

AUSWAHL UND TUNING VON HRTFs

Andreas Silzle
TC Electronic¹

Einleitung

Das Thema dieser Arbeit beschäftigt sich mit der Qualitätsverbesserung von Head Related Transfer Functions (HRTFs, Außenohrübertragungsfunktionen). Bei der Kopfhörerwiedergabe von Lautsprecher-signalen, wie sie normal üblich ist, entsteht eine Imkopflokalisation. Insbesondere bei der Wiedergabe eines fünfkanaligen Signals über Kopfhörer geht sehr viel an Raumeindruck und natürlicher Lokalisation verloren. Ziel der im folgenden beschriebenen Arbeit war es, eine sehr hochwertige fünfkanalige Audiowiedergabe über Kopfhörer zu implementieren. Es war eine ausgezeichnete Außerkopflokalisation für alle Richtungen gefordert. Es sollte die geringstmögliche zusätzliche Klangverfärbung dem Originalsignal hinzugefügt werden. Pilot-hörversuche zeigten, dass auch unter einer größeren Anzahl von an Menschen und Kunstköpfen gemessenen HRTFs kein Satz war, der dem Qualitätsanspruch gerecht werden konnte.

System Definition

Die Idee war es, die fünf Lautsprecherpositionen der standardisierten 5.1 Aufstellung über Kopfhörer zu simulieren. Dafür ist zuerst eine gute Raumsimulation notwendig. Die dann folgenden Außenohrübertragungsfunktionen (HRTFs) erzeugen eine Außerkopflokalisation bei der Kopfhörerwiedergabe und die anschließende invertierte Kopfhörerübertragungsfunktion (IPTF) sorgt für die korrekte Klangfarbe, siehe Abbildung 1.

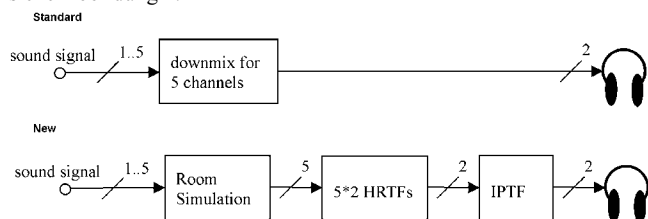


Abbildung 1: Systemdefinition

Problembeschreibung

Die drei genannten Übertragungsblöcke sind notwendig, um die gewünschte Außerkopflokalisation bei Kopfhörerwiedergabe zu erreichen. Sie sollen aber keine zusätzliche Klangfarbenänderung erzeugen. Die Referenzsituation bzgl. der Klangfarbe ist die Standardsituation – eine unprozessierte Wiedergabe des Signals über Kopfhörer. Alle drei Übertragungsblöcke für sich allein erzeugen eine deutliche Klangverfärbung. Insbesondere die HRTFs und IPTFs haben im oberen Frequenzbereich (ab 3kHz) einen ausgeprägten Amplitudengang und sie sind auch noch nach Einführung der Meßmethode mit blockiertem Ohrkanal individuell sehr verschieden, siehe z.B. [1]. Für die angestrebte Applikation war die Verwendung individueller Außenohrübertragungsfunktionen nicht möglich. Es musste deshalb eine mittlere oder typische gefunden werden.

Bei informellen Hörtests wurden die folgenden Schwächen von existierenden HRTFs+IPTFs gefunden. Die Lokalisation ist insbesondere für die frontale Richtung instabil und unpräzise und gegenüber der horizontalen Ebene angehoben, teilweise wird die frontal dargebotene Schallquelle hinten oder gar im Kopf wahrgenommen. Für alle Richtungen in der horizontalen Ebene wurde auch mit der korrekten IPTF eine starke Klangverfärbung gegenüber dem Originalsignal (Standardwiedergabe über Kopfhörer, siehe Abbildung 1) festgestellt.

Lösungsansatz

Es wurde nun die folgende Methode für Auswahl und Tuning von besseren HRTFs und IPTFs angewendet:

1. Auswahl aus einer größeren Anzahl von Übertragungsfunktionen mit direktem Hörvergleich.
2. Tuning dieser Übertragungsfunktionen von einem Tuningexperten bzgl. Klangfarbenminimierung und Lokalisationsoptimierung mit ständiger Hörkontrolle.
3. Überprüfung der gewonnenen Resultate mit einem Hörversuch.

Eine ähnliche Vorgehensweise wurde vom Autor beim Tuning von Mehrkanalaudiokompression (MPEG II) am IRT angewendet, ebenso wie vom Tuningexperten Ulrik Heise für das Einstellen von Raumsimulationen. Da die so gewonnenen Übertragungsfunktionen sehr individuell gefunden wurden, muss mit einem Hörversuch überprüft werden, ob sie allgemeingültig sind.

Parameter der Untersuchung

- 52 menschliche Außenohrkataloge.
- 2 Kunstkopfkataloge.
- Kugelflächenübertragungsfunktionen.
- Mehr als 200 invertierte Kopfhörerübertragungsfunktionen von drei verschiedenen Kopfhörertypen auf über 40 Menschen und Kunstköpfen gemessen.
- Mehr als 15 verschiedene Tonbeispiele (mono, stereo, fünfkanal).
- Mehr als 10 verschiedene Kopfhörer.

Signalverarbeitung

Für die Echtzeitauswahl und das Tuning wurde eine Matlab-Software entwickelt. Für die Signalverarbeitung wurden u.a. benutzt:

- Amplituden- und Phasen-Equalization
- Glättung
- Minimumphasenberechnung

Zur visuellen Kontrolle konnten die Übertragungsfunktionen wahlweise im Zeit- oder Frequenzbereich dargestellt werden, im Frequenzbereich in Amplitude, Phase und Gruppenlaufzeit. Alle Einstellungen konnten abgespeichert werden und ließen so einen Vergleich auf Knopfdruck zu. Nur mit dieser Echtzeitreaktion auf alle Änderungen können die feinen Unterschiede gefunden werden, um den Weg zum angestrebten Ziel zu finden.

Tuning

Die schwierige Aufgabe des Tuningexperten war es:

1. Auswahl einer HRTF, die den besten Kompromiss zwischen Lokalisationsqualität und Klangverfärbung darstellt.
2. Das Tuning dieser HRTF um die Klangverfärbung bzgl. der Referenz zu reduzieren und die Lokalisation zu verbessern.
3. Suche und Tuning einer allgemeingültigen IPTF, die für verschiedene Kopfhörer den besten Kompromiss darstellt.

Test der Ergebnisse mit Hörversuchen

Um die individuellen Ergebnisse des Auswahl- und Tuningprozesses zu verallgemeinern, wurden sie mit einer Reihe von Hörversuchen validiert.

Parameter des Hörversuchs

HRTFs

Quelle	Name
Bester menschlicher Kopf von AUC [2] ²	avh
Kunstkopf Valdemar von AUC	val

¹ Seit 1. Feb. IKA, Ruhr Universität Bochum

² Ausgewählt in einem Lokalisationstest

Kunstkopf Kemar vom MIT gemessen [3]	kem
Getunte AUC HRTFs	atc
Getunte Kemar HRTFs	ktc

Tonsignale

Relativ trockenes Sprachsignal und "normale" Stereo- und Fünfkanaal-Signale.

Raumsimulation

Tests wurden mit und ohne Raumsimulation durchgeführt, um die Qualität der Übertragungsfunktionen allein und mit Raumsimulation zu überprüfen. Es ist klar, dass die Raumsimulation und deren Einstellungen einen großen Einfluss auf das Endresultat haben. Sie waren aber nicht Gegenstand dieser Untersuchung.

Testverfahren

Als Testverfahren wurde ein „double-blind multi-stimulus“ Test gewählt, der an das MUSHRA-Testverfahren angelehnt ist [4].

Klangfarbentest

Parameter

HRTF+IPTF ohne Raumsimulation.

Signal: weibliche Sprache (mono) (SQAM track 49)

Fragestellung

„Ist die Klangfarbe sehr nahe am Referenzsignal, wählen Sie (0). Ist sie schlechter, wählen Sie (-1..-3), ist sie besser, wählen Sie (+1).“³

Ergebnis

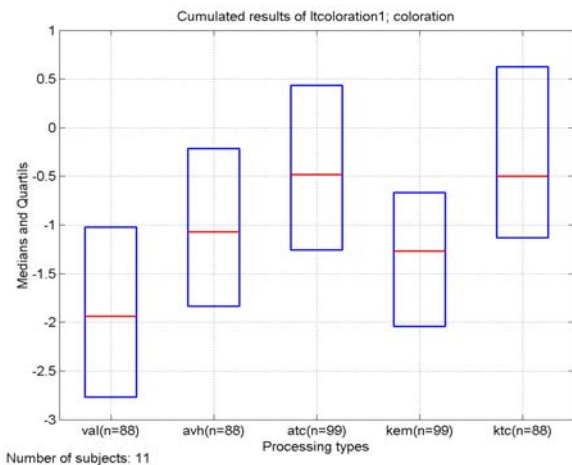


Abbildung 2 Über alle Richtungen gemittelte Klangfarbenergebnisse für Sprache, ohne Raumsimulation. Die Größe n ist die Anzahl der Antworten pro Boxplot.

Die Medianwerte wurden auf signifikante Unterschiede mit dem Kruskal-Wallis-Test getestet, p-value of 0.05.

Getestet	Signifikanter Unterschied zu
val	avh, atc, ktc
avh	val, atc, ktc
atc	val, avh, kem
kem	atc, ktc
ktc	val, avh, kem

Lokalisationstest

Fragestellung

„Ist die Lokalisation an der richtigen Position wählen Sie (5). Besteht ein Fehler von ungefähr 30 Grad wählen Sie (3). Besteht ein Fehler von 90 oder mehr Grad, oder haben Sie eine Imkopflokalisation, wählen Sie (1).“

³ Normalerweise kann in einem Referenztest nichts besser als die Referenz sein, d.h. (0). Aber in Pilottests fanden einige Testpersonen das prozessierte Signal besser als die Referenz. Man kann hier von einer internen Referenz sprechen. Deswegen wurde die Antwortmöglichkeit (+1) eingeführt.

Ergebnis

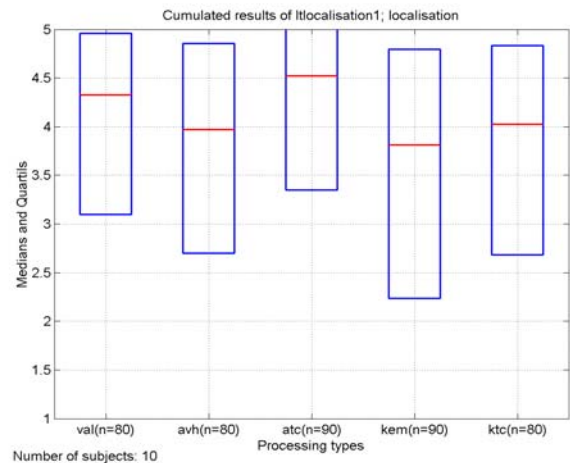


Abbildung 3 Über alle Richtungen gemittelte Richtungsergebnisse für Sprache, ohne Raumsimulation. Die Größe n ist die Anzahl der Antworten pro Boxplot.

Keine signifikanten Unterschiede zwischen den Medianwerten.

Diskussion

Abbildung 2 zeigt, dass die beiden getunten HRTF-Datensätze bzgl. der Klangverfärbung signifikant besser sind als alle anderen. Mit einem Medianwert von -0,5 liegen sie sehr nahe am Original. Abbildung 3 zeigt, dass die Lokalisation von allen HRTF-Datensätzen gut bis sehr gut ist. Der Medianwert des getunten Datensatzes atc ist am höchsten, aber nicht signifikant verschieden von den anderen. Die Medianwerte für den Centerkanal sehen sehr ähnlich aus, nur die unteren Quartilwerte liegen sehr viel niedriger. Es gibt ungefähr einen Anteil von 20% der Testhörer, die eine falsche Lokalisation für die Frontalrichtung haben. Sie wurde durch das Tuning nicht vermindert. Mehr Testergebnisse mit Raumsimulation und mit Stereo- und Fünfkanaalsignalen sind in [5] aufgeführt. Die Ergebnisse für die Frage nach dem Raumeindruck und der Gesamtqualität bei einer Fünfkanaalwiedergabe ähneln Abbildung 2 sehr.

Zusammenfassung

Insgesamt lässt sich sagen, dass eine Auswahl zwischen HRTFs möglich und notwendig ist. Die an Menschen gemessenen Übertragungsfunktionen klingen besser (mehr natürlich) als die der verwendeten Kunstköpfe. Ein Tuning dieser ausgewählten HRTFs einzeln für jede Richtung und selbst für links und rechts ist möglich und resultiert in einer deutlichen Verbesserung bzgl. der Klangverfärbung. Für die Lokalisation von Monoquellen, die sich bei allen Datensätzen auf einem hohen Niveau befindet, konnte keine signifikante Verbesserung mehr erreicht werden, im Gegensatz zu den Attributen Raumeindruck und Gesamtqualität bei Fünfkanaalwiedergabe. Die Übertragungsfunktionen sind im Engage-Algorithmus auf dem System 6000 von TC Electronic implementiert. Sie sind teilweise patentrechtlich geschützt.

References

- [1] Møller, Henrik; et al.; Head-Related Transfer Functions of Human Subjects; JAES, Vol. 43, No. 5, 1995 May
- [2] Møller, Henrik; et al.; Using a Typical Human Subject for Binaural Recording; 100th AES Convention, 1996 May 11-14, Copenhagen, preprint 4157
- [3] Bill Gardner and Keith Martin, HRTF Measurements of a KEMAR Dummy-Head Microphone, 1994; <http://sound.media.mit.edu/KEMAR.html>
- [4] Stoll, Gerhard; Koyamernik; Franc; EBU listening tests on Internet audio codecs; EBU Technical Review; June 2000
- [5] Silzle, Andreas; Selection and Tuning of HRTFs; 112th AES Convention, May 2002, Munich.