

Multimodale Darstellung natürlicher Schallquellen in virtuellen Umgebungen

Dirk Schröder¹, Stefan Reuter¹, Gottfried K. Behler¹, Ingo Assenmacher²

¹ Institut für Technische Akustik, RWTH Aachen, 52066 Aachen, Deutschland

² VR-Gruppe des Rechen- und Kommunikationszentrums, RWTH Aachen, 52074 Aachen, Deutschland

E-Mail: dirk.schroeder@akustik.rwth-aachen.de

Einleitung

In interaktiven, virtuellen Umgebungen ist eine realistische Repräsentation von natürlichen Schallquellen (z.B. Musikern, siehe Abb. 1) notwendig, um beim Nutzer eine Akzeptanz der dargebotenen Szenerie hervorzurufen. Um den präsentierten Szenarien eine gewisse Natürlichkeit zu verleihen, z.B. bei der Teilnahme an einer musikalischen Darbietung in einem virtuellen Konzertsaal, müssen zumindest die beiden dominantesten Sinne, das heißt Hören und Sehen, adäquat stimuliert werden. Je überzeugender diese Stimulation erfolgt, desto leichter verliert der Benutzer das Gefühl sich in einer künstlichen Welt zu befinden (Grad der Immersion).

Während bei der visuellen Komponente eine synchrone Bewegung des Musikers zur gespielten Musik gefordert ist, benötigt man für die akustische Komponente die Verknüpfung der Richtcharakteristiken der Schallquellen mit der raumakustischen Echtzeitsimulation. Speziell im Fall des Spielens von Musikinstrumenten, bei denen die akustische Richtcharakteristik stark über den Frequenzbereich variiert und der Betrachter eine gewisse Bewegungscharakteristik vom Musiker erwartet, wird die Glaubwürdigkeit der Simulation natürlicher Schallquellen durch einen solchen multimodalen Ansatz stark erhöht.



Abbildung 1: Exploration eines virtuellen Konzertsaaes

Messung von Quell-Richtcharakteristiken

In einer ersten Richtcharakteristik-Messreihe wurden verschiedene Musikstücke von Musikern auf 5 verschiedenen Instrumenten (Saxophon, Gitarre, Klarinette, Trompete, Posaune) eingespielt, die dabei mit Hilfe eines 32-kanaligen Messaufbaus aufgenommen wurden. Dieser Messaufbau besitzt die Form eines Pentakisdodekaeder (siehe Abb. 2), wobei in den 32 Ecken jeweils ein Mikrofon

(Sennheiser KE4) installiert ist. Die 32 Mikrofone befinden sich daher im gleichen Abstand zum Mittelpunkt (ca. 2.2m) und liegen somit auf einer gemeinsamen Kugeloberfläche.

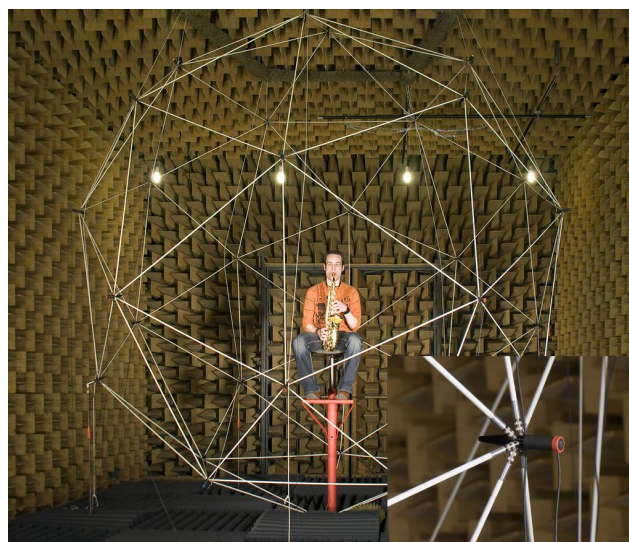


Abbildung 2: Richtcharakteristik-Messaufbau in Form eines Pentakisdodekaeders (32 Messkanäle)

Der Aufbau wurde so aufgestellt, dass sich über und unter dem Mittelpunkt (Elevationswinkel 90° und -90°) jeweils ein Mikrofon befand. Alle weiteren Mikrofone lassen sich so in sechs Ebenen einteilen (Elevationswinkel $53^\circ, 21^\circ, 11^\circ, -11^\circ, -21^\circ, -53^\circ$), die jeweils ein gleichmäßiges Fünfeck bilden. Hierbei liegen die ersten Mikrofone von drei Ebenen ($53^\circ, 11^\circ, -21^\circ$) in Blickrichtung des Musikers (Azimuthwinkel 0°), die der drei anderen Ebenen ($21^\circ, -11^\circ, -53^\circ$) liegen jeweils 36° gegen den Uhrzeigersinn versetzt.

Während der Aufnahme saßen die Musiker auf einem höhenverstellbaren Stuhl, wobei sich das jeweilige Instrument im Mittelpunkt des Aufbaus befand. Zusätzlich wurde für die spätere Auralisation ein weiterer Kanal mit Hilfe eines zusätzlichen Mikrofons (Brüel&Kjaer 4190) aufgenommen, welches außerhalb des Messaufbaus in Hauptabstrahlrichtung des jeweiligen Instruments platziert wurde. Alle Mikrofonsignale wurden zeitgleich mit 33 A/D Wandlern bei $44,1 \text{ kHz}/16 \text{ Bit}$ abgetastet und als unkomprimierte WAV-Dateien abgespeichert.

Dateninterpolation

Für jeden der 32 Kanäle wurden die Terzpegel bestimmt, wobei die Blocklänge der DFT zu 22050 Samples gewählt wurde, um eine ausreichende Frequenzauflösung für die Analyse zu gewährleisten. Die Mittelungslänge bei den Auswertungen wurde variiert, wobei zu beobachten war, dass sich erst mit steigender Mittelungsdauer in allen

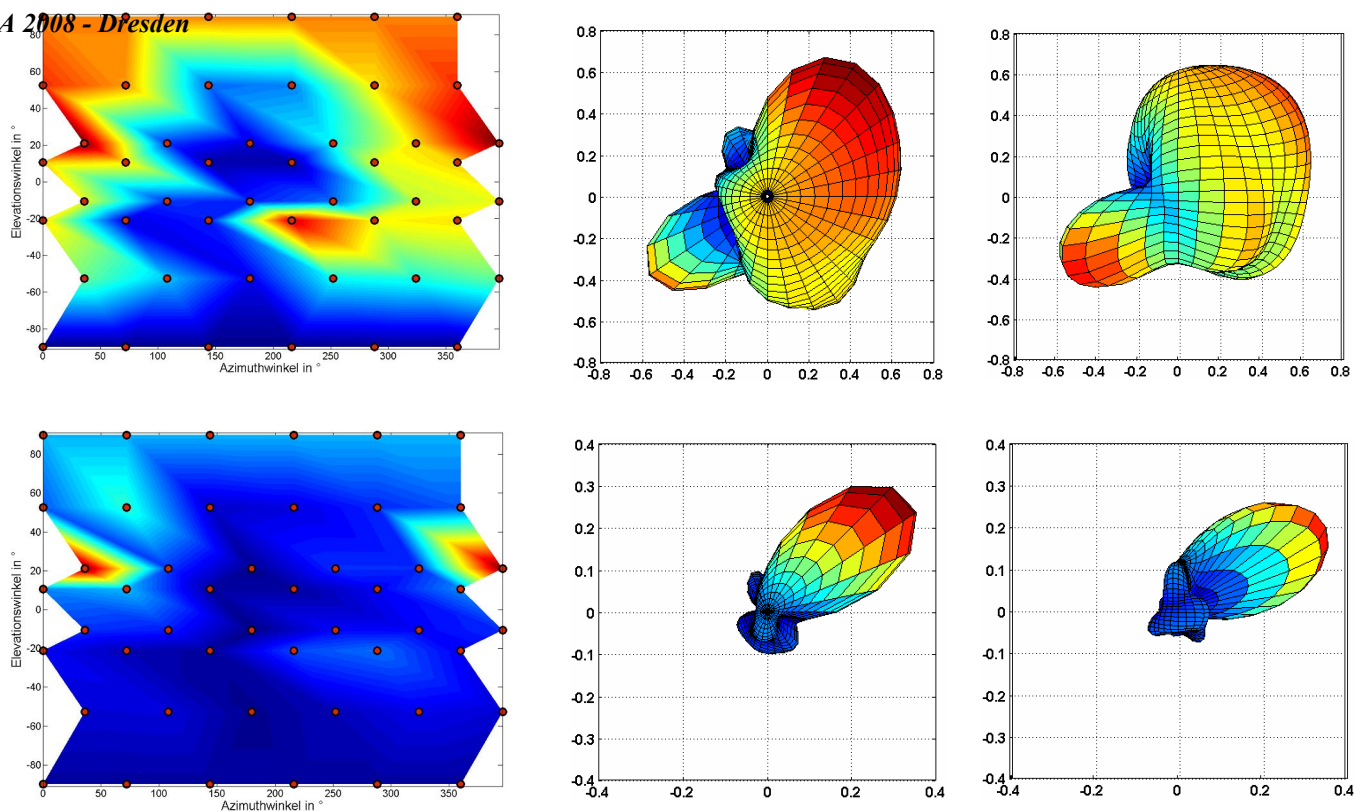


Abbildung 3: Richtcharakteristik eines Saxophons bei 400Hz (obere Reihe) und 6kHz (untere Reihe). Analyselänge 150s. Linke Spalte: gemittelte Rohdaten (Punkte: Messdaten), Mittlere Spalte: Interpolierte Daten (Draufsicht), Rechte Spalte: Interpolierte Daten (Seite)

Terzbändern plausible Richtcharakteristiken berechnen lassen. Anschließend wurden die so berechneten Terz-Pegelwerte der 32 Kanäle interpoliert (Beispiel siehe Abb. 3) und als Datenbank mit einer 5°-Winkelaufösung exportiert, um einen schnellen Datenzugriff während der Echtzeit-Auralisation zu ermöglichen. Dabei wurden zwei verschiedene Interpolationsverfahren verwendet: eine kartesische Spline-Interpolation und ein Sampling-Grid-Verfahren[1] einer Spherical Harmonics Zerlegung (bis 5. Ordnung). Ein Vergleich der beiden Methoden ergab für Musikinstrumente mit starken Konzentrationen in der Richtcharakteristik keine großen Unterschiede. Die Spline-Interpolation verursachte jedoch lokale Sprungstellen und glättete kleine lokale Maxima, wohingegen beim Sampling-Grid-Verfahren diese Probleme nicht auftraten. Es ist daher anzunehmen, dass bei Musikinstrumenten mit komplexeren Richtcharakteristiken (z.B. Violine, Viola, Cello) eine Spline-Interpolation die Ergebnisse zu stark verfälscht.

Motion Tracking

In einem weiteren Schritt wurden die Bewegungen der Musiker mittels eines optischen Tracking-Verfahrens aufgezeichnet. Bei dieser Art des Trackings werden spezielle Marker von Infrarot-Kameras erfasst, wodurch eine Berechnung von Ort und Orientierung der Marker ermöglicht wird. Optische Verfahren besitzen allerdings immer das Problem, dass Marker durch die Person verdeckt werden können, weshalb eine Nachbearbeitung der Tracking-Daten generell nötig ist. Dieses Problem besitzen magnetische Tracking-Verfahren nicht, allerdings würden metallene Instrumenten das Tracking zu stark stören.

Um den Aufwand der Datennachbearbeitung so gering wie möglich zu halten, wurden die Musiker gebeten, sich möglichst synchron zu den vorher eingespielten

Masterspuren zu bewegen. Für die Aufzeichnung der Bewegungsprofile kamen hierbei jeweils 8 Marker zum Einsatz (Kopf, Oberarm, Unterarm, Brust, Becken), die von den optischen Trackern während den Darbietungen erfasst und anschließend als BVH-Datei (Biovision Hierarchy) exportiert wurden. Mit Hilfe der freien Software Blender konnten nun einzelne Bewegungsprofile extrahiert werden, die während der Echtzeit-Simulation frei miteinander kombinierbar sind.

Echtzeit-Simulation

Für die multimodale Echtzeit-Simulation der Musiker wurden zwei Applikationen benutzt. Für den auditiven Stimulus kam dabei die Raumakustik-Software RAVEN[2] zum Einsatz, die momentan am Lehrstuhl für Technische Akustik, RWTH Aachen, entwickelt wird. Auf visueller Seite wurde für die Abbildung virtueller Humanoider die Software VRZula[3] benutzt. Beide Applikation sind in das bestehende VR-Rahmenwerk VISTA[4] integriert.

Literatur

- [1] Pollow, M., Variable Directivity of Dodecahedron Loudspeakers, Diplom-Arbeit, Institut für Technische Akustik, RWTH Aachen, 2007
- [2] Schröder, D., Vorländer, M., Hybrid method for room acoustic simulation in real-time, 19th International Congress on Acoustics (ICA), Madrid, Spain, 2007
- [3] Valvoda, J. T., Virtual Humanoids and Presence in Virtual Environments, Dissertation, Fakultät für Mathematik, Informatik und Naturwissenschaften, RWTH Aachen, 2007
- [4] <http://www.rz.rwth-aachen.de/ca/c/piz/lang/de/>