

Echtzeit-Raumakustiksimulationen für virtuelle Umgebungen

Tobias Lentz¹, Ingo Assenmacher²

¹ RWTH Aachen, Institut für Technische Akustik, 52056 Aachen, Deutschland, Email: tobias.lentz@akustik.rwth-aachen.de

² RWTH Aachen, Rechen- und Kommunikationszentrum, 52056 Aachen, Deutschland, Email: assenmacher@rz.rwth-aachen.de

Einleitung

In der virtuellen Realität werden dem Benutzer künstliche im Rechner generierte Szenen dargeboten. Diese Szenen sind in erster Instanz plastische Objekte, die mit geeigneter 3D Videotechnik dargestellt werden. Um die Immersion, also das „Eintauchen“ in eine solche Szene zu erhöhen ist die Akustik sehr hilfreich, da ein weiteres Sinnesorgan mit in die virtuelle Umgebung einbezogen wird. Komplexere virtuelle Umgebungen sind Räume, in denen dem Benutzer die Möglichkeit gegeben werden soll, in dieser Szene zu agieren bzw. mit bestimmten Objekten zu interagieren. Da in der virtuellen Szene keine Freifeldbedingungen herrschen sollen, bedeutet dies für den akustischen Teil, ebenfalls die Raumakustik zu berechnen und entsprechend der Benutzeraktion anzupassen. Hier ist also eine Echtzeit-Raumakustiksimulation für virtuelle Umgebungen Voraussetzung, um einen zumindest plausiblen Eindruck zu vermitteln.

VR Umgebungen

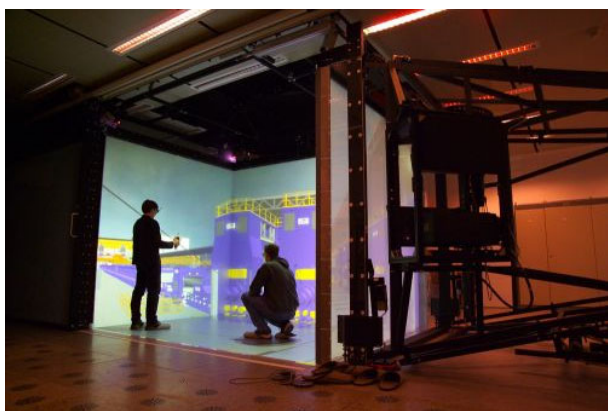


Abbildung 1: Blick in die CAVE zur Erzeugung virtueller Szenen.

Abbildung 1 zeigt die Anlage des Rechen- und Kommunikationszentrums der RWTH Aachen zur Erzeugung einer virtuellen Realität. Diese 5-Seiten Projektionseinrichtung, „CAVE“ genannt, gestattet es dem Benutzer sich auf einer Fläche von 2,7m x 3,6m frei zu bewegen. Die Position wird dabei über ein optisches Trackingsystem erfasst um die stereoskopischen Bilder korrekt zu rendern.

Der korrekte akustische Eindruck wird durch die übersprechkompensierte Wiedergabe binauraler Signale bewerkstelligt, die es ermöglicht, an den Ohren des Benutzers ein Schallereignis mit korrektem räumlichem Bezug zu reproduzieren. Diese dynamische

Übersprechkompensation berechnet anhand der aktuellen Position des Benutzers die entsprechenden Kompensationsfilter. Da in einem solchen VR-System aufgrund der Bildgenerierung immer die aktuelle Position bekannt ist, wird keine zusätzliche Bewegungserfassung benötigt. Der Vorteil einer Darbietung der binauralen Signale über die Übersprechkompensation ist, dass 3 bis 4 Lautsprecher genügen, um eine räumlich korrekte Positionierung der virtuellen Quellen zu erreichen. Weitere Details finden sich unter [2].

Echtzeit-Anforderungen

Zentrale Punkte bei der Betrachtung eines Echtzeit-Systems sind **Update-Rate** und **Latenz**.

Um eine Abschätzung der benötigten Update-Rate durchzuführen wurden die Bewegungen von Benutzern der CAVE für alle sechs Freiheitsgrade geloggt und ausgewertet. Dabei sind hier besonders die Geschwindigkeiten der Kopfdrehung und der Translation in beliebiger Richtung von Bedeutung. Mit diesen Daten kann das „typische“ Benutzerverhalten und die maximale Anforderung an die Update-Rate der raumakustischen Simulation getrennt für die einzelnen Bewegungen bestimmt werden.

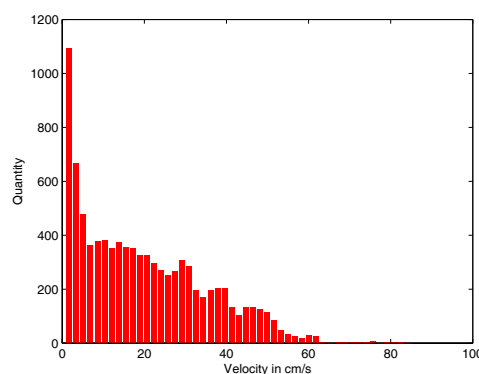


Abbildung 2: Bewegungsprofil der Translation

Aus Abbildung 2 wird deutlich, dass die translatorische Bewegung eines Benutzers in der CAVE hauptsächlich unter ca. 30 - 35cm/s liegt. Spitzenwerte können aber bis 1m/s auftreten. Hieraus lässt sich ableiten, dass die Update-Rate der raumakustischen Berechnungen bei translatorisch bewegtem Empfänger relativ gering gehalten werden kann, da sich der Höreindruck in einem Raum nicht dramatisch über den Weg ändert (siehe auch Untersuchungen von WITEW [1]). Nimmt man beispielsweise für einen Konzertsaal als Grenze für eine Neuberechnung

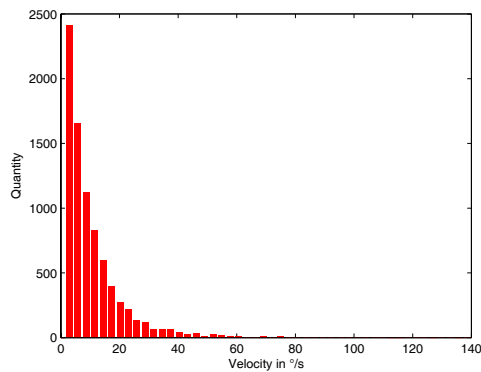


Abbildung 3: Bewegungsprofil der Rotation

der Raumimpulsantwort den halben Sitzabstand von ca. 25cm, muss man die Berechnung laut Bewegungsprofil alle 750ms, um die meisten Bewegungen abzudecken, besser alle 250ms, um auch die Spitzenwerte abzudecken, durchführen. In VR-Systemen besteht allerdings auch die Möglichkeit durch einen Raum zu navigieren. Die Position des Benutzers in der CAVE ändert sich nicht, sondern der Raum wird verschoben. Dies erfordert ebenfalls eine Neuberechnung. Für diese Navigation ist das System dann auf die oben ermittelte Geschwindigkeit von 1m/s zu limitieren.

Abbildung 2 zeigt das Bewegungsprofil der Kopfdrehungen eines Benutzers. Es werden Winkelgeschwindigkeiten bis zu $100^\circ/\text{s}$ erreicht. Die Orientierung des Empfängers im Schallfeld ist wesentlich kritischer, da die einzelnen Reflexionen bei der Berechnung mit den entsprechenden Außenohrimpulsantworten beaufschlagt werden. Dreht der Benutzer den Kopf, werden die Reflexionen aus einer falschen Richtung wahrgenommen und passen schnell nicht mehr zum optischen Eindruck des Raumes. Bei der Überblendung von einer Raumimpulsantwort zur nächsten sollten keine allzu grossen Änderungen auftreten, da sonst Kammerfiltereffekte entstehen. Hier hat sich ein Filterwechsel alle 2° besser alle 1° als nötig erwiesen, was einem Zeitintervall für die Neuberechnung von 20ms bzw. 10ms entspricht.

In diesem Zusammenhang ist auch die Latenz der Signalverarbeitung bzw. der Audioausgabe von hohem Interesse. Da bei einer Kopfdrehung alle 10 - 20ms ein Filterwechsel stattfinden kann, darf die bei PCs übliche blockweise Übergabe der Audiosamples zur Soundkarte eine Blockgrösse von 512 Samples (11.6ms bei 44,1kHz Abtastrate) besser jedoch 256 Samples (5,8ms) nicht überschreiten, berücksichtigt man noch die Laufzeit des Schalls für eine Strecke von ca. 2m. Das gleiche gilt für den Faltungsalgorithmus, der die aktuelle Impulsantwort auf das auszugebende Signal rechnet. Hier muss ebenfalls mit der gleichen Blocklänge wie bei der Soundausgabe gearbeitet werden. Diese segmentierte Faltung ist rechenzeitintensiv, kann aber mit geeigneten Optimierungen auf einem aktuellen PC-System mit bis zu 10s (binaural) betrieben werden.

Zur Erfüllung dieser Anforderungen bietet sich die Nutzung des effizienten Spiegelschallquellenverfahrens auf der Grundlage einer binären Baumstruktur zur Organisation an (siehe SCHRÖDER [4]).

Der Vorteil dieses Verfahrens ist die Möglichkeit viele rechenintensive Operationen in das Preprocessing zu verlegen, sodass eine effizientere Berechnung zur Laufzeit gegeben ist. Weitere Pluspunkte sind die verschiedenen Einsprungpunkte, die der Algorithmus bietet. So ist beispielsweise die zeitkritische Berechnung der Impulsantwort bei der Drehung des Benutzers sehr schnell möglich, da lediglich die Einzel-HRTFs für den Filter in korrekter zeitlicher Abfolge überlagert werden müssen. Da alle Spiegelschallquellen bereits berechnet sind, kann eine hohe Geschwindigkeit für diese Operation erreicht werden. Bei einer mittleren Komplexität des Raumes von ca. 40 Polygonen entstehen und ca. 200 hörbaren Spiegelschallquellen ist die Filterberechnung in einer Zeit von 5ms möglich (PC-System P4 3GHz).

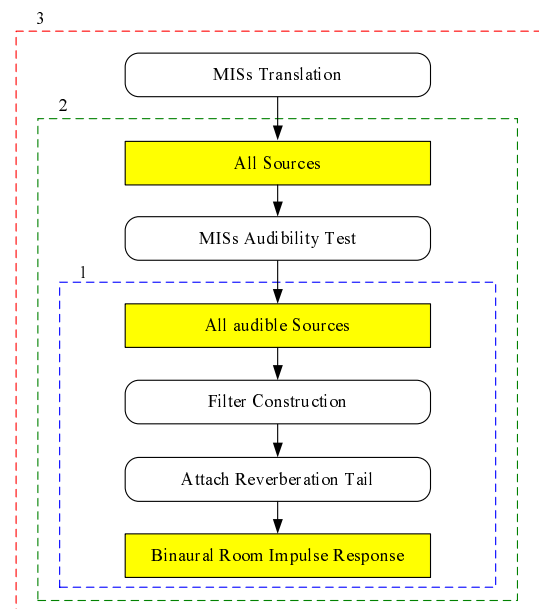


Abbildung 4: Verschiedene Einsprungpunkte: 1) Quelle ist bewegt worden 2) Benutzer hat sich bewegt 3) Orientierung einer Quelle oder des Benutzers hat sich geändert

Literatur

- [1] Witew, I., *Spatial variation of lateral measures in different concert halls*, Proc. 18th ICA Kyoto, Vol4, pp 2949, 2004.
- [2] Lentz, T., Behler, G., *Dynamic Cross-Talk Cancellation for Binaural Synthesis in Virtual Reality Environments*, Audio Engineering Society Convention 117, Preprint Number: 6315, 2004
- [3] Lehnert, H., *Binaurale Raumsimulation: Ein Computermodell zur Erzeugung virtueller auditiver Umgebungen*, Dissertation Ruhr-Universität Bochum, 1992.
- [4] Schröder, D., Lentz, T., *Echtzeit-Implementierung des Spiegelschallquellenverfahrens*, DAGA, 2005.