22. Tonmeistertagung (VDT International Audio Convention), Hannover, 22.-25. Nov. 2002

R. Rabenstein, H. Teutsch,
H. Buchner, W. Herbordt, F. Küch, S. Spors, L. Trautmann

Universität Erlangen-Nürnberg,
Lehrstuhl für Multimediakommunikation und Signalverarbeitung

Raumklangwiedergabe und der MPEG-4 Standard: Das CARROUSO-Projekt

Spatial Sound Reproduction and the MPEG-4 Standard: The CARROUSO-Project

1 Einleitung

Bei den traditionellen Formaten zur mehrkanaligen Wiedergabe wird das zu jedem Lautsprecher gehörige Signal gespeichert bzw. übertragen. Es besteht so eine eindeutige Zuordnung zwischen einer Audiospur auf einem Speichermedium (CD, DVD) und dem entsprechenden Lautsprecher(system). Bei Zweikanal-Stereo bedeutet dies eine Verdopplung der Speicherkapazität gegenüber Mono-Wiedergabe. Diese Erhöhung des Aufwands kann noch leicht hingenommen werden. Bei der Speicherung von Aufnahmen im 5.1 Format werden jedoch bereits Verfahren der Audiokodierung eingesetzt um die erforderliche Speicherkapazität zu begrenzen. Entsprechendes gilt für andere vorgeschlagene Verfahren mit 7+1 oder 10+2 Kanälen.

In den letzten Jahren wurde ein neues vielkanaliges System für die räumliche Wiedergabe komplexer akustischer Szenen geschaffen, die sog. Wellenfeldsynthese. Im Vordergrund steht hier nicht die optimale Gestaltung einzelner Lautsprechersignale, sondern die Erzeugung des gesamten Wellenfelds in einem abgegrenzten Raum. Die technische Umsetzung dieses Konzepts erfordert die streng physikalisch orientierte Herleitung der einzelnen Lautsprechersignale auf der Grundlage der akustischen Wellengleichung [1, 2, 3, 4, 5].

Der Schritt von den momentanen Mehrkanal-Formaten zur Wellenfeldsynthese wirft jedoch erhebliche konzeptionelle Schwierigkeiten auf. Um diese zu lösen und eine technische Realisierung der Aufnahme, Übertragung, Speicherung und Wiedergabe von Vielkanal-Audiomaterial zu erarbeiten wurde auf europäischer Ebene das CARROUSO-Projekt gegründet. Als Werkzeuge dienen dazu Konzepte aus dem MPEG-4 Standard. Dieser Beitrag diskutiert zunächst einige Probleme bei der Realisierung von Vielkanal-Audiosystemen und gibt dann einen kurzen Überblick über die Teile des MPEG-4 Standards, die für Vielkanal-Audio relevant sind. Daran schließt sich eine kurze Darstellung des CARROUSO-Projekts [6] und seiner wichtigsten Entwicklungen an.

2 Probleme bei der Realisierung von Vielkanal-Audiosystemen

Die bisher erprobten Wellenfeldsynthese-Systeme demonstrierten die Eignung dieses Verfahrens zur Wiedergabe von virtueller Akustik, blieben aber zunächst auf die Wiedergabeseite beschränkt. Wie sollen jedoch für dieses Verfahren die Signale für die einzelnen Kanäle aufgezeichnet, gespeichert und übertragen werden, wenn 24, 48, 128 oder mehr unabhängige Lautsprecher zur Wiedergabe verwendet werden? Die Speicherung jedes einzelnen Lautsprechersignals verbietet sich hier aus mehreren Gründen. Zum einen steigt der Speicheraufwand auch bei Verwendung von effizienten Audiokodierverfahren unzulässig an. Zum anderen ist eine Standardisierung von Anzahl und Position der Wiedergabelautsprecher bei diesen Kanalzahlen nicht sinnvoll, da hier die Lautsprecherkonfiguration an den Wiedergaberaum angepasst werden muss und deswegen nicht fest vorgeschrieben werden kann. Schliesslich wäre die Produktion von Tonmaterial mit z.B. 48 Aufnahme- und 128 Wiedergabekanälen wenig praktikabel.

Ein Ausweg besteht darin, die Erzeugung der vielen Lautspechersignale von der Aufnahme- auf die Wiedergabeseite zu verlagern. Es müssen dann nur die einzelnen trockenen Quellensignale übertragen werden, aus denen die Lautsprechersignale erst bei der Wiedergabe erzeugt werden. Eine Speicherung oder Übertragung von ggf. Hunderten von Lautsprechersignalen entfällt damit. Allerdings wirft diese Möglichkeit andere Probleme auf: Wie und nach welchen Kriterien werden die Lautsprechersignale auf der Wiedergabeseite erzeugt? Soll etwa die Tätigkeit des Tonmeisters ebenfalls auf die Wiedergabeseite verlagert werden?

Diese Probleme können durch eine sorgfältige Trennung von kreativen und automatisierbaren Komponenten bei der Produktion von Vielkanal-Aufnahmen gelöst werden. Kreative Komponenten sind beispielsweise die räumliche Positionierung einzelner Quellensignale (z.B. einer Singstimme oder einer Streichergruppe) oder die Gestaltung eines räumlichen Klangeindrucks (z.B. nach der Raumakustik eines realen oder virtuellen Konzertsaals). Diese Komponenten bestimmen die Charakteristik einer Aufnahme und unterliegen der menschlichen Gestaltung. Automatisierbare Komponenten sind z.B. die Berechnung der digitalen Signale für jeden Wiedergabekanal, wie sie im Inneren eines digitalen Mischpults ablaufen.

Die kreativen Komponenten bleiben weiterhin auf der Aufnahmeseite angesiedelt. Sie werden allerdings nicht sofort in die Erzeugung von Lautsprechersignalen umgesetzt, sondern liefern zunächst Steuersignale die zur Wiedergabeseite übertragen werden. Diese Steuersignale geben z.B. die aktuellen Positionen der einzelnen Schallquellen wieder oder charakterisieren die gewünschte Raumakustik durch Raumimpulsantworten oder Wahrnehmungsparameter, wie z.B. Präsenz, Brillianz, oder Klarheit.

Wenn nun die Audiosignale der einzelnen Quellen, ihre Positionsdaten, und Informationen über die Raumakustik auf der Wiedergabeseite vorhanden sind, kann die digitale Berechnung der Lautspechersignale dort automatisch ablaufen. Weitere manuelle Eingriffe sind dann nicht mehr notwendig. Es kann allerdings wünschenswert sein, auf der Wiedergabeseite weitere Informationen über den Wiedegaberaum hinzuzufügen, damit dessen ggf. unerwünschte akustische Eigenschaften kompensiert werden können.

Ein solche Vorgehensweise stellt jedoch hohe Anforderungen an den Umgang mit den verschiedenartigen Datentypen die auf einem Datenträger gespeichert werden. Beim

Auslesen muss zu jedem Zeitpunkt klar sein, ob eine gelesene Zahl ein Abtastwert einer digitalen Audiospur, die räumliche Koordinate einer Quellenposition, oder etwa der Zahlenwert eines Wahrnehmungsparameters ist. Es muss ein Ordnungsschema geben, das diesen Anforderungen gerecht wird ohne jedoch die noch laufende Entwicklung der Wellenfeldsynthese durch zu starre Festlegungen einzuengen.

Ein solches Ordnungsschema stellt der MPEG-4 Standard bereit. Er wurde geschaffen, um die verschiedenen Aspekte audiovisueller Szenen zu beschreiben. Dabei legt der Standard lediglich ein Datenformat fest, das es gestattet die oben erwähnten Datentypen und viele weitere in geordneter Form zu speichern oder zu übertragen. Es wird nicht festgelegt, auf welche Weise diese Daten bei der Aufnahme gewonnen werden, noch wie aus ihnen auf der Wiedergabeseite ein Abbild einer audiovisuellen Szene entstehen kann. Daher enthält der MPEG-4 Standard auch keinerlei Richtlinien über Aufnahmeverfahren und akustische Wiedergabetechniken wie z.B. die Wellenfeldsynthese.

Zur konsequenten Nutzung des MPEG-4 Standards als Ordnungsschema für die Datenvielfalt der Wellenfeldsynthese ist ein erheblicher Entwicklungsaufwand notwendig. Dies betrifft sowohl konzeptionelle Arbeit als auch die Umsetzung vieler Details. Diese Entwicklungsarbeit wird im Rahmen des internationalen Forschungsprojekts CARROUSO geleistet, gefördert von der Europäischen Kommission. Abschnitt 3 beschreibt zunächst kurz den MPEG-4 Standard, der folgende Abschnitt 4 stellt das CARROUSO-Projekt vor.

3 Der MPEG-4 Standard

Im Rahmen der MPEG-4 Standardisierung wurde äußerst umfangreiche Arbeit auf den Gebieten der Video- und Audiocodierung geleistet, so dass es nicht möglich ist, hier eine auch nur annährend komplette Übersicht zu geben. Eine ausführliche Beschreibung des MPEG-4 Standards findet sich z.B. in [7], die hier interessierenden Teile der Audiokodierung sind z.B. in [8, 9, 10] dargestellt.

Frühere Versionen des MPEG-Standards (MPEG-1, MPEG-2) befassten sich mit der effizienten Kodierung von aufgenommenen Videosequenzen oder Audiospuren. MPEG-4 geht hier einen Schritt weiter und versucht, die der Aufnahme zugrundeliegende dreidimensionale Szene zu beschreiben. Aus dieser Szenenbeschreibung kann dann ein gewünschtes Abbild in Form einer Kameraansicht oder einer Mikrofonaufnahme erzeugt werden.

Im Videobereich heißt das, dass der sichtbare Teil einer Szene in seine Komponenten zerlegt wird (Hintergrund und einzelne Objekte des Vordergrunds). Anstelle eines Kamerabilds werden Beschreibungen dieser Objekte gespeichert, aus denen dann wieder eine oder mehrere Ansichten generiert werden. Im Audiobereich gilt das Gleiche: Neben den herkömmlichen Mehrkanal-Formaten wird auch eine strukturierte Betrachtung akustischer Szenen unterstützt (Structured Audio). Das bedeutet, dass auch akustische Szenen in verschiedene Objekte aufgelöst werden können. Dies sind zunächst einzelne Stimmen, Soloinstrumente und Instrumentengruppen. Sie sind durch ihre entsprechenden Audiospuren repräsentiert, aber auch durch geometrische Daten über ihre Position in der Szene. Der “akustische Hintergrund” kann durch Angaben über die Raumakustik charakterisiert werden. Hierfür stehen verschiedene Ansätze zur Verfügung.

Ein physikalischer Ansatz besteht in einer geometrischen Beschreibung der Umgebungsflächen der akustischen Szene. Die Größe und Anordnung dieser Flächen bestimmen die Raumakustik ebenso wie deren Reflexionseigenschaften. Diese Art der physikalischen Charakterisierung verlangt eine genaue Kenntnis der akustischen Umgebung (s. z.B. [11]), die nicht immer vorhanden ist.

Ein anderer Ansatz beruht auf den sogenannten Wahrnehmungsparametern. Diese Parameter haben nicht nur intuitive Bedeutung, sie können auch aus gemessenen Raumimpulsantworten durch Bestimmung gewisser Energieanteile gewonnen werden.

Um audiovisuelle Szenen aus den diversen Objekten wieder zusammensetzen zu können, muss diese Vielzahl von Beschreibungen in einem einheitlichen Datenformat vorliegen. Dies wurde im MPEG-4 Standard mit dem sog. BIFS (Binary Format For Scenes) geschaffen. Der Audioteil dieses Formats (AudioBIFS) erlaubt es auch die mehrkanalige Audioausgabe in einer baumartig gegliederten Struktur aus den einzelnen Quellensignalen zusammenzusetzen. Die einzelnen Äste dieses Baums beinhalten in etwa die Funktion, wie sie auch aus einem Mischpult bekannt sind (Zusammenführung und Verteilung von Signalen, Klangbeeinflussung, Effekte). Die Steuerung geschieht jedoch anhand der im AudioBIFS gespeicherten Steuerparameter.

4 Das CARROUSO Projekt

4.1 Übersicht

Die Beschreibung der Problematik von Vielkanal-Wiedergabe in Abschnitt 1 und die kurze Darstellung der Grundlagen von MPEG-4 in 3 machen klar, dass der MPEG-4 Standard grundsätzlich geeignet ist, um die für eine Vielkanal-Wiedergabe notwendigen Informationen zu transportieren. Allerdings legt der Standard nur das Datenformat für diese räumlichen Informationen fest. Er schreibt aber nicht fest

- wie diese Informationen bei der Aufnahme gewonnen werden,
- wie das technische System zur Vielkanal-Wiedergabe aussieht.

Auf Aussagen zu diesen Themen wird im MPEG-4 Standard bewusst verzichtet, da jede Festlegung die noch laufende technische Entwicklung behindern würde.

Das CARROUSO-Projekt hat sich zur Aufgabe gemacht, die für eine Vielkanal-Übertragung notwendigen Komponenten auf der Aufnahme- und auf der Wiedergabeseite zu entwickeln und die Funktionalität des Gesamtsystems zu demonstrieren. An dieser Entwicklung sind insgesamt zehn Einrichtungen (Universitäten, Forschungseinrichtungen, Unternehmen) aus fünf eurpäischen Staaten beteiligt. Diese Arbeiten werden im Rahmen des 5. Rahmenprogramms der Europäischen Kommission gefördert. Das Acronym CARROUSO steht für “Creating, Assessing, and Rendering in Real Time of High Quality Audio-Visual Environments in MPEG-4 Context”. Weitere Informationen über Details des Projekts finden sich in [6].

4.2 Struktur des CARROUSO Systems

Abbildung 1 zeigt die Struktur des CARROUSO Systems. Es besteht aus den drei Blöcken Aufnahme, Übertragung und Wiedergabe (Recording, Transmission, Rendering).

Der Block Aufnahme umfasst nicht nur qualitativ hochwertige Aufnahme der einzelnen Quellensignale, sondern bei Bedarf auch die Bestimmung der Quellenpositionen. Diese beiden Aufgaben sind in Abbildung 1 unter "Source Recording" zusammengefasst. Dazu kommt noch die Bestimmung der Raumakustik ("Room Parameter Modeling"). Die so gewonnenen Daten werden nach den Vorgaben des MPEG-4 Standards in einen Datenstrom umgewandelt, der auch Video-Daten enthalten kann.

Für die Übertragung werden Audio- und Video-Datenströme in einem Multiplexer zusammengeführt und entweder auf einem Server zur späteren Übertragung abgelegt oder aber gleich übertragen, ggf. nach Umwandlung in ein spezielles Format, z.B. DVB (Digital Video Broadcasting).

Zur Wiedergabe wird dem übertragenen Signal der Audio-Teil entnommen und gemäß dem Standard dekodiert. In einem Audio Compositor werden die einzelnen Quellensignale dann anhand ihrer Position im Raum zu einem Gesamtsignal zusammengefasst, das auch künstlich erzeugte Reflexionen des Aufnahmeraums enthalten kann. Diese Signale werden dann dem Wellenfeldsynthese-System zugeführt ("WFS Rendering"). Allerdings kann eine ungünstige Akustik des Wiedergaberaums den gewünschten Klangeindruck verfälschen. In gewissen Grenzen kann diesem Effekt durch eine Kompensation am Audio Compositor entgegengewirkt werden ("Space Acoustic Compensation"). Die Eingabe der akustischen Eigenschaften des Wiedergaberaums erfordert dann die Mitwirkung des Benutzers ("User interaction").

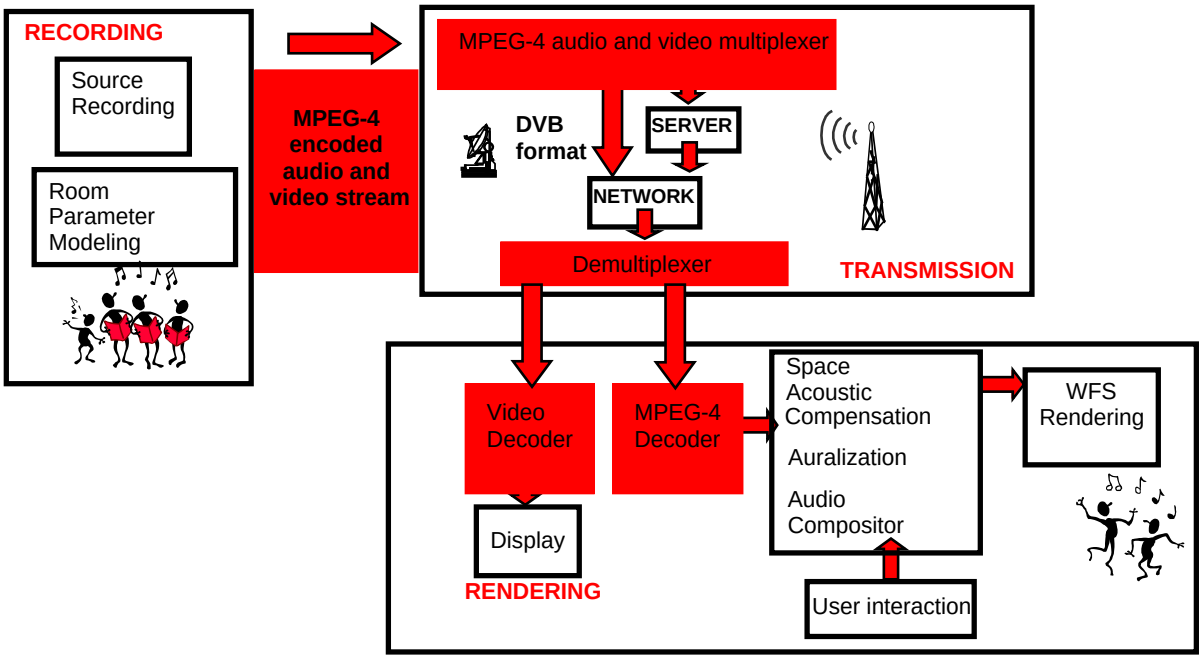

Abbildung 1: *Die Struktur des CARROUSO Systems*

Nach dieser Darstellung der Funktionsblöcke zeigt Abbildung 2 die Anordnung des Aufnahme- und Wiedergabesystems. Die Aufnahme kann z.B. in einem Studio, einem Konzertsaal oder einer Kirche erfolgen (recording room). Das Ziel ist, den Klangeindruck in einem Ausschnitt dieses Raums (gestrichelt gezeichnet) im Wiedergaberaum zu reproduzieren. Dazu werden zunächst die Schallquellen (primary sources) einzeln oder in Grup-

pen aufgenommen (source recording). Weiter werden die Quellenpositionen bestimmt und kennzeichnende Größen der Raumakustik ermittelt, z.B. Raumimpulsantworten (impulse responses).

In Abbildung 2 nicht eingezeichnet sind die gesamte Übertragungskette und der Audio Compositor aus Abbildung 1. Sie vermitteln dem Wellenfeldsynthese-System (WFS) die Daten aus dem Aufnahmeraum, das daraus die Lautsprechersignale für den Wiedergaberaum erzeugt (reproduction room). Auf diese Weise entsteht dort der Klangeindruck des räumlich meist größeren Aufnahmeraums, der in Abbildung 2 gestrichelt als "virtual recording room" eingezeichnet ist. Mit Hilfe des Wellenfeldsynthese-Systems werden die ursprünglichen Schallquellen im Wiedergaberaum als virtuelle Quellen (virtual sources) wiedergegeben. Dies geschieht unabhängig von der Hörerposition an allen Stellen innerhalb der Lautsprecheranordnung (grün schattierter Bereich). Die korrekte Funktion dieses Aufnahme- und Wiedergabesystems erfordert den Einsatz verschiedener neuer Techniken, die in den folgenden Abschnitten beschrieben werden.

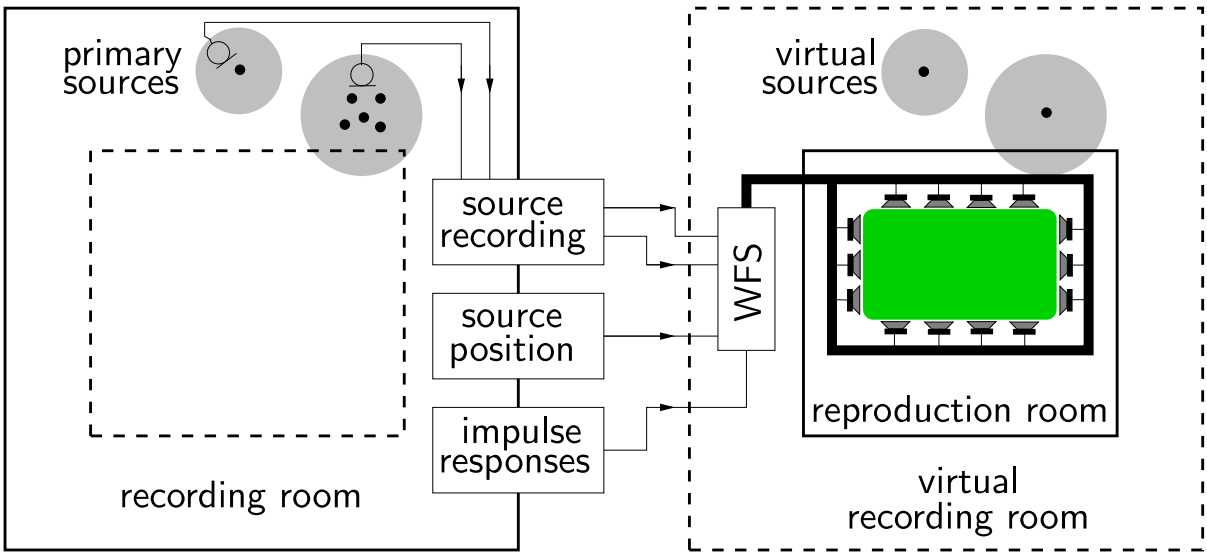

Abbildung 2: *CARROUSO Aufnahme und Wiedergabesystem*

4.3 Aufnahme

Die Aufnahmetechnik der einzelnen Quellensignale unterscheidet sich nicht wesentlich von den bisher verwendeten Verfahren. Es ist lediglich darauf zu achten, möglichst "saubere" Quellensignale zu erhalten, da der räumliche Klangeindruck erst bei der Wiedergabe erzeugt wird.

Neue Techniken müssen dagegen bei der Positionsbestimmung der Schallquellen eingesetzt werden. Abbildung 3 zeigt das Grundprinzip der akustischen Positionsbestimmung. Wenn die Schallquellen eines Sprechers mit zwei

Mikrofonen aufgezeichent werden, dann sind die Entfernungen D_L und D_R vom Sprecher zum linken und zum rechten Mikrofon i.a. verschieden. Daraus resultiert

ein Laufzeitunterschied zwischen den beiden ansonsten sehr ähnlichen Mikrofonsignalen. Dieser Laufzeit kann durch Auswertung der digitalen Signale bestimmt und mit Hilfe

des Mikrofonabstands D_M in den Winkel θ umgerechnet werden, um den der Sprecher versetzt zur Mikrofonachse steht.

Mit nur zwei Mikrofonen ist eine solche Positionsbestimmung aber noch zu ungenau, da damit nur Winkel- aber keine Tiefeninformation zu bestimmen ist. Stattdessen werden Mikrofonarrays mit räumlich verteilten Mikrofonpaaren verwendet. Hier kommt es nicht auf die Klangqualität der aufgezeichneten Signale an, vielmehr liegt die gesuchte Information im Zeitunterschied zwischen den einzelnen Mikrofonkanälen. Aus diesem Grund können hier preisgünstige Elektret-Mikrofone verwendet werden. Für Sprachaufnahmen, z.B. zur Übertragung von Videokonferenzen, ist deren Klangqualität aber auch für Aufnahmezwecke vollkommen ausreichend. Hier werden Mikrofonarrays auch bei der Aufnahme von Quellensignalen eingesetzt, denn sie erlauben auch die Realisierung einer einstellbaren Richtcharakteristik (Beamforming). Sie dient zur Fokussierung auf einzelne, ggf. bewegte Sprecher und zur Unterdrückung von Störgeräuschen. Abbildung 4 zeigt einen Auschnitt aus einem solchen Array.

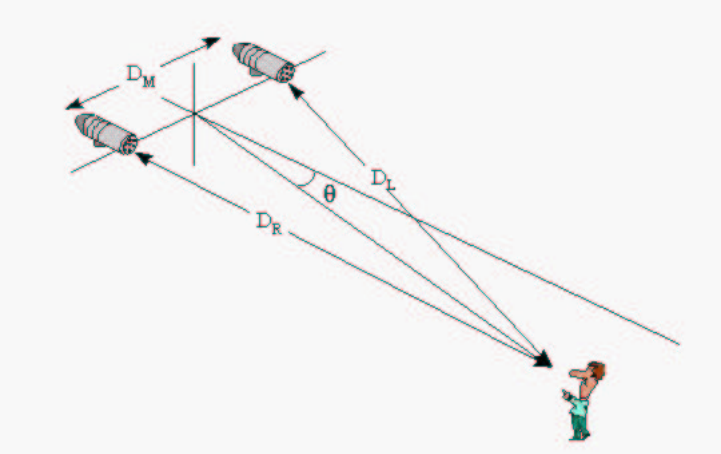

Abbildung 3: *Grundprinzip der akustischen Positionsbestimmung*

Dennoch bieten Mikrofonarrays allein nicht unter allen Umständen eine ausreichend sichere Positionsbestimmung. Als zweites Standbein der Positionsbestimmung können Videosequenzen ausgewertet werden. Bei der Lokalisierung von Personen dient dabei die Hautfarbe als robustes Erkennungsmerkmal. Abbildung 5 zeigt einige Auschnitte aus einer Videosequenz. Der weiße Kasten gibt jeweils Größe und Position des erkannten Gesichts an [12]. Die Ergebnisse der technisch völlig unterschiedlichen Positionsbestimmung mit Mikrofonarrays und aus Videosequenzen können kombiniert werden und ergeben zusammen ein wesentlich verlässlicheres Ergebnis als eine Methode allein [13].

4.4 Wiedergabe

Die Wiedergabe mit einem Wellenfeldsynthese-System erfordert eine umfangreiche digitale Vorverarbeitung, an deren Ende die Signale für die einzelnen Lautsprecher stehen. Wie eingangs erwähnt, werden ja Aufgaben auf die Wiedergabeseite verlegt, die sonst im digitalen Mischpult auf der Aufnahmeseite angesiedelt sind. Im Einzelnen sind folgende Funktionen zu realisieren:

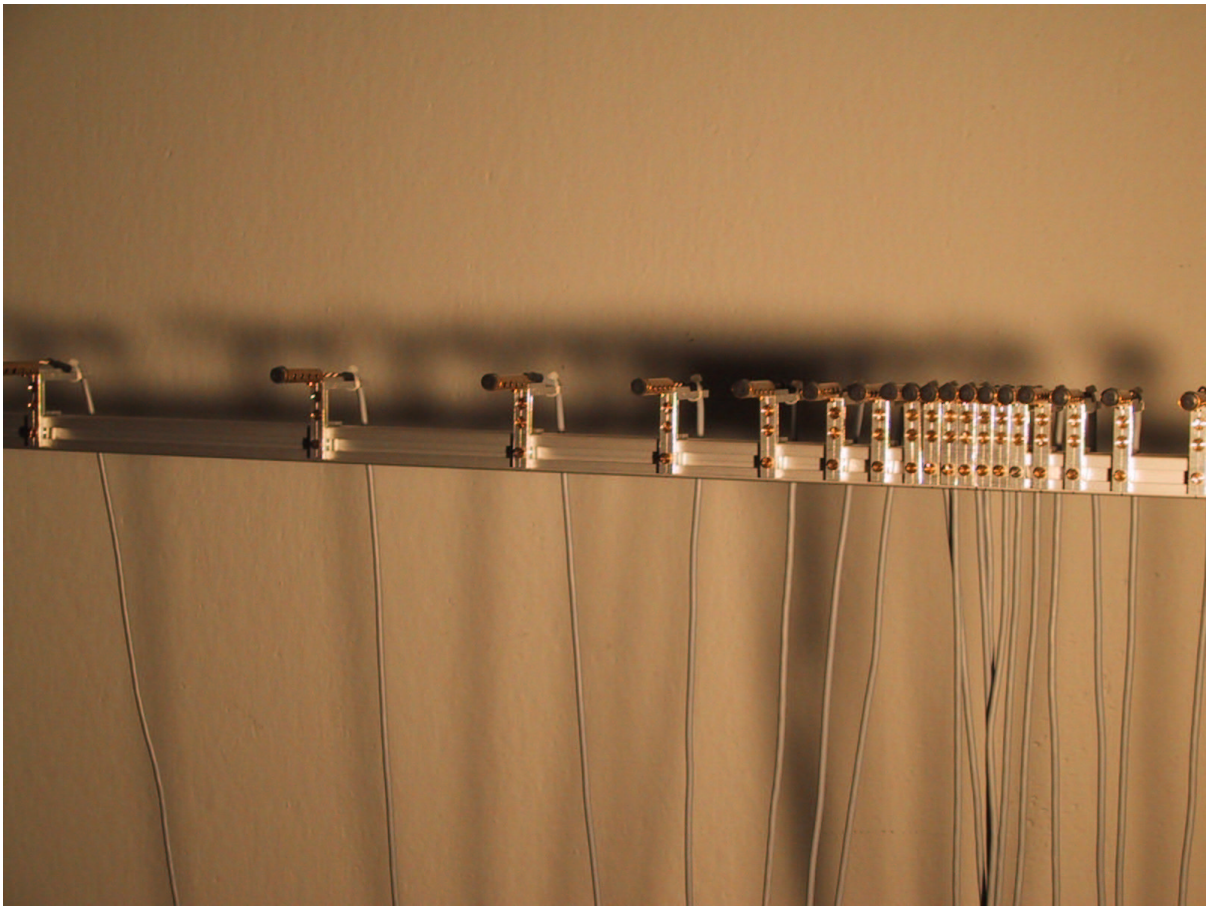

Abbildung 4: *Typischer Aufbau eines Mikrofonarrays für Vielkanalaufnahmen*

- Umrechnung der Positionsdaten jeder Signalquelle in Gewichtungsfaktoren für jeden Lautsprecherkanal. Bei bewegten Schallquellen muss diese Berechnung dynamisch ohne spürbare Verzögerung erfolgen.

- Faltung der trockenen Quellensignale mit den Raumimpulsantworten des Aufnahmeraums. Dabei handelt es sich nicht um Verhallung im Sinne eines Effektgeräts, denn die einzelnen Lautsprechersignale müssen zusammen bei den Hörern einen stimmigen räumlichen Eindruck erzeugen. Die dazu notwendigen Raumimpulsantworten werden, entweder im Aufnahmeraum ermittelt und zum Wiedergaberaum übertragen, oder sie werden bei der Wiedergabe näherungsweise aus geometrischen Daten oder aus Wahrnehmungsparametern berechnet.

- Das akustische Feld im Wiedergaberaum wird nicht allein vom Wellenfeldsynthese-System bestimmt, sondern auch von den im Wiedergaberaum vorhanden Reflexionen. Wenn diese bekannt sind, kann deren Einfluss in gewissen Grenzen durch eine digitale Vorverzerrung der Lautsprechersignale ausgeglichen werden. Diese Vorverzerrung wird auch als Raumkompensation bezeichnet.

- Die Wiedergabe sehr tiefer Frequenzen kann auch einem gemeinsamen Subwoofer überlassen werden, da tiefe Töne nicht zum räumlichen Höreindruck beitragen. Da-

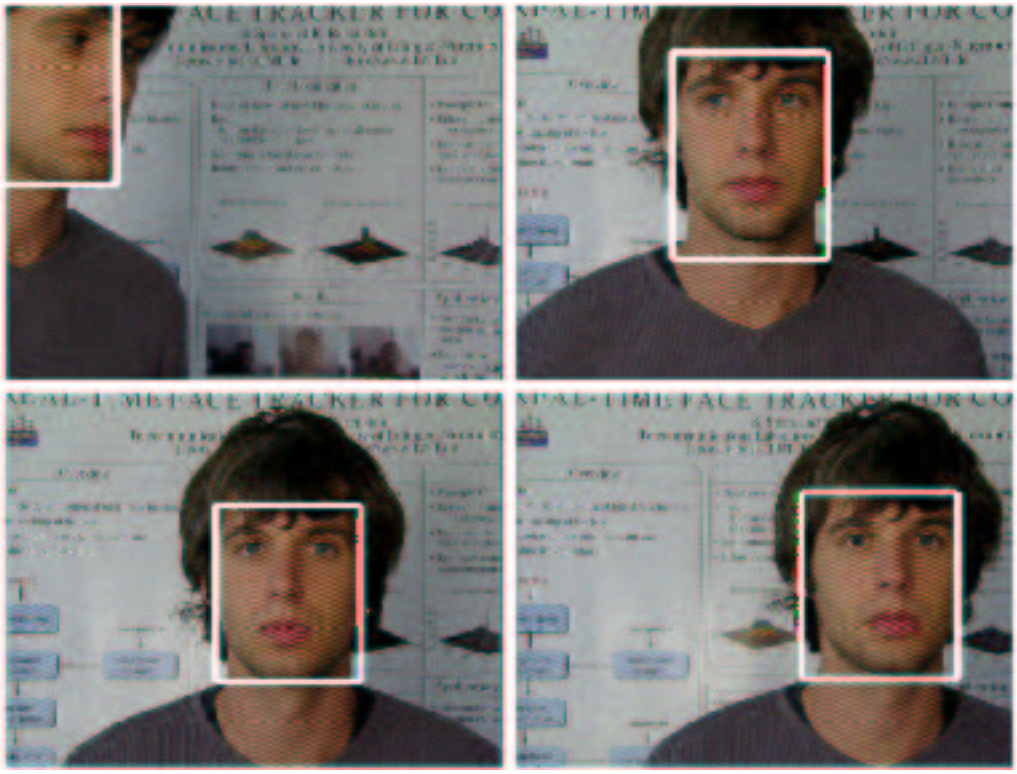

Abbildung 5: *Lokalisierung von Gesichtern in Video-Sequenzen*

zu müssen die entsprechenden Frequenzanteile aus den Lautsprechersignalen entfernt und einem Tieftonkanal zugeführt werden.

Die genannten Aufgaben müssen nicht stufenweise hintereinander ausgeführt werden. Es ist effizienter, die verschiedenen Anforderungen in einem einzigen Verarbeitungsschritt zusammenzufassen. Auf diese Weise bleibt der Rechenaufwand überschaubar.

Abbildung 6 zeigt den Aufbau eines Lautsprecherarrays für eine Wellenfeldsynthese-System. Im Inneren der U-förmigen Anordnung entsteht ein Auschnitt aus dem Schallfeld eines größeren virtuellen Raums. Die Videoprojektion unterstützt den räumlichen Eindruck. Ebenfalls sichtbar ist der Hardware-Aufwand zur Vielkanal-Wiedergabe. Die beiden Rechner auf der rechten Seite übernehmen die Wiedergabe der Videoprojektion von DVD und die Berechnungen für die Wellenfeldsynthese. Unter der Projektionsfläche sind drei helle 19-Zoll-Geräte sichtbar. Die beiden unteren enthalten 24 Verstärker zur Ansteuerung der Lautsprecher. Das obere gehört nicht zur Wiedergabeseite, sondern enthält die Mikrofonvorverstärker und die Digitalisierung des 24-Kanal-Mikrofonarrays aus Abbildung 4 .

Mit diesem Wellenfeldsynthese-System lassen sich auch herkömmliche Mehrkanalformate wie das 5.1 Format wiedergeben. In diesem Fall erzeugt das Lautsprecherarray fünf virtuelle Quellen, deren Aufstellung ohne Rücksicht auf

architektonische Gegebenheiten nach der entsprechenden ITU-Empfehlung [14] gewählt werden kann. Zur Einstellung der virtuellen Positionen dient ein Programm, dessen Bedienoberfläche in Abbildung 7 gezeigt ist. Das umgekehrte U aus 24 Punkten stellt die Lautsprecherpositionen des Wellenfeldsynthese-Systems dar. Die weiter außen

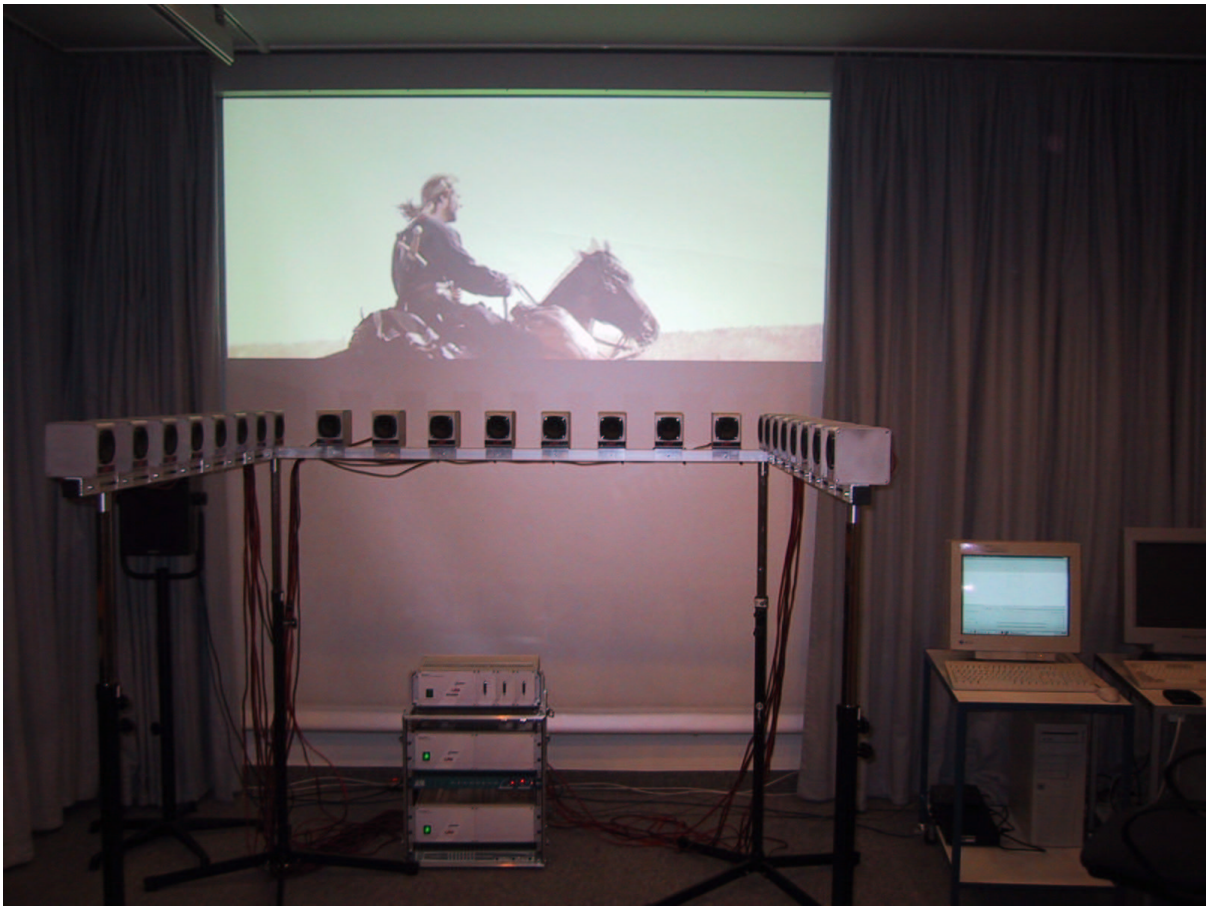

Abbildung 6: *Lautsprecherarray eines Wellenfeldsynthese-Systems*

liegenden Punkte sind die Positionen diskreter Schallquellen. Die Wiedergabe ist jedoch nicht auf ortsfeste Quellen beschränkt. So beschreibt der Halbkreis die Bahn einer Quelle, die sich während der Wiedergabe bewegt.

5 Zusammenfassung

Mit der Möglichkeit, audiovisuelle Szenen in einzelne Objekte zu zerlegen, eröffnet der MPEG-4 Standard neue Möglichkeiten zur Übermittlung, Speicherung und Wiedergabe von Bild- und Tonmaterial. Die technischen Verfahren zur Nutzung dieses Standards müssen aber teilweise noch entwickelt werden. Zu diesem Zweck hat sich das CARROUSO-Projekt die Aufgabe gestellt, ein neuartiges Vielkanal-Wiedergabesystem, die sog. Wellenfeldsynthese, mit Unterstützung des MPEG-4 Standards zu realisieren. Mit den im Standard gebotenen Möglichkeiten lässt sich die gesamte Verfahrenskette von Aufnahme, Übertragung, Speicherung und Wiedergabe erst effizient realisieren. Erste Demonstrationen beweisen sowohl die Leistungsfähigkeit der Wellenfeldsynthese als auch die Vielfältigkeit des MPEG-4 Standards. Auf dieser Grundlage ist zu erwarten, dass sich die Wellenfeldsynthese als neues Verfahren zur Wiedergabe virtueller Akustik etabliert.

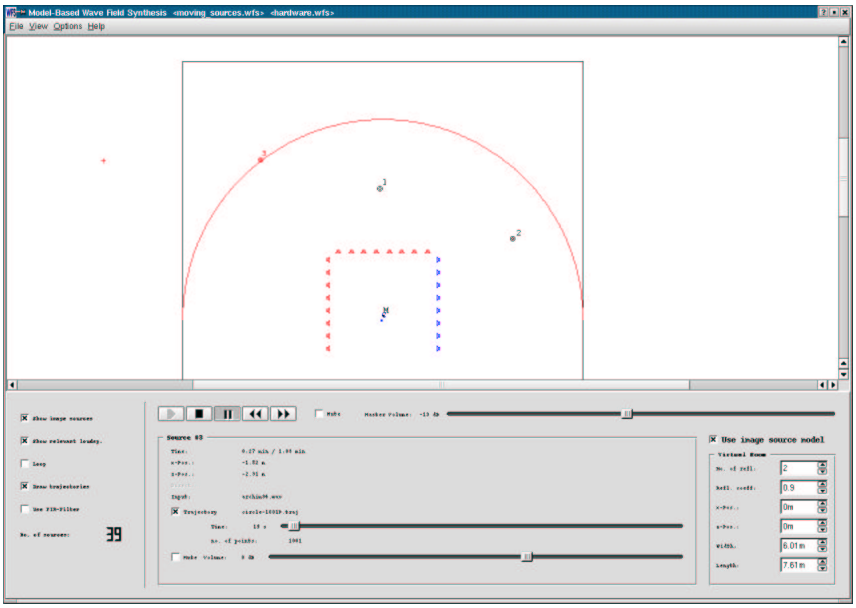

Abbildung 7: *Benutzeroberfläche zur interaktiven Positionierung virtueller Quellen*

Literatur

[1] A.J. Berkhout, "A holographic approach to acoustic control," *Journal of the Audio Engineering Society*, vol. 36, pp. 977–995, December 1988.

[2] E.W. Start, *Direct Sound Enhancement by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1997.

[3] E.N.G. Verheijen, *Sound Reproduction by Wave Field Synthesis*, Ph.D. thesis, Delft University of Technology, 1997.

[4] P. Vogel, *Application of Wave Field Synthesis in Room Acoustics*, Ph.D. thesis, Delft University of Technology, 1993.

[5] D. de Vries, E.W. Start, and V.G. Valstar, "The Wave Field Synthesis concept applied to sound reinforcement: Restrictions and solutions," in *96th AES Convention*, Amsterdam, Netherlands, February 1994, Audio Engineering Society (AES).

[6] "The CARROUSO project," `http://emt.iis.fhg.de/projects/carrouso`.

[7] F.C.N. Pereira and T. Ebrrahimi, *The MPEG-4 Book*, Prentice Hall, Upper Saddle River, 2002.

[8] Riitta Väänänen, "Synthetic audio tools in MPEG-4 standard," in *Proc. 108th AES Convention*. Audio Engeneering Society, February 2000, Preprint 5080.

[9] E. D. Scheirer, R. Väänänen, and J. Houpaniemi, "AudioBIFS: Describing audio scenes with the MPEG-4 multimedia standard," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 237–250, September 1999.

[10] E. D. Scheirer, "The MPEG-4 structured audio standard," in *Proc. Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP'98)*, 1998.

[11] H. Kutruff, *Room Acoustics*, Spon Press, London, 2000.

[12] S.Spors and R.Rabenstein, "A real-time face tracker for color video," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, USA, Salt Lake City, May 2001.

[13] S. Spors, R.Rabenstein, and N.Strobel, "A multi-sensor object localization system," *In Vision, Modelling and Visualization (VMV)*, pp. 19–26, 2001.

[14] ITU, "Recommendation ITU-R BS.1116-1," 1994-1997.