

Auflösung und Interpolation von unterschiedlichen Kopf-über-Torso-Orientierungen in kopfbezogenen Übertragungsfunktionen

Reinhild Roden¹, Fabian Brinkmann¹, Alexander Lindau¹, Stefan Weinzierl¹

¹ Audio Communication Group, TU Berlin, Deutschland, Email: reinhild.rodin@gmx.net

Einleitung

Kopfbezogene Übertragungsfunktionen (HRTFs) liegen meist für verschiedene Schalleinfallrichtungen und eine neutrale Kopf-über-Torso-Orientierung vor. Werden Kopffrotationen eines Hörers simuliert, entsprechen diese deshalb einer gemeinsamen Bewegung von Kopf und Torso, wohingegen der Kopf natürlicherweise unabhängig über dem Torso gedreht wird. Da Unterschiede zwischen HRTFs mit konstanter und variabler Kopforientierung hörbar sind [1], aber die Messung von Daten mit unterschiedlicher Kopf-über-Torso-Orientierung sehr zeitaufwändig ist, wird in der aktuellen Studie untersucht, inwieweit sich HRTFs mit verschiedenen Kopforientierungen unhörbar interpolieren lassen.

Frühere Studien zur Interpolation von HRTFs unterschiedlicher Quellpositionen haben gezeigt, dass (a) eine Spline-Interpolation von Betrag und Phase im Frequenzbereich der Interpolation der Impulsantwort überlegen ist [2], und dass (b) ein in 4°-Schritten aufgelöster HRTF Datensatz für die meisten Quellpositionen eine perzeptiv artefaktfreie Interpolation erlaubt [3]. Inwieweit dies auch für die Interpolation von Kopf-über-Torso-Orientierungen gilt, sollte mit der vorliegenden Studie untersucht werden.

Interpolation

Bei der Wahl der zur Interpolation verwendeten HRTFs wurden zwei Ansätze verfolgt: Zum einen wurden für die Interpolation HRTFs mit identischer Torso-zu-Quell- aber unterschiedlicher Kopf-zu-Quell-Orientierung verwendet („Kopf-Interpolation“), zum anderen HRTFs mit identischer Kopf-zu-Quell- aber unterschiedlicher Torso-zu-Quell-Orientierung („Torso-Interpolation“). Letzteres erschien vielversprechend, da für diesen Fall die durch die Pinna verursachte hochfrequente Feinstruktur der HRTFs aufgrund der identischen Kopf-zu-Quell-Orientierungen besser erhalten bleibt. Darüber hinaus wurde mit Blick auf rechenkritische Anwendung ein Nearest-Neighbour-Ansatz getestet, sowie – in Anlehnung an Hartung et al. [2] und Minnaar et al. [3] – (a) die Interpolation mittels Splines und Inverse-Distance-Weighting (linear) durchgeführt, (b) sowohl kopfbezogene Impulsantworten (HRIRs) als auch HRTFs getrennt nach Betrags- und Phasenfrequenzgang interpoliert und (c) minimal- und originalphasige HRIRs verwendet. Vor der Interpolation originalphasiger HRIRs wurde bei zehnfacher Abtastrate die interaurale Zeitdifferenz (ITD) entfernt, separat interpoliert und nach der Interpolation wieder hinzugefügt. Dies geschah mittels fractional delay [5] und ITD-Schätzung durch (a) onset detection (-6 dB

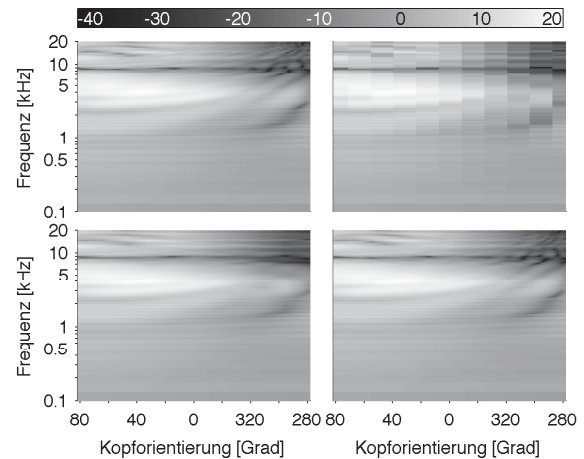


Abbildung 1: Exemplarische Betragsspektren [dB] für das rechte Ohr bei frontaler Quelle und 16° Interpolationsintervall. Oben links: Referenz, oben rechts: Nearest Neighbour, unten links: Kopf-Interpolation, unten rechts: Torso-Interpolation.

unterhalb Max.) und (b) Kreuzkorrelation. Demzufolge wurden für 17 zu testende Algorithmen die HRTFs von 23 Interpolationsintervallen [1, 2..10, 12..30, 35..45°] auf die Referenzauflösung von 0.5° interpoliert. Hierfür wurden HRTFs für sechs typische sowie kritische Quellpositionen ([Azimuth°; Elevation°]: [0;0] (frontal), [90;0] (links), [0;90] (oben), [45;0], [315;30], [315;-30]) und hypothetisch unerschwinglich aufgelöste (0.5°) Kopf-über-Torso-Orientierungen zwischen ±82° gemessen. Die Messung erfolgte mit dem Kopf-und-Torso-Simulator FABIAN [4], dessen computergesteuertes Halsgelenk hochpräzise in drei Freiheitsgraden über dem Torso bewegt werden kann.

Physikalische Evaluation

Zum Vergleich der Algorithmen und um eine Vorauswahl für den Hörversuch treffen zu können, wurden spektrale Differenzen zwischen interpolierten HRTFs und der Referenz bestimmt. Diese sind beispielhaft für ein Interpolationsintervall von 16° bei frontaler Quellposition in Abb. 1 gezeigt. Neben typischen Umschaltartefakten für den Nearest-Neighbour-Algorithmus wird erwartungsgemäß ersichtlich, dass die hochfrequente Feinstruktur in den HRTFs für die Torso-Interpolation nahezu perfekt erhalten bleibt, wohingegen sie bei der Kopf-Interpolation vermischt.

Für eine bessere Vergleichbarkeit der Algorithmen und Quellpositionen untereinander wurde ein spektrales Einzahl-Fehlermaß entwickelt, das die Fehlerenergie

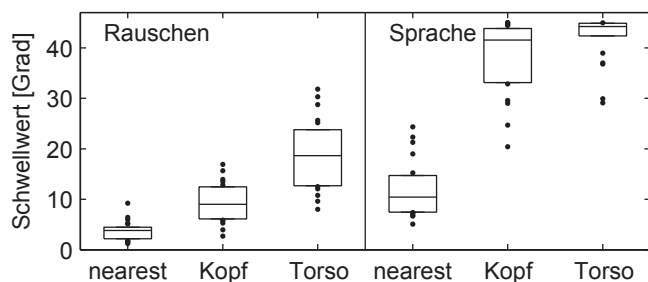


Abbildung 2: Schwellwerte (Mediane und Interquartil-abstände) für 25 Versuchspersonen und 6 Konditionen.

zwischen Referenz und interpolierten HRTFs über 39 auditorische Filterbänder mittelt und für beide Ohren addiert [1]. Erwartungsgemäß zeigten sich die kleinsten Fehler für die Torso-Interpolation sowie für frontale und elevierte Quellen. Größere Fehler wurden für die Kopf-Interpolation, den Nearest-Neighbour-Algorithmus sowie seitliche und negativ elevierte Quellen beobachtet. Hingegen hatte die Wahl der Interpolante (Spline, linear), der Interpolationsdomäne (HRIR, HRTF) und der ITD-Schätzung (Onset, Kreuzkorrelation) einen geringen Einfluss.

Perzeptive Evaluation

In einem adaptiven 3AFC (ZEST) Experiment wurde für 25 Probanden (6 Frauen, 19 Männer, Altersmedian 27, meist musikalisch/akustisch geschult und hörversuchserfahren) das ebenmerkliche Interpolationsintervall bestimmt, das an der 50%-Diskriminationsschwelle zwischen Referenz und interpolierten HRTFs angesetzt wurde. Die kopfhörerbasierte [6] Echtzeitauralisation erfolgte mittels dynamischer Binauralsynthese für Kopforientierungen zwischen $\pm 82^\circ$ und eine kritische Quellposition [315;-30]. Unter Verwendung zweier Stimuli (männl. Sprache, rosa Rauschen) in Kombination mit drei Interpolationsalgorithmen (Nearest Neighbour; Kopf-Interpolation / HRIR-basiert / Spline; Torso-Interpolation / HRTF-basiert / linear) wurden durch jeweils 20 Trials 3x2 Schwellen bestimmt. Verwendet wurden jeweils die Algorithmen, die in der physikalischen Evaluation die geringsten Fehler aufwiesen. Die erhobenen Daten wurden anschließend in einer 2-faktoriellen Varianzanalyse (ANOVA) für vollständige Messwiederholung ausgewertet.

Abb. 2 zeigt die Schwellwerte der 25 Probanden für die sechs Konditionen. Die ANOVA weist den Faktor 'Stimulus' als hochsignifikant aus ($F(1,24) = 431.4$; $p < .001$; $\eta_p^2 = .947$), wobei für den kritischen Stimulus Rosa Rauschen kleinere Schwellen gefunden wurden als für Sprache. Der Faktor 'Algorithmus' erweist sich ebenfalls als hochsignifikant ($F(2,48) = 276.0$; $p < .001$; $\eta_p^2 = .920$, Posthoc-Paarvergleiche). Für den Nearest-Neighbour-Ansatz liegen die Ergebnisse teils nahe an der Referenz, während die günstigsten Schwellen für die Kopf-Interpolation ermittelt wurden.

Für eine bessere Interpretation der Ergebnisse wurden kumulierte Schwellwertverteilungen geschätzt [7],

ausgewählte Perzentile sind in Tab. 1 dargestellt. Die tatsächlichen Perzentile für Kopf- und Torso-Interpolation und Sprache sind dabei aufgrund eines Deckeneffektes wahrscheinlich sogar noch etwas höher als die Schätzwerte aus Tab. 1.

Tabelle 1: Ausgewählte Perzentile der geschätzten kumulierten Schwellwertverteilungen.

Perzentil	Rauschen			Sprache		
	nearest	Kopf	Torso	nearest	Kopf	Torso
50 %	4,4	10,5	20,8	12,7	41,1	43,6
25 %	3,3	7,9	15,5	9,5	35,6	41,4
5 %	1,5	4,6	9,6	5,7	23,7	31,7

Diskussion

In der vorliegenden Studie wurde die Interpolation von Kopf-über-Torso-Orientierungen in HRTFs untersucht. Dabei lieferte, wie erwartet, eine Interpolation über HRTFs mit identischer Kopf-zu-Quell- aber unterschiedlicher Torso-zu-Quell-Orientierung („Torso-Interpolation“) die besten Ergebnisse, da hier hochfrequente Pinna-Cues auch bei größeren Interpolationsintervallen erhalten bleiben. Für 95% der Hörenden liegt die ebenmerkliche Schwelle bei 10° oder höher, wohingegen eine Interpolation über HRTFs mit identischer Torso-zu-Quell- aber unterschiedlicher Kopf-zu-Quell-Orientierung („Kopf-Interpolation“) mit 5° weniger gutmütig abschneidet. Ein Nearest-Neighbour-Ansatz sollte für kritische, breitbandige Stimuli nicht verwendet werden, auch wenn er bei Sprache bis zu einem Interpolationsintervall von 6° zu guten Ergebnissen führen kann.

Literatur

- [1] Brinkmann, F.; Roden, R.; Lindau, A.; Weinzierl, S.: Audibility of head-above-torso orientation in head-related transfer functions. In: *Forum Acusticum*. Krakau, Polen. 2014 (angenommener Beitrag).
- [2] Hartung, K.; Braasch, J.; Sterbing, S.J.: Comparison of different methods for the interpolation of head-related transfer functions. In: *16th AES-Convention*, S. 319–329, Rovaniemi, Finland, 1999.
- [3] Minnaar, P.; Plogsties, J.; Christensen, F.: Directional resolution of head-related transfer functions required in binaural synthesis. *J Audio Eng Soc.* 53(10):919–929, 2005.
- [4] Lindau, A.; Hohn, T.; Weinzierl, S.: Binaural resynthesis for comparative studies of acoustical environments. In: *122th AES Convention*. Wien, Österreich. 2007.
- [5] Laakso, T.I.; Välimäki, V.; Karjalainen, M.; Laine, U.K.: Splitting the unit delay. *IEEE Sig Proc Mag*, 13(1):30-60. 1996.
- [6] Erbes, V.; Schultz, F.; Lindau, A.; Weinzierl, S.: An extraural headphone system for optimized binaural reproduction. In: *DAGA 2012*. Darmstadt, Germany, S. 313–314. 2012.
- [7] Żychaluk, K.; Foster, D.H.: Model-free estimation of the psychometric function. *Attention, Perception, & Psychophysics*. 71(6):1414-1425. 2009.