

# **ADAPTIVE HÖRSCHWELLENMESSUNG**

Christian Kaernbach

*Universität Leipzig, Institut für Allgemeine Psychologie*

In der audiologischen Praxis wird noch häufig die manuelle Audiometrie eingesetzt, bei der es selbst bei gut ausgebildetem Personal noch zu Versuchsleitereffekten kommen kann. Zusätzlich kann es verfahrensbedingt zu Kriterieneffekten seitens der Patienten kommen. Automatische adaptive Verfahren zur Hörschwellenmessung, die diese Fehlerquellen vermeiden, sind in der Wissenschaft weit verbreitet, aber auf Grund fehlender apparativer Voraussetzungen in vielen Praxen nicht realisierbar. Die zunehmende Computerisierung in allen Bereichen der apparativen Medizin wird dazu führen, daß künftige Audiometer als Programm auf einem Computer realisiert sein werden. Dies bietet die Chance, auch in der audiologischen Praxis automatische adaptive Verfahren zur Hörschwellenmessung einzusetzen. Der vorliegende Beitrag stellt ein einfaches automatisches Verfahren zur adaptiven Hörschwellenmessung vor, das in dieser Form im Labor eingesetzt wird, aber auch in die audiologische Praxis Eingang finden könnte. Dabei können solche Verfahren nicht nur zur Messung von Reintonhörschwellen eingesetzt werden, sondern auch auf Sprachaudiometrie und ähnliche Felder erweitert werden.

## **EINLEITUNG**

In der klassischen Audiometrie geht es um das Vermessen der Hörschwelle mit psychophysischen Methoden. In der Praxis kommt noch meist die manuelle Audiometrie zum Einsatz. Auch bei gut ausgebildetem Personal lassen sich Versuchsleitereffekte nicht ausschließen. Der nächste Abschnitt behandelt Versuchsleitereffekte am Beispiel des „klugen Hans“. Dabei wird deutlich, daß es selbst skeptischen Wissenschaftlern unmöglich ist, non-verbale Signale zu unterdrücken, die zu einer Verfälschung des Untersuchungsergebnisses führen können. Der darauf folgende Abschnitt behandelt die von den Patienten ausgehenden Kriterienprobleme. Dann werden klassische adaptive Verfahren vorgestellt, wie sie im Labor üblich

sind, und schließlich ein besonders für die audiologische Praxis geeignetes Verfahren dargestellt und mit anderen Verfahren verglichen.

### VERSUCHSLEITEREFFEKTE: DER „KLUGE HANS“



**Abb. 1.** Der kluge Hans bei einer Demonstration seines Könnens.

Der „kluge Hans“ war ein Pferd, dem ungewöhnliche Intelligenzleistungen nachgesagt wurden. Der Berliner Mathematiklehrer Wilhelm von Osten (1838-1909) hatte das Pferd 1888 gekauft und (nebst einer Katze und einem Bär) versucht, dem Tier das Rechnen beizubringen. Beim Pferd hatte er Erfolg. Zunächst hatte er ihm die Zahlen Eins bis Neun mit Hilfe von Kegeln beigebracht, schließlich, indem er mit Kreide die Zahl an die Tafel schrieb. Das Pferd antwortete mit der richtigen Zahl von Hufklopfen. Nach zwei Jahren konnte das Pferd einfache arithmetische Aufgaben lösen. Wilhelm von Osten zog mit missionarischem Eifer mit seinem Pferd von Auftritt zu Auftritt. Er nahm kein Geld für die Vorführungen.

Schnell sprachen sich die Fähigkeiten des Pferdes herum. Die Wissenschaftler blieben skeptisch, konnten aber während über 15 Jahren keinen Betrug nachweisen. Hatte das Pferd anfänglich Schwierigkeiten, Aufgaben in Abwesenheit seines Herrn zu lösen, so konnte es bald auch ohne ihn die gestellten Fragen korrekt beantworten.

Im Jahre 1904 stellte der Direktor des Berliner Psychologischen Instituts, Professor Carl Stumpf eine 13-köpfige Kommission zusammen, die das Phänomen untersuchen sollten. Die Zusammensetzung der Kommission hatte er sehr breit angelegt: ein Zirkustrainer, ein Zoologe, ein Tierarzt, und auch ein Politiker waren darunter. Stumpf ging von der Annahme aus, daß, wenn Betrug vorläge, es in einer derartig breit angelegten Kommission jemanden gäbe, der dies herausfinden würde. Die Kommission konnte das Rätsel nicht lösen.

Erst sein Student Oskar Pfungst löste das Rätsel (Pfungst, 1907). Das Pferd war in der Tat ungewöhnlich begabt, wenn auch anders, als sich das sein Besitzer vorstellte. Es reagierte auf unwillentlich gegebene nonverbale Signale seitens der Zuhörer. Diese mußten nicht „mitspielen“, ja, konnten sogar der Sache skeptisch gegenüberstehen: wenn das Pferd die kritische Zahl von Hufschlägen erreichte, richteten sich die Zuhörer unmerklich ein klein wenig auf. Das reichte für das Pferd als Signal, mit dem Klopfen aufzuhören. Erst wenn kein Zuschauer die korrekte Lösung kannte, war das Pferd nicht mehr in der Lage, die richtige Antwort zu geben. Pfungst übte diese nonverbalen Signale, bis er in der Lage war, das Pferd Zahlen klopfen zu lassen, zu denen es gar keine Aufgabe bekommen hatte. Die „richtige“ Zahl war eben nur Pfungst bekannt.

Nonverbale Kommunikation zwischen Versuchsleiter und Versuchsperson kann bei Messungen aller Art das Ergebnis verfälschen: Wenn der Versuchsleiter ein bestimmtes Ergebnis erwartet, kann sich dieses Ergebnis auch einstellen, selbst wenn die Tatsachen ganz anders sind.

#### Das gilt selbstverständlich auch für die Hörschwellenmessung:

Wenn der Versuchsleiter eine bestimmte Form der Hörschwelle erwartet, sei es, weil im Rahmen einer Verlaufskontrolle eine Besserung erwartet wird, sei es, weil ein berufstypisches Audiogramm erwartet wird, dann kann diese Erwartung von sensiblen Patienten unwillentlich ausgewertet werden und sich – evtl. entgegen den tatsächlichen Gegebenheiten – das erwartete Audiogramm einstellen. Dies wird erleichtert durch den Umstand, daß bei der manuellen Audiometrie eine Variante der Ja/Nein-Aufgabe eingesetzt wird, was zu den bekannten Kriterienproblemen seitens der Patienten führt (siehe nächster Abschnitt).

Man glaube ja nicht leichtfertig, Versuchsleitereffekte bewußt kontrollieren zu können. Wilhelm von Osten war kein Betrüger - er glaubte fest an die Fähigkeiten seines Pferdes. Auch die Kommission unter Leitung des angesehenen Leiter des Berliner Psychologischen Instituts, Professor Carl Stumpf, war trotz aller kritischer Grundhaltung nicht davor gefeit, als wissen-

de Versuchsleiter die Ergebnisse des Versuchs zu beeinflussen. Oskar Pfungst hatte erst nach ausführlichem Training seine nonverbalen Signale so unter Kontrolle, daß er falsche Botschaften vermitteln konnte. Das beweist nicht, daß es ihm möglich gewesen wäre, gar keine Botschaft zu übermitteln: Techniken zum „Austricksen“ des Polygraphen („Lügendetektors“) basieren nicht auf dem Unterbleiben einer Erregung an kritischen Stellen, sondern auf dem willentlichen Hervorrufen von Erregungen an unkritischen Stellen. Man kann also wohl mit einer Flut irrelevanter nonverbaler Signale die ungewollten nonverbalen Signale überdecken, setzt damit aber die Messung einer Fülle von nicht kontrollierten Einflüssen aus.

Die manuelle Audiometrie ist nicht leicht vor Versuchsleitereffekten zu schützen. Eine Mindestmaßnahme wäre, den Patienten den Sichtkontakt zum Versuchsleiter (vgl. Abbildung 2) zu nehmen. Aber selbst dann ist das Einpegeln der dargebotenen Intensitäten eine andere Art der Kommunikation, und hier die Erwartungshaltung des Versuchsleiters auszublenden ist nicht einfach. Ein besserer Ansatz wäre es, das Einpegeln der Intensitäten automatisch vorzunehmen. Das ist der Ansatz der adaptiven Verfahren, die in diesem Beitrag vorgestellt werden. Aber zunächst müssen wir uns einem ebenso wichtigen Problem widmen: Kriterieneffekten, die von den Patienten ausgehen können.



**Abb. 2.** Sichtkontakt bei der klassischen manuellen Audiometrie.

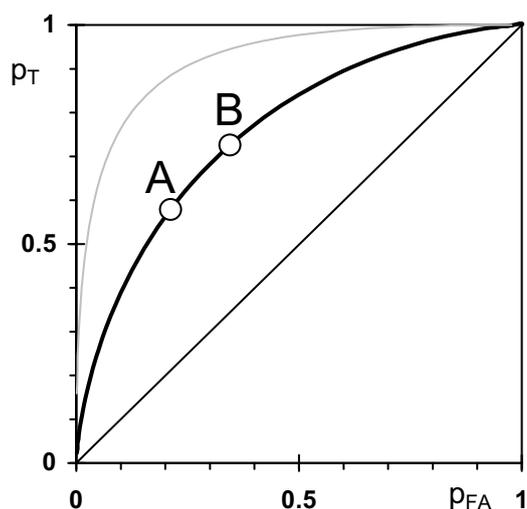
### **KRITERIENEFFEKTE BEI JA/NEIN-AUFGABEN: DIE RECEIVER OPERATING CHARACTERISTICS (ROC)**

Es ist in der Psychophysik seit langem bekannt, daß bei der Entscheidung für die Antwort „Ja“ („ich habe das Signal gehört“) oder „Nein“ („ich habe das Signal nicht gehört“) nicht nur sensorische Einflüsse (Hörvermögen), sondern auch sogenannte strategische Faktoren (man könnte sagen: das „Hörenwollen“) eine große Rolle spielen. Diese Zusammenhänge sind gut untersucht in der sogenannten Signalentdeckungstheorie (*Signal Detection Theory*, SDT, siehe Green & Swets, 1966, Macmillan & Creelman, 1991, Kaernbach, 1991b).

Die strategischen Faktoren rühren daher, daß die Versuchsperson sich ein Kriterium zurechtlegen muß, ab wann sie mit „Ja“ antwortet. Denn jede Situation ruft ein unterschiedliches Ausmaß von Sicherheit hervor, ob das Signal wirklich da war oder nicht. Die Wahrscheinlichkeit für die Antwort „Ja“ wird dabei stark von der Wahl dieses Kriterium abhängen. Abbildung 3 erläutert dies anhand der *Receiver Operating Characteristics* (ROC).

Versuche zur SDT werden in der Regel mit 50% Signalversuchen und 50% Kontrollversuchen durchgeführt. In der Praxis der Hörschwellenmessung liegen meistens ausschließlich Signalversuche vor: nur selten wird der Versuchsleiter die Angaben des Patienten überprüfen, indem gar kein Signal eingespielt wird. Aber auch ohne Kontrollversuche wird dessen Kriterium die Wahrscheinlichkeit für die Antwort „Ja“ und damit den gefundenen Schwellenwert drastisch beeinflussen. Ohne Kontrollversuche fehlt aber jegliche Aussage über das Kriterium des Patienten, und so ist es nicht möglich, diese Einflüsse herauszurechnen.

Dabei liegt seitens der Versuchsperson bzw. des Patienten keine Täuschungsabsicht vor. Selbst bei bester Kooperationsbereitschaft können Kriterieneffekte nicht vermieden werden. Denn man kann ja nicht auf das Festlegen eines Kriteriums verzichten: für irgendein Kriterium muß man sich entscheiden. Dies wird sogar oft von seiten der Versuchspersonen als unangenehm und belastend empfunden: es ist nicht einfach, so ein Kriterium zu finden und einigermaßen konstant zu halten. Und es ist völlig unmöglich, es „objektiv“ zu wählen, so daß Daten über Versuchspersonen hinweg bzw. zwischen Patienten vergleichbar wären. Verfahren, die in die audiologische Praxis Eingang finden sollen, müssen also nicht nur durch Automatisierung Versuchsleitereffekte vermeiden, sondern auch durch Auswahl einer geeigneten Aufgabe Kriterienprobleme seitens der Patienten ausschließen.



**Abb. 3.** *Receiver Operating Characteristics (ROC).* Der Versuchsperson wird in der Hälfte der Fälle ein schwaches Signal geboten, in der anderen Hälfte der Fälle kein Signal (Kontrollversuche). Sie muß antworten, ob sie das Signal wahrgenommen hat oder nicht (Ja/Nein-Aufgabe). Dargestellt ist die Wahrscheinlichkeit, bei Vorliegen eines Signals mit „Ja“ zu antworten (Treffer-Wahrscheinlichkeit  $p_T$ ) über der Wahrscheinlichkeit, bei Kontrollversuchen mit „Ja“ zu antworten (Falsch-Alarm-Wahrscheinlichkeit  $p_{FA}$ ). Eine positive Leistung liegt vor, wenn  $p_T$  größer ist als  $p_{FA}$ , der Datenpunkt also links oberhalb der Diagonalen liegt. Das ist bei Punkt A der Fall. Aber auch Punkt B liegt oberhalb der Diagonalen, und kann von der Versuchsperson bei gleicher Signalstärke und Hörvermögen erreicht werden. Er entspricht einer anderen strategischen Einstellung, einer größeren Hörbereitschaft, mit mehr Treffern, aber auch mehr falschen Alarmen. Die Gesamtheit der Punkte, die von der Versuchsperson bei dieser Signalstärke und diesem Hörvermögen erzielt werden können, liegt auf der schwarz eingezeichneten ROC-Kurve. Bei einem stärkeren Signal oder besserem Hörvermögen würde das Verhalten der Versuchsperson durch die grau eingezeichnete höhere ROC-Kurve wiedergegeben werden. Entscheidend dafür, ob die Versuchsperson Punkt A oder irgend einen anderen Punkt auf der für dieses Signal gültigen ROC-Kurve einstellt, ist das Kriterium, anhand dessen die Versuchsperson entscheidet, wann sie „Ja“ sagen soll.

### **KLASSISCHE ADAPTIVE VERFAHREN IM LABOR**

Das einfachste und älteste adaptive Verfahren ist das 1946 vorgestellte simple up-down Verfahren von George von Békésy. Dabei wird dem Probanden ein Ton einer bestimmten Intensität vorgespielt, und er zeigt an, ob er diesen gehört hat oder nicht. Bei einer positiven Antwort wird die Intensität um eine Stufe reduziert, bei einer negativen Antwort um eine Stufe erhöht. Nach einer bestimmten Anzahl von Laufrichtungsänderungen bricht man das Verfahren ab und ermittelt diejenige Intensität, bei der mit einer Wahrscheinlichkeit von 50% angegeben wird, daß der Ton gehört wird.

Bei diesem Verfahren wird wie auch bei der manuellen Audiometrie eine Ja/Nein-Aufgabe eingesetzt. Dies führt zu einer großen Variabilität der Ergebnisse zwischen verschiedenen Versuchstagen und zwischen Versuchspersonen. Es ist also nicht ausreichend, das Einpegeln der dargebotenen Intensitäten automatisch vornehmen zu lassen. Zwar sind dann Versuchsleitereffekte ausgeschlossen, aber die Erwartung der Patienten kann immer noch einen großen Einfluß ausüben.

Um Kriterieneffekte zu vermeiden, werden heute in adaptiven Verfahren meist sogenannte kriterienfreie Aufgaben eingesetzt. Die häufigste Realisierung von Aufgaben diese Typus sind 2IFC (2-intervals forced-choice) Aufgaben, bei denen die Versuchsperson sich entscheiden muß, in welchem von zwei Intervallen der Ton war. Es wird also z.B. ein Licht eingeschaltet, das zunächst grün und dann rot leuchtet. Der Ton wird entweder zum grünen oder zum roten Licht gegeben, und die Versuchsperson muß entscheiden, bei welchem Licht der Ton zu hören war. Sie braucht dazu kein Kriterium zu bilden, sondern kann die beiden Intervalle gegeneinander vergleichen. Das bedeutet in der Regel bereits einen Komfortgewinn für die Versuchsperson.

Der Terminus "forced choice" bezieht sich dabei darauf, daß die Versuchsperson sich zwischen den beiden Intervallen entscheiden muß, auch wenn sie meint, das Signal gar nicht (oder in beiden Intervallen) gehört zu haben. Dabei erzielt sie selbst bei vollkommener Unsicherheit über die Antwort eine Trefferwahrscheinlichkeit von mindestens 50%, bei guter Hörbarkeit des Signals dagegen von 100%. Ziel des adaptiven Verfahrens muß es nun sein, denjenigen Intensitätswert zu bestimmen, bei dem mit einer Wahrscheinlichkeit von 75% die richtige Antwort gegeben wird.

Eine einfache Variation des simple up-down Verfahrens, das diesem Aufgabentypus angepaßt ist, stellt das weighted up-down Verfahren (Kaernbach, 1991a) dar. Dabei wird die Intensität nach jeder richtigen Antwort um eine Stufe gesenkt, und nach jeder falschen Antwort um drei Stufen erhöht. Man kann sich leicht klarmachen, daß dieses Verfahren einen Inten-

sitätswert einstellen wird, bei dem es zu 75% richtigen Antworten kommt: wenn auf drei richtige Antworten eine falsche Antwort kommt, wird die Intensität nicht verändert. Nach einer gewissen Anzahl von Versuchen wird meist der Mittelwert über die dargebotenen Intensitäten (unter Ausschluß der Anfangsphase, wo meist überschwellige Intensitäten dargeboten wurden) als Schätzwert für die Hörschwelle genommen.

Die Kombination von 2IFC und weighted up-down Verfahren ist eine zuverlässige und laborerprobte Methode, die Schwellenintensität zu ermitteln. Sie vermeidet sowohl Versuchsleitereffekte als auch Kriterienprobleme. Für den Einsatz in der audiologischen Praxis sollte sie allerdings modifiziert werden. Der folgende Abschnitt beschreibt ein adaptives Verfahren, daß speziell an die Bedürfnisse der Audiologie angepaßt wurde. Im Anschluß daran wird ein Experiment beschrieben, bei dem dieses Verfahren mit zwei anderen audiometrischen Verfahren verglichen wird.

### **EIN ADAPTIVES VERFAHREN FÜR DIE AUDIOMETRIE**

Ein Verfahren, das sich im Labor bewährt hat, muß noch lange nicht gut für den Einsatz in der audiologischen Praxis sein. Zwei Eigenheiten klassischer Laborverfahren zur adaptiven Schwellenmessung bedürfen der Anpassung an die Bedürfnisse des klinischen Alltags:

- Das Vergleichen von zwei Intervallen fordert den Patienten eine hohe Aufmerksamkeit und Koordinationsleistung ab.
- Das Verfahren der „erzwungenen Wahl“ (forced choice) ist in der klinischen Praxis kaum vermittelbar.

Darüber hinaus sollte ein für die Praxis taugliches Verfahren wenigstens ansatzweise auf die Problematik von Meßfehlern und Ausreißern eingehen. Diese Aspekte werden im folgenden einzeln behandelt.

#### **Eine einfache kriterienfreie Aufgabe**

Fordert man einen Patienten zu erhöhter Konzentration auf eine auditive Aufgabe auf, wird er nicht selten die Augen schließen. Das ist für ein 2IFC Experiment ungünstig, denn die visuelle Modalität wird oft genutzt, um die Dauer der beiden Intervalle zu kennzeichnen, z.B. über die Farbe eines Lichtes. Aber auch der Vergleich der beiden Intervalle stellt für Personen mit geringer Gedächtnisspanne bereits ein Problem dar. Daher ist es wünschenswert, daß eine Aufgabe für ein automatisches adaptives Verfahren zwar kriterienfrei ist, dabei aber doch mit nur einem Intervall auskommt.

In unseren Untersuchungen haben wir daher folgende Aufgabe verwendet: der Proband muß entscheiden, ob ein kurzer (200 ms) oder ein langer (600 ms) Ton vorlag. In Verbindung mit einer sinnfälligen Tastatur (mit einer kurzen und einer langen Taste) hat sich herausgestellt, daß diese Aufgabe gut vermittelbar ist.

### **Die Möglichkeit, seiner Unsicherheit Ausdruck zu verleihen**

Eine forced-choice Aufgabe ist im Labor problemlos einsetzbar. Man erklärt der Versuchsperson, daß ihre Werte auch dann, wenn sie eigentlich zu unsicher ist, um eine Antwort zu geben, häufig immer noch besser als zufällig sind, und daß die Auswertung dieser Daten für den Versuchsleiter wichtig ist. In der audiologischen Praxis ist es den Patienten hingegen schwierig zu vermitteln, daß sie auch dann eine Antwort geben müssen, wenn sie nichts gehört haben. Ganz abgesehen von dem Erklärungsaufwand, der nicht in jeder Situation leistbar ist, kann es trotz guter Erklärungen seitens des Versuchsleiters zu Vorbehalten beim Patienten kommen. Es kann der Eindruck entstehen, diese Untersuchung sei nicht seriös, da sie auf dem (erzwungenen) Raten des Patienten basiere, und daher komme es auch gar nicht so sehr auf eine konzentrierte Mitarbeit an. Nicht zuletzt kann es dabei sogar zu einem Vertrauensverlust gegenüber dem Arzt kommen.

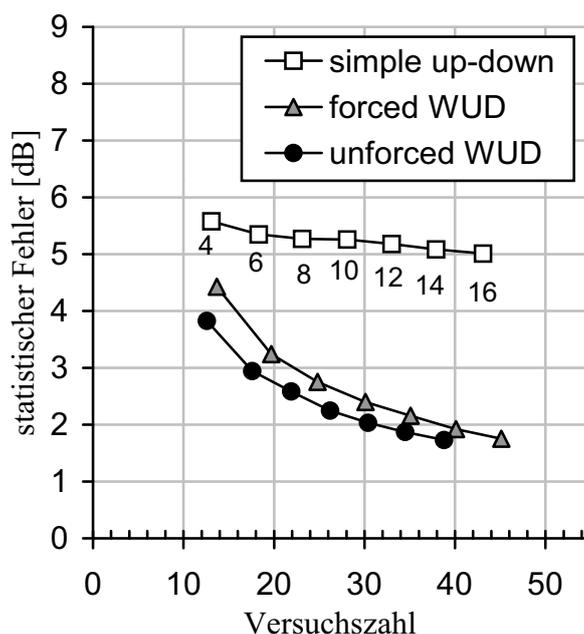
Daher wurde ein Verfahren auf der Basis einer unforced-choice Aufgabe entwickelt (Kaernbach, 2001), das es zuläßt, daß die Probanden anzeigen, daß sie sich nicht für eine der beiden Alternativen entscheiden möchten. Obwohl für die Klinik entwickelt, hat sich dieses unforced weighted up-down genannte Verfahren auch im Labor als theoretisch und empirisch anderen klassischen Labor-Verfahren überlegen erwiesen (siehe Abbildung 4). In unserem Labor verwenden wir diese neue Methode nun fast ausschließlich, wenn es um das Messen von Wahrnehmungsschwellen geht.

### **Heuristiken zum Erkennen von Ausreißern**

Im Labor ist es nicht tragisch, wenn einmal ein Meßwert auf Grund statistischer Streuung weit vom tatsächlichen Schwellenwert abweicht. Solche Abweichungen treten meist in beide Richtungen auf und werden durch die Mitteung über Versuchspersonen aufgefangen.

In der audiologischen Praxis ist es hingegen untragbar, wenn z. B. die Anpassung einer Hörhilfe oder das Verschreiben eines Medikamentes von solchen Meßfehlern abhängen. Daher ist es unerläßlich, zusätzlich zu den Meßdaten weitere Kriterien einfließen zu lassen. Ein ganz wichtiges Krite-

rium ist dabei die Erfahrung des Versuchsleiters, der Ausreißer unter den Meßwerten häufig daran erkennt, daß sie nicht plausibel sind.



**Abb. 4.** Experimenteller Vergleich von drei automatischen adaptiven Verfahren im Labor: simple up-down (Békésy), forced weighted up-down und unforced weighted up-down (Kaernbach, 2001). Mit zunehmender Versuchszahl wird für alle drei Verfahren der statistische Fehler kleiner. Dabei erreicht allerdings das unforced weighted up-down Verfahren bei gleicher Versuchszahl die kleinsten statistischen Fehler.

Ein automatisches Verfahren kann Plausibilitätsprüfungen nur ansatzweise durchführen. Das sollte es aber, schon um den Versuchsleiter dazu zu ermutigen, seinen eigenen Sachverstand einzusetzen. Ansonsten könnte der Nimbus eines vom Computer gesteuerten Verfahrens dazu führen, daß die so ermittelten Werte unkritisch übernommen werden. Daher wurde als zusätzliche Maßnahme ein System von Heuristiken entwickelt, das nach Abschluß aller Messungen mit gewissen Plausibilitätskriterien versucht, besonders unplausible Meßwerte zu ermitteln, und diese ein zweites Mal vermeißt.

## VERGLEICH VON AUTOMATISCHER UND MANUELLER AUDIOMETRIE

### Methodik

Bei 88 Studenten wurden je 6 Audiogramme erhoben, und zwar je zwei Audiogramme für je drei Methoden. Jedes Audiogramm bestand aus 14 Meßwerten, wobei auf jedem Ohr die Schwellen bei 1, 2, 3, 4, 6, 8 und 0.5 kHz in dieser Reihenfolge gemessen wurden. Die drei Verfahren waren:

- manuelle Audiometrie mit dem Audiometer AT409 der Firma Auritec und mit studentischen Versuchsleitern, die eine Woche lang auf einer audiometrischen Station eingewiesen worden waren,
- ein automatisches Grenzverfahren, das im AT409 implementiert ist,
- sowie das oben beschriebene unforced weighted up-down Verfahren, das auf einem PC implementiert wurde.

Die Audiogramme wurden in 6 Wochen erhoben (ein Audiogramm pro Woche), wobei für jeden Teilnehmer eine bestimmte Reihenfolge der drei Verfahren festgelegt wurde (alle 6 möglichen Permutationen wurden gleich häufig eingesetzt). Diese Reihenfolge wurde in den ersten drei Wochen angewandt, und in den folgenden drei Wochen wiederholt.

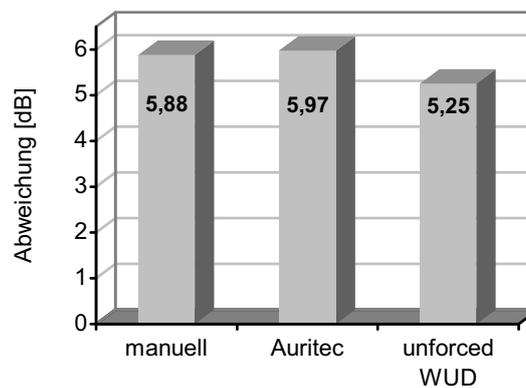
Beim unforced weighted up-down Verfahren wurde eine Messung nach 8 Laufrichtungswechseln beendet. Bis zum ersten Laufrichtungswechsel betrug die Schrittweite 10 dB, danach 5 dB. Bei falschen Antworten wurde ein auditives Feedback gegeben. Die dritte Antwortmöglichkeit wurde mit einem Timeout realisiert. Gab der Proband drei Sekunden lang keine Antwort, so wurde angenommen, daß er auch keine Antwort geben wollte, und es wurde eine Intensitätsanpassung von einem Schritt nach oben (das Mittel der Intensitätsanpassungen nach einer richtigen und nach einer falschen Antwort) vorgenommen. Die Dauer des Timeout wurde im Verlaufe des Verfahrens adaptiv an das Reaktionszeitmuster des Probanden angepaßt: Sie betrug 150% der mittleren Reaktionszeit.

### Ergebnisse

Als Reliabilitätsparameter wurde ermittelt, wie genau die Audiogramme der jeweils ersten Sitzung in der zweiten Sitzung reproduziert werden konnten. Dazu wurde die Wurzel aus der mittleren quadratischen Abweichung gebildet, ein Wert, der in dB angibt, wie hoch der Erwartungswert für die Abweichung ist. Abbildung 5 zeigt die Durchschnittswerte der Abweichung für die drei Verfahren. Dabei wurden beim unforced weighted up-down Ver-

fahren zunächst keine Heuristiken eingesetzt. Dennoch erweist sich dieses Verfahren als reliabler als die beiden anderen Verfahren.

Für ein automatisches Verfahren ist es wichtig, ob dieses mit den unterschiedlichen Reaktionszeitmustern der Probanden zurecht kommt. Das lässt sich an der Zahl der Abbrüche ablesen. Von 176 Audiogrammen pro Verfahren wurde das unforced weighted up-down Verfahren 2-mal abgebrochen, das automatische Grenzverfahren der Firma Auritec 22-mal. Die manuelle Audiometrie hingegen kam immer zu einem Ergebnis. Die Abbrüche beim unforced weighted up-down Verfahren sind auf eine Probandin mit multipler Sklerose zurückzuführen, die mit der Tastatur nicht zurecht kam. In diesem Fall ist sicherlich eine manuelle Audiometrie angebracht.



**Abb. 5.** Abweichung zwischen erster und zweiter Messung für drei verschiedene Verfahren. Der Unterschied zwischen unforced weighted up-down und den anderen beiden Verfahren ist signifikant. Diese unterscheiden sich hingegen nicht signifikant.

Die Dauer der Messung betrug 5 Minuten für die ersten beiden Verfahren, und 12 Minuten für das unforced weighted up-down Verfahren. Hinzu kamen bei diesem Verfahren noch vorab 3 Minuten automatischer Instruktion über den Computer.

In einer zweiten Messung mit 78 Studenten wurde die Auswirkung von Heuristiken auf das unforced up-down Verfahren studiert. Bei der Auswertung der Daten wurden unplausible Werte automatisch und zum Teil neu vermessen. Dabei wurden als Kriterien verwandt:

- Eine Abweichung vom Normaudiogramm von mehr als 30 dB,
- eine Abweichung von den benachbarten Frequenzen bzw. von der entsprechenden Frequenz auf dem anderen Ohr von mehr als 20 dB,
- sowie eine Überschreitung der üblichen Tracklänge um mehr als 25%.

Für jeden dieser Fälle wurden Unplausibilitätswerte vergeben, und diese wurden addiert. Bis zu drei der so ermittelten unplausibelsten Meßwerte wurden bei Überschreiten einer gewissen Schwelle automatisch wiederholt. Die Kriterien sowie deren Gewichte und die Schwelle wurden zunächst ad hoc festgelegt. Dabei ergab sich eine Reduzierung der Abweichung zwischen Erst- und Zweitmessung um 22% bei gleicher Versuchsdauer.

## DISKUSSION

Das unforced weighted up-down Verfahren erweist sich als eine stabile und zuverlässige automatische adaptive Methode, die in der audiologischen Diagnostik zum Einsatz kommen kann. Sie ist seitens der Patienten kriteriums-unabhängig, frei von Versuchsleitereffekten, und spart Versuchsleiterzeit, da dieser die Patienten nur zum Audiometer führen und die Instruktion starten muß. Eine gute Ausbildung als Audiometrist wird damit keineswegs überflüssig, da in besonders gelagerten Fällen auf die manuelle Audiometrie nicht verzichtet werden kann, und da die audiometrische Expertise weitaus mehr umfaßt als das korrekte Erfassen von Routine-Audiogrammen. Ein solches Verfahren kann als Ergänzung und Entlastung der Audiometristin für unproblematische Routine-Audiogramme angesehen werden und stellt ihre Zeit frei für andere, anspruchsvollere Aufgaben. Gleichzeitig werden Routine-Audiogramme in einer sehr zuverlässigen Weise erhoben.

Ein wichtiger Aspekt bei einem in die Klinik einzuführenden Verfahren ist der Komfort für den Patienten. Dieser ist beim unforced weighted up-down Verfahren hoch im Vergleich zu den anderen getesteten Laborverfahren. So beklagen sich die Versuchspersonen beim simple up-down (Békésy) Verfahren über die Willkür, die sie beim Festlegen des Kriteriums walten lassen können. Beim forced-choice Verfahren beklagen sie sich über den Zwang zur Entscheidung, auch wenn keine Entscheidungsgrundlage gegeben ist. Beim unforced weighted up-down Verfahren haben sie hingegen die Antwortmöglichkeiten, die sie sich wünschen: Eine Aufgabe, an der sie sich selbst testen können, aber auch die Möglichkeit, sich dieser Aufgabe bei mangelnder Entscheidungsgrundlage zu entziehen.

Als Nachteil des Verfahrens kann der erhöhte Zeitaufwand für die Patienten gesehen werden. Zukünftige Bemühungen werden dahin gehen, das Verfahren weiter zu optimieren, um einen geringeren Zeitbedarf zu erzielen. So sollen z. B. die ad hoc gewählten Parameter der Heuristiken anhand der gewonnenen Daten optimiert werden, um geringere Fehler bei gleichem Zeitaufwand, bzw. den gleichen Fehler mit weniger Zeitaufwand zu erzielen. Außerdem soll durch Hinzunahme einer dritten Taste („ich weiß es nicht“) als von den Patienten wählbare zusätzliche Alternative zum Timeout eine Beschleunigung erreicht werden. Als Ziel wird angestrebt, weniger als 6 dB Fehler in max. 10 Minuten Meßzeit zu erreichen. Im übrigen muß eine längere Patientenmeßzeit nicht notwendig als Mangel angesehen werden, da die Patienten evtl. die Praxis mit dem berechtigten Gefühl verlassen, besonders gründlich untersucht worden zu sein. Außerdem steht es dem Versuchsleiter frei, durch die Wahl der Parameter des Verfahrens einen größeren Fehler in Kauf zu nehmen, um Patientenzeit zu sparen, oder ein besonders genaues Audiogramm aufzunehmen und dabei etwas mehr Patientenzeit zu investieren. Bei einer voll ausgebuchten audiometrischen Abteilung spielt neben der Patientenzeit natürlich auch die Apparatezeit eine Rolle.

Die relativ geringen Unterschiede zu dem mit Versuchsleitereffekten behafteten und vom Patientenkriterium abhängigen manuellen Verfahren, die bei dieser Untersuchung zu Tage traten, können an der hohen Motivation der beteiligten studentischen Versuchspersonen gelegen haben. Diese, sowie die ebenfalls hochmotivierten Versuchsleiter, wollten wissenschaftlich genaue Audiogramme erstellen. Hinzu kam das Fehlen einer Fremdmotivation (z. B. das „Hörenwollen“ von Patienten als Nicht-Wahr-Haben-Wollen entgegen des tatsächlichen, eingeschränkten Hörvermögens), die einer objektiven Erhebung der Hörschwelle hätte entgegenstehen können. Daher ist es ein wichtiges Anliegen, das Verfahren in der audiologischen Praxis an Patienten zu testen. Dazu müssen allerdings erst die apparativen Voraussetzungen geschaffen werden.

Ein besonderer Vorteil eines computergestützten automatisierten Verfahrens ist die Erweiterbarkeit auf andere Arten audiologischer Tests. So kann das Verfahren leicht auf Sprachaudiometrie erweitert werden, aber auch Tests zur Zeitverarbeitung, zur Diskrimination (z.B. Tonhöhen oder Lautstärken), zur Stereoortung, oder zum auditiven sensorischen Gedächtnis lassen sich mühelos in die Architektur eines solchen Computertests einbauen. Dies eröffnet neue Perspektiven der Diagnostik, die es gestattet, relevante behaviorale Daten in unkomplizierter Art und Weise über ein vertrautes Interface zu erheben.

### DANKSAGUNG

Ich möchte mich bedanken bei Frank Neutzler, der das Audiometerprogramm geschrieben hat, bei Susanne Lamm und Thomas Rigotti, die als studentische Versuchsleiter hervorragende Arbeit geleistet haben. Mein besonderer Dank gilt der Firma Auritec GmbH, Hamburg, die die hier vorgestellten Untersuchungen unterstützt hat.

### LITERATURVERZEICHNIS

- Green, D. M., Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Kaernbach, C. (1991a). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49, 227-229.
- Kaernbach, C. (1991b). Poisson signal detection theory, *Perception & Psychophysics*, 50, 498-506.
- Kaernbach, C. (2001). Adaptive threshold estimation with unforced-choice tasks. *Perception & Psychophysics*, 63, 1377-1388.
- Pfungst, O. (1907). *Das Pferd des Herrn von Osten (Der Kluge Hans). Ein Beitrag zur experimentellen Tier- und Menschen-Psychologie*. Leipzig: Johann Ambrosius Barth.
- Macmillan, N. A., Creelman, C. D. (1991) *Detection theory: A user's guide*. New York: Cambridge University Press.