

Von der Simulation zur Realität

Vergleich gemessener und berechneter HRTFs

Felix Klooss¹, Christoph Reuter², Isabella Czedik-Eysenberg^{2,3}, Anja-Xiaoxing Cui²

¹ Universität Wien, MediaLab

² Universität Wien, Musikwissenschaftliches Institut

³ Universität Osnabrück, Institut für Musikwissenschaft und Musikpädagogik

Hintergrund

Unter Head-Related Transfer Function (HRTF) versteht man eine Funktion, die die Übertragung von einer Schallquelle zum Gehörgang in Abhängigkeit von der Einfallrichtung beschreibt. Diese Änderungen lassen sich in Interaural Time Differences (ITDs), Interaural Level Differences (ILDs) und Spectral Cues (SCs) kategorisieren [1].

Den größten Einfluss auf die Übertragung nehmen die individuelle Ohr- und Kopfform [2]. Somit hat jede Einfallrichtung für jeden Menschen eine individuelle Ausprägung dieser Charakteristika, anhand derer ihr Gehirn die Richtungswahrnehmung konstruiert. Wird ein Audiosignal mit einer individualisierten HRTF einer Einfallrichtung gefiltert und der entsprechenden Person dargeboten (z.B. über entzerrte Kopfhörer), so nimmt diese das Signal als aus dieser Richtung kommend wahr [3].

HRTFs können gemessen oder berechnet werden [4]. Herkömmliche Messungen sind aufwendig, da eine geeignete Messapparatur groß und teuer ist und die zu vermessende Person physisch anwesend sein muss. Ungewollte Kopfbewegungen werden meist kontrolliert, um die Messung bei zu großer Abweichung zu wiederholen, jedoch ist eine gradgenaue Messung nicht garantiert [5]. Physikalische Berechnungen sind eine wertvolle Alternative, da sie in Abwesenheit der zu vermessenden Person und ohne unwillkürliche Bewegungen vorgenommen werden können. Der dafür notwendige Scanvorgang der Kopf- und Ohrform ist fehleranfällig und nicht trivial [vgl. 6].

Ziele

Es werden HRTFs, die auf demselben digitalen Modell beruhen, auf drei verschiedene Arten ermittelt und anschließend miteinander verglichen. Für physikalische Berechnungen wird die kommerzielle Software Comsol (C) [7] und die Open-Source-Software Mesh2HRTF (M2H) [8] verwendet. Messungen werden hingegen mit einer von der Universität Wien patentierten Messmethode (M) [9] an einem 3D-gedruckten Kunstkopf durchführt.

Methode

Um Abweichungen des Scanvorgangs auszuschließen, wird ein bereits digitalisierter Kopf verwendet und 3D-gedruckt. Die durch den Druck entstehenden Abweichungen sind gering. Der gedruckte Kopf wird auf den Messapparat (M) gesetzt, der diesen vor einer Schallquelle ausrichtet und den gesamten Messvorgang automatisiert durchführt.

Beide physikalischen Berechnungen basieren auf der Randlelementmethode, wobei M2H die Burton-Miller-Kollokation gekoppelt mit der mehrstufigen schnellen Multipolmethode verwendet. Zum leichteren Vergleich werden alle drei Ergebnisse ins Sofa-Format gebracht. Sowohl M als auch M2H erzeugen dieses Format nativ. Der Datenexport für C wird über Matlab automatisiert und in Python unter Zuhilfenahme von Numpy, Scipy und Sofar in das Sofa-Format überführt. Zusätzlich wird für die weitere Verarbeitung, Auswertung und Darstellung Matplotlib und Pyfar verwendet.

Die Messungen mit M finden in einer akustisch unbehandelten Umgebung statt, um deren Möglichkeiten innerhalb einer herkömmlichen Arbeitsumgebung zu demonstrieren. Um störende Raumeinflüsse zu minimieren, werden alle Impulsantworten mit einem 4 ms breiten Blackman-Harris-Fenster zentriert über dem Sample mit dem höchsten Amplitudenbetrag gefiltert. Es werden anschließend die Directional Transfer Functions (DTFs, [10]) aller HRTFs (C, M2H & M) extrahiert. Zur Ermittlung der Magnitudendifferenzen werden diese nach der ERB (Equivalent rectangular bandwidth)-Skala je Messmethode zusammengefasst. Alle Vergleiche beziehen sich auf die Horizontalebene mit einer Winkelauflösung von einem Grad.

Ergebnisse

Alle drei Methoden führen zu sehr ähnlichen Ergebnissen. Mittels M und C erhaltene DTFs weisen etwas weniger Magnitudendifferenzen auf, als DTF_M verglichen mit DTF_{M2H} und DTF_{M2H} verglichen mit DTF_C. Besonders die für die Lokalisation dominante, der Schallquelle zugewandten Seite zeigt hohe Übereinstimmungen der Magnituden der über ERBs gemittelten Frequenzbins.

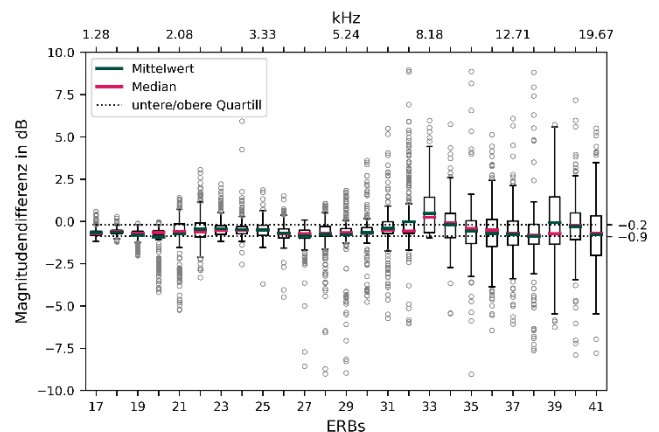


Abbildung 1: Boxplots der Magnitudendifferenz zwischen DTF_M und DTF_C pro ERB von 0 – 359° in dB.

50% der Magnitudendifferenzen der gesamten Horizontalebene zwischen DTF_M und DTF_C von 1.2 kHz bis 20.8 kHz liegen innerhalb -0.9 dB und -0.2 dB (Abb. 1) bzw. zwischen DTF_M vs. DTF_{M2H} innerhalb -1.7 dB und -0.3 dB. Mit steigender Frequenz werden zunehmend Ausreißer (graue Kreise) sichtbar.

Wird hingegen die der Schallquelle abgewandte Seite aus der Analyse ausgeschlossen, so liegen 50% der Magnitudendifferenzen zwischen DTF_M vs. DTF_C innerhalb 0 dB und 0.4 dB (Abb. 2a) bzw. zwischen DTF_M vs. DTF_{M2H} innerhalb -0.6 dB und 0.0 dB (Abb. 2)

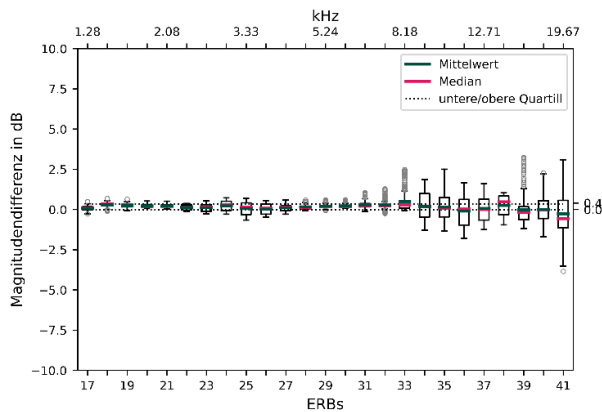


Abbildung 2a: Boxplots der Magnitudendifferenz zwischen DTF_M und DTF_C pro ERB von 0 – 180° in dB.

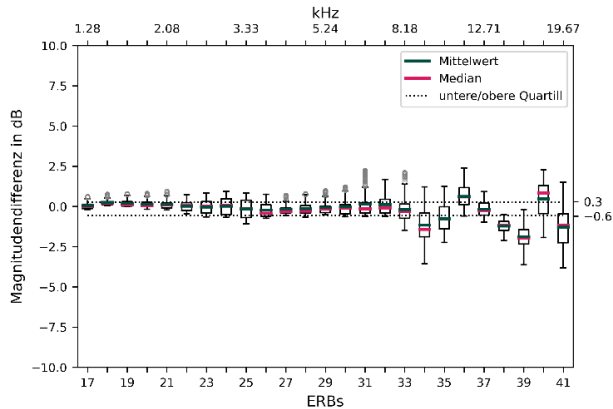


Abbildung 2b: Boxplots der Magnitudendifferenz zwischen DTF_M und DTF_{M2H} pro ERB von 0 – 180° in dB.

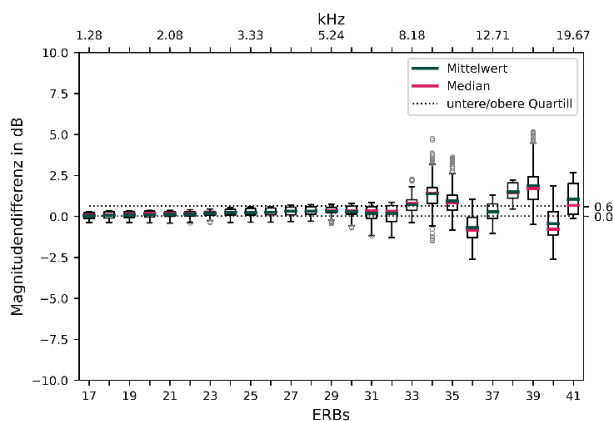


Abbildung 2c: Boxplots der Magnitudendifferenz zwischen DTF_{M2H} und DTF_C pro ERB von 0 – 180° in dB.

Abbildung 3 zeigt Magnitudengänge der DTFs für unterschiedliche Einfallsrichtungen nach Messmethode. Die durch C, M2H und M erhaltenen DTFs weisen auf der Schallquelle zugewandten Seite hohe Übereinstimmungen auf (Abb. 3a, b). Auf der abgewandten Seite sind optisch deutliche Differenzen auszumachen. In Abhängigkeit von der Einfallsrichtung schwanken dort die Magnituden stark, wie sich in Abbildung 3c und 3d bei lediglich einem Grad Winkelabweichung zeigt.

Trotz der tendenziell höheren Übereinstimmung zwischen DTF_M und DTF_C im Vergleich zu DTF_M und DTF_{M2H} sind Instabilitäten in DTF_C auszumachen, die sich in DTF_{M2H} nicht beobachten lassen. Diese machen sich in einer Kerbe in Abbildung 3a und ebenso in der Heatmap der Abbildung 4a in ähnlichem Frequenzbereich bemerkbar.

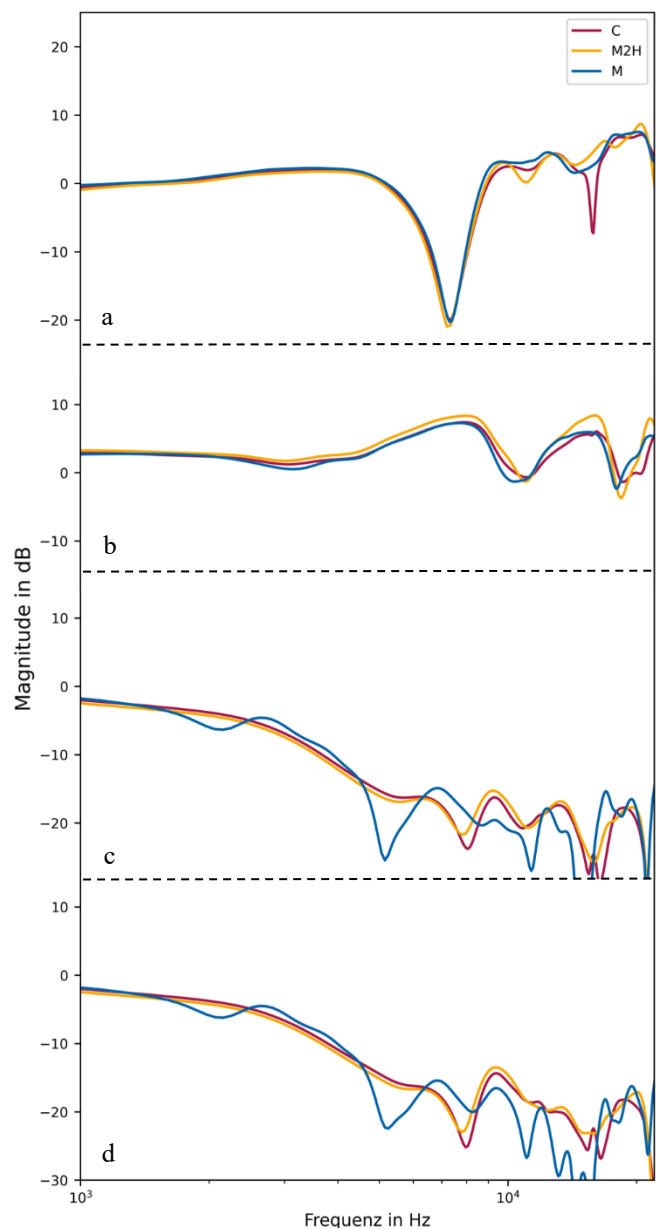


Abbildung 3a–d: Magnitudengänge der DTF_C , DTF_{M2H} und DTF_M für unterschiedliche Einfallsrichtungen auf der Horizontalebene. a: 0°, b: 90°, c: 270°, d: 271°

Für eine bessere optische Übersicht der Magnitudengänge in Abhängigkeit des Einfallswinkels in der Horizontalebene

breitet **Abbildung 4** die Magnituden nach Frequenz und Einfallswinkel auf. Auch hier lässt sich der hohe Übereinstimmungsgrad beobachten.

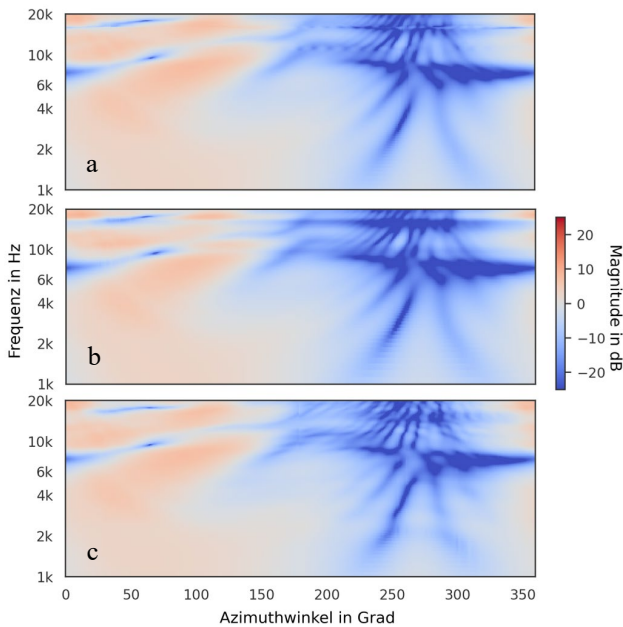


Abbildung 4a–c: Magnitudengänge DTF_C (a), DTF_{M2H} (b) und DTF_M (c) nach Einfallswinkel in der Horizontalebene.

Conclusio

Sowohl **C** und **M2H** als auch **M** erzeugen DTFs mit hoher Übereinstimmung größtenteils unterhalb der Wahrnehmungsschwelle. Da im Schallschatten kleinste Winkeländerungen große Auswirkungen haben, zeigen sich dort auch die deutlichsten Differenzen in den DTFs, deren perzeptueller Einfluss jedoch als gering angenommen wird.

Ohne weitere Untersuchungen kann nicht festgestellt werden, welche der Methoden näher an den DTFs einer realen Person liegt. Festzuhalten sind jedoch die hohe Flexibilität und Komplexität von **C**, die in diesem Zusammenhang mehr Nach- als Vorteile bringen. Erst in Verbindung mit Matlab können notwendige Exporte automatisiert werden. **M2H** ist im Gegensatz dazu wesentlich benutzerfreundlicher, liefert Ergebnisse schneller und ist kostenfrei zugänglich.

M bietet vor allem Vorteile, wenn komplexe weitere Systeme (z.B. Hörgeräte, Kopfhörer) in die Messung eingebunden werden. Es können beliebige Messpositionen (z.B. nach der Lebedev-Quadratur) in beliebiger Auflösung angesteuert werden. Besonders bei wenigen Messpunkten ist **M** gegenüber **C** und **M2H** deutlich schneller, da diese auf dem Prinzip der Reziprozität beruhen.

Literatur

- [1] Hebrank, J. & Wright, D. (1974). Spectral cues used in the localization of sound sources on the median plane. *The Journal of the Acoustical Society of America*, 56(6), 1829–1834.
- [2] Musicant, A. D. & Butler, R. A. (1984). The influence of pinnae-based spectral cues on sound localization. *The Journal of the Acoustical Society of America*, 75(4), 1195–1200.

- [3] Wightman, F. L. & Kistler, D. J. (1989). Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2), 858–867.
- [4] Katz, B. F. (2001). Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *The Journal of the Acoustical Society of America*, 110(5 Pt 1), 2440–2448.
- [5] ARI HRTF-Database detailed description (2023, 3. März). Austrian Academy of Science. <https://www.oew.ac.at/en/isf/das-institut/software/hrtf-database/detailed-description>
- [6] sdx-lv (2023, 3. März). 3D scanning. Sourceforge. https://sourceforge.net/p/mesh2hrtf/wiki/Basic_tutorial_3d_scanning/
- [7] COMSOL Multiphysics®, www.comsol.com. COMSOL AB, Stockholm, Sweden.
- [8] Ziegelwanger, H. & Kreuzer, W. & Majdak, P. (2015). MESH2HRTF: An Open-Source software Package for the Numerical calculation of head-related transfer functions.
- [9] Klooss, F. (2022). Robot for measuring HRTFs (Application No. 22208557.3). European Patent Office.
- [10] Middlebrooks, J. C. (1999). “Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency,” *The Journal of the Acoustical Society of America*, 106, 1493–1510.