verfahrensablaufs. Sämtliche Ereignisse wie die Präsentation der Stimuli und die Aufnahme der Versuchspersonenantworten werden in ihrer zeitlichen Abfolge koordiniert. Für bestimmte Testverfahren sind dabei während des Versuchs Entscheidungen über die Auswahl der Stimuli, z.B. unter Berücksichtigung vorheriger Versuchspersonenantworten, erforderlich. Dies gilt z.B. für die adaptiven psychophysischen Verfahren (s. Abs. 4.1).

Die Software muss über geeignete Schnittstellen sowohl mit dem Benutzer als auch mit der Wiedergabeperipherie kommunizieren können. Während der Durchführung eines Hörversuchs ist der Benutzer in der Regel⁴¹ durch die Versuchsperson gegeben. Die Benutzerschnittstelle soll eine grafische Benutzeroberfläche enthalten, über welche die Ausgabe von Informationen und Instruktionen und die Eingabe von Daten erfolgen. Für die Bedienung bzw. Anzeige soll die Standardperipherie eines Personal-Computers, d.h. Maus, Tastatur und Monitor, verwendet werden. Die Präsentation der Stimuli soll auf verschiedene Weisen möglich sein. Entweder soll eine Audiodatei direkt über eine Soundkarte wiedergegeben oder externe Ton- und Bildquellen durch OSC⁴²-Befehle angesteuert werden können. Auf letztere Weise soll auch die Anbindung an Umgebungen für binaurale Echtzeitfaltung und Wellenfeldsynthese ermöglicht werden.

Nach der Aufnahme der Versuchspersonenantworten kann für einige Testverfahren eine Verarbeitung dieser notwendig sein um die gewünschten quantitativen oder qualitativen Daten zu erheben. So ist für die Ermittlung der Schwellwertschätzung bei einigen adaptiven psychophysischen Verfahren (s. Abs. 4.1) eine statistische Auswertung erforderlich. Im Fall der RGT ist die Möglichkeit für eine manuelle Bearbeitung der Konstrukte durch den Versuchsleiter vorzusehen (s. Abs. 4.3.2.1). Derartige Funktionen oder Editiermöglichkeiten sollen ebenfalls Bestandteil der Software sein. Die so erhaltenen Daten sollen gespeichert und im Anschluss an die Versuchsdurchführung nach der Statistiksoftware *SPSS* exportiert werden können.

Damit die Durchführung eines Hörversuchs möglich ist, müssen vorher Versuchsbedingungen definiert und Einstellungen bezüglich des verwendeten Testverfahrens vorgenommen werden können. Hierfür soll dem Versuchsleiter eine grafische Benutzeroberfläche zu Verfügung stehen.

Sämtliche beschriebenen Funktionen sollen sich zunächst auf die in Kapitel 4 beschriebenen Testverfahren erstrecken. Die Software soll jedoch in ihrer Struktur um zusätzliche Testverfahren erweiterungsfähig sein.

⁴¹Eine Abweichung hiervon stellt z.B. die Durchführung der RGT mittels Interviewer (s. Abs. 4.3.2.1) dar.

⁴²Abkürzung für *Open-Sound-Control*; OSC ist ein Kommunikationsprotokoll welches den Informationsaustausch zwischen Computern und multimedialer Peripherie in Echtzeit über Netzwerk ermöglicht.

5.3 Darstellung des entwickelten Programmkonzepts

Das entwickelte Programmkonzept soll nun dargestellt werden. Abbildung 5.1 auf Seite 78 soll den Inhalt des folgenden Textes schemenhaft wiedergeben und dessen Verständnis durch eine grafische Veranschaulichung unterstützen. Die verschiedenen Bausteine des Konzepts werden in der Abbildung durch Schlagworte bezeichnet, welche im Text an den Stellen mit inhaltlichem Bezug in Klammern genannt werden.

Die Bedienung der Software erfolgt grundsätzlich auf zwei verschiedenen Ebenen, welche einerseits aus einem System von grafischen Benutzeroberflächen und andererseits aus einem direkten Datenzugriff bestehen. Der Großteil aller Funktionen ist über das GUI⁴³-System erreichbar, welches sich wiederum anhand der zwei grundlegend verschiedenen Benutzergruppen untergliedern lässt.

Die erste Kategorie von Anwendern wird durch den Versuchsleiter bzw. den Experimentator einer Untersuchung gebildet. Der zugehörige Teil des GUI-Systems ist über das *Hauptfenster* des Programms zu erreichen, welches bereits Bestandteil von diesem ist. Innerhalb dieses Systemabschnittes können drei grundsätzlich verschiedene Aufgabenfelder bearbeitet werden. Zum einen können Einstellungen vorgenommen werden, welche weniger mit den methodischen Aspekten einer Untersuchung in Zusammenhang stehen, als vielmehr die Konfiguration des Programmes im Hinblick auf technische Parameter betreffen und somit auf dessen grundsätzliches Verhalten, z.B. im Hinblick auf die Formatierung von Ausgabedaten oder die Interaktion mit externer Peripherie, Einfluss nehmen (vgl. *Programm-Setup*). Der zweite Aufgabenbereich (vgl. *Versuchsreihe-Setup*) erstreckt sich über sämtliche Einstellungen welche im Zuge der Einrichtung einer Versuchsreihe, d.h. der Gesamtheit der einer bestimmten Untersuchung zugeordneten Versuchsdurchführungen, vorzunehmen sind. Hierzu zählt die Verwaltung und Konfiguration der Abschnitte, welche innerhalb eines Hörversuchs durchgeführt werden sollen. Jedem Versuchsabschnitt ist dabei ein bestimmtes Testverfahren zuzuordnen. Für das jeweilige Testverfahren sind eine Reihe von Einstellungen vorzunehmen, wofür jeweils verfahrensspezifische GUIs zu Verfügung stehen. Da die vorgeschalteten Benutzeroberflächen von diesen weitestgehend unabhängig sind, kann das System an dieser Stelle zu einem späteren Zeitpunkt um zusätzliche Testverfahren erweitert werden. Als weitere Einstellung im Kontext einer Versuchsreihe können mehrere sogenannte Sitzungen definiert werden (vgl. *Session-Setup*), welche

⁴³ GUI ist die Abkürzung für *Graphical-User-Interface* welche im Allgemeinen gleichbedeutend mit dem deutschen Begriff der *grafischen Benutzeroberfläche* verwendet wird.

jeweils Teilmengen von Versuchsdurchläufen repräsentieren, die unter in gleicher Weise definierten Versuchsbedingungen durchgeführt werden sollen. Die Zuordnung der einzelnen Versuchsbedingungen zu den verschiedenen Sitzungen erfolgt auf der Ebene der zuvor beschriebenen GUIs für die Konfiguration der Testverfahren, da die hierbei existierenden Wahlfreiheiten von letzteren abhängig sind. Unterschiedliche Versuchsbedingungen können zum gegenwärtigen Zeitpunkt dadurch realisiert werden, dass für die Messeinheiten eines Testverfahrens unterschiedliche Stimuli verwendet, deren Position in der Darbietungsreihenfolge variiert oder sonstige Merkmale der Ablaufsteuerung und Datenverarbeitung (z.B. Adaptions- und Schätzalgorithmus) verändert werden. Die dritte Gruppe der im Zuge der Einrichtung einer Versuchsreihe vorzunehmenden Einstellungen ist durch die Anlage der Stimuli (vgl. *Stimulus-Setup*) gegeben. Da diese auf einer globalen, d.h. den Testverfahren übergeordneten Ebene erfolgt und jedem Stimulus eine innerhalb der gesamten Versuchsreihe eindeutige ID zugeordnet ist, kann jeder Stimulus in jedem Versuchsabschnitt bzw. für jedes Testverfahren verwendet werden, wodurch unnötiger Aufwand durch mehrmaliges Anlegen vermieden wird. Das dritte Aufgabenfeld, welches durch den Experimentator über das GUI-System zu bearbeiten ist, betrifft das *Setup* und die *Kontrolle* der *Versuchsdurchführungen* selbst. Unmittelbar vor einem Hörversuch müssen einzelne Versuchsparameter spezifiziert werden. Dies sind die Eingabe einer Versuchspersonen-ID und eines Schlüssels für die Kodierung der auszuwählenden Sitzung. Weitere Kontrollaufgaben wie das Starten des Versuchs, das Löschen einzelner erhobener Daten und gegebenenfalls die Ausführung einzelner Versuchsabschnitte können von diesem Teil des GUI-Systems ausgeführt werden.

Nach dem Start eines Hörversuchs durch den Experimentator wird der zweite Teil des GUI-Systems geöffnet, welcher von der Versuchsperson zu bedienen ist. Die Ausgangsbasis hierfür stellt ein Fenster (vgl. *Basisschirm*) dar, welches den gesamten Bildschirm überdeckt und während der Versuchsdurchführung einen neutralen Hintergrund bietet. Vor und nach der Durchführung eines Hörversuchteils werden hierauf Informationen über den aktuellen Versuchsstatus und außerdem Instruktionen abgebildet. Von diesem Fenster aus gelangt die Versuchsperson durch Betätigung entsprechender Kontrollelemente auf das verfahrensspezifische GUI, welches der Aufnahme der Versuchspersonenantworten dient und dessen Gestalt vom jeweiligen Testverfahren abhängt. Auch hier ist die Programmstruktur offen für Erweiterungen, da die Gestaltung des übergeordneten *Basisschirms* vom jeweiligen Testverfahren unabhängig ist.

Neben der Verwendung der grafischen Benutzeroberflächen kann – wie bereits ein-

gangs gesagt wurde – die Bedienung der Software durch den Experimentator auch über einen direkten Zugriff auf Daten erfolgen. In einigen Fällen (in der Abb. als *optional* gekennzeichnet) stellt dies eine Alternative dar, durch welche u.U. das Arbeitstempo erhöht werden kann. So können die Datensätze von Programm- und Versuchsreiheneinstellungen bei Bedarf jeweils in ihrer Gesamtheit kopiert werden. Dies entspricht der in Software häufig vorhandenen Möglichkeit der Erstellung, Speicherung und späteren Verwendung von benutzerdefinierten Formaten. Für die übrigen Daten ist jedoch der direkte Zugriff unbedingt erforderlich. Dies wurde konzeptionell so vorgesehen, da – im Gegensatz zu einer Einbindung dieser in eine grafische Benutzeroberfläche, soweit dies im Einzelnen überhaupt möglich ist – eine größere Flexibilität erreicht werden kann, wenn die einzelnen Daten in ihrem spezifischen Dateiformat gespeichert werden können. Die zu einer Versuchsreihe gehörenden Daten werden in einem gemeinsamen Versuchsreihenordner abgelegt und bestehen aus den verwendeten Audio-Files, Daten zur Entschlüsselung anderer Daten⁴⁴ (vgl. *Datenschlüssel*), den Log-Files, den für den Import in eine Statistik-Software in geeigneter Weise formatierten empirischen Daten (vgl. *Daten-Export-Sheets*) und – für bestimmte Testverfahren – grafisch aufbereiteten Daten (vgl. *Plots*).

⁴⁴Als Beispiel sei hier die Entschlüsselung der einer bestimmten Sitzungskennziffer zugeordneten Versuchsbedingungen genannt.

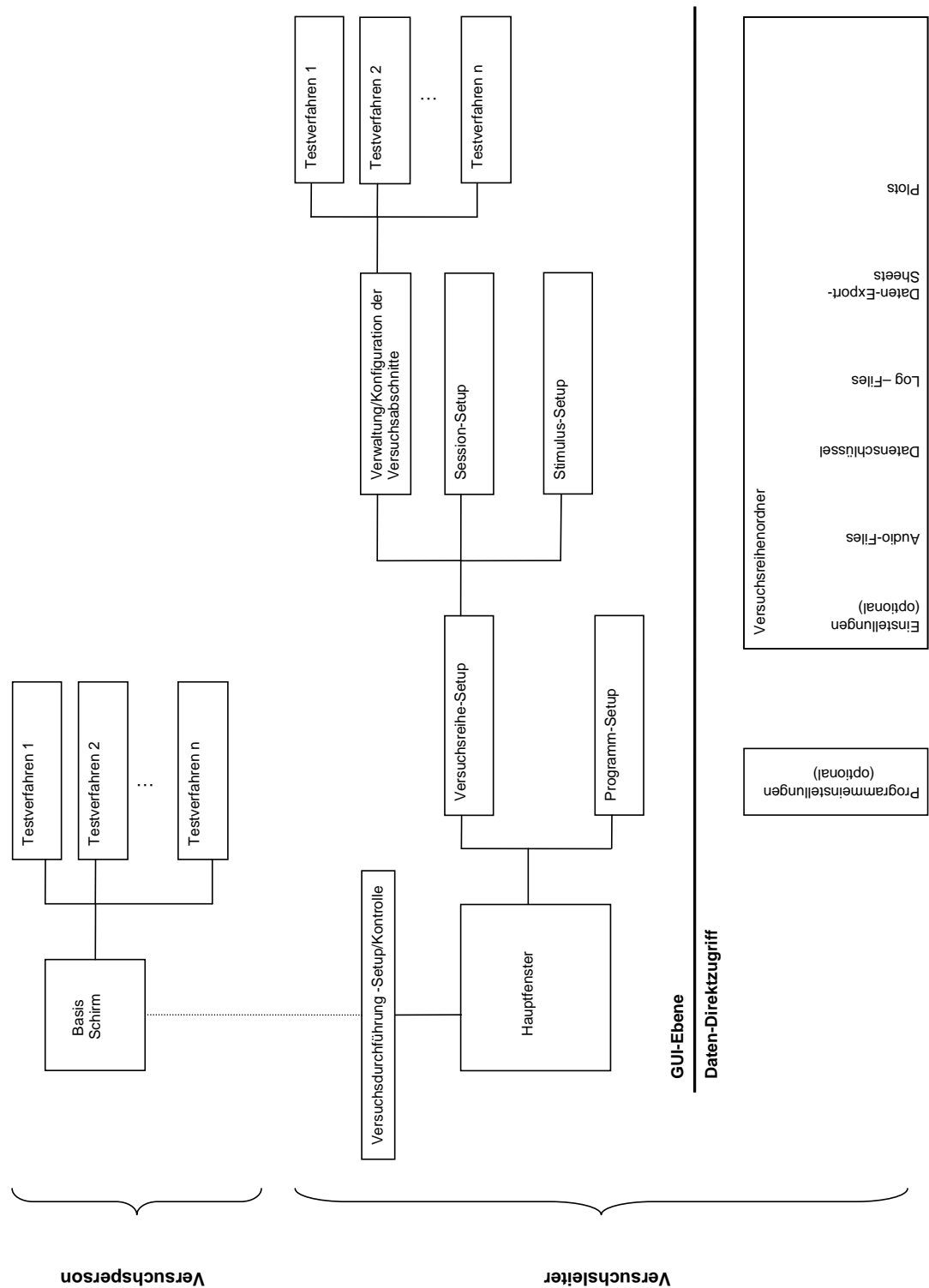


Abbildung 5.1: Anwenderseitige Programmstruktur der Softwarebibliothek

6 Implementierung der Adaptionsmechanismen und Schätzmethoden

6.1 Die MATLAB-Funktionen im Überblick

Die Adaptionsmechanismen und Schätzmethoden der in Abschnitt 4.1 dargestellten adaptiven psychophysischen Verfahren werden unter MATLAB in die folgenden Funktionen umgesetzt, deren Quellcode in Anhang A.1 zu finden ist: *staircase1d1u.m*, *staircase2d1u.m*, *staircase3d1u.m*, *PEST.m*, *mvPEST.m*, *MLBayes.m*. Die ersten drei sind Implementierungen des Staircase-Verfahrens mit den Adaptionsregeln 1-Down/1-Up, 2-Down/1-Up und 3-Down/1-Up. *PEST.m* und *mvPEST.m* basieren auf dem PEST-Verfahren in der originalen und der beschriebenen modifizierten Fassung nach Findlay (1978). Die Funktion *MLBayes.m* basiert auf der Implementierung des Best-PEST-Verfahrens nach Lieberman und Pentland (1982) unter Einbeziehung von optionalen Erweiterungen, welche mehreren ähnlichen adaptiven Verfahren entnommen wurden, die entweder auf einer Parameterschätzung nach der Maximum-Likelihood-Methode oder der Bayesschen Statistik beruhen.⁴⁵ So kann für die Initialisierung des Verfahrens anstelle der von den Autoren des Best-PEST vorgenommenen Definition impliziter Trials, eine gaußförmige oder gleichförmige A-priori-Likelihood vorgegeben werden (vgl. Abs.4.1.7.3). A-priori-Information kann auf Wunsch von der abschließenden Schätzung der Schwelle ausgeschlossen werden. Für die Parameterschätzung können entweder der Modalwert oder das arithmetische Mittel herangezogen werden. Letzteres wird in den Verfahren von Emerson (1986), King-Smith et al. (1994) und Treutwein (1997) verwendet. Für die Terminierung kann alternativ zur Vorgabe einer festen Anzahl an Trials ein dynamisches Kriterium gewählt werden (vgl. Abs. 4.1.7.2). Als aktuelles Konfidenzintervall wird dabei ein sogenanntes *symmetrisches zweiseitiges bayessches Wahrscheinlichkeitsintervall* nach dem Vorbild der Implementierung⁴⁶ des YAAP-Verfahrens von Treutwein (1997) berechnet.

⁴⁵Diese beiden Gruppen adaptiver Verfahren unterscheiden sich vor allem im Hinblick auf Theorie und Terminologie (vgl. Treutwein, 1995, Textbox 4). Da der bayessche Ansatz nicht auf ein spezielles Maß der zentralen Tendenz festgelegt ist und konzeptionell A-priori-Information einschließt, kann die Maximum-Likelihood-Methode formal als Sonderfall dessen betrachtet werden.

⁴⁶Es lag der Modula-2-Code in der Version 3.0 vom 1.12.1993 vor.

Die implementierten Funktionen ermitteln nach jedem Aufruf den Wert der Reizstärke für die darauffolgende Darbietung und geben nach Terminierung des Verfahrens die abschließende Schätzung der Schwelle zurück. Die Reizstärken bisheriger Darbietungen und die zugehörigen Antworten werden in Form eines Vektors übergeben. Außerdem benötigen die Funktionen die Eingabe von verfahrensspezifischen Parameterwerten, von Informationen bezüglich der Terminierung, des verwendeten Antwort-Paradigmas und des Reizstärkenbereichs (Range). Eine Auflistung der Eingabe- und Ausgabeparameter im Einzelnen findet sich in Abschnitt 6.3. Die Einbindung der Funktionen in die Ablaufsteuerung und Benutzeroberfläche der zu erstellenden Softwarebibliothek erfolgt im Rahmen der Masterarbeit von André Włodarski (Włodarski, 2008).

6.2 Aspekte der praktischen Umsetzung

Die relevanten Publikationen der Verfahren lassen an einigen Stellen Gestaltungsspielräume offen, da dort deren Beschreibung meist von einem konzeptionellen Standpunkt aus erfolgt und Aspekte der praktischen Umsetzung nicht immer vollständig behandelt werden. Zudem bedürfen manche dieser Aspekte grundsätzlich keiner eindeutigen Festlegung, sondern können durch den Entwickler unter Berücksichtigung von Erfordernissen des Nutzungskontexts entschieden werden. Selbst wenn derartige Details der Implementierung durch die Autoren der Verfahren bereits festgelegt worden sind, kann eine Veränderung dieser u.U. dennoch sinnvoll sein, vorausgesetzt dass hierdurch das Verfahren in seiner Funktionsweise unberührt bleibt. Im Folgenden soll dargelegt werden, wie derartige Freiheiten für die vorliegenden Implementierungen genutzt wurden.

Diskretisierung der Reizstärke Sämtliche in Abschnitt 4.1 beschriebenen adaptiven Verfahren verwenden eine über den gesamten Reizstärkenbereich konstante Auflösung, welche durch die minimale Schrittweite⁴⁷ gegeben ist. Durch die Wahl einer geeigneten Einheit der Reizstärke ist im Fall parametrischer Verfahren sicherzustellen, dass die Form der psychometrischen Funktion möglichst gut durch das zugrundegelegte Modell beschrieben wird (vgl. Abs. 4.1.4). Außerdem wird für die nichtparametrischen Verfahren an verschiedenen Stellen gefordert, dass gleiche Differenzen von Reiz-

⁴⁷Mit minimaler Schrittweite ist an dieser Stelle die kleinste Schrittweite gemeint, welche noch für Platzierungen von Reizstärkenwerten herangezogen wird. Diese soll an dieser Stelle abgegrenzt werden von anderen noch kleineren Schrittweiten, welche nach Beendigung eines Verfahrens aufgerufen werden (vgl. *stopping step size* bei PEST, Anm. 18 auf S. 47)

stärkenwerten annähernd gleichen Differenzen der zugehörigen Wahrnehmungsgröße entsprechen sollten (vgl. etwa Cornsweet, 1962). Derartige Annahmen sind jedoch im Vorfeld durch den Experimentator zu überprüfen und können in vielen Fällen durch eine geeignete Transformation der Reizstärkenwerte, wie z.B. durch Logarithmierung, erfüllt werden. Für die Implementierung der Verfahren ergibt sich daher lediglich die genannte Notwendigkeit der gleichmäßigen Aufteilung des in einer geeigneten Einheit (bzw. Hilfseinheit) vorgegebenen Reizstärkenbereichs. In der Forschungspraxis am Fachgebiet hat sich bereits die Vorgehensweise gefestigt, den Reizstärkenbereich auf ganze Zahlen $\{1, 2, \dots, Range\}$ abzubilden, mit welchen innerhalb der Software gerechnet wird. Entsprechend sind auch die Implementierungen in dieser Arbeit ausgelegt.

Verhalten der nichtparametrischen Verfahren (Staircase/PEST) an den Rändern des Reizstärkenbereichs Bei den nichtparametrischen adaptiven Verfahren Staircase und PEST kann der Fall eintreten, dass das Regelwerk des Adaptionsmechanismus ein Überschreiten der Ränder des Reizstärkenbereichs verlangt. Der weitere Verlauf des Verfahrens in einem solchen Fall wird in den betreffenden Publikationen nicht diskutiert. Die Implementierungen in dieser Arbeit sind derart angefertigt, dass dann der Reizstärkenwert des Randes für die darauffolgende Darbietung gewählt wird. Außerdem wird im Fall der beiden PEST-Implementierungen bei jeder Reizplatzierung am Rande des Range, der Zähler für die Schritte in eine gegebene Richtung zurückgesetzt. Auf diese Weise werden nachfolgende Verdopplungen der Schrittweite „im Hintergrund“ (d.h. ohne dass sich die Reizstärke tatsächlich ändern kann) vermieden, welche sonst bei nochmaligen Entscheidungen des Signifikanztests für einen Reizstärkenwechsel über den Rand hinaus auftreten könnten. Es wird somit in einem solchen Fall verhindert, dass sich der Track beim nächsten Richtungswechsel unangemessen weit vom Rand wegbewegt.

Berechnung der Likelihood-Funktion (Maximum-Likelihood-/Bayes-Verfahren)

Für die Erstellung eines Adaptions- und Schätzalgorithmus nach dem Vorbild des Best-PEST-Verfahrens konnte auf von den Autoren veröffentlichten Quellcode (s. Lieberman und Pentland, 1982), angefertigt in APPLESOFT BASIC, zurückgegriffen und dieser teilweise als Vorlage verwendet werden. Die erstellte Implementierung weist jedoch neben den im vorherigen Abschnitt genannten Erweiterungen die folgenden zusätzlichen Unterschiede auf. Zum einen wird zugunsten einer konsistenten Datenstruktur der Softwarebibliothek die Likelihood-Funktion bei jedem Aufruf der MATLAB-Funktion von neuem berechnet. Die bisher dargebotenen Reizstärkenwerte

und die zugehörigen Antworten werden der Funktion als Argument übergeben. Auf eine Logarithmierung der Likelihood-Funktion wird verzichtet und statt dessen eine fortlaufende Normierung auf die Fläche unter dem Funktionsgrafen durchgeführt (vgl. Abs. 4.1.7.1). Um die Rechenzeit unter diesen Gegebenheiten möglichst gering zu halten, wurde die iterative Berechnung der Funktionswerte durch eine Vektoroperation ersetzt.

6.3 Beschreibung der Eingabe- und Ausgabeparameter

Im Folgenden werden die Eingabe- und Ausgabeparameter der erstellten MATLAB-Funktionen tabellarisch aufgeführt und in ihrer Bedeutung jeweils kurz beschrieben. Deren Reihenfolge in der Auflistung entspricht der Reihenfolge in welcher sie übergeben bzw. zurückgegeben werden.

6.3.1 staircaseXd1u.m (X=1, 2, 3)

Ausgabeparameter

x_next	Integer	Reizstärkenwert für die Darbietung im nächsten Trial, einer X-Down/1-Up-Regel folgend.
x_est	Double	Schätzung der Schwelle und gleichzeitig Indikator für Terminierung (x_est = 0: das Verfahren wurde noch nicht terminiert, d.h. es liegt noch keine Schätzung vor).

Eingabeparameter

x_hist	Integer-Array	Enthält die Reizstärkenwerte sämtlicher bisher dargebotener Stimuli.
r_hist	Integer-Array	Enthält die zugehörigen Antworten (+1: positiv, -1: negativ).
range	Integer	Legt den Bereich [1, range] möglicher Reizdarbietungen (Range) fest, welcher in Schritten von 1 aufgelöst wird.
nrmax	Integer	Maximale Anzahl an Umkehrungen (Reversals) des Tracks. Bei Erreichen dieser Zahl wird das Verfahren terminiert.
halvings	Integer-Array	Enthält die Indizes der Umkehrpunkte, bei denen Halbierungen der Schrittweite erfolgen sollen.
numlastr	Integer	Anzahl der letzten aufeinanderfolgenden Reversals, welche in die Schätzung der Schwelle eingehen sollen.

6.3.2 PEST.m und mvPEST.m

Ausgabeparameter

x_next	Integer/ Double	Reizstärkenwert für die Darbietung im nächsten Trial. (Nach dem letzten Funktionsaufruf ist dieser Wert gleich der Schätzung der Schwelle in x_est.)
x_est	Double	Schätzung der Schwelle und gleichzeitig Indikator für Terminierung (x_est=0: das Verfahren wurde noch nicht terminiert, d.h. es liegt noch keine Schätzung vor).

Eingabeparameter

x_hist	Integer- Array	Enthält die Reizstärkenwerte sämtlicher bisher dargebotener Stimuli.
r_hist	Integer- Array	Enthält die zugehörigen Antworten (+1: positiv, -1: negativ).
range	Integer	Legt den Bereich [1, range] möglicher Reizdarbietungen (Range) fest, welcher in Schritten von 1 aufgelöst wird.
dx_ini	Integer	Anfangswert der Schrittweite (Zweierpotenz).
dx_max	Integer	Maximale Schrittweite, welche während des Verfahrens zu keinem Zeitpunkt (auch nicht im ersten Schritt) überschritten wird (Zweierpotenz).
P_t	Double	Konvergenzniveau des Tracks (aus [0, 1]).
W	Double	Deviation-Limit für Signifikanztest. (<i>nur für PEST.m</i>)
M	Integer	Parameter M. (<i>nur für mvPEST.m</i>)

6.3.3 MLBayes.m

Ausgabeparameter

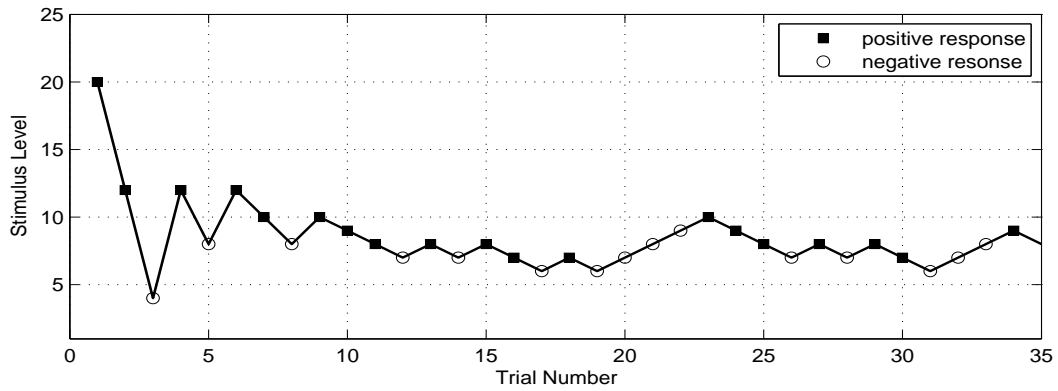
x_next	Integer	Reizstärkenwert für die Darbietung im nächsten Trial (beinhaltet A-priori-Information, wenn diese spezifiziert ist).
x_est	Integer/ Double	Abschließende Schätzung der Schwelle (hiervon kann auf Wunsch die A-priori-Information ausgeschlossen werden, s. Parameter <i>exclude</i>). $x_est = 0$: das Verfahren wurde noch nicht terminiert. Wenn <i>averagetype</i> ==2 gewählt ist, kann der Parameter auch reelle Werte aus [1, range] enthalten.
finalLike	Double- Array	Enthält die Likelihood-Funktion der abschließenden Schätzung.

Eingabeparameter

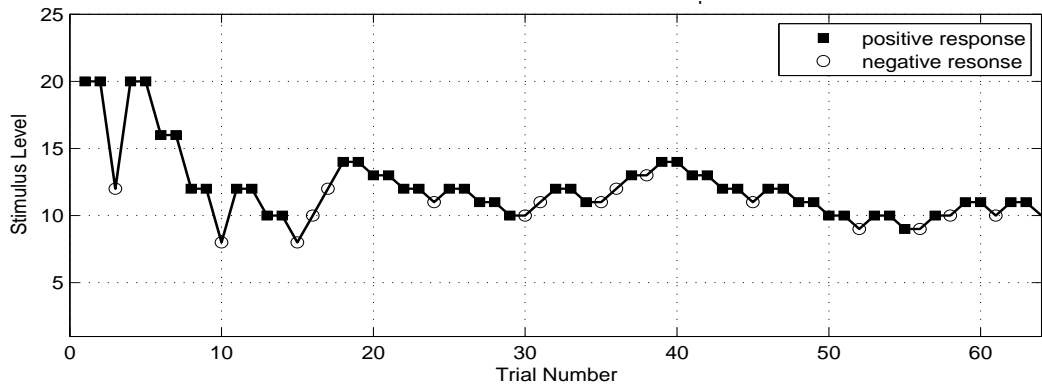
x_hist	Integer-Array	Enthält die Reizstärkenwerte sämtlicher bisher dargebotener Stimuli.
r_hist	Integer-Array	Enthält die zugehörigen Antworten (+1: positiv, -1: negativ).
range	Integer	Legt den Bereich [1, range] möglicher Reizdarbietungen (Range) fest, welcher in Schritten von 1 aufgelöst wird.
beta	Double	Spread-Parameter.
lambda	Double	Lapsing-Rate (aus [0, 1-gamma]).
n_or_gamma	Integer/ Double	Festlegung des Antwortparadigmas (Werte größer oder gleich 2: Anzahl n der Intervalle beim nAFC-Paradigma, Werte aus [0, 1]: False-Alarm-Rate γ für ja/nein-Paradigma).
priortype	Integer	Gibt an, auf welche Art A-priori-Information berücksichtigt werden soll (1: implizite Trials nach Lieberman und Pentland (1982), 2: gaussförmige A-priori-Likelihood, sonst: konstante A-priori-L. mit erster Reizdarbietung in der Mitte des Range).
mu	Double	Mittelwert der gaussförmigen A-priori-Likelihood (wird nur für <i>priortype</i> == 2 verwendet).
sigma	Double	Standardabweichung der gaussförmigen A-priori-Likelihood (wird nur für <i>priortype</i> == 2 verwendet).
averagetype	Integer	Gibt an, nach welcher Methode die Mittelwertbildung erfolgen soll (1: Modalwert, d.h. Maximum-Likelihood-Methode, 2: arithmetischer Mittelwert).
exclude	Integer	Gibt an, ob A-priori-Information von der Berechnung der abschließenden Schätzung ausgeschlossen werden soll (1: Ausschluss und sonstige: kein Ausschluss).
termtype	Integer	Gibt an, auf welche Weise das Verfahren terminiert werden soll (0: Terminierung per vorgegebenem Konfidenzintervall, ganzzahlige Werte größer als 1: Terminierung nach einer fest vorgegebenen Anzahl an Trials; letztere entspricht dann gleichzeitig dem Parameterwert.)
ConCoeff	Double	Konfidenzkoeffizient (Werte aus [0, 1], nur für <i>termtype</i> == 0 erforderlich).
CI_max	Integer	Obere Schranke für das Konfidenzintervall, bei deren Erreichen das Verfahren terminiert wird. Werte müssen positiv und sollten sinnvoller Weise ganzzahlig und kleiner als der Range sein (nur für <i>termtype</i> == 0 erforderlich).

6.4 Test durch Simulation

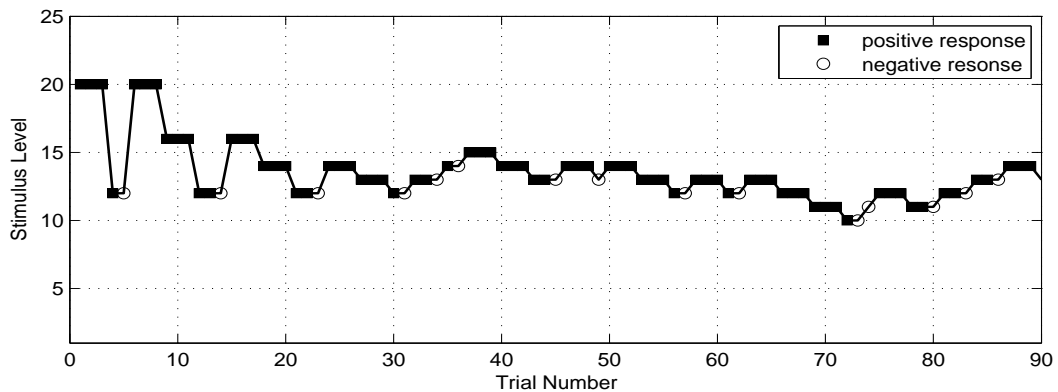
Die Funktionalität der MATLAB-Funktionen wird u.a. durch Simulation einzelner Verfahrensdurchläufe und grafische Ausgabe der zugehörigen Reizstärkenverläufe überprüft. Hierfür wurden die Skripte *simulation_staircase.m*, *simulation_PEST.m* und *simulation_MLBayes.m* angefertigt, welche im Anhang A.2 zu finden sind. Die Versuchsperson wird durch ein Beobachtermodell nachgebildet, welches im Rahmen der Magisterarbeit von Stefanie Otto (Otto, 2008) erstellt wurde. Die psychometrische Funktion des Beobachters wird dabei durch die logistische Funktion modelliert (vgl. Gl. (4.5) auf S. 38). Abbildungen 6.1–6.3 zeigen exemplarisch Reizstärkenverläufe, welche aus den Simulationen hervorgegangen sind. Dem Maximum-Likelihood-/Bayes-Verfahren wurde dabei das Best-PEST für die drei verschiedenen wählbaren Varianten der Initialisierung zugrundegelegt. Als globale, d.h. vom konkreten Verfahren unabhängige Einstellungen, wurden die Wahl des 3AFC-Paradigmas, eine „wahre“ Schwelle von $\vartheta = 10$ im Wendepunkt der psychometrischen Funktion, einen „wahren“ Spread-Parameter von $\beta = Range/10$ und eine „wahre“ Lapsing-Rate von $\lambda = 0$ verwendet. Für die nichtparametrischen Verfahren wurde ein Reizstärkenbereich $\{1, 2, \dots, Range\}$ mit $Range = 25$ festgelegt. Für die Simulationen der drei Varianten des Best-PEST-Verfahrens wurde mit $Range = 50$ jeweils ein doppelt so großer Bereich angenommen. Die Beschriftungen der Abbildungen enthalten die jeweiligen Schwellwertschätzungen und Beschreibungen der verfahrensspezifischen Einstellungen. Die erhaltenen Daten stellen jeweils das Ergebnis eines einzigen Simulationsdurchlaufs dar und sind daher nicht geeignet um Aussagen über die Güte eines Verfahrens zu treffen.



(a) 1-Down/1-Up-Verfahren. „Wahre“ Schwelle (umgerechnet⁴⁸ auf Konvergenzniveau $P_t = 0.5$) $\vartheta_{0.5} \approx 7.25$; Schätzung der Schwelle $\hat{\vartheta}_{0.5} = 7.5$.

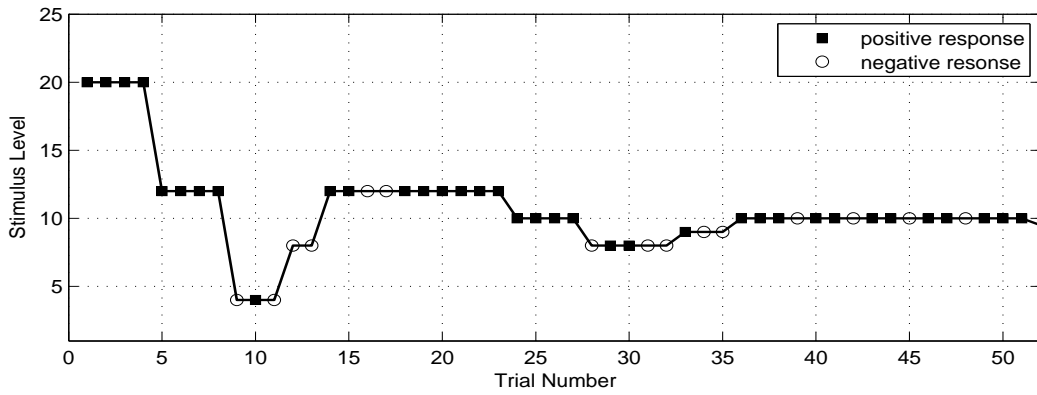


(b) 2-Down/1-Up-Verfahren. „Wahre“ Schwelle (umgerechnet⁴⁹ auf Konvergenzniveau $P_t = 0.707$) $\vartheta_{0.707} \approx 10.61$; Schätzung der Schwelle $\hat{\vartheta}_{0.707} = 10.25$.

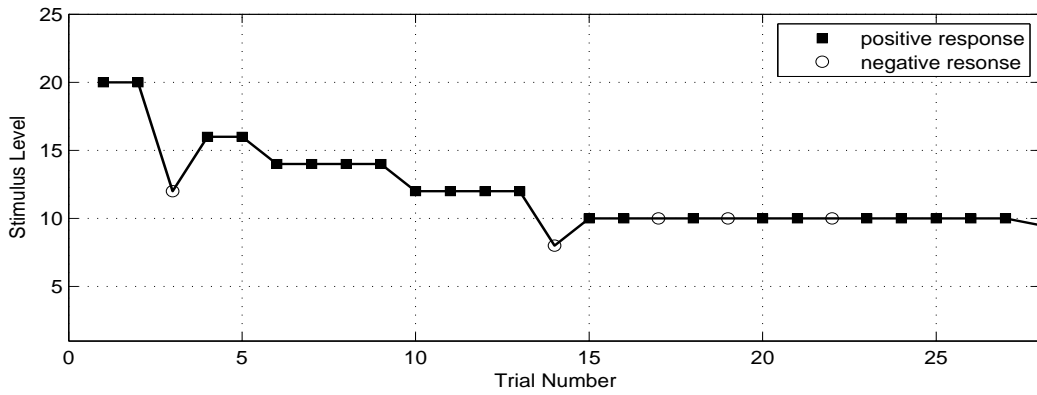


(c) 3-Down/1-Up-Verfahren. „Wahre“ Schwelle (umgerechnet⁵⁰ auf Konvergenzniveau $P_t = 0.794$): $\vartheta_{0.794} \approx 12.01$; Schätzung der Schwelle $\hat{\vartheta}_{0.794} = 11.75$.

Abbildung 6.1: Simulationen der Tracks nach dem Staircase-Verfahren für die drei implementierten Adaptionsregeln. Einstellungen: $Range = 25$, Startwert der Reizstärke $x_1 = 20$, Halbierungen der Schrittweite nach dem 2., 4. und 6. Reversal, Terminierung nach 20 Reversals, Mittelung über die letzten vier Reversals. Weitere, verfahrensunabhängige Parameter werden im Text auf Seite 85 aufgeführt.



(a) **Original PEST-Verfahren nach Taylor und Creelman (1967).** Wahl des Deviation-Limit: $W = 1$; „wahre“ Schwelle $\vartheta = 10$; Schätzung der Schwelle $\hat{\vartheta} = 9.5$.



(b) **More-Virulent-PEST-Verfahren nach Findlay (1978).** Wahl des Parameters: $M = 12$; „wahre“ Schwelle $\vartheta = 10$; Schätzung der Schwelle $\hat{\vartheta} = 9.5$.

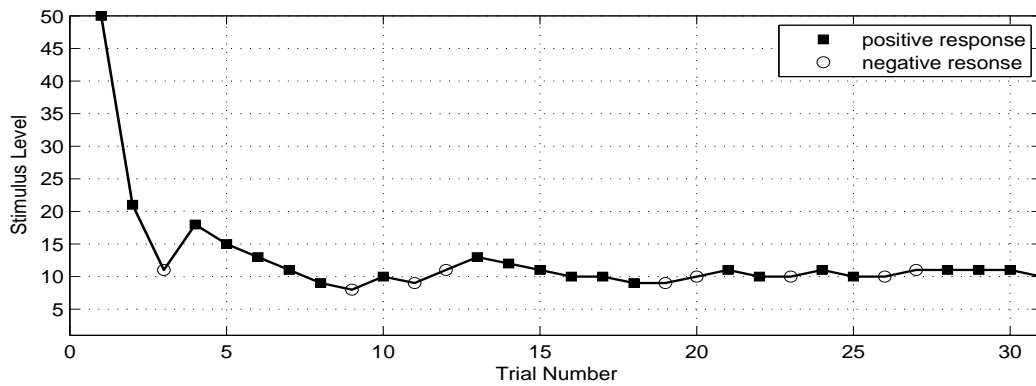
Abbildung 6.2: Simulationen der Tracks für die zwei implementierten PEST-Verfahren. Einstellungen: $Range = 25$, Startwert der Reizstärke $x_1 = 20$, Startwert der Schrittweite $\Delta x_1 = 8$ und Konvergenzniveau $P_t = 0.67^{51}$. Weitere, verfahrensunabhängige Parameter werden im Text auf Seite 85 aufgeführt.

⁴⁸Durch Umstellen von Gleichung (4.5) auf Seite 38 ergibt sich: $x = -\beta \cdot \ln \left(\frac{1-\gamma-\lambda}{\psi-\gamma} - 1 \right) + \theta$. Mit $\hat{\psi} = P_t$ und $\theta = \vartheta$ erhält man somit in x den Wert der durch das Konvergenzniveau P_t definierten Schwelle ϑ_{P_t} .

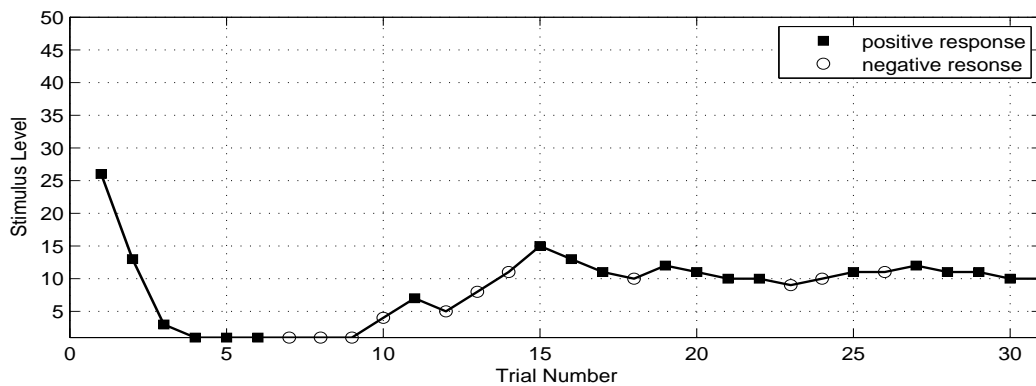
⁴⁹Siehe Anmerkung 48.

⁵⁰Siehe Anmerkung 48.

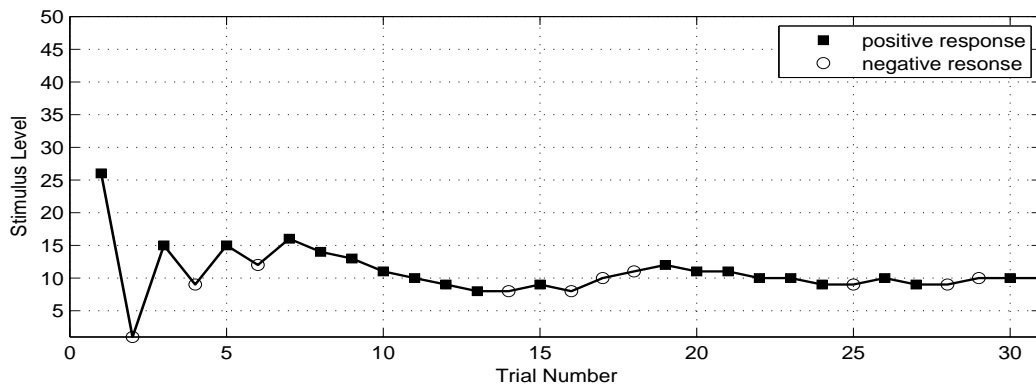
⁵¹Dies entspricht dem (auf zwei Dezimalstellen gerundeten) Ordinatenwert im Wendepunkt der psychometrischen Funktion des Beobachters bei Verwendung des 3AFC-Paradigmas.



(a) **Best-PEST-Verfahren** – Initialisierung durch Festlegung impliziter Trials nach Lieberman und Pentland (1982). „Wahre“ Schwelle $\vartheta = 10$; Schätzung der Schwelle $\hat{\vartheta} = 10$.



(b) **Best-PEST-Verfahren** – Initialisierung durch Vorgabe einer Normalverteilungsdichte $\mathcal{N}(x; \mu, \sigma)$ mit $\mu = 25.5$ und $\sigma = 35$. „Wahre“ Schwelle $\vartheta = 10$; Schätzung der Schwelle $\hat{\vartheta} = 10$.



(c) **Best-PEST-Verfahren** – Initialisierung durch Vorgabe einer konstanten Verteilungsdichte und Wahl des Anfangswertes in der Mitte des Reizstärkenbereichs. „Wahre“ Schwelle $\vartheta = 10$; Schätzung der Schwelle $\hat{\vartheta} = 10$.

Abbildung 6.3: Simulierte Reizstärkenverläufe nach dem Best-PEST-Verfahren für drei Varianten der Initialisierung. Einstellungen: $Range = 50$, $\beta = Range/10$, $\lambda = 0$, $\gamma = 1/n = 1/3$, Terminierung nach 30 Trials. Weitere, verfahrens-unabhängige Parameter werden im Text auf Seite 85 aufgeführt.

7 Entwurf der grafischen Benutzeroberflächen

Für den Entwurf der grafischen Benutzeroberflächen werden zunächst von Hand Skizzen angefertigt und diese anschließend durch Verwendung des Matlab-GUI-Editors *GUIDE*⁵² unter Durchführung einiger Modifikationen in Quellcode umgesetzt. Im Hinblick auf eine möglichst hohe Gebrauchstauglichkeit sollte bei diesem Vorgang die Orientierung an softwareergonomischen Grundsätzen der Dialoggestaltung angestrebt werden, wie sie in der Norm DIN EN ISO 9241-110 festgelegt sind. Im Folgenden werden diese aufgelistet und jeweils in ihrer Bedeutung kurz angerissen.

Aufgabenangemessenheit: Die Benutzeroberfläche sollte dem Benutzer in dem Maße Informationen anzeigen und zu Eingaben anleiten, als dass die volle Funktionalität des Back-End ausgenutzt werden kann. Unnötiger Informationsaustausch sollte vermieden werden.

Selbstbeschreibungsfähigkeit: Der Benutzer sollte durch die Benutzeroberfläche in ausreichendem Maße über den aktuellen Arbeitsschritt und die von ihm vorzunehmenden Eingaben informiert werden. Letztere sollten durch Rückmeldungen quittiert werden.

Erwartungskonformität: Der Aufbau des Dialogs sollte in Bezug auf Arbeitsschritte und Informationsdarstellung sowohl in sich möglichst konsistent sein, als auch mit der Erwartung des Benutzers übereinstimmen, welche sich aus Nutzungskontext und Erfahrungen mit ähnlicher Software ergibt. Insbesondere ist dabei auf die Verwendung einer dem Nutzungskontext entlehnten Terminologie zu achten.

Steuerbarkeit: Der Ablauf des Dialogs sollte durch den Benutzer gesteuert werden können. Insbesondere sollten Aktionen abgebrochen und ausgeführte Schritte zurückgenommen werden können.

Fehlertoleranz: Undefinierte Systemzustände aufgrund von Fehleingaben sollten vermieden und durch einen entsprechenden Mechanismus abgefangen werden, wel-

⁵²Abkürzung für *Graphical-User-Interface-Development-Environment*.

cher mit dem Benutzer kommuniziert und ihm eine Korrekturmöglichkeit anbietet.

Lernförderlichkeit: Die Benutzeroberfläche sollte den Benutzer beim Erlernen der Bedienung unterstützen. U.a. kann dies durch die Bereitstellung einer Benutzerdokumentation, verschiedener Hilfsfunktionen und eines möglichst konsistenten und transparenten Bedienkonzepts erreicht werden.

Individualisierbarkeit: Dialogeigenschaften sollten an die individuellen Ansprüche des Benutzers angepasst werden können, soweit dies nicht im Widerspruch zu den übrigen Grundsätzen steht.

Bei der Entwicklung der Benutzeroberflächen konnten im Rahmen dieser Arbeit nicht alle Grundsätze in gleichem Maße Berücksichtigung finden, da für die Umsetzung einiger Forderungen zu viel Zeit benötigt worden wäre. Es musste daher weitestgehend auf die Einrichtung von Möglichkeiten der Individualisierung, Fehlerbehandlungsmechanismen und Hilfsfunktionen verzichtet werden. Andererseits ergeben sich aus dem Nutzungskontext zusätzliche Anforderungen für die Gestaltung jener Oberflächen, welche dem Dialog mit der Versuchsperson während der Durchführung eines Hörversuchs zugrundeliegen. Zur Reduktion von Störeinflüssen (vgl. Abs. 3.1.3) sollten diese zum einen möglichst neutral gestaltet werden. Das Layout und die dargestellten Informationen bzw. Instruktionen sollten möglichst das Antwortverhalten der Versuchsperson nicht beeinflussen. Zum anderen sollte die in der Regel geringe Erfahrung der Versuchsperson im Umgang mit der Software berücksichtigt werden. Hierfür ist auf eine besondere Verständlichkeit von Informationen und Instruktionen und die Vermeidung von Fehlbedienungsmöglichkeiten (s. *Fehlertoleranz*) zu achten.

8 Zusammenfassung und Ausblick

Für die Durchführung von Hörversuchen ist am FG Audiokommunikation der TU Berlin auf Software zur automatisierten Kontrolle des Versuchsablaufs nicht mehr zu verzichten. Es besteht der Wunsch nach einem Programm mit einheitlichem Bedienkonzept und Hardwareanforderungen, welches gleichzeitig möglichst viele verschiedene Testverfahren abdeckt bzw. diesbezüglich bei Bedarf erweiterbar ist. Im Rahmen dieser Arbeit wurde ein Programmkonzept entwickelt, welches diesen Ansprüchen gerecht werden soll. Hierfür wurde zunächst eine Analyse des Nutzungskontexts und der Anforderungen durchgeführt, welche experimentelles Arbeiten im Allgemeinen und die Durchführung einer im Vorfeld durch das FG getroffenen Auswahl von Testverfahren an eine anwenderseitige Programmstruktur stellen. Zusammen mit André Wlodarski konnte ein System von grafischen Benutzeroberflächen erstellt werden. Hinsichtlich der Implementierung des Back-Ends konnte die Umsetzung der Adaption- und Schätzmechanismen in MATLAB-Funktionen übernommen werden. Schließlich wurde noch eine Benutzerdokumentation der Softwarebibliothek angefertigt.

Es war das Ziel der beiden Masterarbeiten eine Software zu entwickeln, welche im eingangs definierten Sinne eine möglichst hohe Gebrauchstauglichkeit aufweist. Als ein etabliertes, dahingehend optimiertes Vorgehen wurde der benutzerorientierte Gestaltungsprozess nach Norm DIN EN ISO 13407 vorgestellt. Ein wesentliches Element von diesem ist die Auswertung von Gestaltungslösungen unter Einbeziehung des Benutzers. Eine derartige Evaluation sollte noch durchgeführt werden, um anhand der hierbei gewonnenen Informationen Verbesserungen an der Benutzerschnittstelle vornehmen zu können.

Ein möglicher Anknüpfungspunkt für Weiterentwicklungen des Anwenderkonzepts soll an dieser Stelle bereits genannt werden. Die Anlage der Stimuli erfolgt derzeit manuell über eine grafische Benutzeroberfläche. Der Selbstversuch ergibt, dass für eine größere Anzahl an Stimuli, wie sie etwa im Kontext adaptiver Verfahren häufig benötigt wird, eine beträchtliche Zeit eingeplant werden muss. Hierbei würde es möglicherweise helfen einen Automatismus zu integrieren, welcher die Stimulusdefinitionen anhand vorgegebener Parameter vornimmt.

Literaturverzeichnis

- Anderson, A. J. (2003). Utility of a dynamic termination criterion in the ZEST adaptive threshold method. *Vision Research*, 43, Nr. 2, S. 165–170.
- Bech, S.; Zacharov, N. (2006). *Perceptual Audio Evaluation*. John Wiley & Sons, Ltd., Chichester.
- Berg, J. (2002). *Systematic evaluation of perceived spatial quality in surround sound systems*. Dissertation, School of Music, Luleå University of Technology.
- Berg, J. (2005). OPAQUE - a tool for the elicitation and grading of audioquality attributes. In: *Proceedings of The 118th AES Convention*. Barcelona.
- Berg, J.; Rumsey, F. (1999). Spatial attribute identification and scaling by repertory grid technique and other methods. In: *AES 16th International Conference Munich*, S. 51–56. Audio Engineering Society.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Springer, Berlin, 6. Aufl.
- Bortz, J.; Döring, N. (2006). *Forschungsmethoden und Evaluation*. Springer, Berlin, 4. Aufl.
- Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. *American Journal of Psychology*, 75, S. 485–491.
- Danielsson, M. (1991). Repertory Grid Technique. Forschungsbericht, Luleå University of Technology.
- DIN EN ISO 13407 (2000). Benutzer-orientierte Gestaltung interaktiver Systeme.
- DIN EN ISO 9241-11 (1999). Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Teil11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze.
- DIN EN ISO 9241-110 (2006). Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten, Teil110: Grundsätze der Dialoggestaltung.
- Dixon, W. J.; Mood, A. M. (1948). A Method for Obtaining and Analyzing Sensitivity Data. *Journal of the American Statistical Association*, 43, S. 109–126.
- Emerson, P. L. (1986). A quadrature method for bayesian sequential threshold estimation. *Perception & Psychophysics*, 39, S. 381–383.
- Epting, F. R.; Suchman, D. I.; Nickson, K. J. (1971). An evaluation of elicitation procedures for personal constructs. *British Journal of Psychology*, 62, S. 513–517.

- Findlay, J. M. (1978). Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, 23, Nr. 2, S. 181–185.
- Fransella, F.; Bannister, D. (1977). *A manual for repertory grid technique*. Academic press, London.
- Gescheider, G. A. (1997). *Psychophysics: The Fundamentals*. Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, 69, Nr. 6, S. 1763–1769.
- Hansen, M. (2008). PsyLab Documentation, Version 2.1. Institut für Hörtechnik und Audiologie, Fachhochschule OOW, Oldenburg.
- Harvey, jr., L. O. (1986). Efficient estimation of sensory thresholds. *Behavior Research Methods, Instruments & Computers*, 18, Nr. 6, S. 623–632.
- Hastings, N. A.; Peacock, J. B. (1975). *Statistical Distributions*. Butterworth & Co (Publishers) Ltd, London.
- Hofstätter, P. R. (1980). *Psychologie*, Bd. 6 von *Das Fischer Lexikon*. Fischer, Frankfurt am Main. Neuausgabe 622.–628. Tsd.
- Kelly, G. A. (1955). *The Psychology of Personal Constructs*. Norton, New York.
- King-Smith, P. E.; Grigsby, S. S.; Vingrys, A. J.; Benes, S. C.; Supowit, A. (1994). Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Evaluation and Practical Implementation. *Vision Research*, 34, Nr. 7, S. 885–912.
- Kirk, R. E. (1982). *Experimental Design*. Procedures for the Behavioral Sciences. Brooks/Cole Publishing Company, Pacific Grove.
- Klein, S. A. (2001). Measuring, estimating and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63, Nr. 8, S. 1421–1455.
- Laback, B. (2006). Psychoakustik II. Schwerpunkt: Experimentelle Audiologie. Skript zur Vorlesung. Institut für Schallforschung, Österreichische Akademie der Wissenschaften, Wien.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63, Nr. 8, S. 1279–1292.
- Leek, M. R.; Watson, C. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical Society of America*, 76, Nr. 4, S. 1037–1044.
- Lehmann, G. (2002). *Statistik. Eine Einführung*. Spektrum Akademischer Verlag, Heidelberg; Berlin.
- Levitt, H. (1968). Testing for Sequential Dependencies. *The Journal of the Acoustical Society of America*, 43, Nr. 3, S. 65–69.

- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, Nr. 2, S. 467–477.
- Lieberman, H.; Pentland, A. (1982). Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods & Instrumentation*, 14, Nr. 1, S. 21–25.
- Madigan, R.; Williams, D. (1987). Maximum-Likelihood Psychometric Procedures in Two-Alternative Forced-Choice: Evaluation and Recommendations. *Perception & Psychophysics*, 42, Nr. 3, S. 240–249.
- Maempel, H.-J. (2001). Die Bedeutung der Klanggestaltung für das Hörerurteil. Bericht über eine empirische Untersuchung zu Produktion und Rezeption von Popmusik. In: *Bericht über die 21. Tonmeistertagung Karlsruhe 2000*, S. 953–969. Verband Deutscher Tonmeister, Saur, München.
- Maempel, H.-J. (2007). Perzeptive Messung und Evaluation. Skript zur Lehrveranstaltung ‚Kommunikationstechnik-Labor (WS 06/07)‘. FG Audiokommunikation, TU Berlin.
- Malmstrom, E. J.; French, G. M. (1963). Scale-Symmetry and the Semantic Differential. *The American Journal of Psychology*, 76, Nr. 3, S. 446–451.
- Mann, I. T.; Phillips, J. L.; Thompson, E. G. (1979). An examination of methodological issues relevant to the use and interpretation of the semantic differential. *Applied Psychological Measurement*, 3, S. 213–229.
- Martz, H. F.; Waller, R. A. (1982). *Bayesian reliability analysis*. John Wiley & Sons, Ltd., New York.
- Marvit, P.; Florentine, M.; Buus, S. (2003). A comparison of psychophysical procedures for level-discrimination thresholds. *Journal of the Acoustical Society of America*, 113, Nr. 6, S. 3348–3361.
- Merten, K. (1983). *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis*. Westdeutscher Verlag, Opladen.
- Mond, H.-J. (2007). *Implementierung einer graphischen Benutzeroberfläche zur Steuerung eines Wellenfeldsynthesystems*. Magisterarbeit, FG Audiokommunikation, Institut für Sprache und Kommunikation, Fakultät I, TU Berlin.
- Osgood, C. E.; Suci, G. J.; Tannenbaum, P. H. (1967). *The measurement of meaning*. University of Illinois Press, Urbana. Neuauflage der Originalausgabe von 1957.
- Otto, S. (2008). *Vergleichende Simulation adaptiver, psychometrischer Verfahren zur Schätzung von Wahrnehmungsschwellen*. Magisterarbeit, FG Audiokommunikation, Institut für Sprache und Kommunikation, Fakultät I, TU Berlin.
- Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics*, 28, Nr. 4, S. 377–379.

- Pokorny, M. A. (1998). Using maximum-likelihood adaptive methods to estimate difference thresholds and points of subjective equality. *Behavior Research Methods, Instruments, & Computers*, 30, Nr. 4, S. 624–636.
- Sarris, V. (1990). *Erkenntnisgewinnung und Methodik*, Bd. 1 von *Methodologische Grundlagen der Experimentalpsychologie*. Reinhardt, München; Basel.
- Sarris, V. (1992). *Versuchsplanung und Stadien*, Bd. 2 von *Methodologische Grundlagen der Experimentalpsychologie*. Reinhardt, München; Basel.
- Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, 16, S. 209–222.
- Scheer, J. W. (1992). Psychologie der persönlichen Konstrukte und Repertory Grid-Technik. Ein idiographischer Ansatz in klinischer und medizinischer Psychologie. In: Schmidt, L. R. (Hrsg.), *Psychologische Aspekte medizinischer Maßnahmen. Jahrbuch der medizinischen Psychologie*, Bd. 7, S. 273–290. Springer, Heidelberg.
- Schmitt, M. (1990). Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung. In: Albert, D.; Pawlik, K.; Stapf, K.-H.; Stroebe, W. (Hrsg.), *Lehr- und Forschungstexte Psychologie*, Bd. 36. Springer-Verlag, Berlin; Heidelberg.
- Shadish, W. R.; Cook, T. D.; Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, Boston.
- Taylor, M. M.; Creelman, C. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 41, Nr. 4, S. 782–787.
- Treutwein, B. (1995). Adaptive Psychophysical Procedures. *Vision Research*, 35, Nr. 17, S. 2503–2522.
- Treutwein, B. (1997). YAAP: yet another adaptive procedure. *Spatial Vision*, 11, Nr. 1, S. 129–134.
- Treutwein, B.; Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, 61, Nr. 1, S. 87–106.
- Wald, A. (1947). *Sequential Analysis*. John Wiley & Sons, Inc., New York.
- Watson, A. B.; Pelli, D. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, Nr. 2, S. 113–120.
- Westermann, R. (2000). *Wissenschaftstheorie und Experimentalmethodik. Ein Lehrbuch zur Psychologischen Methodenlehre*. Hogrefe, Göttingen.
- Wetherill, G. (1963). Sequential Estimation of Quantal Response Curves. *Journal of the Royal Statistical Society*, 25, S. 1–48.
- Wetherill, G. B.; Levitt, H. (1965). Sequential Estimation of Points on a Psychometric Function. *British Journal of Mathematical and Statistical Psychology*, 18, S. 1–10.

- Wlodarski, A. (2008). *Erstellung einer Softwarebibliothek für Hörversuche – Implementierung und Strukturierung des Quellcodes*. Masterarbeit, FG Audiokommunikation, Institut für Sprache und Kommunikation, Fakultät I, TU Berlin.
- Zuberbühler, H. J. (2002). Rapid Evaluation of Perceptual Thresholds. The Best-Pest Calculator: A web-based application for non-expert users. <http://www.psychophysics.ethz.ch/tools>.

Abbildungsverzeichnis

3.1	Übersicht über Techniken zur Kontrolle von Störvariablen in experimentellen/quasiexperimentellen Untersuchungen	20
3.2	Übersicht über Testverfahren der auditiv-perzeptiven Forschung . . .	26
3.3	Beispiel einer grafischen Benutzeroberfläche zur Sample-Familiarisation unter Matlab	29
4.1	Logistische Funktion für verschiedene Werte des Spread-Parameters .	37
4.2	Modellierung der psychometrischen Funktion für verschiedene Antwortparadigmen mittels logistischer Funktion	37
4.3	Ausschnitt eines möglichen Reizstärkenverlaufs für ein einfaches Staircase-Verfahren	41
4.4	Ausschnitt eines möglichen Reizstärkenverlaufs für ein 3-Down/1-Up-Staircase-Verfahren	43
4.5	Ausschnitt eines nach dem PEST-Verfahren aufgenommenen Reizstärkenverlaufs	45
4.6	Likelihood-Funktionen und Reizstärkenverlauf für die ersten 4 Trials eines Runs	54
4.7	Verwendete Adjektivpaare einer Untersuchung in der Audioevaluation	62
4.8	Polaritätsprofil aus einer Untersuchung in der Audioevaluation	62
4.9	Grid aus einer Untersuchung in der Audioevaluation	70
5.1	Anwenderseitige Programmstruktur der Softwarebibliothek	78
6.1	Simulierte Reizstärkenverläufe nach dem Staircase-Verfahren für die drei implementierten Adaptionsregeln	86
6.2	Simulierte Reizstärkenverläufe für die zwei implementierten PEST-Verfahren	87
6.3	Simulierte Reizstärkenverläufe nach dem Best-PEST-Verfahren für drei Varianten der Initialisierung	88

Tabellenverzeichnis

4.1	Übersicht der ausgewählten Adaptionenregeln und ihre zugehörigen Konvergenzniveaus; Ratekorrektur für nAFC-Paradigma	44
-----	---	----

A.1 MATLAB-Funktionen

Staircase 1-Down/1-UP

103

```

else
    change = 0; % bleibt gleich
end

% Reversal?
if (change ~= 0) && (latest_change ~=0) && (latest_change ~= change)
    nr = nr + 1;
    x_r(nr) = x_hist(trial-1);
    latest_change = change;
    if isempty(find(halvings==nr))==0
        numhalv = numhalv + 1;
    end
elseif (change ~= 0) && (latest_change == 0)
    latest_change = change;
end

end% Ende Analyse

%
%               * * * * *
%               Nächster Schritt

% Adaptionsregel: Entscheidung über die Richtung des nächsten Schrittes
% (1-Down/1-UP)
if r_hist(Ntrials) == 1
    this_change = -1;
else
    this_change = +1;
end

% Reversal?
if (this_change ~= latest_change) && (latest_change ~=0)
    nr = nr + 1;
    x_r(nr)=x_hist(Ntrials);
    if isempty(find(halvings==nr))==0
        numhalv = numhalv + 1;
    end
end

% Terminierung?
if nr >= nrmax
    termflag = 1;
end

% Berechnung der Größe des nächsten Schrittes
% x_hist(Ntrials)--> x_hist(Ntrials + 1)
dx = round(dx_ini/(2^numhalv));

% Berechnung des nächsten Reizstärkenwertes
x_next = x_hist(Ntrials) + this_change * dx;

% Verhalten an den Rändern
if x_next > range
    x_next = range;
end

if x_next < 1
    x_next = 1;
end

%
%               * * * * *
%               Berechnung der Schwellenschätzung

```



```
if termflag == 1
    x_est = mean(x_r((nrmax-numlastr+1):nrmax));
end
```

Staircase 2-Down/1-UP

```
function [x_next, x_est] = staircase2d1u(x_hist, r_hist, range, nrmax, halvings)
    ( , numlastr)
```

```
% Parameter
Ntrials = length(x_hist); % Anzahl bisher durchgeführter Trials
numhalv_total = length(halvings); % Anzahl der insgesamt vorgesehenen
    ( Halbierungen
dx_ini = 2^numhalv_total; % Anfangswert der Schrittweite

% Initialisierungen
latest_change = 0; % Indikator für zuletzt erfolgte Reizstärkenänderung
                    % (0: noch keine Änderung im bisherigen Verlauf,
                    % 1: Erhöhung, -1: Verminderung)
nr = 0; % Zähler für Reversals des Tracks
numhalv = 0; % Zähler für aktuell zu erfolgende Halbierungen
termflag = 0; % Indikator für Terminierung des Verfahrens mit dem aktuellen
                % Trial, d.h. für das Erreichen der maximalen Anzahl nrmax an
                % Reversals. (1: beenden, 0: nicht beenden)
x_r = zeros(nrmax, 1); % Speicherung der Reizstärkenwerte bei sämtlichen
    ( Reversals
T = 1; % Zähler für Trials bei gleichbleibender Reizstärke
x_est = 0;
```

```
%
%
% * * * * *
% Analyse des bisherigen Tracks
```

```
for trial = 2:Ntrials

    % Reizstärkenwechsel?
    if x_hist(trial) ~= x_hist(trial-1)
        T=1;
        % Art des Reizstärkenwechsels
        if x_hist(trial) > x_hist(trial-1)
            change = 1; % Erhöhung
        else
            change = -1; % Verminderung
        end

    else
        change = 0; % bleibt gleich
        T=T+1;
    end

    % Reversal?
    if (change ~= 0) && (latest_change ~= 0) && (latest_change ~= change)
        nr = nr + 1;
        x_r(nr) = x_hist(trial-1);
        latest_change = change;
        if isempty(find(halvings==nr))==0
            numhalv = numhalv + 1;
        end
    elseif (change ~= 0) && (latest_change == 0)
        latest_change = change;
```

```

        end

    end

%
%                               * * * * *
%                               Nächster Schritt

% Adaptionsregel: Entscheidung über die Richtung des nächsten Schrittes
% (2-Down/1-UP)
if (Ntrials>=2)
    if (T>=2) && (r_hist(Ntrials) == 1) && (r_hist(Ntrials-1) == 1)
        this_change = -1;
    elseif (r_hist(Ntrials) == -1)
        this_change = +1;
    else
        this_change = 0;
    end
else
    this_change = 0;
end

% Reversal?
if (this_change ~= 0) && (this_change ~= latest_change) && (latest_change ~=0)
    nr = nr + 1;
    x_r(nr)=x_hist(Ntrials);
    if isempty(find(halvings==nr))==0
        numhalv = numhalv + 1;
    end
end

% Terminierung?
if nr >= nrmax
    termflag = 1;
end

% Berechnung der Größe des nächsten Schrittes
% x_hist(Ntrials) --> x_hist(Ntrials + 1)
dx = round(dx_ini/(2^numhalv));

% Berechnung des nächsten Reizstärkenwertes
x_next = x_hist(Ntrials) + this_change * dx;

% Verhalten an den Rändern
if x_next > range
    x_next = range;
end
if x_next < 1
    x_next = 1;
end

%
%                               * * * * *
%                               Berechnung der Schwellenschätzung

if termflag == 1
    x_est = mean(x_r((nrmax-numlastr+1):nrmax));
end

Staircase 3-Down/1-UP

function [x_next, x_est] = staircase3d1u(x_hist, r_hist, range, nrmax, halvings)
    ( , numlastr)

```

```

% Parameter
Ntrials = length(x_hist); % Anzahl bisher durchgeführter Trials
numhalv_total = length(halvings); % Anzahl der insgesamt vorgesehenen 2
    Halbierungen
dx_ini = 2^numhalv_total; % Anfangswert der Schrittweite

% Initialisierungen
latest_change = 0; % Indikator für zuletzt erfolgte Reizstärkenänderung
    % (0: noch keine Änderung im bisherigen Verlauf,
    % 1: Erhöhung, -1: Verminderung)
nr = 0; % Zähler für Reversals des Tracks
numhalv = 0; % Zähler für aktuell zu erfolgende Halbierungen
termflag = 0; % Indikator für Terminierung des Verfahrens mit dem aktuellen
    % Trial, d.h. für das Erreichen der maximalen Anzahl nrmax an
    % Reversals. (1: beenden, 0: nicht beenden)
x_r = zeros(nrmax, 1); % Speicherung der Reizstärkenwerte bei sämtlichen 2
    Reversals
T = 1; % Zähler für Trials bei gleichbleibender Reizstärke
x_est = 0;

%
%
% * * * * *
% Analyse des bisherigen Tracks

for trial = 2:Ntrials

    % Reizstärkenwechsel?
    if x_hist(trial) ~= x_hist(trial-1)
        T=1;
        % Art des Reizstärkenwechsels
        if x_hist(trial) > x_hist(trial-1)
            change = 1; % Erhöhung
        else
            change = -1; % Verminderung
        end

    else
        change = 0; % bleibt gleich
        T=T+1;
    end

    % Reversal?
    if (change ~= 0) && (latest_change ~=0) && (latest_change ~= change)
        nr = nr + 1;
        x_r(nr) = x_hist(trial-1);
        latest_change = change;
        if isempty(find(halvings==nr))==0
            numhalv = numhalv + 1;
        end
    elseif (change ~= 0) && (latest_change == 0)
        latest_change = change;
    end

end

%
%
% * * * * *
% Nächster Schritt

% Adaptionsregel: Entscheidung über die Richtung des nächsten Schrittes
% (3-Down/1-UP)
if (Ntrials>=3)

```

```

        if (T>=3) && (r_hist(Ntrials) == 1) && (r_hist(Ntrials-1) == 1) && (r_hist(Ntrials-2) == 1)
            this_change = -1;
        elseif r_hist(Ntrials) == -1
            this_change = +1;
        else
            this_change = 0;
        end
    else
        this_change = 0;
    end

% Reversal?
if (this_change ~= 0) && (this_change ~= latest_change) && (latest_change ~= 0)
    nr = nr + 1;
    x_r(nr)=x_hist(Ntrials);
    % ggf. Halbierung der Schrittweite
    if isempty(find(halvings==nr))==0
        numhalv = numhalv + 1;
    end
end

% Terminierung?
if nr >= nrmax
    termflag = 1;
end

% Berechnung der Größe des nächsten Schrittes
% x_hist(Ntrials) --> x_hist(Ntrials + 1)
dx = round(dx_ini/(2^numhalv));

% Berechnung des nächsten Reizstärkenwertes
x_next = x_hist(Ntrials) + this_change * dx;

% Verhalten an den Rändern
if x_next > range
    x_next = range;
end
if x_next < 1
    x_next = 1;
end

%
% * * * * *
% Berechnung der Schwellenschätzung

if termflag == 1
    x_est = mean(x_r((nrmax-numlastr+1):nrmax));
end

```

PEST

```

function [x_next, x_est] = PEST(x_hist, r_hist, range, dx_ini, dx_max, P_t, W)

% Parameter
Ntrials = length(x_hist); % Anzahl bisher durchgeführter Trials

% Initialisierungen
latest_change = 0; % Kodierung für die zuletzt erfolgte Reizstärkenänderung
% (-1: Erniedrigung, +1: Erhöhung, 0: es ist im bisherigen Verlauf noch keine Reizstärkenänderung eingetreten.
doubling = 0; % merkt sich, wenn die Schrittweite verdoppelt, d.h. ner um 1

```

```

        % verringert wird
rule4 = 0; % Indikator für das Vorhandensein einer Verdopplung der Schritt-
        % weite vor der zuletzt erfolgten Richtungsumkehr (siehe Regel 4)
termflag = 0; % Indikator für Terminierung (1: Terminierung erfolgt,
        % 0: keine Terminierung)
ner = 0; % Initialisierung der number of effective reversals (ner)
nstep = 0; % zählt bei Reizstärkenänderungen alle Schritte in eine
        % gegebene Richtung

```

```

%-----
% Analyse des Tracks & Reizstärkenwert für die nächste Darbietung
%-----

```

```

% Initialisierung des Signifikanztests
decision = 1;

```

```

for trial = 1:Ntrials

```

```

    %-----
    % Signifikanztest
    %-----

```

```

    % T zählt die Anzahl der Trials, die bei einer gleichbleibenden
    % Reizstärke durchgeführt wurden
    % N_C zählt die positiven Antworten bei einer gleichbleibenden
    % Reizstärke

```

```

    % Reset des Tests wenn im vorherigen Trial Entscheidung für Änderung
    % der Reizstärke
    if decision ~= 0
        T = 0;
        N_C = 0;
    end

```

```

    T = T + 1; % Zählen sämtlicher Trials bei derselben Reizstärke
    if r_hist(trial) == +1
        N_C = N_C + 1; % Zählen positiver Antworten bei derselben Reizstärke
    end

```

```

    E = P_t * T; % erwartete Anzahl positiver Antworten nach T bei gleicher
        % Reizstärke durchgeführten Trials
    N_bu = E + W; % obere Grenze
    N_bl = E - W; % untere Grenze

```

```

    % Der eigentliche Kern des Tests
    % Die Richtung der Reizstärkenänderung wird durch die Variable decision
    % ausgedrückt:
    %   decision = +1: Erhöhung
    %   decision = -1: Verminderung
    %   decision = 0: keine Änderung

```

```

    if (N_C < N_bu) && (N_C > N_bl)
        x_next = x_hist(trial);
        decision = 0;
        continue % springe zum nächsten Trial, wenn keine Änderung der
            % Reizstärke erfolgt
    elseif N_C >= N_bu
        decision = -1; % Verminderung
    end

```

```

elseif N_C <= N_bl
    decision = +1; % Erhöhung
end

%-----
% Regeln zur Änderung der Schrittweite
%-----

% 1. Bei jeder Richtungsumkehr erfolgt eine Halbierung der
% Schrittweite, d.h. 'ner' (number of effective reversals) erhöht sich
% um 1

if (latest_change ~= 0) && (latest_change ~= decision)

    % hat es vor dieser Richtungsumkehr eine Verdopplung der Reizstärke
    % gegeben? (bedeutsam für Regel 4)
    if doubling == 1
        rule4 = 1; % ja
    else
        rule4 = 0; % nein
    end

    ner = ner + 1; % Halbierung der Schrittweite
    nstep = 0; % Zurücksetzen des Zählers der Schritte in eine Richtung
    doubling = 0;
end

nstep = nstep + 1;

% 2. Im zweiten Schritt in eine gegebene Richtung bleibt die Schritt-
% weite und somit auch 'ner' unverändert.

if nstep == 2
    doubling = 0;
end

% 3. Der vierte und die darauf folgenden Schritte in eine gegebene
% Richtung sind jeweils doppelt so groß wie der ihnen vorher gehende.

if nstep >= 4
    ner = ner - 1;
    doubling = 1;
end

% 4. Die Größe des dritten Schrittes in eine gegebene Richtung hängt von
% den Schritten ab, welche zur letzten Richtungsumkehr führten: Wenn der
% Schritt, welcher dieser vorausging, aus einer Verdopplung der Schritt-
% weite hervorgegangen war, dann wird die Schrittweite nicht verdoppelt
% (sondern beibehalten). Falls dieser allerdings nicht das Ergebnis ei-
% ner Verdopplung war, so wird die Schrittweite verdoppelt.

if nstep == 3
    if rule4 == 1 % es hat eine Verdopplung gegeben
        % Schrittweite wird beibehalten
        doubling = 0;
    else % es hat keine Verdopplung gegeben
        ner = ner - 1; % die Schrittweite wird verdoppelt
        doubling = 1;
    end
end
end

```

```
latest_change = decision; % Setze den Indikator für die Richtungsumkehr

%-----
% Berechnung der aktuellen Schrittweite dx
%-----
dx = dx_ini/(2^ner);

% Begrenzung der Schrittweite durch die maximale Schrittweite dx_max
if dx > dx_max
    dx = dx_max;
end

%-----
% Terminierung ja/nein?
%-----

if dx < 1
    termflag = 1;
else
    termflag = 0;
    dx = round(dx); % soll den Fall abfangen, dass keine Zweierpotenz
                    % für dx_ini bzw. dx_max eingegeben wurde
end

%-----
% Berechnung des nächsten Reizstärkenwertes
%-----

if decision == 1 % Erhöhung der Reizstärke
    x_next = x_hist(trial) + dx;
    if x_next >= range
        x_next = range; % Setze auf Rand, wenn Rand erreicht/überschritten
        nstep = 0; % Am Rand werden die Schritte nicht weitergezählt,
                  % sodass keine Verdopplungen der Schrittweite
                  % resultieren
    end
end

if decision == -1 % Erniedrigung der Reizstärke
    x_next = x_hist(trial) - dx;
    if x_next <= 1
        x_next = 1;
        nstep = 0; % s.o.
    end
end

end

%-----
% Ausgabe der abschließenden Schwellenschätzung (falls terminiert)
%-----

if termflag == 1
    x_est = x_next;
else
    x_est = 0;
end
```

More-Virulent-PEST

```
function [x_next, x_est] = mvPEST(x_hist, r_hist, range, dx_ini, dx_max, P_t, M)
( )

% Parameter
Ntrials = length(x_hist); % Anzahl bisher durchgeführter Trials

% Initialisierungen
latest_change = 0; % Kodierung für die zuletzt erfolgte Reizstärkenänderung
                    % (-1: Erniedrigung, +1: Erhöhung, 0: es ist im bis-
                    % herigen Verlauf noch keine Reizstärkenänderung ein-
                    % getreten.
doubling = 0; % merkt sich, wenn die Schrittweite verdoppelt, d.h. ner um 1
               % verringert wird
rule4 = 0; % Indikator für das Vorhandensein einer Verdopplung der Schritt-
            % weite vor der zuletzt erfolgten Richtungsumkehr (siehe Regel 4)
termflag = 0; % Indikator für Terminierung (1: Terminierung erfolgt,
              % 0: keine Terminierung)
ner = 0; % Initialisierung der number of effective reversals (ner)
nstep = 0; % zählt bei Reizstärkenänderungen alle Schritte in eine
            % gegebene Richtung

%-----
% Analyse des Tracks & Reizstärkenwert für die nächste Darbietung
%-----

% Initialisierung des Signifikanztests
decision = 1;

for trial = 1:Ntrials

    %-----
    % Signifikanztest
    %-----

    % T zählt die Anzahl der Trials, die bei einer gleichbleibenden
    % Reizstärke durchgeführt wurden
    % N_C zählt die positiven Antworten bei einer gleichbleibenden
    % Reizstärke

    % Reset des Tests wenn im vorherigen Trial Entscheidung für Änderung
    % der Reizstärke
    if decision ~= 0
        T = 0;
        N_C = 0;
    end

    T = T + 1; % Zählen sämtlicher Trials bei derselben Reizstärke
    if r_hist(trial) == +1
        N_C = N_C + 1; % Zählen positiver Antworten bei derselben Reizstärke
    end

    % *** Findlays 2. Modifikation:
    % Das Deviation-Limit ist eine Funktion von ner
        if ner > 1
            W = ner / 2;
        else
            W = 0.5;
        end
    end
end
```



```
end

E = P_t * T; % erwartete Anzahl positiver Antworten nach T bei gleicher
              % Reizstärke durchgeführten Trials
N_bu = E + W; % obere Grenze
N_bl = E - W; % untere Grenze

% Der eigentliche Kern des Tests
% Die Richtung der Reizstärkenänderung wird durch die Variable decision
% ausgedrückt:
%   decision = +1: Erhöhung
%   decision = -1: Verminderung
%   decision = 0: keine Änderung

if (N_C < N_bu) && (N_C > N_bl)
    x_next = x_hist(trial);
    decision = 0;
    continue % springe zum nächsten Trial, wenn keine Änderung der
              % Reizstärke erfolgt
elseif N_C >= N_bu
    decision = -1; % Verminderung
elseif N_C <= N_bl
    decision = +1; % Erhöhung
end

%-----
% Regeln zur Änderung der Schrittweite
%-----

% *** Findlays 1. Modifikation:
% erreicht T den Wert M oder jedes ganzzahlige Vielfache davon,
% so wir ner um 1 erhöht ***
ner = ner + floor(T/M);

% 1. Bei jeder Richtungsumkehr erfolgt eine Halbierung der
% Schrittweite, d.h. 'ner' (number of effective reversals) erhöht sich
% um 1

if (latest_change ~= 0) && (latest_change ~= decision)

    % hat es vor dieser Richtungsumkehr eine Verdopplung der Reizstärke
    % gegeben? (bedeutsam für Regel 4)
    if doubling == 1
        rule4 = 1; % ja
    else
        rule4 = 0; % nein
    end

    ner = ner + 1; % Halbierung der Schrittweite
    nstep = 0; % Zurücksetzen des Zählers der Schritte in eine Richtung
    doubling = 0;
end

nstep = nstep + 1;

% 2. Im zweiten Schritt in eine gegebene Richtung bleibt die Schritt-
% weite und somit auch 'ner' unverändert.
```

```

if nstep == 2
    doubling = 0;
end

% 3. Der vierte und die darauf folgenden Schritte in eine gegebene
% Richtung sind jeweils doppelt so groß wie der ihnen vorher gehende.

if nstep >= 4
    ner = ner - 1;
    doubling = 1;
end

% 4. Die Größe des dritten Schrittes in eine gegebene Richtung hängt von
% den Schritten ab, welche zur letzten Richtungsumkehr führten: Wenn der
% Schritt, welcher dieser vorausging, aus einer Verdopplung der Schritt-
% weite hervorgegangen war, dann wird die Schrittweite nicht verdoppelt
% (sondern beibehalten). Falls dieser allerdings nicht das Ergebnis ei-
% ner Verdopplung war, so wird die Schrittweite verdoppelt.

if nstep == 3
    if rule4 == 1 % es hat eine Verdopplung gegeben
        % Schrittweite wird beibehalten
        doubling = 0;
    else % es hat keine Verdopplung gegeben
        ner = ner - 1; % die Schrittweite wird verdoppelt
        doubling = 1;
    end
end

latest_change = decision; % Setze den Indikator für die Richtungsumkehr

%-----
% Berechnung der aktuellen Schrittweite dx
%-----
dx = dx_ini/(2^ner);

% Begrenzung der Schrittweite durch die maximale Schrittweite dx_max
if dx > dx_max
    dx = dx_max;
end

%-----
% Terminierung ja/nein?
%-----

if dx < 1
    termflag = 1;
else
    termflag = 0;
    dx = round(dx); % soll den Fall abfangen, dass keine Zweierpotenz
                    % für dx_ini bzw. dx_max eingegeben wurde
end

%-----
% Berechnung des nächsten Reizstärkenwertes
%-----

if decision == 1 % Erhöhung der Reizstärke
    x_next = x_hist(trial) + dx;
    if x_next >= range

```

```

        x_next = range; % Setze auf Rand, wenn Rand erreicht/überschritten
        nstep = 0; % Am Rand werden die Schritte nicht weitergezählt,
                    % sodass keine Verdopplungen der Schrittweite
                    % resultieren
    end
end

if decision == -1 % Erniedrigung der Reizstärke
    x_next = x_hist(trial) - dx;
    if x_next <= 1
        x_next = 1;
        nstep = 0; % s.o.
    end
end

end

%-----
% Ausgabe der abschließenden Schwellenschätzung (falls terminiert)
%-----

if termflag == 1
    x_est = x_next;
else
    x_est = 0;
end
end

```

Maximum-Likelihood/Bayes

```

function [x_next, x_est, finalLike] = MLBayes(x_hist, r_hist, range, beta, 2
    ( lambda, n_or_gamma, priortype, mu, sigma, averagetype, exclude, termtyp, 2
    ( ConCoeff, CI_max)

% Parameter
Ntrials=length(x_hist); % Anzahl aller bisherig tatsächlich durchgeführten
                        % Trials
% Welches Antwortparadigma liegt vor? Wie groß ist dementsprechend die
% Ratewahrscheinlichkeit/False-Alarm-Rate GAMMA zu wählen?
if (n_or_gamma >=0) && (n_or_gamma <=1) % ja/nein-Paradigma
    paradigm = 0;
    gamma = n_or_gamma;
elseif (n_or_gamma >=2) % nAFC-Paradigma
    paradigm = 1;
    n = n_or_gamma;
    gamma = 1/n;
else
    error('You have chosen an invalid response paradigm');
end

% Initialisierung
PosProb = zeros(range, 1); % Vektor für die Wahrscheinlichkeit einer positi-
                        % ven Antwort
cleanLike = ones(range, 1); % "saubere" Likelihood-Funktion als Rechteckver-
                        % teilung
aprioriLike = ones(range, 1); % Initialisierung der a-priori-Likelihood als
                        % Rechteckverteilung (zunächst)
x_est = 0; % Abschließende Schätzung

% Berechnung der psychometrischen Funktion und Speicherung in einem Vektor

```

```

PsyFunc = zeros(2*range, 1);
for x=1:2*range
    PsyFunc(x) = gamma + ((1-gamma-lambda)/(1+exp(-(x-range)/beta)));
end

%-----
% A-priori-Information
%-----

switch priortype

    case 1 % 1. Variante: Initialisierung per impliziter Trials nach
        % Lieberman & Pentland (1982)

        % Definition der impliziten Trials
        if paradigm == 0 % ja/nein-Paradigma
            x_im(1)=1;
            r_im(1)=-1;
            x_im(2)=range;
            r_im(2)=+1;
        end

        if paradigm == 1 % nAFC-Paradigma
            for i=1:n
                x_im(i)=1;
                r_im(i)=-1;
            end
            r_im(1)=+1; % genau eine positive Antwort
        end

        % Berechnung der a-priori-Likelihood
        for i=1:length(x_im)
            % Berechnung des Vektors mit den Auftretenswahrscheinlichkeiten für 2
            %   ↳ eine
            % positive Antwort bei Darbietung x_im(i) unter Annahme 2
            %   ↳ verschiedener
            % Schwellen theta = x, für x aus [1:range]

            PosProb(1:range) = PsyFunc(range+x_im(i)-1:-1:x_im(i));

            % Berechnung für negative Antwort
            NegProb = 1 - PosProb;

            if r_im(i)==1 %falls positive Antwort
                aprioriLike=aprioriLike.*PosProb;
            else %falls negative Antwort
                aprioriLike=aprioriLike.*NegProb;
            end
            % Normierung auf Maximum
            aprioriLike=aprioriLike/max(aprioriLike);

        end

    case 2 % 2. Variante: Initialisierung des Verfahrens per Annahme einer
        % Gauss-Funktion für die A-priori-Likelihood (vgl. z.B. QUEST)

        temp = normpdf(1:range, mu, sigma);
        aprioriLike = aprioriLike.*temp';
        aprioriLike = aprioriLike/max(aprioriLike); % Normierung auf Maximum

```

```

        otherwise % 3. Variante: Initialisierung des Verfahrens per 2
            & Gleichverteilung
            % - kein weiterer Code nötig -

    end

%-----
% Berechnung der Likelihood-Funktion(en)
%-----

%
%*****

% Berechnung der "sauberen" Likelihood-Funktion, welche ausschließlich auf
% den tatsächlich durchgeführten Trials beruht

for i=1:Ntrials
    % Berechnung des Vektors mit den Auftretenswahrscheinlichkeiten für eine
    % positive Antwort bei Darbietung x_hist(i) unter Annahme verschiedener
    % Schwellen theta = x, für x aus [1:range]

    PosProb(1:range) = PsyFunc(range+x_hist(i)-1:-1:x_hist(i));

    % Berechnung für negative Antwort
    NegProb = 1 - PosProb;

    if r_hist(i)==1 %falls positive Antwort
        cleanLike=cleanLike.*PosProb;
    else %falls negative Antwort
        cleanLike=cleanLike.*NegProb;
    end

    % Normierung auf die Fläche unter dem Funktionsgraphen
    cleanLike=cleanLike/sum(cleanLike);

end

%
%*****

% Berechnung der gesamten Likelihood-Funktion, welche auch
% die a-priori-Information mit einschließt

Like = aprioriLike.*cleanLike;
Like = Like/sum(Like); % Normierung auf die Fläche unter dem Funktions-
% grafen

%
%*****

% Berechnung der Likelihood-Funktion, welche für die abschließende
% Schätzung der Schwelle verwendet werden soll

if exclude == 1
    finalLike = cleanLike;
else
    finalLike = Like;
end

finalLike = finalLike/sum(finalLike); % Normierung auf die Fläche unter dem
% Grafen

```

```
% -----
% Berechnung des nächsten Reizstärkenwertes x_next
% -----

switch averagetype

    case 1 % Maximum-Likelihood-Methode

        % wähle den (oder die) Stimulus/-i im Maximum der Likelihood-Funktion
        x_ml = find( Like == max( Like ) );
        % wähle einen Stimulus bei Gleichwahrscheinlichkeit mehrerer 2
        % Alternativen
        x_next = round(mean(x_ml));

    case 2 % Arithmetischer Mittelwert

        x=1:range;
        x_mean = sum(x' .* Like);
        x_next = round(x_mean);
    otherwise
        error('You have choosen an invalid averaging method');

end

% -----
% Terminierung
% -----

if termtype == 0 % Terminierung wenn eine vorgegebene obere Grenze für das
                % Konfidenzintervall unterschritten wurde

% Die folgende Berechnung des aktuellen Konfidenzintervalls erfolgt wie be-
% schrieben bei Treutwein, 1995, Gl. 23 und umgesetzt im Modula-2-Code zu
% YAAP, YAAP_mod.odc, 1993/12/01, Prozedur "ProbabilityInterval"

    i_low=1; % Index für Summation entspr. des "unteren Integrals"
    i_up=range; % Index Summe "oberes Integral"
    lowerProb = 0;
    upperProb = 0;
    k = (1 - ConCoeff)/2;
    while ( ( lowerProb < k ) || ( upperProb < k ) ) && ( i_low <= i_up )
        if lowerProb < k
            lowerProb = lowerProb + finalLike(i_low);
            i_low = i_low + 1;
        end
        if upperProb < k
            upperProb = upperProb + finalLike(i_up);
            i_up = i_up - 1;
        end
    end

    CI = i_up - i_low;

    if CI <= CI_max
        termflag = 1;
    else
        termflag = 0;
    end
end
```

```
elseif termtype >= 1 % Terminierung nach einer vorgegebenen Anzahl an Trials

    if length(x_hist) == termtype
        termflag = 1;
    else
        termflag = 0;
    end
else
    error('You have chosen an invalid value for termtype');
end

% -----
% Berechnung der abschließenden Schätzung x_est
% -----
if termflag == 1
    switch averagetype

        case 1 % Maximum-Likelihood-Methode

            % wähle den (oder die) Stimulus/-i im Maximum der Likelihood- )
            % ( Funktion
            x_ml = find( finalLike == max( finalLike ) );
            % wähle einen Stimulus bei Gleichwahrscheinlichkeit mehrerer )
            % ( Alternativen
            x_est = round(mean(x_ml));

        case 2 % Arithmetischer Mittelwert

            x=1:range;
            x_mean = sum(x' .* finalLike);
            x_est = x_mean;

    end
end
```

A.2 Skripte für den Funktionstest durch Simulation

Staircase-Verfahren

```
%-----
% Simulation eines Tracks des Staircase-Verfahrens mit den Adaptionsregeln
% 1-Down/1-Up (Variante 1), 2-Down/1Up (Variante 2) oder 3-Down/1Up (Vari-
% ante 3). Halbierungen der Schrittweite erfolgen bei vorgegebenen Umkehr-
% punkten (Reversals). Die Schätzung der Schwelle erfolgt durch Mittelung
% über eine vorgegebene Anzahl letzter aufeinanderfolgender Reversals.
%
% Die 4 Stellen im Code, an welchen jeweils eine Variante auskommentiert
% werden muss, sind mit "--->" gekennzeichnet.
%
% (S.Ciba, 07/11/2008)
%-----

clc; clear all; close all;

% globale Simulationsparameter
n = 3; % Antwortparadigma
range = 25; % Wertbereich für Reizstärke
RealThres = 10; % "wahre" Schwelle (Wendepunkt der psychometrischen Funktion)
RealSpread = range/10; % "wahrer" Spread-Parameter
RealLaps = 0; % "wahre" Lapsing-Rate

% Berechnung der "wahren" psychometrischen Funktion des Beobachters
RealPsyFunc = zeros(range, 1);
for x=1:range
    RealPsyFunc(x) = 1/n + ((1-1/n-RealLaps)/(1+exp(-(x-RealThres)/RealSpread)))
    ( );
end

% Verfahrensparameter Staircase Xdown/1up (X = 1, 2, 3)
x_hist(1) = round(0.8*range); % Anfangswert der Reizstärke
nrmax = 20; % Anzahl der Reversals für die Terminierung
halvings = [2; 4; 6]; % Reversals für Halbierungen der Reizstärke
numlastr = 4; % letzte Reversals für Mittelwertbildung

Ntrials_max = 100; % zusätzliches Terminierungskriterium (Wenn das Kriterium
                    % greift und das Verfahren vorzeitig beendet wird,
                    % liefert dieses jedoch keine Schwellenschätzung. Diese
                    % muss dann "von Hand" berechnet werden).

% Initialisierungen
trial = 1;
x_est = 0;

% Simulationsdurchlauf
while (x_est==0) && (trial <= Ntrials_max)

    % Erzeuge eine Antwort des Beobachters basierend auf der aktuellen
    % Antwortwahrscheinlichkeit
    ResProb = RealPsyFunc(x_hist(trial));
    Resarray = [ones(round(ResProb*100),1); -1*ones(round((1-ResProb)*100),1)];
    Randpoint = round(rand(1)*(length(Resarray)-1))+1;
    r_hist(trial) = Resarray(Randpoint);

    % Ermittlung der Reizstärke für die nächste Darbietung
    %--->
    % 1. Variante: 1-Down/1-Up
```



```

    [x_hist(trial+1), x_est] = staircase1d1u(x_hist, r_hist, range, nrmax, 2
        & halvings, numlastr); % 1Down/1Up
    % 2. Variante: 2-Down/1-Up
    [x_hist(trial+1), x_est] = staircase2d1u(x_hist, r_hist, range, nrmax, 2
        & halvings, numlastr); % 2Down/1Up
    % 3. Variante: 3-Down/1-Up
    [x_hist(trial+1), x_est] = staircase3d1u(x_hist, r_hist, range, nrmax, 2
        & halvings, numlastr); % 3Down/1Up

    trial = trial + 1;
end

% Anzahl der benötigten Trials
Ntrials = trial-1;

% Plotting

% Psychometrische Funktion des Beobachters
figure(1)
plot(1:range, RealPsyFunc);
title('psychom. Fkt. des Beobachters');
axis([1 range 0 1]);
grid on

%Track
% Antwortvektor - trennen der positiven und negativen Antworten (für plot)
trial_pos = find(r_hist+1);
x_pos = x_hist(trial_pos); % zugehörige Reizstärkenwerte
trial_neg = find(1-r_hist);
x_neg = x_hist(trial_neg);

figure(2)
plot(1:Ntrials+1, x_hist, '-', 'Color', 'k', 'LineWidth', 1.25);
%--->
% 1. Variante (s.o.)
%title('Track - Staircase 1-Down/1-Up');
% 2. Variante (s.o.)
title('Track - Staircase 2-Down/1-Up');
% 3. Variante (s.o.)
%title('Track - Staircase 3-Down/1-Up');

hold on
posh = plot(trial_pos, x_pos, 'square', 'MarkerFaceColor', 'k', 'MarkerEdgeColor', 'k');
negh = plot(trial_neg, x_neg, 'o', 'MarkerEdgeColor', 'k');
axis([0 (Ntrials+1) 1 range]);
xlabel('Trial Number');
ylabel('Stimulus Level');
grid on
h=legend([posh, negh], 'positive response', 'negative response', 'Location', 'NorthEast');
legend('boxon');
set(h, 'FontSize', 10);
hold off

set(gca, 'Units', 'normalized', 'Position', [0.1 0.5 0.8 0.4]);

% Berechnung der "wahren" Schwelle, welche dem Konvergenzniveau der
% gewählten Adaptionsregel entspricht
%--->
% 1. Variante (s.o.)
%P_t=0.5;

```

```
% 2. Variante (s.o.)
P_t=0.707;
% 3. Variante (s.o.)
%P_t=0.794;

TrueRealThres = -RealSpread*log((1-RealLaps-1/n)/(P_t-1/n)-1) + RealThres;
text(Ntrials*0.4, -0.4*range, ['Threshold_Estimate:', num2str(x_est)]);
text(Ntrials*0.4, -0.5*range, ['Real_Threshold(adapted_to_target_probability):',
    num2str(TrueRealThres)]);
%-->
% 1. Variante (s.o.)
%print ('-dpdf', '-r600', 'track_stairc_1d1u');
% 2. Variante (s.o.)
print ('-dpdf', '-r600', 'track_stairc_2d1u');
% 3. Variante (s.o.)
%print ('-dpdf', '-r600', 'track_stairc_3d1u');
```

PEST-Verfahren

```
%-----
% Simulation eines Tracks des PEST-Verfahrens im Original (Variante 1) oder
% in der modifizierten Fassung (Variante 2) nach Findlay
%
% Die 3 Stellen im Code, an welchen jeweils eine Variante auskommentiert
% werden muss, sind mit "-->" gekennzeichnet.
%
% (S.Ciba, 07/11/2008)
%-----

clc; clear all; close all;

% globale Simulationsparameter
n = 3; % Antwortparadigma
range = 25; % Wertbereich für Reizstärke
RealThres = 10; % "wahre" Schwelle (Wendepunkt der psychometrischen Funktion)
RealSpread = range/10; % "wahrer" Spread-Parameter
RealLaps = 0; % "wahre" Lapsing-Rate

% Berechnung der "wahren" psychometrischen Funktion des Beobachters
RealPsyFunc = zeros(range, 1);
for x=1:range
    RealPsyFunc(x) = 1/n + ((1-1/n-RealLaps)/(1+exp(-(x-RealThres)/RealSpread)))
end

% Verfahrensparameter PEST/More-Virulent-PEST nach Findlay(1978)
P_t = 0.67; % Konvergenzniveau (Target-Probability)
dx_ini = 8; % Anfangswert der Schrittweite (sollte sinnvoller Weise eine
    Zweierpotenz sein)
dx_max = dx_ini; % Maximalwert der Schrittweite (vgl. dx_ini)
x_hist(1) = round(0.8*range); % Anfangswert der Reizstärke

W = 1; % Deviation-Limit für (PEST)
M = 12; % für More-Virulent-PEST

Ntrials_max =100; % zusätzliches Terminierungskriterium

% Initialisierungen
x_est = 0;
trial = 1;

% Simulationsdurchlauf
```

```

while (x_est == 0) && (trial <= Ntrials_max)

    % Erzeuge eine Antwort des Beobachters basierend auf der aktuellen
    % Antwortwahrscheinlichkeit
    ResProb = RealPsyFunc(x_hist(trial));
    Resarray = [ones(round(ResProb*100),1); -1*ones(round((1-ResProb)*100),1)];
    Randpoint = round(rand(1)*(length(Resarray)-1))+1;
    r_hist(trial) = Resarray(Randpoint);

    % Ermittlung der Reizstärke für die nächste Darbietung nach PEST
    %-->
    % 1. Variante: original Verfahren (Taylor und Creelman)
    [x_hist(trial+1), x_est] = PEST(x_hist, r_hist, range, dx_ini, dx_max, P_t, 2
    & W);
    % 2. Variante: More-Virulent-PEST (Findlay)
    [x_hist(trial+1), x_est] = mvPEST(x_hist, r_hist, range, dx_ini, dx_max, 2
    & P_t, M);

    trial = trial + 1;
end

% Anzahl der benötigten Trials
Ntrials = trial-1;

%Plotting

% Psychometrische Funktion des Beobachters
figure(1)
plot(1:range, RealPsyFunc);
title('psychom. Fkt. des Beobachters');
axis([1 range 0 1]);
grid on

%Track
% Antwortvektor - trennen der positiven und negativen Antworten (für plot)
trial_pos = find(r_hist+1);
x_pos = x_hist(trial_pos); % zugehörige Reizstärkenwerte
trial_neg = find(1-r_hist);
x_neg = x_hist(trial_neg);

figure(2)
plot(1:Ntrials+1, x_hist, '-', 'Color', 'k', 'LineWidth', 1.25);
%-->
% 1. Variante (s.o.)
title('Track - PEST (original)');
% 2. Variante (s.o.)
%title('Track - More-Virulent-PEST');

hold on
posh = plot(trial_pos, x_pos, 'square', 'MarkerFaceColor', 'k', '2
    & MarkerEdgeColor', 'k');
negh = plot(trial_neg, x_neg, 'o', 'MarkerEdgeColor', 'k');
axis ([0 (Ntrials+1) 1 range]);
xlabel('Trial Number');
ylabel('Stimulus Level');
grid on
h=legend( [posh, negh], 'positive response', 'negative resonse', 'Location','2
    & NorthEast');
legend('boxon');
set(h, 'FontSize', 10);
hold off

```

```
set(gca, 'Units', 'normalized', 'Position', [0.1 0.5 0.8 0.4]);
text(Ntrials*0.4, -0.4*range, ['Threshold_Estimate: ', num2str(x_est)]);
text(Ntrials*0.4, -0.5*range, ['Real_Threshold: ', num2str(RealThres)]);
%--->
% 1. Variante (s.o.)
print ('-dpdf', '-r600', 'track_PEST');
% 2. Variante (s.o.)
%print ('-dpdf', '-r600', 'track_mvPEST');
```

Maximum-Likelihood-/Bayes-Verfahren

```
%-----
% Simulation eines Tracks eines ML/Bayes-Verfahrens (basierend auf dem
% Best-PEST-Verfahren und einigen Erweiterungen).
%
% parametrisierbare Merkmale:
% - Initialisierung entweder per Vorgabe impliziter Trials nach Lieberman
% & Pentland (1982), per Vorgabe einer gaussförmigen (vgl. Watson & Pelli,
% 1983) oder konstanten/rechteckförmigen A-priori-Likelihood.
% - Wahl der nächsten Reizdarbietung und Schätzung der Schwelle entweder
% durch Wahl des Maximums (ML) oder des arithm. Mittelwertes der
% (normierten) Likelihood-Funktion (auch A-posteriori-Dichte) (vgl.
% King-Smith, 1994)
% - Terminierung erfolgt entweder per fest vorgegebener Anzahl an Trials oder
% per Konfidenzintervall (vgl. Treutwein, 1995, 1997).
% - Möglichkeit des Ausschlusses der A-priori-Information für die
% abschließende Schätzung der Schwelle (vgl. Watson & Pelli, 1983)
%
%
% (S.Ciba, 14/11/2008)
%-----

clc; clear all; close all;

% globale Simulationsparameter
n = 3; % Antwortparadigma (nAFC)
RealGamma = 1/n;
range = 25; % Wertebereich für Reizstärke
RealThres = 10; % "wahre" Schwelle (Wendepunkt der psychometrischen Funktion)
RealSpread = range/10; % "wahrer" Spread-Parameter
RealLaps = 0; % "wahre" Lapsing-Rate

% Berechnung der "wahren" psychometrischen Funktion des Beobachters
RealPsyFunc = zeros(range, 1);
for x=1:range
    RealPsyFunc(x) = RealGamma + ((1-RealGamma-RealLaps)/(1+exp(-(x-RealThres)/2
        ( RealSpread))));
end

% Verfahrensparameter ML/Bayes
beta = range/10; % Spread-Parameter
lambda = 0; % Lapsing-Rate
mu = 1+((range-1)/2); % Festlegung des Mittelwerts der gaussförmigen
    % a-priori Verteilung
sigma = 0.7*range; % Festlegung der Streuung der a-priori-Verteilung
ConCoeff = 0.95;
CI_max = range*0.2; % obere Schranke des Konfidenzintervalls bei Terminierung. )
    ( Aufgrund der ganzzahligen Auflösung des Range und der symmetrischen Lage )
    ( des Konfidenzintervalls sind für die ML-Methode nur ganzzahlige Werte )
    ( sinnvoll.
```

```
% ---> Verschiedene Möglichkeiten für Initialisierung, Platzierung, )
    ( Terminierung und Schätzung (s.o.)
priortype = 3; % Art der A-priori-Formung: 1 für implizite Trials (L & P, 1982) )
    ( , 2 für gaussförmige Verteilung, sonst: gleichförmige Verteilung
averagetype = 1; % Art der Berechnung des zentralen MaBes (1: Modalwert (ML), )
    ( 2: arithm. Mittelwert (Bayes))
exclude = 0; % A-priori-Information von der abschließenden Schätzung )
    ( ausschließen (=1) oder nicht (sonst)
termtype = 30; % Art der Terminierung: 0: Konfidenzintervall, >=1: vorgegebene )
    ( Anzahl an Trials, welche gleichzeitig durch termtype übergeben wird

% Simulationsdurchlauf

% Startwert beziehen
x_hist=[];
r_hist=[];
[x_hist(1), x_est, finalLike] = MLBayes(x_hist, r_hist, range, beta, lambda, n, )
    ( priortype, mu, sigma, averagetype, exclude, termtype, ConCoeff, CI_max);

% Initialisierung
trial = 0;

while x_est == 0
    trial = trial + 1;

    % Erzeuge eine Antwort des Beobachters basierend auf der aktuellen
    % Antwortwahrscheinlichkeit (vgl. Magisterarbeit S.Otto)
    ResProb = RealPsyFunc(x_hist(trial));
    Resarray = [ones(round(ResProb*100),1); -1*ones(round((1-ResProb)*100),1)];
    Randpoint = round(rand(1)*(length(Resarray)-1))+1;
    r_hist(trial) = Resarray(Randpoint);

    % Ermittlung der Reizstärke für die nächste Darbietung
    [x_hist(trial + 1), x_est, finalLike] = MLBayes(x_hist, r_hist, range, beta )
        ( , lambda, n, priortype, mu, sigma, averagetype, exclude, termtype, )
        ( ConCoeff, CI_max);

end
Ntrials = trial; % Anzahl der durchgeführten Trials

% Plotting

% Psychometrische Funktion des Beobachters
figure(1)
plot(1:range, RealPsyFunc);
title('psychom. Fkt. des Beobachters');
axis([1 range 0 1]);
grid on

% final Likelihood-Funktion
figure(2)
plot(finalLike);
title('Finale Likelihood-Funktion');

% Track
% Antwortvektor - trennen der positiven und negativen Antworten (für plot)
trial_pos = find(r_hist+1);
x_pos = x_hist(trial_pos); % zugehörige Reizstärkenwerte
trial_neg = find(1-r_hist);
x_neg = x_hist(trial_neg);
```

```

figure(3)
plot(1:Ntrials+1, x_hist, '--', 'Color', 'k', 'LineWidth', 1.25);
%--->
% 1. Variante: Best-PEST mit impliziten Trials nach L & P (1982)
%title('Track - Best-PEST (implicit trials)');
% 2. Variante: Best-PEST mit gaussförmiger A-priori-Likelihood
%title('Track - Best-PEST (gaussian prior)');
% 3. Variante Best-PEST mit gleichförmiger A-priori-Likelihood
%title('Track - Best-PEST (uniform prior)');
% Plot ohne spezifisches Label
% title('Track - ML/Bayes');

hold on
posh = plot(trial_pos, x_pos, 'square', 'MarkerFaceColor', 'k', 'MarkerEdgeColor', 'k');
negh = plot(trial_neg, x_neg, 'o', 'MarkerEdgeColor', 'k');
axis([0 (Ntrials+1) 1 range]);
xlabel('Trial Number');
ylabel('Stimulus Level');
grid on

h=legend([posh, negh], 'positive response', 'negative response', 'Location','NorthEast');
legend('boxon');
set(h, 'FontSize', 10);

hold off
set(gca, 'Units', 'normalized', 'Position', [0.1 0.5 0.8 0.4]);
text(Ntrials*0.4, -0.4*range, ['Threshold Estimate:', num2str(x_est)]);
text(Ntrials*0.4, -0.5*range, ['Real Threshold:', num2str(RealThres)]);

%--->
% 1. Variante (s.o.)
%print('-dpdf', '-r600', 'track_bestPEST_implicittrials');
% 2. Variante (s.o.)
%print('-dpdf', '-r600', 'track_bestPEST_gaussianprior');
% 3. Variante (s.o.)
%print('-dpdf', '-r600', 'track_bestPEST_uniformprior');
% Dateiname ohne spezifische Informationen
%print('-dpdf', '-r600', 'track_MLBayes');

```

Anhang B: Benutzerdokumentation der Softwarebibliothek

