

Kleinste wahrnehmbare Unterschiede in der HRTF-Rekonstruktion aus Hauptkomponenten auf Basis von anthropometrischen Abmessungen

Shaima'a Doma, Hark Braren und Janina Fels

Lehr- und Forschungsgebiet für Medizinische Akustik, Institut für Technische Akustik, RWTH Aachen, 52074 Aachen, Deutschland

Email: {shaimaa.doma; hark.braren; janina.fels}@akustik.rwth-aachen.de

Einleitung

Bei der individuellen Binauralsynthese werden virtuelle Schallquellen mit Hilfe der persönlichen Außenohrübertragungsfunktion (engl. Head-Related Transfer Function, HRTF) erzeugt. Diese spiegelt die Interaktion der am Gehörgangseingang einfallenden Schallwellen mit der individuellen Körper-, vor allem Kopf-, Torso- und Pinnageometrie wider. Ansätze zur HRTF-Individualisierung versuchen die individuelle HRTF auf Basis von begrenzten Informationen über den Probanden zu approximieren. Um unterschiedliche Individualisierungsalgorithmen direkt miteinander zu vergleichen, wurde im Lehr- und Forschungsgebiet für Medizinische Akustik ein Demonstrationstool entwickelt, das die Ergebnisse dreier Verfahren in Echtzeit auralisiert. Schieberegler ermöglichen dem Benutzer eine schrittweise Änderung von (u.a. anthropometrischen) Eingangsparametern. Damit soll eine perzeptiv für den Benutzer möglichst passende HRTF gefunden, sowie ein Herantasten an kleinste wahrnehmbare Unterschiede ermöglicht werden. Die Frage nach einer geeigneten Schrittweite für die Variation von Körperabmessungen führte zur hiesigen Studie. Untersucht wird die Hörbarkeit von Veränderungen an den Abmessungen, konkret im Falle der Individualisierung mittels Hauptkomponentenanalyse (PCA). In erster Linie wird die erforderliche Genauigkeit beim Erfassen der Messdaten bestimmt, weiterhin ist von Interesse, ob unterschiedliche Effekte bei einer Verkleinerung bzw. Vergrößerung der Körperdimensionen auftreten, und ob die am Probanden gemessenen individuellen Werte das Ergebnis beeinflussen. Eine frühere Untersuchung zur Wahrnehmbarkeit von ITD-Änderungen wird zum Vergleich gezogen.

HRTF-Rekonstruktion

Im Mittelpunkt dieser Studie ist die Rekonstruktion von HRTF-Betragsspektren aus Hauptkomponenten (engl. Principal Components, PCs). Die Daten von 47 Probanden aus der ITA HRTF-Datenbank [1] liegen der Hauptkomponentenanalyse (PCA) zu Grunde, die die gemeinsamen Varianzen zwischen den 47 individuell gemessenen HRTFs ausnutzt. Um die Eingangsdatenmenge zu vergrößern, werden zusätzlich die gespiegelten Daten mitberücksichtigt und nur die linken Ohren der 94 Datensätze analysiert. Neben den PCs ergibt sich ein Satz von Gewichten, um die linken HRTFs der Probanden in der Datenbank als gewichtete Summe von (wenigen) PCs darzustellen. Die linken HRTFs werden anschließend für das rechte Ohr gespiegelt, was zu

Tabelle 1: Sechs anthropometrische Abmessungen für die Schätzung der PC-Gewichte: $\mu \pm \sigma$ beschreibt die Verteilung der Parameter in der ITA HRTF-Datenbank (Mittelwert und Standardabweichung), Δ die parameterabhängige Schrittweite für den Hörversuch.

	Abk.	$\mu \pm \sigma$ [mm]	Δ [mm]
Kopfbreite	w	70.9 ± 3.3	0.33
Pinna-Offset (nach unten)	du	18.2 ± 7.0	0.70
Kopftiefe (nach vorne)	df	103.9 ± 5.6	0.56
Kopfhöhe	h	133.2 ± 6.3	0.63
Pinnabreite	$d6$	28.3 ± 3.0	0.30
Cavum conchae - Tiefe	$d8$	14.7 ± 1.6	0.16

symmetrischen HRTFs führt. Durch anthropometrische Abmessungen (s. Abb. 1) können zusätzlich die für datenbankexterne Probanden benötigten Gewichte geschätzt werden. Multiple Regressionsanalyse (MRA) wird dafür auf die bekannten, aus MRT-Scans extrahierten anthropometrischen Daten der HRTF-Datenbank und die dazugehörigen berechneten Gewichte angewendet. Das sich ergebende lineare Modell benutzt einen beschränkten Satz an Kopf- und Ohrabmessungen, hier sechs, deren Werte in der Datenbank die geringste Korrelation untereinander aufweisen (s. Tabelle 1). Dadurch wird Kollinearitätsproblemen vorgebeugt, die das Regressionsmodell verzerren können [2].

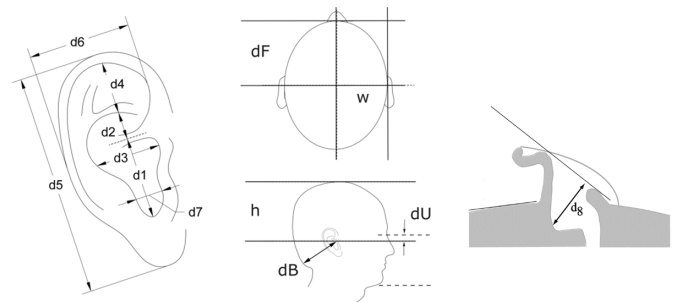


Abbildung 1: Definitionen der Abmessungen von Kopf und Pinna entsprechend der ITA-Datenbank (Auszug).

Neben dem rekonstruierten Betragsspektrum ist die Phase der HRTF von Bedeutung. Relevant für die Wahrnehmung sind zum einen die monauralen Phasenanteile, die durch Hilbert-Transformation aus dem Amplitudengang berechnet werden können. Zum anderen wird die Phasenkomponente der interauralen Zeitdifferenz (ITD) nachgebildet, beispielsweise mithilfe

des *Analytischen Ellipsoidmodells* [3] nach Bomhardt und Fels, das die Kopfbreite, -tiefe und -höhe für die Berechnung des Phasenunterschieds einsetzt.

Versuchsaufbau

Das Ziel der Studie ist es, über einzelne anthropometrische Abmessungen und deren wahrnehmbaren Einfluss auf Betragsspektren von HRTFs Aussagen zu treffen. Dafür wird jeweils ein zu untersuchender Parameter variiert und die weiteren fünf Regressionsparameter des MRA-Modells konstant gehalten. Ebenfalls konstant bleiben die Eingangsparameter zum ITD-Modell. Dadurch wird garantiert, dass bemerkbare Unterschiede rein auf Änderungen am realen Betragsspektrum und an der damit einhergehenden Änderung der Minimalphase zurückzuführen sind. Als Referenzmaße werden die individuellen Abmessungen der Probanden benutzt, wobei für die Pinnaparameter über den rechten und linken Wert gemittelt wird. Vier Testgrößen (w , df , $d6$, $d8$) werden ausgewählt. Eine Vergrößerung und eine Verkleinerung gegenüber der Referenzgröße werden separat untersucht, und somit 8 Testblöcke durchgeführt, zuzüglich eines Trainingsblocks mit der Kopfhöhe h als veränderliche Variable. Die Abweichung von der Referenz wird durch ein QUEST-Verfahren [4] adaptiv in diskreten Schritten eingestellt. Relativ zur Verteilung der jeweiligen Körperdimension in der ITA HRTF-Datenbank wird eine parameter-spezifische Schrittweite definiert: $\Delta_i = \sigma_i/10$, wobei σ_i für die Standardabweichung der Körperabmessung i steht.

Das adaptive QUEST-Verfahren ist so parametrisiert, dass eine maximale Abweichung um $3.5 \sigma_i$ vom individuellen Referenzmaß des Probanden zugelassen wird. Im 3-AFC-Verfahren werden Probanden stets aufgefordert, ein erneutes Abspielen der drei Stimuli zu starten, ehe sich die Möglichkeit zum Antworten bietet (s. Abb. 2).

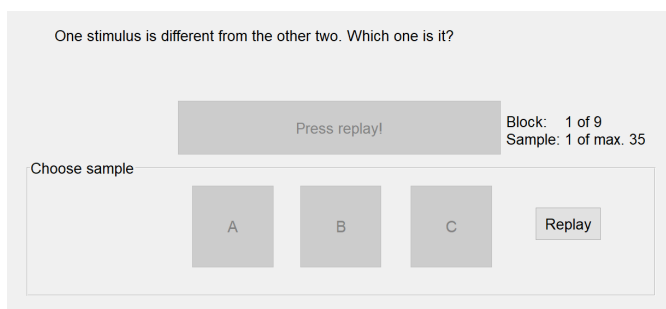


Abbildung 2: Benutzeroberfläche des 3AFC-Tests.

Die Richtungen der virtuellen Quellen bei der Wiedergabe sind (nach einer Vorauswahl auf der Horizontalebene) so gewählt, dass der Einfluss des jeweiligen Parameters auf die HRTF maximiert wird. Anhand der Regressionskoeffizienten im MRA-Modell kann der Einfluss abgeschätzt werden, womit sich die Richtungen 30° für df und $d8$, sowie 90° für w und $d6$ ergeben. Ein Vorversuch bestätigt die höhere Relevanz der Parameter in den genannten Richtungen, durch eine höhere Sensitivität auf Veränderung ihrer Werte. Wegen der Symmetrie

der verwendeten HRTFs können die virtuellen Quellen spiegelbildlich bei -30° und -90° platziert werden, um den Vorteil der rechten Seite auszunutzen. Ein dreifacher rosa Rauschpuls, bestehend aus jeweils 200 ms langen Rauschintervallen und 150 ms langen Pausen, dient als Rohsignal, das mit der Inversen der individuellen Kopfhörerübertragungsfunktion (HpTF) [5] der Probanden sowie der rekonstruierten HRTF multipliziert wird. Der Schalldruckpegel wird auf einen Maximalwert von 60 ± 2 dB kalibriert.

Insgesamt werden 24 Probanden (7 w. & 17 m.) getestet, die teils der ITA HRTF-Datenbank angehören, und teils individuelle 3D-Kopf-Modelle mittels Streifenprojektion erstellt bekommen, aus denen jeweils die personenbezogenen Körperabmessungen extrahiert werden. Da für die Rekonstruktion jeweils der Mittelwert der Abmessungen des linken und rechten Ohres verwendet wird, werden Probanden ausgeschlossen, die für Parameter $d6$ bzw. $d8$ überdurchschnittlich asymmetrische Werte besitzen. Als weiteres Ausschlusskriterium dient eine Hochtonaudiometrie, bei der eine geringe Unterschreitung der 20 dBHL-Linie für Frequenzen $f > 8$ kHz entsprechend Abb. 3 für die Teilnahme toleriert wird. Somit können die laut Middlebrooks [6] relevanten Frequenzanteile der HRTF bis 13 kHz noch hinreichend gut wahrgenommen werden.

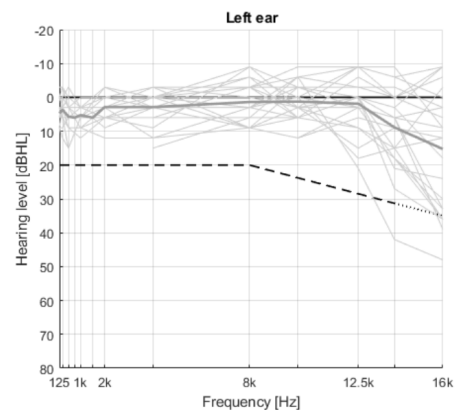


Abbildung 3: Audiogramme der 24 Teilnehmer (linkes Ohr) und deren Mittelwert (dunkelgrau). Die gestrichelte Linie definiert eine tolerantere Grenze für intaktes Hörvermögen.

Ergebnisse

Als kleinste wahrnehmbare Unterschiede (engl. just noticeable differences, JNDs) werden die 50%-Schwellenwerte aus den psychometrischen Funktionen abgelesen, die sich nach der jeweils letzten QUEST-Iteration eines Testblocks ergeben. Die Werte liegen zunächst als Vielfache der parameterabhängigen Schrittweite vor und lassen sich direkt in absolute Abweichungen (in mm) umrechnen. Einflüsse der untersuchten Körperdimensionen (w , df , $d6$, $d8$), der Vergrößerung oder Verkleinerung (up vs. $down$), sowie die Aufteilung in Kopf- und Pinnaparameter ($head$ vs. $pinna$) werden im Folgenden analysiert.

Nur bei der Kopfbreite w zeigt sich ein signifikanter Unterschied zwischen up und $down$, mit einer höheren Wahrnehmungsschwelle für absolute Vergrößerungen.

Mit einem Mittelwert der 50%-Schwelle von 3.56 mm besteht bei der Pinna-Tiefe $d8$ die höchste Sensitivität auf eine absolute Änderung (s. Abb. 4), die wiederum der höchsten Anzahl an relativen Schritten Δ_{d8} entspricht.

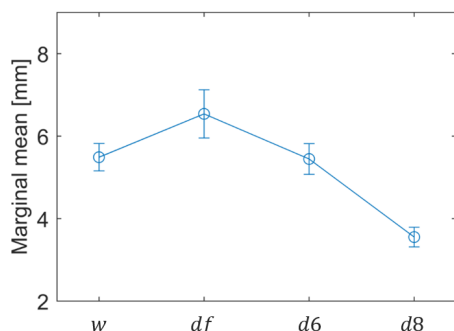


Abbildung 4: Randmittelwerte der 50%-Schwellen als absolute Abweichungen von der Referenz in mm.

In Abb. 5 sind die Randmittelwerte dargestellt, bei Normierung auf die individuellen Referenzmaße der Probanden. Hier wird die Abweichung von der Referenz als Skalierung interpretiert. Parameter der Gruppe *head* zeigen eine höhere Sensitivität auf prozentuale Änderungen der Messdaten, während Parameter der Gruppe *pinna* generell niedrigere absolute Schwellenwerte besitzen.

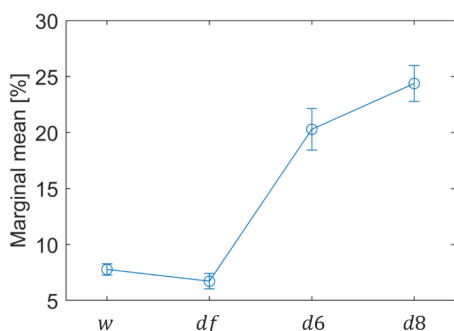


Abbildung 5: Randmittelwerte der 50%-Schwellen, normiert auf die jeweiligen individuellen Körperabmessungen der Probanden.

Ein weiteres Ergebnis ist die starke ($\sigma = -0.755$) bzw. mittelstarke ($\sigma = -0.6421$) Korrelation zwischen der individuellen Kopftiefe df der Probanden und der Sensitivität auf deren absolute Veränderungen. In Abb. 6 ist die Verteilung für eine Vergrößerung von df dargestellt. Absolute Schwellenwerte von Probanden mit einer größeren Kopftiefe sind für beide Variationsrichtungen tendenziell niedriger als für Probanden mit kleineren Kopftiefen.

Diskussion

Die Mittelwerte der absoluten 50%-Schwellen liegen zwischen 3.31 und 7.35 mm. Absolute Schwellen liefern eine Aussage über die beim Individualisierungsverfahren erforderliche Messgenauigkeit, um das Ergebnis der rekonstruierten HRTFs nicht zu verfälschen. Weiterhin kann eine geeignete Größenordnung für die diskreten Schritte der Schieberegler im Demonstrationstool entnommen werden. Hier fiel die Entscheidung auf das

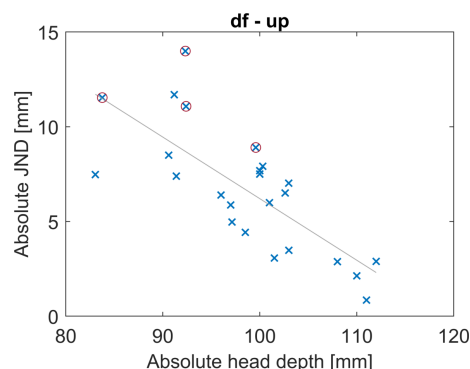


Abbildung 6: Eine starke negative Korrelation ($\sigma = -0.755$) zeigt sich zwischen absoluten JND-Werten bei Vergrößerung der Kopftiefe und dem individuell gemessenen Referenzmaß. Probanden mit einer größeren Kopftiefe sind sensitiver auf deren positive Abweichung von der Referenz. Mit roten Kreisen markierte Probanden wurden von der statistischen Analyse ausgeschlossen, aufgrund von Ausreißern in mindestens einem Testblock.

25%-Quantil, um für geschätzte 75% der Benutzer bei Bewegung des Reglers um einen Schritt keinen detektierbaren Unterschied zu bewirken. Somit kann nahe genug an die JNDs herangetastet werden. Die abgerundeten Werte sind Tabelle 2 zu entnehmen. Der Durchschnittswert von 3 mm wird für nicht untersuchte Körpermaße übernommen.

Tabelle 2: Abgerundete 25%-Quantile der absoluten Schwellenwerte können als geeignete Schrittweiten zur Variation der Körperabmessungen im Demo-Tool angenommen werden.

Parameter	w	df	d6	d8
Schrittweite [mm]	3	3	4	2

Die bemerkbaren Unterschiede kommen durch Veränderungen am reellwertigen Betragsspektrum einer HRTF, aus dem zusätzlich der Minimalphasenanteil der HRTF berechnet wird, zustande. Interessant ist die Gegenüberstellung von bekannten JND-Werten für andere Komponenten der HRTF. Für lateralen Schalleinfall bestimmte Bomhardt [7] den kleinsten wahrnehmbaren Unterschied in ITDs zu $80 \mu\text{s}$. Wird der Kopf in einem vereinfachten Modell als schallharte Kugel angenommen, lässt die Umlenkung um den Kopf (s. Abb. 7) auf eine entsprechende Änderung der Kopfbreite w schließen, die zum selben Phasenversatz führt:

$$\Delta w_{\text{jnd}} = \frac{c * \Delta \text{ITD}}{\pi} \approx 8.66 \text{ mm} . \quad (1)$$

Der Mittelwert von $\Delta w = 5.49 \text{ mm}$ für die Wahrnehmungsschwelle, der sich im hiesigen Hörversuch ergibt, liegt unter dem in Gl. 1 berechneten Wert. Das Ergebnis suggeriert einen leichter wahrnehmbaren Einfluss der Kopfbreite w auf die monauralen spektralen Eigenschaften, als auf den ITD-Anteil der Phase.

Die Angabe der JND-Schwellenwerte als Vielfache der parameterspezifischen Schrittweiten lässt auf die Wahr-

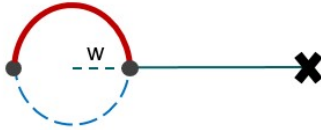


Abbildung 7: Vereinfachte Darstellung des Schallweges um den Kopf als Auslenkung entlang der Kugelfläche bei lateralem Einfall.

scheinlichkeit schließen, bei Nutzung nicht-individueller Werte einen Unterschied zu bemerken. Ein Durchschnittsprüfung würde beispielweise bei einer Schwelle von 15 Schritten (entsprechend $1.5 \sigma_i$) für die meisten Werte aus der Datenbank keine Änderung gegenüber der mit eigenen Werten synthetisierten HRTF erkennen. Im Fall der Kopftiefe könnte die Wahl einer etwas überdurchschnittlichen Größe für mehr Probanden geeignet sein, weil Probanden mit einer kleineren Kopftiefe weniger sensitiv auf Abweichungen vom eigenen (korrekten) Maß sind.

Fazit

In diesem Hörversuch wurden die kleinsten wahrnehmbaren Änderungen an vier Körperabmessungen untersucht, die in die Rekonstruktion von HRTFs aus Hauptkomponenten einfließen. Darunter waren zwei Kopf- und zwei Pinnaabmessungen. Ein signifikanter Unterschied zwischen den Schwellen für Verkleinerung und Vergrößerung ergab sich nur für die Kopfbreite, die eine höhere Sensitivität auf eine Verkleinerung aufweist. Im Gegensatz zu den Kopfparametern zeigten sich für die beiden Pinnaparameter kleinere Schwellenwerte, die wiederum einer weitaus größeren prozentualen Änderung entsprechen. Des Weiteren ergab sich eine hohe Korrelation zwischen einer großen individuellen Kopftiefe und einer hohen Sensitivität auf die entsprechende Veränderung dieser.

Die aufgeführten Ergebnisse sind nicht als globale Indikatoren für die Sensitivität auf Änderungen einzelner Körperabmessungen zu verstehen, da deren Relevanz stark vom jeweils gewählten Regressionsmodell abhängt. Auf MRA-Modelle bestimmter HRTF-Richtungen haben einzelne Parameter sogar einen vernachlässigbaren Einfluss, wie sich im Vorversuch bestätigte. Die im Hörversuch eingesetzten Richtungen maximieren die Sensitivität und stellen somit das Worst-Case-Verhalten dar. Alleine durch die Regressionskoeffizienten lassen sich die JND-Werte aber nicht vorhersagen. Des Weiteren beschränkt sich die Untersuchung auf die Veränderung einzelner Variablen, was vorerst die Frage nach zusätzlichen Effekten bei mehr als einem nicht-individuellen Wert offen lässt. Weitere Untersuchungen könnten diesen Einfluss untersuchen und alternative Möglichkeiten für MRA-Modelle erproben. Eine tolerantere Einschätzung für JNDs könnte sich bei Verwendung realitätsnäherer Rohsignale, anstelle der breitbandigen Rauschpulse, die die spektralen Unterschiede zwischen HRTFs besonders stark betonen, ergeben.

Literatur

- [1] Bomhardt, R.; de la Fuente Klein, M. and Fels, J.: A high-resolution head-related transfer function and three-dimensional ear model database. Proceedings of Meetings on Acoustics 172ASA (2016)
- [2] Gilbert W. Stewart: Collinearity and least squares regression. Statistical Science (1987)
- [3] Bomhardt, R.; Lins, M. and Fels, J.: Analytical ellipsoidal model of interaural time differences for the individualization of head-related impulse responses. Journal of the Audio Engineering Society 64 (2016)
- [4] Watson, A. B. and Pelli, D. G.: QUEST: A Bayesian adaptive psychometric method. Perception & psychophysics 33.2 (1983)
- [5] Masiero, B. and Fels, J.: Perceptually robust headphone equalization for binaural reproduction. Audio Engineering Society Convention 130 (2011)
- [6] Middlebrooks, J. C.: Individual differences in external-ear transfer functions reduced by scaling in frequency. The Journal of the Acoustical Society of America 106.3 (1999)
- [7] Bomhardt, R.: Anthropometric Individualization of Head-Related Transfer Functions. PhD thesis. RWTH Aachen University (2017)