

der IEM CUBE – ein periphones (Re-)Produktionssystem

Johannes M. Zmölnig*, Winfried Ritsch†, Alois Sontacchi‡
Institut für Elektronische Musik und Akustik,
Kunstuniversität Graz, Österreich

Abstract

Traditionelle Mehrkanal-Wiedergabesysteme werden hauptsächlich für die Wiedergabe von pantophonen Schallfeldern eingesetzt. Voll-periphone Wiedergabe ist in der Vergangenheit hauptsächlich durch fehlende Rechenleistung zur Manipulation einer großen Anzahl von Audiokanälen sowie durch die benötigten aufwendigen Lautsprecher-Aufstellungen begrenzt gewesen. In den letzten Jahren ist jedoch digitale Hardware schnell genug geworden, um jenen Anforderungen gerecht zu werden. Am IEM (Institut für Elektronische Musik und Akustik, Graz, Ö) wurde daher ein mittelgroßer Aufführungsraum für periphone elektro-akustische Musik eingerichtet. Dieser ist mit einer Hemisphäre aus 24 Lautsprechern ausgestattet, wodurch die Wiedergabe von dreidimensionalen Schallfeldern nach den Prinzipien von ambisonic mit mindestens dritter Ordnung ermöglicht wird. Dafür wurde ein lineares 3D-Mix-System entwickelt. Dieses System kann als Produktions-Werkzeug, das ein virtuelles periphones Schallfeld in eine Menge an ambisonic-Kanälen abbildet, als Wiedergabe-Umgebung zur Reproduktion von 3D-Schallfeldern aus einer solchen Menge an ambisonic-Kanälen, aber auch als Live-Instrument, welches die freie Positionierung und Bewegung einer Anzahl an virtuellen Schallquellen im Raum in Echtzeit erlaubt, verwendet werden.

1 Einführung

Eine der größten Herausforderungen der Elektroakustischen Musik ist bis heute die Komposition des Raumes. Obwohl die Prinzipien der periphonen (Re-)Produktion von Schallfeldern seit Jahrzehnten bekannt sind, konnten diese in Echtzeitumgebungen bis jetzt nicht eingesetzt werden, da die Rechenleistung nicht ausreichte, um die für periphone Produktion und Reproduktion notwendige mehrkanalige Signalverarbeitung zu betreiben.

In den letzten Jahren ist nun endlich so schnelle digitale Hardware auf den Markt gekommen, so dass mehrkanalige digitale Signalverarbeitung in Echtzeit auf der Basis von handelsüblichen Personal Computern realisiert werden kann.

* zmoelnig@iem.kug.ac.at

† ritsch@iem.kug.ac.at

‡ sontacchi@iem.kug.ac.at

Ein System zur Positionierung von virtuellen periphonen Quellen sowohl in Konzertsituationen (“live”-Positionierung) als auch in der Postproduktion ist entwickelt worden: *abcde*, an Ambisonic Based Coding and Decoding Environment.

2 Software-Design

Aufgrund der rapiden Entwicklung von immer schnellerer Hardware, sollen PC-basierte Systeme so skalierbar sein, dass auch zukünftige Hardware sinnvoll genutzt werden kann.

Wenn man von (mehrkanaliger) Audio-Verarbeitung spricht, bedeutet dies, dass mit schnellere Hardware entweder die Anzahl der gleichzeitig verarbeitbaren Kanäle oder die Genauigkeit des Systems erhöht wird.

Diese Skalierbarkeit sollte jedoch möglichst transparent für die Benutzenden sein. Sinnvollerweise trennt man daher das “Frontend” (also z.B. das User-Interface) vom “Backend” (der eigentlichen Audio-Engine, in der eine Anzahl von virtuellen Quellen durch Signalmanipulation über eine Lautsprecheranordnung wiedergegeben wird).

Diese Architektur erlaubt es, Bedienoberflächen zu definieren, die genau die – und nur die – Funktionalität zur Verfügung stellen, die von den Benutzenden gebraucht wird. Unabhängig davon kann die Audio-Engine skaliert werden, sodass ein Kompromiss zwischen Anforderungen zur Problemlösung und Hardware-Beschränkungen gefunden wird.

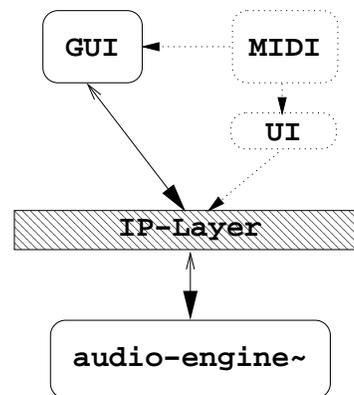


Abbildung 1: Software-Struktur des periphonen Mixers *abcde*

Frontend und Backend werden über ein Ethernet-basiertes Übertragungsprotokoll verbunden. Dadurch können Audio-Engine und User-Interface auf verschiedenen Computern ausgeführt werden, um die Rechenleistung in vernünftigen Grenzen zu halten.

3 ambisonic

Die Aufgabe der Audio-Engine ist es, eine Anzahl von virtuellen (Mono-)Quellen so auf einer Lautsprecheranordnung wiederzugeben, dass das Auditorium der Illusion erliegt, sich im "originalen" Schallfeld¹ zu befinden.

Die Eingangssignale müssen derart manipuliert werden, dass man Lautsprecher-Signale erhält, die diesen Eindruck erwecken.

Weiters ist es oft wünschenswert, eine Repräsentation des periphonen Schallfeldes aufzunehmen, die eine möglichst einfache Wiedergabe erlaubt. Die einfache Aufzeichnung der Lautsprecher-Kanäle ist insofern ungünstig, da diese Mehrkanal-Aufzeichnung an eine bestimmte Lautsprecher-Anordnung gebunden ist, und außerdem die Anzahl der Übertragungskanäle recht hoch werden kann.

Ein Übertragungsformat das ein Maximum an Information über das Schallfeld bei einem gleichzeitigen Minimum an tatsächlichen Übertragungskanälen wird daher gesucht.

Der *ambisonic*-Ansatz (Fellgett 1975), (Bamford 1995) erfüllt diese Bedingungen. Die Idee hinter diesem Ansatz ist, ein existierendes (oder virtuelles) Schallfeld, welches aus ebenen Wellen zusammengesetzt ist, in eine unendlich lange Fourier-Bessel-Reihe, die so genannten *sphärischen Harmonischen*, zu entwickeln. Diese Reihenentwicklung muss aus praktischen Gründen an einem bestimmten Punkt abgebrochen werden, wodurch die Grad der Annäherung an das Ausgangsschallfeld, die Ordnung des *ambisonic*-Systems, definiert wird. Im hier behandelten periphonen Fall ergibt sich die Anzahl der Übertragungskanäle für die Ordnung M des Systems als $L = (M + 1)^2$.

Um einen Satz an Übertragungskanälen für eine ebene Welle aus der Richtung $[\varphi, \vartheta]^T$ zu erhalten, wird das Quellensignal mit Richtcharakteristiken (den "sphärischen Harmonischen", die sich prinzipiell aus Linearkombinationen von trigonometrischen Termen ergeben) bewertet, Mehrere Sätze von solchen *ambisonic*-Kanälen können überlagert werden.

Um die Lautsprecher-Kanäle zu erhalten, wird eine inverse Enkodierung verwendet, die garantiert, dass die Fourier-Bessel-Entwicklung von analysiertem und synthetisiertem Schallfeld bis zur gegebenen Ordnung übereinstimmen.

Der *ambisonic*-Ansatz garantiert, dass eine beliebige Anzahl von virtuellen Quellen durch eine fixe Zahl

¹Wobei "original" nicht unbedingt "real" bedeuten muss.

von Übertragungskanälen ausreichend beschrieben werden können, und dass die Übertragungskanäle völlig unabhängig vom Lautsprecherlayout bei der Wiedergabe sind.

Encoder (der virtuellen Quellen) und Decoder (in Lautsprecher-Kanäle) der *ambisonic*-Repräsentation eines Schallfeldes sind also völlig unabhängig von einander, solange ein Konsens über die verwendete Codierungsvorschrift herrscht.²

4 Audio-Engine: Produktion

Die Aufgabe des Encoders ist es, eine Anzahl von virtuellen Quellen mit Positionsinformation in eine ausreichende Repräsentation des resultierenden Schallfeldes umzurechnen.

4.1 Richtungspositionierung

Die Aufgabe der Richtungspositionierung über $[\varphi, \vartheta]^T$ wird vollständig durch die *ambisonic*-Enkodierung erfüllt.

Um möglichst große Flexibilität zu erreichen, sind die Codierungsvorschriften vollständig als Plugins implementiert, wodurch sie einfach erweitert, verändert oder ganz neu definiert werden können.

Bewegung der Einfallsrichtung $[\frac{\partial\varphi}{\partial t}, \frac{\partial\vartheta}{\partial t}]^T$ erhält man, indem einfach zwischen aufeinanderfolgenden Positionen interpoliert wird.

4.2 Entfernung und Bewegung

Der *ambisonic*-Ansatz basiert auf der Annahme, dass ein Schallfeld nur aus Ebenen Wellen aufgebaut ist. Diese Annahme ist natürlich für die allermeisten (wenn nicht alle) "natürlichen" Schallfelder nicht zulässig.

Anpassungen der *ambisonic*-Codierungsvorschriften, die dem Rechnung tragen, sind von (Malham und Furse 1999) vorgeschlagen worden. Da zum Beispiel nahe Quellen größer erscheinen als entfernte, können sie nicht mehr so scharf lokalisiert werden. Diese Unschärfe erreicht man, wenn der omnidirektionale Übertragungskanal (der Kanal 0-ter Ordnung) gegenüber den Kanälen höherer Ordnung, die mehr Richtungsinformation tragen, stärker gewichtet wird.

Weiters erreichen entfernte Quellen das Ohr (aufgrund der endlichen Ausbreitungsgeschwindigkeit des Schalles in der Luft) später und klingen (aufgrund des Dämpfungsverhalten der Luft) dumpfer als Quellen aus der Nähe der Zuhörenden.

Dieser Dämpfungseffekt folgt einem $\frac{1}{r}$ -Gesetz und kann durch einen Tiefpass erster Ordnung approximiert werden.

²Es gibt historisch bedingt verschiedene Codierungsvorschriften, die zwar prinzipiell übereinstimmen, sich jedoch in konstanten Multiplikationsfaktoren unterscheiden, z.B. um eine gleichmäßige Aussteuerung aller Kanäle zu erreichen.

Die Verzögerung durch die endliche Schallgeschwindigkeit ist in statischen und/oder reflexionsfreien Umgebungen ohne Bedeutung.

Erlaubt man jedoch Reflexionen, so werden die Laufzeitunterschiede dieser Reflexionen vom Hörorgan zur Entfernungsbewertung hinzugezogen.

Doppler-Effekt Bewegt sich eine Schallquelle schnell genug zum Ohr hin und dann vom Ohr weg, so lässt sich eine Tonhöhenänderung beobachten. Der so genannte "Doppler-Effekt" ergibt sich aus der Überlagerung der statischen Ausbreitungsgeschwindigkeit des Schalles und der Geschwindigkeit der Quelle relativ zum Ohr. Bewegt sich die Quelle auf das Ohr zu, so liegen Bäuche und Knoten der Druckwelle enger zusammen als wenn die Entfernung Quelle-Ohr konstant bliebe. Dasselbe geschieht, wenn das Signal (aufgrund der endlichen Ausbreitungsgeschwindigkeit des Schalles) verzögert wird, und sich die Verzögerungszeit stetig ändert (Begault 1991).

Daher kann durch ein variables, stetig interpolierendes Verzögerungsglied der Eindruck einer sich schnell bewegenden Schallquelle erzeugt werden.

4.3 Räumlichkeit

Neben der Positionierung der virtuellen Quellen im Raum ist es von großer Bedeutung, den akustischen Raum, in dem die Quellen positioniert sind, zu definieren. Der akustische Raum wird hauptsächlich über die auftretenden Reflexionen bestimmt.

Erste Reflexionen Die ersten Reflexionen von den Wänden, werden einzeln wahrgenommen und vom Hörapparat benutzt, um sowohl die Entfernung der Schallquellen wie auch die Entfernung der reflektierenden Wände abzuschätzen. Sie sind daher von großer Wichtigkeit für die Wahrnehmung des Raumes. Allerdings ist ihre Simulation relativ rechenaufwendig, da jede Reflexion eigens als virtuelle Spiegelquelle gerechnet werden muss. Im einfachen Falle eines quaderförmigen Raumes ergeben sich für jede virtuelle Quelle zusätzliche sechs Spiegelquellen für die allerersten Reflexionen.

Der Rechenaufwand lässt sich durch Verwendung von kugelförmigen Räumen drastisch minimieren. Wenn sich das Ohr im Zentrum eines solchen Raumes befindet, reduziert sich die Anzahl der ersten Reflexionen auf eins.

Zusätzlich kommt diese erste Reflexion noch aus der gleichen Richtung wie die Originalquelle. Es genügt daher, die Quelle und die Spiegelquelle getrennt entfernungs- und richtungsbewerten. Danach wird eine Richtungs-bewertung der Überlagerung von Quelle und Spiegelquelle durchgeführt.

Da kugelförmige Räume eher unüblich sind, ergibt diese Vereinfachung kein allzu gutes Raumbild. Allerdings entsteht ein guter Eindruck von der Entfernung der Schallquelle bei sehr geringem Rechenaufwand. Für

bessere Näherungen der ersten Reflexionen, können einfach neue virtuelle Quellen an der Position der Spiegelquellen definiert werden, wodurch der Eindruck von lokalisierbaren Reflexionen entsteht.

Verhallung Während die ersten Reflexionen für jede einzelne Schallquelle getrennt wahrgenommen werden, werden Reflexionen höherer Ordnung schnell zu dicht und chaotisch um geschlossen berechnet zu werden.

Es genügt daher, anstatt der Verhallung jeder einzelnen Quelle nur einen Mixdown aller Quellen zu verhallen – in diesem Fall werden also die überlagerten *ambisonic*-Kanäle verhallt. Da die Richtungsinformation des Diffusfeldes ziemlich grob ist, reicht es, das verhallte Schallfeld nur mit *ambisonic* niederer Ordnung zu verhallen.

Da es viele gut klingende, einfach zu benutzende Hallgeräte als Hardwarelösungen am Markt gibt, kann man solche externen Geräte in den Software-Encoder einschleifen.

5 Audio-Engine: Reproduktion

Der Ausgang des Encoders ist eine vollständige Repräsentation des periphonen Schallfeldes im *ambisonic*-Format. Die Decoder-Einheit reproduziert das Schallfeld in einem der *ambisonic*-Enkodierung inversen Prozess. Daher teilen sich Decoder und Encoder das Plugin, das die verwendete Codierungsvorschrift definiert.

Da der Dekodierungsprozess vom verwendeten Lautsprecher-Setup abhängt, wird diese Anordnung dem Decoder über ein Konfigurationsfile mitgeteilt. Die Decoder-Einheit kann also ein *ambisonic*-Schallfeld auf praktisch jedem Lautsprecher-Layout wiedergeben.

Es ist möglich, das gesamte Schallfeld zu rotieren, um die "vorne"-Richtung der tatsächlichen Orientierung des Auditoriums anzupassen.

Basic ambisonic Dekodierung funktioniert nur innerhalb eines kleinen Sweet-Spots gut. Dies ist für Produktionsumgebungen wünschenswert, wo der/die Ton-technikerIn an einem kleinen, klar definierten Ort sitzt, wo er/sie das bestmögliche Reproduktions-Ergebnis erhält.

Für den Fall, dass ein periphones Schallfeld für ein großes Publikum reproduziert werden soll (zum Beispiel bei einem Konzert), werden sich die meisten Zuhörenden außerhalb des Sweet-Spots befinden. Es ist daher notwendig die Sweet-Area der guten Wiedergabe auf Kosten der sehr guten Wiedergabequalität im originalen kleinen Sweet-Spot auszudehnen. Die größte Sweet-Area kann durch die so genannte *in-phase*-Dekodierung erreicht werden.

Um den Kompromiss zwischen großem Wiedergabegebiet und guter Qualität zu optimieren, kann man zwischen den beiden Dekodierungsalgorithmen überblenden.

6 Frontends

Normalerweise, sollen sich die Benutzenden über die zu verwendende Codierungsvorschrift, Lautsprecheranordnung und andere Interna der Audio-Engine keine Gedanken machen müssen. Vielmehr soll es möglich sein, intuitiv eine Anzahl von virtuellen Quellen zu positionieren.

6.1 Ein einfaches Interface

Die Möglichkeit jede virtuelle Quelle einzeln positionieren zu können, kann sich manchmal als nachteilig erweisen, insbesondere dann, wenn das System häufige “Standardsituationen”, wie historische Stereo- oder Quadrophonie-Aufnahmen, über ein beliebiges Lautsprecherarray wiedergeben soll. Mit Sicherheit ist es unpraktisch, wenn jedesmal, wenn eine 5.1-Aufnahme wiedergegeben werden soll, sechs Kanäle von Hand positioniert werden müssen.

Ein einfaches Interface ermöglicht es, solche Standardsettings zu speichern und bei Bedarf durch Klicken eines Buttons abzurufen.

Dadurch können einfach beliebige in Lautsprecher-Kanäle dekodierte Mehrkanalaufnahmen wiedergegeben werden.

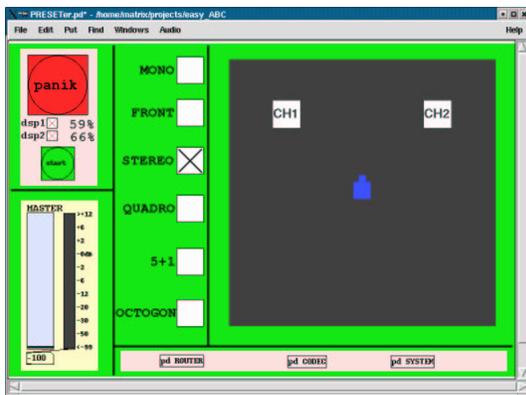


Abbildung 2: ein einfaches Interface für den 3D-Mixer *abcde*

6.2 Ein voll peripherer Mixer

Die Stärke eines peripheren Mixers liegt natürlich nicht in der Wiedergabe von historischen Mehrkanalaufnahmen, sondern in der Möglichkeit der freien Positionieren und Bewegung von virtuellen Quellen.

Daher wird ein Mixer-Interface benötigt, das die volle Kontrolle über diese Parameter erlaubt. Diese Parameter sind $[r, \varphi, \vartheta]$, da sphärische Koordinaten die natürlichen Koordinaten von *ambisonic* Systemen sind.

Da die Mixer-Applikation keinen eingebauten Sequenzer hat, können nur statische (nicht-bewegte) Quellen abgebildet werden.



Abbildung 3: ein periphones Mixer-Interface für *abcde*

Allerdings kann der Eindruck von bewegten Schallquellen erzeugt werden, da die Audio-Engine zwischen zeitlich aufeinanderfolgenden Positionen interpoliert.

Für kontinuierliche Bewegungen ist es notwendig, die Engine mit einem konstanten Strom von Positionsdaten für jede Quelle zu versorgen. Dieser Datenstrom entsteht, wenn am User-Interface ein Controller verändert wird oder über einen externen Sequenzer.

Interfaces zu anderer Software Zwei verschiedene Interfaces für externe Sequenzer stehen zur Verfügung. Diese Interfaces kommunizieren mit dem (graphischen) User Interface, anstatt direkt mit der darunterliegenden Audio-Engine, um die benutzte Hardware total transparent zu halten.

- **MIDI:** Zur Kommunikation mit den meisten (Soft- und Hardware) Sequenzern ist ein bi-direktionales MIDI-Interface implementiert. Dadurch können Positions-Tracks von bis zu 32 virtuellen Quellen aufgenommen und wiedergegeben werden.
- **OSC/FUDI:** Ein unbeschränkte Anzahl von Quellen kann über eine Ethernet-Verbindung, die entweder auf das Open-Sound-Control- oder das (ähnliche) FUDI-Protokoll aufsetzt, positioniert werden. Dadurch können komplexe Bewegungspfade in Echtzeit von Software à la Max, pure-data und SuperCollider generiert werden.

7 Lautsprecher-Layout

Obwohl die Decoder-Einheit der Audio-Engine *ambisonic*-Schallfelder auf praktisch jedem Lautsprecher-Layout wiedergeben kann, können die besten Ergebnisse auf regelmäßigen Aufstellungen erreicht werden. Während es einfach ist, regelmäßige 2dimensionale Vierecke zu berechnen, gilt dies nicht für den 3dimensionalen Fall.

Weiters ist es aus architektonischen Gründen oft unmöglich, die Lautsprecher auf einer ungefähren Kugeloberfläche zu positionieren.

Das menschliche Hörorgan hört zwar periphon, allerdings ist die Auflösung in der Horizontalebene viel

besser als in der Vertikalen.

Da die Anzahl der Wiedergabe-Kanäle immer beschränkt ist, macht es daher Sinn, mehr Lautsprecher in der Horizontalebene aufzustellen als in der dritten Dimension.

Das Layout einer Halbkugel, die in mehrere horizontale Lautsprecher-Ringe aufgeteilt ist, ist eine einfache und gute Anordnung.

Im *IEM CUBE* wurden 24 Lautsprecher in drei solchen Ringen angeordnet, wobei der unterste Ring aus zwölf Lautsprechern besteht, um eine maximale Lokalisationsschärfe in der Horizontalebene zu erreichen. Der mittlere Ring besteht aus acht Lautsprechern, und die restlichen vier Lautsprecher bilden den obersten Ring. (siehe Abb.4)

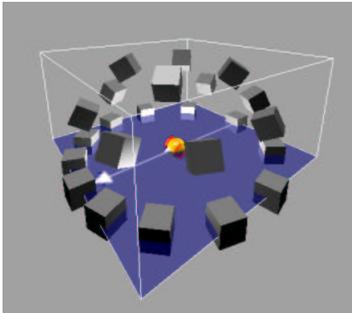


Abbildung 4: Lautsprecher-Layout im *IEM CUBE*

8 Conclusio

Ein peripheres Produktions- und Reproduktionssystem für die Verwendung in Echtzeit-Umgebungen ist gezeigt worden.

Die Software ist so skalierbar und konfigurierbar, dass sie in beinahe jeder Umgebung lauffähig ist.

Eine High-End Realisierung des Systems wurde im *IEM CUBE* geschaffen, einem mittelgroßen Konzertsaal zur Wiedergabe von *ambisonic*-Schallfeldern von mindestens dritter Ordnung.

Die Audio-Engine wurde ursprünglich auf zwei PIII-800MHz Computern realisiert, wodurch Live-Enkodierung von bis zu 24 virtuellen Quellen, die voll positionierbar und stetig bewegbar sind, ermöglicht wird.

Für noch höhere Ansprüche kann die Enkodierungseinheit auf einem hochperformanten PC ausgeführt werden, wodurch etwa 50 individuelle virtuelle Quellen in Echtzeit in ein *ambisonic*-Schallfeld dritter Ordnung abgebildet werden können.

Ein Instrument wurde geschaffen, das einer größeren Zuhörerschaft in typischen Konzertsituationen ermöglicht, Kompositionen des Raumes zu erleben.

Literatur

- Bamford, J. S. (1995). An Analysis of Ambisonic Sound Systems of First and Second Order. Master's thesis, University of Waterloo; Waterloo, Ontario, Canada.
- Begault, D. R. (1991). Challenges to the Successful Implementation of 3-D Sound. In *J. Acoust. Soc. Am.*
- Daniel, J. (2000). *Representation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia.* Ph. D. thesis, Université Paris 6, Paris, France.
- Fellgett, P. (1975). Ambisonics Part One: General System Description. *Studio Sound* 17, 20–40.
- Höldrich, R. R. und A. Sontacchi. Konzepte zur Schallfeldsynthese und Schallfeldreproduktion. In *Proceedings des FA für Akustik, Jahrestagung der ÖPG, Graz.*
- Malham, D. G. (1992). Experience with Large Area 3D Ambisonic Sound Systems. In *Proc. Inst. Acoust.*
- Malham, D. G. und R. Furse (1999). the Furse-Malham Set. http://www1.york.ac.uk/inst/mustech/3d_audio/secondor.html.
- Zmölning, J. M. (2002). Entwurf und Implementierung einer Mehrkanal-Beschallungsanlage. Diplomarbeit, Institut für Elektronische Musik und Akustik, Kunstuniversität Graz.