

6. Diskussion

Mit Verwendung des WHF Juniors lässt sich ein erhebliche Zeitreduktion erzielen. Die Verminderung der Untersuchungszeit um mehr als die Hälfte erleichtert es den Kindern, bis zum Ende konzentriert in diesem Test mitzuarbeiten. Diese Verringerung ist auf die Verwendung weniger Reize zurückzuführen. Geller (16) erwähnte 1984 zur optimalen Zeitausnutzung in der klinischen Routine eine minimale Präsentation von 5 Reizen. In diesem Falle wurden 5 Reize pro Frequenz skaliert.

6.1 Erwachsene

6.1.1 Pegellautheitsfunktionen bei erwachsenen Probanden

In der vorliegenden Untersuchung sind Steigungswerte der Pegellautheitsfunktionen im Bereich von 0,52 dB/KU sowohl in der konventionellen als auch in der verkürzten Juniorversion über die verschiedenen Frequenzen gemittelt gefunden worden. Diese Werte sind mit den in der Literatur beschriebenen Steigungen vergleichbar. Moser (58) beschreibt Steigungen zwischen 0,47 dB/KU und 0,57 dB/KU (**Tabelle 8**). Auch Launer et al. (48) berichten 1994 von einer Steigung der Pegellautheitsfunktion bei Normalhörenden von $m = 0,53 \pm 0,13$. Meister et al. (55) beschreiben 1996 Steigungen bei Normalhörenden um 0,5 KU/dB. 1997 finden sie Steigungen von 0,41 – 0,43 dB/KU (500 und 2000Hz). Im Vergleich dazu finden sich in den von der Oldenburger Gruppe entwickelten Kindertests Steigungswerte von 0,45 - 0,49 dB/KU (53).

Die Steigungen im kurzen wie im langen Test reproduzieren die in der Literatur beschriebenen Ergebnisse. Im einzelnen existieren Unterschiede zwischen den beiden Testversionen, gekennzeichnet durch eine geringfügige Abweichung in der Steilheit der Funktion, die aber keine systematisch andere Bewertung der Einzelwerte der neuen Testversion erkennen lassen. In Anbetracht der interindividuellen Schwankungen liegen die Abweichungen im Toleranzbereich. Auch Meister (56) 1997 findet beim Vergleich der Ergebnisse von Lang- und Kurztest (Oldenburger Kindermethode) im Kurztest steilere Pegellautheitsfunktionen.

Die höhere Bewertung der Einzelantworten in einigen Frequenzen wird nicht als Lerneffekt erklärt. Lerneffekte führen nach Kießling (40) 1993 zu signifikant höheren Skalierungen der Einzelwerte im zweiten Durchgang, da eine Übung des Versuchsablaufes zu einer Verbesserung der Skalierungsqualität führt. Zu dem finden Meister et al. (54) 1995 keine

Lerneffekte bei vorheriger Orientierung der Probanden über den zu erwartenden Messbereich. Nach Heller (27) 1985 weist auch Hohmann (30) 1993 darauf hin, dass die Orientiertheit einen hoch signifikanten Einfluss auf das Antwortverhalten der Patienten hat. Bei orientiertem Verfahren ergibt sich eine deutlich geringere Standardabweichung der Steigung der Pegellautheitsfunktion (31). Die Probanden verteilen die angebotenen Pegel auf die gesamte Skala. Auch Kießling (37) 1996 erwähnt diesen Punkt.

Die in der vorliegenden Untersuchung belegten niedrigeren Bewertung der Einzelwerte in der kurzen Messung sind durch die Lerneffekte nicht zu erklären, da die Teilnehmer vorher über den Untersuchungsbereich informiert worden sind. Zudem ist die Reihenfolge der beiden Testverfahren randomisiert worden. Auch eine Erklärung durch einen Sequenzeffekt ist fraglich. Bei Untersuchungen im freien Schallfeld zeigte sich 1995 bei Hellbrück et al. (26) ein deutlicher Sequenzeffekt, d.h., die Skalierung war abhängig von der Lautstärke des vorhergehenden Rauschens. Heller (28) 1991 postulierte deshalb keine Folgen von Tonhöhen und Reizen mit einer ähnlichen Lautheitsempfindung hintereinander zu präsentieren, um das absolute Urteilen nicht mit dem Lautheitsvergleich zu vermischen. Für die Replizierbarkeit der Messung wäre aber die gleiche Reihenfolge der Stimuli erforderlich, um die Sequenzeffekte zu eliminieren. Bei der Auswahl der Pegel ist, wie in der Methodik erwähnt, auf große und sehr geringe Pegelschwankungen bei direkt aufeinander folgenden Pegeln verzichtet worden.

6.1.2 Varianz der Einzelantworten der erwachsenen Probanden

Bei Erwachsenen zeigen sich bei der langen Version Standardabweichungen zwischen 1,64 KU und 7,99 KU bei den untersuchten 7 Frequenzen. Die Ergebnisse sind aufgrund der unterschiedlichen Methoden mit den Literaturangaben schwer vergleichbar. Die Werte deuten aber auf eine ähnliche Abweichung hin. So ermitteln Hellbrück et al. (22) 1981 bei Untersuchungen mit dem WHF, dass in 50% aller Fälle die interindividuelle SD kleiner als 3 dB ist (80 % <6 dB). Dies sei auf Unterschiede der Personen zurückzuführen. 1985 zeigen Hellbrück und Moser (23) eine Standardabweichung der Mittelwerte von 5 Skalenteilen. Auch Loesch (50) erwähnt 1986 bei Untersuchungen eine von 5 Skalenteilen als mittlere interindividuelle Abweichung bei Normalhörenden. Baumann (3) beschreibt 1997 interindividuelle Schwankungen um $7,1 \pm 2,3$ KU.

Intraindividuelle Standardabweichung (Retest)

Allen et al. (1) 1990 geben beim Test-Retestvergleich einer Person eine Abweichung von 5 bis 7 dB an. Die größten Unterschiede seien in den niedrigen Frequenzen und unabhängig vom Pegel zu finden. Die Ergebnisse werden als reproduzierbar klassifiziert. Pluinage (69) beschreibt 1989 mit der Methode der LGOB eine Test-Retest-Abweichung von ± 3 dB. Er führt die Schwankungen auf die randomisierte Reihenfolge zurück. Auch Ricketts et al. (73) zeigten 1996 eine Abweichung zwischen 3,3 und 4,1 dB (Korrelationen 0,98 - 0,99). Cox et al. (8) finden mit den Contour-Test eine Reliabilität von 6 dB. Robinson et al. (74) zeigen 1996 zwischen Normalhörenden und Schwerhörigen ähnliche intraindividuelle Schwankungen im Durchschnitt von 5,4 dB. Rasmussen et al. (72) belegen 1998 bei den mittleren Kategorien wie leise und mittellaut eine größere Ungenauigkeit (7-8 dB) im Vergleich zu 5 dB bei den anderen Kategorien.

Untersuchungen mit dem Oldenburger Versuchsaufbau:

Bei Untersuchungen zeigen Kießling et al. (40) 1993 normalverteilte Ergebnisse mit einer Standardabweichung von 5 (jüngere Menschen) - 7 (ältere Menschen) KU beim direkt folgenden Retest. Kießling bewertet die Skalierungsergebnisse als hinreichend stabil. Die Reliabilität sei unabhängig von der Frequenz, dem Hörverlust, dem Geschlecht und dem Alter des Probanden. Meister et al. (54) berichten hingegen 1995 bei einem Test-Retest innerhalb von 30 min. von Abweichungen im Mittel von 2,9 KU.

Untersuchungen mit dem WHF:

Hellbrück et al. (22) teilen 1981 die Erkenntnis mit, die intraindividuelle SD betrage die Hälfte von der interindividuellen Abweichung (1,5 - 3 dB). 1995 stellen Hellbrück et al. (26) wie Pluinage (69) fest, dass die Replizierbarkeit eine gleiche Reihenfolge der Stimuli voraussetze. Man solle von einer Standardabweichung von \pm einer halben verbalen Kategorie, also 5 KU, ausgehen. Heller (27) erwähnt 1985 Untersuchungen von Nowak (64) 1985, der bei Versuchen mit KU-Verfahren eine Gesamtstreuung von 3,67 KU und bei Versuchswiederholung eine intraindividuelle Streuung von 2,48 KU gefunden hat. Intraindividuelle Unterschiede von 2,6 – 4,3 KU finden sich auch bei Baumann (3). Hellbrück

(20) belegt 1983 bei einem nach 14 Tagen durchgeführten Retest eine Standardabweichung von 5 Skalenteilen.

Zur Bestimmung der intraindividuellen Reliabilität werden in der vorliegenden Untersuchung bei den Einzelpersonen die Differenzen gebildet. Bei Erwachsenen beträgt diese im Mittel 3,01 KU. Die intraindividuelle Schwankung ist deutlich geringer als die interindividuelle Standardabweichung. Zur Betrachtung einer systematischen Verzerrung werden auch die Differenzen zwischen den beiden Testverfahren bestimmt. Die Differenz liegt im Mittel bei 3,95 bzw. 3,96. Auch der Unterschied zwischen den Testverfahren liegt somit unter der interindividuellen Unterschiedsgrenze. Auffallend ist hier aber eine deutlich höhere Abweichung um 1 KU vom kurzen Test im Vergleich zum langen als zwischen den beiden kurzen Tests. Eine mögliche Erklärung ist die Simplifizierung der Tonreize, die wiederum keine derart genaue Auflösung von den Probanden verlangt. Insgesamt sind die Untersuchungsergebnisse denen in der Literatur ebenbürtig. Die Schwankungsbreite bleibt aber bei diesem subjektivem Verfahren groß. Kießling (40) stellt dazu 1993 grundsätzlich die Neigung der Versuchspersonen fest, niedrige Lautheitsurteile abzugeben, um so einen Handlungsspielraum im oberen Bereich zu behalten. Diesem Verhalten sei durch Orientierung über den zu erwartenden Messbereich nur partiell entgegenzuwirken. Auch Haubensak (19) beschreibt 1992 dieses Problem (Sequenzeffekte). Nach einem relativ lauten Ton wird die Lautheit eines nachfolgenden Tones überschätzt, wohingegen nach einem leisen Ton eine relative Unterschätzung erfolgt. Diese Misseinschätzung macht sich bis zu 5 Darbietungsfolgen später noch bemerkbar. Intraindividuelle Unterschiede können natürlich auch Sequenzeffekte vortäuschen. So kann eine Versuchsperson z.B. am Anfang der Messung anders skalieren als zum Ende der Messung.

6.2 Kinder

6.2.1 Pegellautheitsfunktionen bei Kindern

Kugler et al. (47) beschreiben 1995 bei Kindern im Alter von 6 - 8 Jahren Steigungen der Pegellautheitsfunktionen, die deutlich über denen der Erwachsenen liegen, im Mittel bei 0,64 dB/KU bei den vier untersuchten Frequenzen. Bei dieser Untersuchung zeigt sich eine mittlere Steigung von 0,62 dB/KU im langen Test und 0,60 dB/KU beim kurzen Test bei den Schulkindern. 21 der untersuchten Schul- und Vorschulkinder sind zu einem Retest einbestellt worden. Die Steigungen betragen im Mittel bei der langen Version 0,62 dB/KU und bei den

kurzen Messungen 0,62 und 0,64 dB/KU. Diese Werte entsprechen den der Literatur zu entnehmenden Daten (38). Bei den Kindern im Vorschulalter steigt der Graph im Mittel beim WHF mit 0,56 dB/KU, im Juniortest mit 0,62 dB/KU. Die von Kugler beschriebenen Werte können im WHF1 nicht gefunden werden und bestätigen sich erst im Juniortest. Die Steigungen im langen Test ähneln eher den Werten bei den erwachsenen Probanden und werden durch eine Überforderung der Kinder erklärbar. Zu dem wirken die wenigen, sich deutlich voneinander unterscheidende Reize des Juniortests weniger verwirrend. Den Kindern fällt es leichter, sich für eine der vorgegebenen Lautheitskategorien zu entscheiden. Außerdem lassen die fünf Teddys keine feineren Abstufungen zu. Das gleiche Phänomen, das bei den Erwachsenen zu einer geringeren Differenzierung der Antworten führt, könnte bei den Kindern Gegenteiliges bewirken, nämlich ein eindeutiges Antwortverhalten. Mit dem Oldenburger Versuchsaufbau werden von Meister et al. (56, 57) 1997 und 1998 bei acht 6 - 8-jährigen Kindern in den Frequenzen von 500 und 2000 Hz Skalierungen anhand von Stoff-Elefanten vorgenommen. Hierbei zeigten sich Steigungen von 0,46 - 0,48 dB/KU (2000/500 Hz).

6.2.2 Varianz der Einzelantworten der Kinder

In dieser Untersuchung umschließen die Zahlenwerte der Standardabweichungen des WHF die des Junior-Hörfelds (2,68 - 14,25 KU im WHF1 und 3,81 - 8,45 KU im Junior1 bei Schulkindern). Deutlich wird aber die Prominenz der höheren SD in der Originalversion im Vergleich zum Juniortest. Kugler (46) nimmt 1995 Standardabweichungen bei Kindern von 2 bis 10 KU wahr und erklärt dies mit einer vermehrten Anzahl an Ausreißern (45). Auch Meister et al. (56) 1997 benennen eine höhere Schwankungsbreite der Skalierungen der Kinder als bei Erwachsenen. Die Unterschiede zwischen den beiden Testversionen beschreiben die in der Literatur ermittelten Daten. In der Literatur fehlen bisher Daten zum Vergleich der Ergebnisse mit Vorschulkindern. Tepřt et al. (82) finden 1994 bei 30 hörgesunden Kindern eine gute Reproduzierbarkeit, deren errechnete Standardabweichung bei $\pm 7,53$ Skalierungseinheiten liegt. Die an den 21 untersuchten Kinder ermittelte Test-Retest-Reliabilität liegt zwischen 6 und 8 KU.

Diese Ergebnisse der Kinder im allgemeinen in dieser Untersuchung widerlegen die Annahme von Gescheider (17) 1988, dass eine geringere Anzahl der Stimuli zu einer erhöhten Streuung

führt. Sie lassen vielmehr vermuten, dass die Kinder mit dem langen Test überfordert sind und sichere Angaben bei einer kürzeren Messzeit machen. Dazu geben Gérard et al. (18) 1990 zu bedenken, dass Kinder generell Schwierigkeiten bei der Wahrnehmung von Intensitätsunterschieden haben. Die hohen intellektuellen Fähigkeiten, die Kindern in einer Untersuchung wie der Lautheitsskalierung abverlangt werden, sollten die Möglichkeit einer Überforderung der Kinder mit einer zu großen Anzahl an Aufgaben nicht in Vergessenheit geraten lassen. 1983 beschreiben Roche et al. (75), dass die Standardfehler mit der Häufigkeit der Untersuchung sinken. Als Übungseffekt bezeichnen sie, dass die Kinder mit den Untersuchungsgeräten vertraut werden, bei der Untersuchung mitzuhelfen lernen, ihre Motivation ändern und auch ihr Bewusstsein für das Hören schärfen.

Zum weiteren Vergleich der Werte fehlen in dieser Altersgruppe die entsprechenden Daten in der Literatur.

6.3 Statistische Tests

Die Werte dieser Untersuchung im Vergleich zu Mosers 1996 ermittelten Daten verdeutlichen eine hoch signifikante Übereinstimmung der Ergebnisse. Steigungswerte und Achsenabschnitte weisen nur geringfügige Abweichungen auf. Die leichten Unterschiede liegen innerhalb einer SD. Die Einzelantworten und Steigungen lassen eine Einordnung als gleichwertige Untersuchung zu.

Bei Berechnung der Fläche werden Unterschiede deutlich. Es zeigt sich keine signifikante Übereinstimmung bei der Korrelation mit dem Intraclass-Koeffizienten und dem Wilcoxon-Test. Die Intraclass-Koeffizienten der Flächenwerte der beiden Testversionen der Schulkinder sind aber nur geringfügig schlechter als die der Erwachsenen und bestätigen damit Collins (7) Untersuchungen 1989 mit AME, die eine ähnliche Korrelation der Wiederholungsmessung bei Erwachsenen und Kindern aufzeigte.

Die erstmals in dieser Arbeit verwendeten Untersuchungen der Ergebnisse durch Flächenwerte eignen sich zwar zur einfachen Erfassung des Hörfeldes, sie stellen aber in der Praxis den Arzt vor den wenig anschaulichen Vergleich von Zahlenwerten, deren Aussagekraft geringer ist als die der graphischen Darstellung. Zwar scheint die Genauigkeit durch den Zahlenwert erhöht zu sein, jedoch ist deren Nutzen in der klinischen Routine beschränkt. Die Lautheitsskalierung, ein Verfahren mit einer Standardabweichung von 10 % der Antwortmöglichkeiten, tut sich schwer, den Validitätskriterien der Statistik zu genügen.

So wird bei zunächst offensichtlicher Kongruenz der Daten, nach Berechnung der Pegellautheitsflächen ein signifikant divergierendes Ergebnis sichtbar. Die Flächenberechnung sowie die Erfassung des Intraclass-Koeffizienten und des Wilcoxon-Tests eignen sich nicht für die Überprüfung der Äquivalenz zweier subjektiver Testverfahren. Unter diesem Aspekt sind auch die mangelnden Signifikanzen der beiden Verfahren zu sehen. Deutlich wird die Ähnlichkeit der Verfahren durch die vergleichbaren Kurvenverläufe (**Abb. 6, 7, 8, 9**). Auch sieht man in den **Abbildungen 11**, dass die Flächenwerte dicht beieinander liegen.

6.4 Verwendungshäufigkeiten

Bei allen Versuchsteilnehmern wird ein unterschiedliches Antwortverhalten in den beiden Testversionen deutlich. Die Simplifizierung der zu skalierenden Kategorien bietet eine Erklärungsmöglichkeit. Beim WHF ist eine maximale Feinabstimmung gefragt, die eine hohe Differenzierung von den Probanden verlangt. Dies fordert den Erwachsenen heraus und erhöht seine Konzentration. Bei Kindern hingegen kann eine zu detaillierte Untersuchung schnell zu Unlust führen. So zeigt sich vor allem bei der langen Testversion eine deutliche Bevorzugung der extremen Kategorien. Die Kinder entscheiden sich also im Zweifel lieber für ein Extrem. Eine mögliche Überforderung durch die feine Differenzierung lässt sie die leichte Variante wählen. Auch eine motivationsabhängige Gewichtung ist eine mögliche Erklärung. Wenn ein Kind die Lust an der Messung verliert, entscheidet es fast wahllos höchstens noch grob unterscheidend. Hellbrück et al. (25) sprechen 1997 von einer Dreibereichsstruktur, in der Kinder die Lautstärkekontinua einteilen (laut, mittel und leise). Dies bildet die Basis für die Verschiebung zu den Extrema Auch Meister (51) findet 2000 eine Bevorzugung zweier Kategorien (z.B. laut und leise) vor allem durch die Kinder im Alter von 4- 5 Jahren.

Im Allgemeinen wird die Aufschlüsselung der einzelnen Kategorien nach Verwendungshäufigkeiten nicht unbedingt als Validitätsaussage beschrieben. So weist Ricketts (73) 1996 in seinen Untersuchungen bei der Einführung die Probanden darauf hin, dass manche Kategorien öfter benutzt und andere gar nicht verwendet werden könnten. Ähnliches sagen Allen et al. (1)1990.