Design und Realisierung von virtueller Akustik für ein Augmented-Reality-Labor

Florian Völk, Stefan Kerber, Hugo Fastl, Stefan Reifinger
AG Technische Akustik, Lehrstuhl für Mensch-Maschine-Kommunikation, TU München
Email: florian.voelk@mytum.de

Einleitung

Seit vielen Jahren ist die Möglichkeit bekannt, dem menschlichen Hörsystem die Existenz von real nicht vorhandenen Schallquellen vorzutäuschen, indem Quellsignale mit binauralen Raumimpulsantworten gefaltet und mittels Kopfhörern dargeboten werden (vgl. [1], [2]). Erfasst man ständig Position und Orientierung des Anwenders, ist es möglich, das System dynamisch zu gestalten und dadurch die erreichbare Realitätsnähe weiter zu steigern (siehe [3]). Die ständig wachsende Leistungsfähigkeit der für ein solches Vorgehen notwendigen Hardware erschließt neue bzw. erweiterte Anwendungsfelder, vor allem durch die zunehmende Echtzeitfähigkeit. Virtuelle Szenen mit mehreren, falls nötig bewegten Schallquellen können generiert und während des laufenden Betriebes manipuliert werden.

Im vorliegenden Beitrag wird die Konzeption und Realisierung eines nach dem beschriebenen Prinzip arbeitenden Systems zur Erzeugung von bis zu acht ruhenden oder bewegten virtuellen Schallquellen in der Horizontalebene vorgestellt. Dieses wurde insbesondere zur Ergänzung von Augmented-Reality-Umgebungen (AR-Umgebungen) um eine akustische Komponente entwickelt. Generierte virtuelle Szenen sind daher interaktiv modifizierbar. Das System wird im folgenden als *virtuelle Akustik* bezeichnet.

Aufbau des Systems

Um möglichst vielfältige Anwendungsmöglichkeiten für psychoakustische Experimente sowie eine einfache Integrierbarkeit in vorhandene AR-Umgebungen zu gewährleisten, wurde die virtuelle Akustik modular geplant und realisiert. Abbildung 1 zeigt die Soft- und Hardwarekomponenten sowie die wichtigsten Kommunikationspfade des Systems.

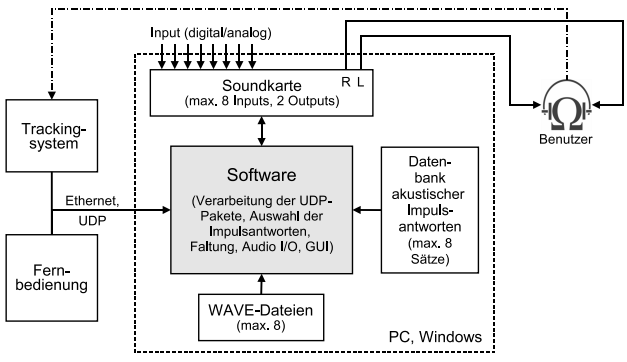

Abbildung 1: Komponenten und Kommunikationspfade der kopfhörerbasierten, PC-gestützten virtuellen Akustik.

Die einzelnen Module werden von einer zentralen Softwarekomponente (grau hinterlegt) gesteuert, die auf einem PC (gestrichelte Linie) unter den neueren Versionen von Microsoft Windows läuft. Zusätzlich zur Steuerung des Systems übernimmt diese Software auch die gesamte notwendige Audiosignalverarbeitung (siehe Abschnitt *Signalverarbeitung*). Virtuelle Schallquellen können in der durch die Oberkanten der Gehörgänge und Unterkanten der Augenhöhlen des Benutzers definierten Horizontalebene erzeugt werden. Entweder erfolgt die Positionierung über die graphische Bedienoberfläche (siehe Abb. 2) oder mittels eines definierten Befehlssatzes über Ethernetpakete (UDP/IP). Durch wiederholtes Setzen der Position können auf letztgenannte Weise auch bewegte Quellen synthetisiert werden.

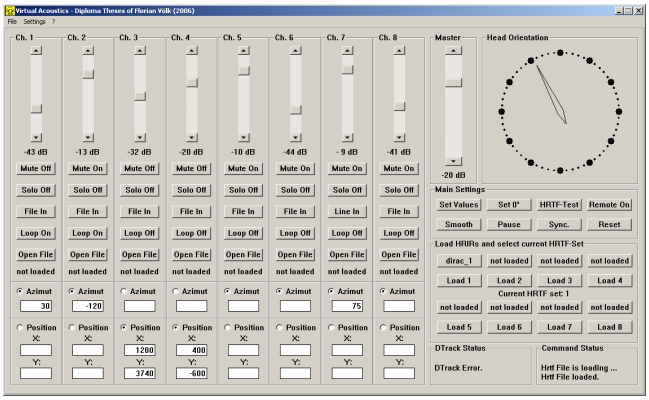

Abbildung 2: Screenshot der lokalen Bedienoberfläche. Das System läuft unter den neueren Windows-Versionen.

Zur Ermittlung von Position und Orientierung des Benutzers wird markerbasiertes Kameratracking verwendet (A.R.T. GmbH, siehe [5]). Dazu wurde auf dem benutzten Kopfhörer ein reflektierender Marker angebracht, der von sechs Infrarotkameras beleuchtet und erfasst wird. Ein zentraler PC berechnet aus den Informationen der Kamerabilder ständig die dreidimensionale Lage des Markers und sendet diese über Ethernet (UDP/IP) an die Software der virtuellen Akustik. Diese wählt auf Grundlage der vom Trackingsystem empfangenen Daten und der vom Benutzer gewählten Quellpositionen für jede Schallquelle die zur aktuellen Hörsituation passenden Impulsantworten aus einer Datenbank, faltet diese mit dem zugehörigen Audiosignal und gibt die Faltungsprodukte über Soundkarte und Kopfhörer aus.

Schallsignale ohne Richtungsinformationen können von maximal acht digitalen oder analogen Mono-Eingängen der Soundkarte oder als Mono-Dateien im Wave-Format von der Festplatte gelesen werden. Binaurale Raumim-

pulsantworten müssen als Sätze in die Datenbank geladen werden, die pro Ohr 360 Impulsantworten (eine pro Winkelgrad in der Horizontalebene) umfassen. Maximal acht solcher Sätze können gleichzeitig geladen werden, wobei der aktuell zu verwendende online wählbar ist. Die Einschränkung auf in der Horizontalebene befindliche Schallquellen resultiert aus einem an die vorgesehene Anwendung angepassten Kompromiss zwischen erzielbarer Authentizität der Simulation und dafür notwendigem Aufwand. Signalverarbeitung und Trackingsystem sind in der Lage, auch die vertikale Richtung zu berücksichtigen, die Datenbank der Impulsantworten und das Zugriffsverfahren müssten dafür aber erweitert werden.

Latenzkriterium

Wesentlich für die realitätsnahe Erzeugung virtueller Schallquellen durch Faltung mit binauralen Impulsantworten ist laut [4], dass die Verzögerung, die zwischen einer Änderung der Kopfposition und dem Anliegen der korrekten Ohrsignale am Trommelfell auftritt, 85 ms nicht überschreitet. Diese Latenz Δt setzt sich hier zusammen aus der Zeit t_{track}, die das Trackingsystem für eine Positionsermittlung benötigt, der Dauer der Netzwerkübertragung der Positionsdaten zur Software t_{net} und der Verzögerung durch die Signalverarbeitung t_{sig}:

$$\Delta t = t_{track} + t_{net} + t_{sig} \tag{1}$$

Laut Herstellerangaben benötigt das Trackingsystem für eine Positionsermittlung weniger als 40 ms, die ermittelten Daten werden mit einer Frequenz von 60 Hz, also mit einer maximalen zusätzlichen Verzögerung von etwa 16.7 ms über das Netzwerk versandt. Die virtuelle Akustik wurde so konzipiert, dass die Systemlatenz im worst case einen Wert von 80 ms unterschreitet, um die gewünschte Funktionalität nach [4] in jedem Fall sicherzustellen. Daher muss nach Gleichung 1 gelten:

$$t_{sig} < 80\,\text{ms} - 56.7\,\text{ms} = 23.3\,\text{ms} \tag{2}$$

Die Signalverarbeitung wurde aus diesem Grund so entworfen, dass die Zeit t_{sig}, um die ein zu bearbeitendes Datensample verzögert wird, kleiner als 23.3 ms ist.

Signalverarbeitung

Um die Latenz bei der Ein- und Ausgabe der Sounddaten möglichst gering zu halten, wurde ASIO, eine plattformübergreifende Treiber- und Schnittstellenarchitektur (Steinberg, vgl. [6]), für die Kommunikation zwischen Betriebssystem und Soundkarte gewählt. Bei der verwendeten Abtastrate von $f_A = 44.1\,\text{kHz}$ entspricht die gesetzte Obergrenze von 23.3 ms etwa 1027 Datensamples. Da ASIO die Ein- und Ausgabe als Doppelpuffer organisiert, d. h. während ein Puffer wiedergegeben wird, muss der andere mit Daten gefüllt werden, konnte mit einer Puffergröße von k = 512 Samples das Latenzkriterium erfüllt werden. Vorraussetzung dafür ist, dass die notwendigen Schritte der Signalverarbeitung so schnell durchgeführt werden können, dass immer neue Audiodaten zur Verfügung stehen, wenn vom Soundkartentreiber für einen Puffer Daten angefordert werden. Die Signalverarbeitung wurde deshalb als Faltung nach einem modifizierten Overlap-Save-Verfahren implementiert (vgl. dazu [7]), um die rechtzeitige Befüllung der ASIO-Puffer sicherzustellen. Das Overlap-Save-Verfahren, eine partitionierte Faltung, die als komplexe Multiplikation im Frequenzbereich ausgeführt wird, wurde derart verändert, dass während der Ausführung die Signale der unter Umständen an verschiedenen Positionen im Raum befindlichen virtuellen Schallquellen für jeweils ein Ohr zusammengefasst werden können. Nach der Faltung stehen also zwei Datenblöcke zur Verfügung, deren Daten einen zeitlichen Abschnitt der die komplette virtulle akustische Szene beschreibenden Ohrsignale repräsentieren. Die weitere Verarbeitung der Puffer, also die Ausgabe der Samples über die Soundkarte, übernimmt deren ASIO-Treiber.

Zusammenfassung

Im vorliegenden Beitrag wurde die Entwicklung und Realisierung eines Systems vorgestellt, mit dem sich virtuelle akustische Umgebungen mit bis zu acht verschiedenen ruhenden oder bewegten Schallquellen in der Horizontalebene generieren lassen. Dazu werden die Quellsignale mit vorher gemessenen oder mit synthetisierten binauralen Raumimpulsantworten gefaltet, deren Auswahl durch ein Trackingsystem dynamisch an die jeweils aktuelle Hörsituation angepasst erfolgt. Das System wurde modular mit einer zentralen Softwaresteuerung konzipiert und lässt sich online steuern. Damit ist es sowohl zur Erweiterung visueller AR-Umgebungen um eine akustische Komponente als auch zur Durchführung psychoakustischer Lokalisationsexperimente geeignet.

Literatur

[1] Blauert, J.: Spatial Hearing, The Psychophysics of Human Sound Localization. The MIT Press, Cambridge, Massachusetts, London, England, Rev. Ed., 1996

[2] Haferkorn, F. und W. Schmid: System zur Erzeugung von vorgebbaren Hörereignisorten. Fortschritte der Akustik, DAGA '96 (1996), DEGA, Oldenburg, 366-367

[3] Horbach, U., R. Pellegrini, P. Mackensen und G. Theile: Design and Applications of a Data-based Auralization System for Surround Sound. Preprints of the 106th AES Convention (1999)

[4] Mackensen, P.: Auditive Localization. Head Movements, an Additional Cue in Localization. Dissertation, Technische Universität Berlin, 2004

[5] A.R.T. GmbH, URL: `http://www.ar-tracking.de`

[6] Steinberg Media Technologies GmbH: ASIO 2.1, Documentation, Release #1, 2005

[7] Torger, A. und A. Farina: Real-Time Partitioned Convolution for Ambiophonics Surround Sound. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (2001)