

# Evaluierung des virtuellen Kopfhörers für die dynamische binaurale Synthese

Jan Sokoll<sup>1</sup>, Tobias Lentz<sup>1</sup>, Ingo Assenmacher<sup>2</sup>

<sup>1</sup> *Institut für Technische Akustik, RWTH Aachen, 52056 Aachen, Deutschland, Email: Jan.Sokoll@akustik.rwth-aachen.de*

<sup>2</sup> *Rechen- und Kommunikationszentrum, RWTH Aachen, 52056 Aachen, Deutschland, Email: assenmacher@rz.rwth-aachen.de*

## Einleitung

Eine Virtuelle Realität (VR) ist eine computergenerierte Umgebung, die dem Benutzer die Möglichkeit der Echtzeitinteraktion mit der präsentierten Szene gibt. Eine wichtige Eigenschaft der VR ist die dreidimensionale (3D) und multimodale Schnittstelle zwischen Benutzer und Computer. Zusätzlich zur visuellen Komponente sollten andere menschliche Sinne wie Akustik und Haptik in die Interaktion integriert werden, um eine möglichst große Immersion in die virtuelle Welt zu erreichen. Eine sehr einfache 3D Optik kann schon auf einem herkömmlichen Fernseher mit einer entsprechenden Brille dargeboten werden, die korrekte Reproduktion von 3D Audio ist hingegen viel komplizierter. Es gibt diverse Ansätze um das Problem 3D Audio realistisch darzubieten zu lösen. Am Institut für Technische Akustik (ITA) der RWTH Aachen wird seit mehr als 10 Jahren eine dynamische Übersprechkompensation (Cross-Talk Cancellation, CTC) entwickelt um binaurale Signale an einem so genannten „Sweet Spot“ reproduzieren zu können und diesen nachzuführen. 3D Audio kann damit sehr realistisch für einen Benutzer wiedergegeben werden. Dieses System wird auch als virtueller Kopfhörer [1] bezeichnet und soll nun evaluiert werden.

Für eine objektive Analyse des Systems muss die Kanaltrennung der CTC gemessen werden. Dazu wurde eine statische CTC [2] aufgebaut, um die bestmögliche Kanaltrennung als Referenz zu ermitteln. Dann wurden die Eigenschaften der dynamischen CTC untersucht. Für eine subjektive Analyse müssen Hörversuche durchgeführt werden. Es wurden die Lokalisierungseigenschaften des virtuellen Kopfhörers und der binauralen Synthese für einen Benutzer in einer VR Umgebung evaluiert. Dabei wurden sowohl nur akustische Stimuli als auch eine Kombination aus akustischen und visuellen Stimuli verwendet.

Seit Juni 2004 ist im Rechen- und Kommunikationszentrum der RWTH Aachen eine CAVE (Cave Automatic Virtual Environment) installiert. Diese CAVE ist eine stereoskopische Projektionseinheit, mit vier Wänden und dem Boden als Projektionsflächen. Zehn Projektoren mit UXGA Auflösung und polarisiertem Licht werden benutzt um 3D Bilder einem Benutzer darzubieten. Abbildung 1 zeigt ein Bild der RWTH CAVE. Der virtuelle Kopfhörer und die binaurale Synthese sind in dieses System integriert, so dass eine volle 360° virtuelle Realität mit einer optischen und einer akustischen Komponente kreiert werden kann [3].

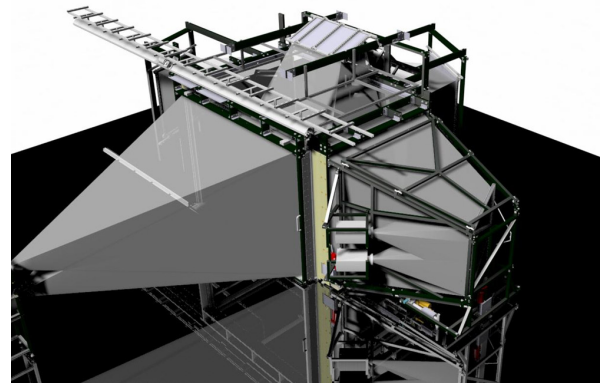


Abbildung 1: Die CAVE der RWTH Aachen

## Messungen

Der wichtigste Faktor für eine ausreichende Qualität der CTC ist die Kanaltrennung am Ohr des Hörers. Abbildung 2 Kurve b) zeigt die Kanaltrennung für eine individuell eingemessene statische CTC. Die späten Reflexionen wurden hierbei abgeschnitten und die Impulsantwort wurde mit einem 1/3 Oktaven Fenster geglättet. Die Filterlänge wurde auf 1024 Taps begrenzt, um einen Vergleich mit dem dynamischen System zu ermöglichen. Abbildung 2 Kurve a) zeigt die berechnete Kanaltrennung, Kurve c) die der dynamischen CTC. Im Unterschied zum statischen System werden bei der dynamischen CTC keine individuellen Filter benutzt sondern die Filter mit HRTFs aus einer Datenbank online berechnet.

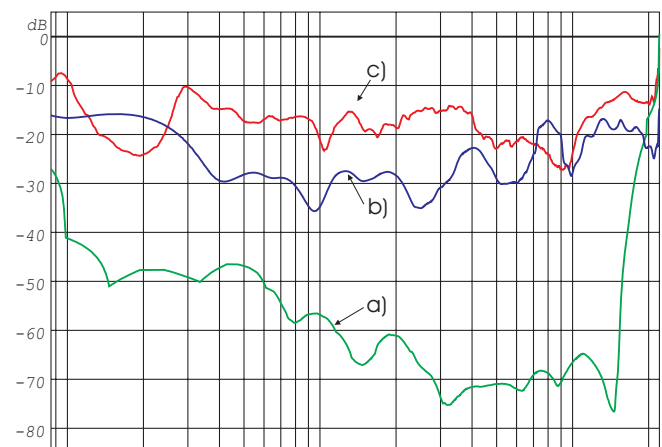
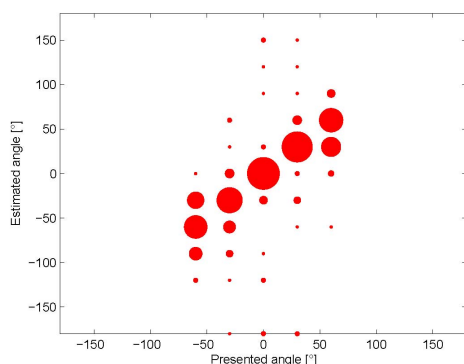


Abbildung 2: Kanaltrennung bei einer Filterlänge von 1024 taps: Kurve a): berechnete Kanaltrennung, Kurve b): statische CTC, Kurve c): dynamische CTC

## Hörversuche

Bis jetzt wurde nur die technische Leistung der Übersprechkompensation evaluiert. Mit den Messungen ist eine objektive Bewertung der Kanaltrennung möglich. Um das System aber genauer zu untersuchen, müssen Hörversuche durchgeführt werden. Die subjektiven Lokalisationseigenschaften des virtuellen Kopfhörers in Verbindung mit der binauralen Synthese können so bestimmt werden.

Die Hörversuche wurden in der speziellen Umgebung der CAVE durchgeführt. Durch die reflektierenden Wände wird die Lokalisationsleistung verringert, da die Kanaltrennung an den Ohren des Benutzers nicht mehr optimal ist. Die Lokalisationsleistung bei der Benutzung von „echten“ Kopfhörern wurde von [4] untersucht. In einer multi-sensorischen Umgebung wie der CAVE helfen akustische Stimuli, welche nicht im Sichtfeld des Benutzers liegen, bei der Orientierung. In diesem Zusammenhang ist die vorne/hinten Vertauschung ein wichtiger Faktor. Aus diesem Grund wurden die Stimuli beim ersten Hörversuch hinter dem Benutzer platziert.

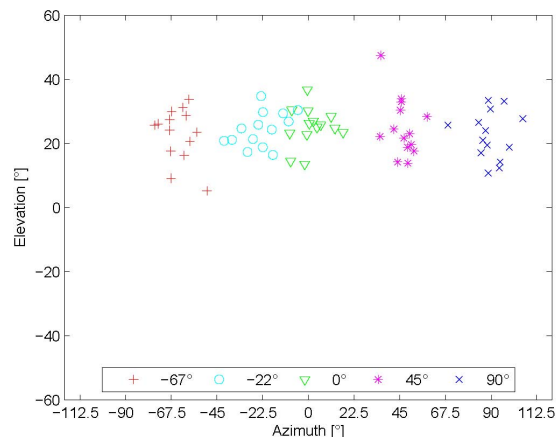


**Abbildung 3:** Lokalisierungsergebnisse für die virtuellen Quellen in der Horizontalebene hinter dem Benutzer.

Eine komplette Drehung des Körpers war nicht erlaubt, die Bewegung des Kopfes dagegen sehr wohl. Die Lokalisation in der Azimuthebene war sehr gut, vorne/hinten Vertauschungen traten kaum auf. Die Lokalisation in der Medianebene war dagegen sehr schwer. Abbildung 3 zeigt die dargebotenen und empfundenen Azimutwinkel dieses Hörversuchs.

In einem weiteren Hörversuch wurden die akustischen Stimuli durch visuelle Stimuli ergänzt. Der Benutzer konnte diesen Test selbständig mittels eines Flight-Sticks durchführen und durfte sich frei in der CAVE zu bewegen. Auch hier war die Lokalisation in der Horizontalebene sehr gut, in der Medianebene schwer. Abbildung 4 zeigt die Lokalisierungsergebnisse für die Quellen mit einer Elevation von  $30^\circ$ .

In einem letzten Test konnten die Benutzer eine kleine virtuelle Welt namens „Deep Forest“ erkunden, die mit diversen interaktiven Elementen versehen war. Dieser Test zeigte, dass die Lokalisation von akustischen Stimuli auch in der Medianebene sehr gut funktioniert sobald entsprechende visuelle Stimuli vorhanden sind.



**Abbildung 4:** Ergebnisse des Hörversuchs mit simultan präsentierten akustischen und visuellen Stimuli für Quellen mit einer Elevation von  $30^\circ$ .

## Zusammenfassung / Ausblick

In diesem Paper wurden die Eigenschaften eines 3D-Audio-Wiedergabesystems basierend auf dynamischer Übersprechkompensation in Kombination mit binauraler Synthese bezüglich objektiver und subjektiver Parameter evaluiert. Aufgrund von nicht individualisierten HRTFs für die Filterberechnung bei der dynamischen CTC ist die Kanaltrennung natürlich nicht so gut wie die des statischen Systems, aber immer noch befriedigend. Während der Hörversuche wurden rein akustische Stimuli als auch eine Kombination aus akustischen und optischen Stimuli dargeboten. Die Lokalisierungseigenschaften in der Horizontalebene sind sehr gut, in der Medianebene aber leider noch nicht ausreichend. Der Test in einer „echten“ virtuellen Realität hat gezeigt, dass die Wiedergabe von 3D Audio in einer komplexen virtuellen Szene sehr gut funktioniert. Die Ergebnisse dieser Evaluierung werden Einfluss auf die Weiterentwicklung des virtuellen Kopfhörers und der binauralen Synthese nehmen.

## Literatur

- [1] LENTZ, T. and BEHLER, G. Dynamic cross-talk cancellation for binaural synthesis in virtual reality environments. Proceedings of the 117th Audio Engineering Society Convention San Francisco, USA, 2004.
- [2] ATAL, B.S. and SCHRÖDER, M. R. Journal of the AES, 1967.
- [3] ASSENMACHER, I., LENTZ, T., KUHLEN, T. and VORLÄNDER, M. Integrating real-time binaural acoustics into vr applications. In Virtual Environments 2004, Eurographics ACM SIGGRAPH Symposium Proceedings, pages 129–136, June 2004.
- [4] BEGAULT, D. R. Challenges to the successful implementation of 3-d sound. Journal of the Audio Engineering Society, 39(11):864–870, 1991.