Kommunikationsakustik: Instrumentelle Analyse und Synthese auditiver Szenen

Jens Blauert
Ruhr-Universität Bochum, D-44780 Bochum
http://www.ika.ruhr-uni-bochum.de

Es wird im Folgenden vorgeschlagen und begründet, die Bezeichnung „Kommunikationsakustik“ generell für diejenigen Bereiche der Akustik zu verwenden, die der Informationstechnik und Informatik nahe stehen. Nach einem Kurzüberblick über die einschlägigen Forschungsgebiete des Institutes für Kommunikationsakustik der Ruhr-Universität Bochum wird dann auf zwei Gebiete exemplarisch eingegangen, die in besonderer Weise kennzeichnend für Kommunikationsakustik sind, nämlich die instrumentelle Analyse und die instrumentelle Synthese auditiver Szenen. Auf beiden Gebieten zeigt sich, dass außer akustischen und auditiven Phänomenen kognitive und multimodale Phänomene zu berücksichtigen sind. Zukünftige kommunikationsakustische Systeme werden generell in zunehmenden Maße entweder wissensbasierte und multimodale Komponenten enthalten, oder die kommunikationsakustischen Systeme werden als Subsysteme in komplexere Systeme eingebettet sein, die solche Komponenten enthalten. Dieser technologischen Entwicklungstrend wird das Bild der Kommunikationsakustik künftig maßgebend prägen.

.

1. Vorbetrachtung

Seit der Erfindung der Vakuumtriode zu Anfang des 20. Jahrhunderts und der damit eröffneten Möglichkeit, Niederfrequenzverstärker zu entwickeln, hat die Akustik einen stürmischen Entwicklungssprung erfahren. Endlich konnten Erfindungen in großem Stile angewendet werden, die schon seit einiger Zeit in den Schubladen lagen - z.B. die Magnettonaufzeichnung nach Poulsen (1889) und die Lichttonaufzeichnung nach Ruhmer (1901).

Seit ca. 1920 entwickelte sich der Rundfunk, gefolgt von der Beschallungstechnik mit Lautsprechern (ab ca. 1924) und dem Tonfilm (ab ca. 1928). Aus der vorwiegend physikalisch orientierten (reinen) Akustik wurde durch die Verbindung mit der Elektrotechnik eine Ingenieurwissenschaft mit großen Anwendungspotenzial. Der Begriff Elektroakustik entstand.

Ab etwa 1965 hielten die Computer Einzug in die Akustik. Die sich hierdurch ergebenden Möglichkeiten der Signalverarbeitung und -speicherung führten zu einem weiteren enormen Entwicklungsschub der Akustik, der alle Fachgebiete innerhalb der Akustik erfasst hat und dessen Ende bis heute noch unabsehbar ist. Nachdem die ingenieurmäßig orientierte Akustik im 20. Jahrhundert ein Ehe mit der Elektrotechnik eingegangen war, findet sie sich nun in einer Beziehung wieder, zu der digitale Signalverarbeitung und Informatik als wichtige Partner hingekommen sind.

Der Einzug der Computer hat natürlich auch das Bild der Elektrotechnik einschneidend gewandelt, was sich u.a. an der folgenden terminologischen Entwicklung ablesen lässt: Bis ca. 1950 unterteilte man die Elektrotechnik in die zwei großen Bereiche Starkstromtechnik und Schwachstromtechnik, nach 1950 wurden dann dafür die Bezeichnungen Elektrische Energietechnik und Elektrische Nachrichtentechnik üblich. In jüngster Zeit ist es nun aber überraschenderweise so, dass das Wort Elektrotechnik im Bewusstsein der Öffentlichkeit vorwiegend für die Elektrische Energietechnik steht. Die nachrichtentechnische Sparte der Elektrotechnik heißt heute Informationstechnik. Fast alle deutschen Fakultäten für Elektrotechnik haben sich inzwischen entsprechend umbenannt.

Die Elektroakustik im o.g. Sinne steht – bedingt durch ihr Anwendungsspektrum – der Informationstechnik näher als der Energietechnik. Besonders innig sind die Verflechtungen mit einem wichtigen Teilgebiet der Informationstechnik, nämlich der Kommunikationstechnik. Folgerichtig haben wir deshalb in Bochum seit etwa 1976 zur Kennzeichnung der informationstechnischen

Sparte der Akustik die Bezeichnung Elektroakustik durch die Bezeichnung Kommunikationsakustik ersetzt – inzwischen wurde auch der Institutsname angepasst.

Der Bezeichnungswechsel war auch deshalb angezeigt, weil Elektroakustik heute in der Regel in einem engeren Sinne verstanden wird als noch vor einigen Jahren. Man meint damit heute vorwiegend das Gebiet, das sich speziell mit der elektroakustischen Energiewandlung befasst.

Dass die Akustik sehr viel mit Kommunikation zu tun hat, ist eine Binsenweisheit. Wir alle wissen, dass die zwischenmenschliche Kommunikation wesentlich über den akustischen Kanal (Sprache/Gehör) abläuft. Behinderungen im auditiven Sinnesbereich sind deshalb besonders schwerwiegend. Wer nicht gut hört, ist in der Gefahr der sozialen Isolation.

In gewisser Weise ist die Bezeichnung Kommunikationsakustik eine Tautologie, denn „Akustik“, ein im 17. Jahrhundert entstandenes Kunstwort, leitet sich bekanntlich von dem altgriechischen Verb für „hören“ (ἀκούειν...ak'u:ɪn) ab. Wenn wir dennoch hier das Wort Kommunikation voranstellen, so hat das zwei Gründe, einen sachlichen und einen politischen.

Der sachliche Grund ist, dass es natürlich Schalle und mit diesen verbundene Hörwahrnehmungen gibt, die nicht der Kommunikation dienen - und somit auch Gebiete der Akustik, die nicht oder nicht direkt Kommunikationsakustik sind. Der politische Grund besteht darin, das Kommunikationsakustik heute für viele junge Leute viel attraktiver klingt als schlicht Akustik. In der augenblicklichen Situation, in der ein gefährlicher Mangel an Studentinnen und Studenten der Natur- und Ingenieurwissenschaften herrscht, ist dies ein wesentlicher und legitimer Grund.

Im Übrigen ist die Kommunikationsakustik als isolierte Fachwissenschaft an wissenschaftlichen Hochschulen nicht mehr zu halten. Im Verbund der Gebiete, die zusammen die modernen Informations- und Kommunikationstechnik vertreten, hat sie jedoch eine unverzichtbare Funktion, insbesondere an Fakultäten für Elektrotechnik und Informationstechnik. Akustische (auditive) Komponenten finden sich eingebettet in einer Vielzahl von modernen Informations- und Kommunikationsystemen.

Mit anderen Fachgebieten zusammenzuarbeiten fällt Akustikern zum Glück nicht schwer. Sie sind daran gewöhnt, denn die Akustik ist schon wegen ihrer Mittlerposition zwischen Schall und Hören eine inhärent interdisziplinäre Wissenschaft. Die Kommunikationsakustik ist u.a. in besonderer Weise dafür qualifiziert, Wissen über den Menschen in seiner Rolle als Informationsquelle und –senke zu ermitteln und dieses für Anwendung in der Informationstechnik bereitzuhalten. Die hierbei verwendeten Untersuchungsverfahren lassen sich in vielen Fällen auch auf andere, d.h. nichtakustische Sinnesmodalitäten übertragen.

2. Das Forschungsspektrum des Bochumer Institutes

Aus dem Bochumer Institut für Kommunikationsakustik sind seit Beginn seiner Tätigkeit in Jahre 1974 bisher mehr als 40 Doktorarbeiten und mehr als 500 Studien- und Diplomarbeiten hervorgegangen. Die wissenschaftlichen Arbeitsgebiete sind kommunikationsakustisch geprägt und zielen überwiegend auf technologische Anwendungen. Zur Zeit sind drei Hochschullehrer(innen) am Institut tätig, und zwar Prof. Dr.-Ing. Herbert Hudde, PD Dr. phil. Ute Jekosch sowie der Autor. Die Forschungsaktivitäten sind in vier Projektgruppen gegliedert, die folgende Bezeichnungen tragen (Näheres auf der Webseite des Institutes):

(1) Auditorische Signalverarbeitung und Binauraltechnik
(2) Simulation und Virtuelle Umgebungen
(3) Technische Audiologie und Messtechnik
(4) Sprachkommunikationstechnologien und Produktgeräusche

In diesem Beitrag wird in Folgenden exemplarisch auf zwei Gebiete eingegangen, die vorwiegend den Projektgruppen (1) und (2) zuzuordnen sind und im besonderen Maße kennzeichnend für die Kommunikationsakustik sind. Es handelt sich um die instrumentelle Analyse und die instrumentelle Synthese von auditiven Szenen.

3. Audio-Übertragungssysteme

Eine wichtige Klasse von Systemen der Kommunikationsakustik sind die Audio-Übertragungssysteme. Diese Systeme erlauben es, auditive Szenen über Zeit und Raum zu übertragen, d.h. sie ermöglichen es einem oder mehreren Zuhörern, etwas, das sie an einem Ort zu einer Zeit hören könnten, an einem anderen Ort und/oder zu einer anderen Zeit zu hören. Klassische Systeme dieser Art sind z.B. Rundfunk oder Telefon.

Die wesentlichen Elemente solcher Systeme sind: Schallempfänger (ein oder mehrere Mikrophone), ggf. Mittel zur Signalbearbeitung und Speicherung, ein oder mehrere Übertragungskanäle (inkl. Kodierer und Dekodierer) sowie Schallsender (Lautsprecher oder Kopfhörer). Falls weitgehend authentische Übertragung gewünscht wird, so greift man gerne auf sog. binaurale Systeme zurück. Als Schallempfänger dient dann ein sog. Kunstkopf.

Da in modernen Systemen dieser Art die Signale in digital kodierter Form vorliegen, ergibt sich das in Bild 1 gezeigte Prinzipschaltbild. Das System enthält also Möglichkeiten zur Audio-Signalverarbeitung und -speicherung (Die Übertragungskanäle sind der Einfachheit wegen hier nicht explizit eingezeichnet worden.)

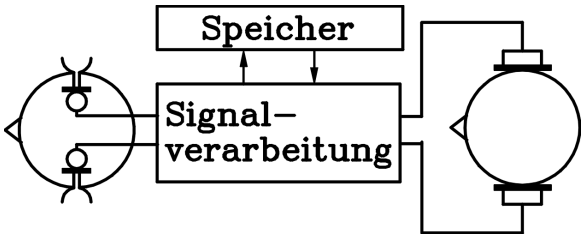

Bild 1: Prinzipbild eines Audio-Übertragungssystems

Es ist nicht das Ziel dieses Aufsatzes, Audio-Übertragungssysteme in Einzelnen zu diskutieren. Das oben gezeigte Schema eignet sich aber sehr gut dazu, deutlich zu machen, was unter Analyse und Synthese von auditiven Szenen zu verstehen ist.

Zunächst zur Analyse: Der Zuhörer (auf der rechten Seite des Bildes) hört dem zu, was ihm von dem System akustisch dargeboten wird. In seinem Wahrnehmungsraum formiert sich eine auditive Szene, die er je nach Situation und Aufgabenstellen perzeptiv und mental analysiert. Die Analyse kann sich z.B. darauf beziehen, Sprecher zu orten und sich auf einen bestimmten zu konzentrieren, um diesen besser zu verstehen. Sie kann aber z.B. auch dazu dienen, die Qualität eines Produktgeräusches in Funktionszusammenhang zu bewerten.

Forschungsziel der instrumentellen Analyse auditiver Szenen ist es nun, die perzeptiven und mentalen Leistungen des Zuhörers zu verstehen, zu modellieren und instrumentell - d.h. in der Regel durch Software-Algorithmen - nachzubilden. Ein System zur instrumentellen Szenenanalyse hat somit das in Bild 2 gezeigte Prinzipbild.

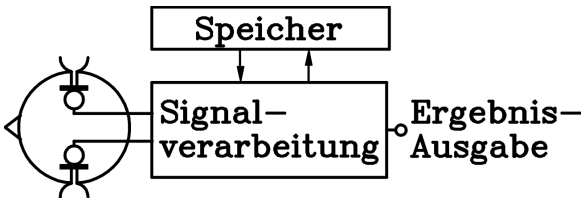

Bild 2: Prinzipbild zur instrumentellen Analyse auditiver Szenen

Entsprechend gelangt man zum Prinzipbild der instrumentellen Synthese auditiver Szenen (Bild 3). Hierbei wird nicht die Empfängerseite, sondern die Sendeseite des Übertragungssystems modelliert und instrumentell nachgebildet - und zwar auch hier in der Regel durch Software-Algorithmen. Die Modellierung kann sich auf die Quellsignale der Schallquellen und/oder die Schallausbreitung (von den Quellen bis zu den Trommelfellen der Zuhörer) erstrecken. Gegebenenfalls sind Rückwirkungen von Aktionen des Zuhörers auf die Schallsignale an seinen Ohren zu berücksichtigen.

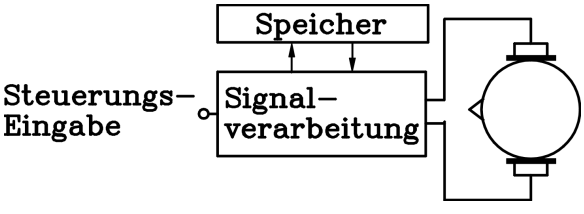

Bild 3: Prinzipbild zur instrumentellen Synthese auditiver Szenen

4. Analyse auditiver Szenen

Für Verfahren zur instrumentellen Analyse auditiver Szenen (engl.: computational auditory scene analysis, CASA) besteht ausgeprägter technologischer Bedarf. Das Gebiet ist deshalb z.Zt. international ein Schwerpunkt der Forschung.

Wichtige Anwendungsfelder sind Systeme zur Identifikation und Ortung von Schallquellen, u.a. unter akustisch schwierigen Bedingungen, wie sie in Mehrquellensituationen, in reflexionsbehafteten Umgebungen und bei Anwesenheit von Störschall gegeben sind - z.B. für akustische Überwachungs- oder Navigationssysteme. Weiterhin: Systeme zur Trennung und Klangentfärbung von konkurrierenden Schallquellensignalen (sog. cocktail-party processoren), die z.B. als front-end für Hörgeräte und robuste Spracherkenner benötigt werden.

Außerdem gilt für die Modellierung vieler Erkennungs- und Zuordnungsaufgaben in auditiven Bereich, dass es günstig oder sogar unumgänglich ist, zunächst eine Szenenanalyse vorzuschalten, z.B. bei Raumakustik-Analysesystemen, Systemen zur instrumentellen Qualitätsanalyse von Lautsprache und von Produktgeräuschen.

In diesem Zusammenhang sind auch die sog. „content filter“ erwähnenswert, die in der Zukunft immer wichtiger werden, um audiovisuelle Programmmaterialien für Archivierungs- und Suchzwecke automatisch inhaltlich zu analysieren und zu kodieren (vgl. die von ISO/IEC vorgeschlagene, content-bezogene MPEG 7-Kodierung).

Sofern die Verfahren zur auditiven Szenenanalyse auf dem Vorbild des menschlichen auditorischen Systems beruhen, haben die zugrundeliegenden Modelle in der Regel in etwa die im Folgenden dargestellte Struktur -. wobei hier nur die Grundzüge wiedergegeben werden können. Es sei jedoch angemerkt, dass es für einige auditive Analysesaufgaben auch nichtbiologische Ansätze gibt, auf die hier gar nicht eingegangen wird (z.B. gesteuerte microphon arrays oder die sog. blind source separation).

Die Modelle sind binaural aufgebaut, d.h. sie erhalten als Eingangssignale solche, die den Schallsignalen am linken und rechten Ohr eines Zuhörers entsprechen. Solche Signale können z.B. von einem Kunstkopfaufnahmesystem geliefert werden. Als erster Verarbeitungsschritt folgt oftmals - wenn überhaupt - je Kanal eine leichte Bandpassfilterung, die die Funktion des Mittelohres nachbildet.

Die Ausgangssignale des linken bzw. rechten Mittelohres werden dann jeweils einem Modul zugeleitet, der die Funktion des Innenohrs repräsentiert, nämlich: spektrale Zerlegung in spektral benachbarte Bandpasskomponenten, automatische Aussteuerungskontrolle und eine Art von Analog-nach-Digital-Wandlung. Während jedes Innenohrmodul nur ein Eingangssignal erhält, liefert es jedoch mehrere Ausgangssignale ab, nämlich eines pro Bandpassbereich. Diese Ausgangssignale sind, je nach Modellkonzept, Folgen von Impulsen (neural spikes) wechselnder zeitlicher Dichte oder aber ein Signal, das die Zeitfunktion der Spikehäufigkeit angibt.

In nächsten Schritt werden nun die von linken und vom rechten Innenohr ankommende Signale frequenzbandweise einem Modul zugeführt, der interaurale Laufzeitdifferenzen zwischen den linken und rechten Signalen ermittelt. Das hierzu verwendete Berechnungsverfahren beruht zumeist auf der Bildung einer interauralen Kreuzkorrelationsfunktion pro Bandpassbereich – oder ähnlichen Verfahren. Mehrere unkorrelierte Schallquellen die räumlich getrennt angeordnet sind, führen deshalb zumeist zu unterschiedlichen Gipfeln in der Kreuzkorrelationsfunktion. Die Gipfel können durch spezielle kontrastverstärkende Verarbeitungsschritte (z.B. contralateral inhibition) deutlich hervorgehoben werden. Aus Aus deren Lage und Form können dann die unterschiedlichen Quellen identifiziert und ihre räumliche Staffelung in seitlicher Richtung ermittelt werden.

Zusätzlich zur Analyse der interauralen Laufzeitdifferenzen erfolgt in der Regel eine Analyse der interauralen Pegeldifferenzen. Das Ergebnis dieser Analyse kann dann z.B. dazu verwendet werden, die Kreuzkorrelationsfunktionen sinnvoll zu modifizieren. Es sei im Übrigen darauf hingewiesen, dass wir auch mit nur einem Ohr hören kön-

nen. In den Modellen werden deshalb zumeist parallel zu den binauralen sog. sogenannte monaurale Verarbeitungswege vorgesehen.

Als Ergebnis der geschilderten, insoweit rein signalgetriebenen (bottom-up) Vorverarbeitung erhält man schließlich ein zeitlich veränderliches Aktivitätsmuster mit den drei Dimensionen Lage des Frequenzbandes, seitliche Auslenkung und Intensität. Dieses Aktivitätsmuster muss dann weiter verarbeitet werden.

Eine Reihe von Aufgaben der auditiven Szenenanalyse sind aufgrund dieses binauralen Aktivitätsmusters durch weiterführende Bottom-up-Verarbeitung lösbar. So gelingt es auf dieser Grundlage z.B. recht gut, mehrere Schallquellen in nicht zu halliger Umgebung zu orten und zu verfolgen. Auch die Trennung und Klangentfärbung vom konkurrierenden Sprachsignal gelingt in solchen akustisch „trockenen“ Umgebungen erstaunlich gut - die Leistungen des Menschen werden sogar übertroffen.

Leider fallen insbesondere die Trennungsleistungen der Bottom-up-Algorithmen rapide ab, sobald ein stärkerer Anteil reflektierten Schalls vorliegt. Auch die Berücksichtigung weiterer Merkmale der Quellschalle über die interauralen Differenzen hinaus – bei konkurrierenden Sprachsignalen z.B. deren unterschiedliche Grundfrequenzen und Obertonreihen – hat bisher zu keiner durchgreifenden Verbesserung dieser Situation geführt.

Es scheint, dass die instrumentelle Analyse auditiver Szenen bisheriger Art an einem Wendepunkt angelangt ist. Weitere Fortschritte erfordern die Einbeziehung von zusätzlichem Wissen über die zu analysierenden Szenen.

Die Einbeziehung von Wissen ist aber nur durch wesentliche Erweiterungen der Modellstruktur möglich. Den signalgesteuerten Signalverarbeitungsschritten können z.B. hypothesengetriebene Schritte (top-down processing) auf Signal- und/oder Symbolniveau folgen. Der Beweis für die Tüchtigkeit solcher Strukturen ist im Bereich der Technologie der Spracherkennung bereits erbracht worden

Bild 4 gibt eine Überblick über die vorgeschlagene Gesamtarchitektur. Auf das binaurale Aktivitätsmuster folgt ein Stufe, die eine erste Segmentierung und Kategorisierung vornimmt und so eine - fehlerbehaftete - symbolische Repräsentation des binauralen Aktivitätsmusters produziert. Diese Repräsentation wird dann auf einer „Wandtafel (black board)“ abgelegt und von unterschiedlichen Expertenmodulen evaluiert.

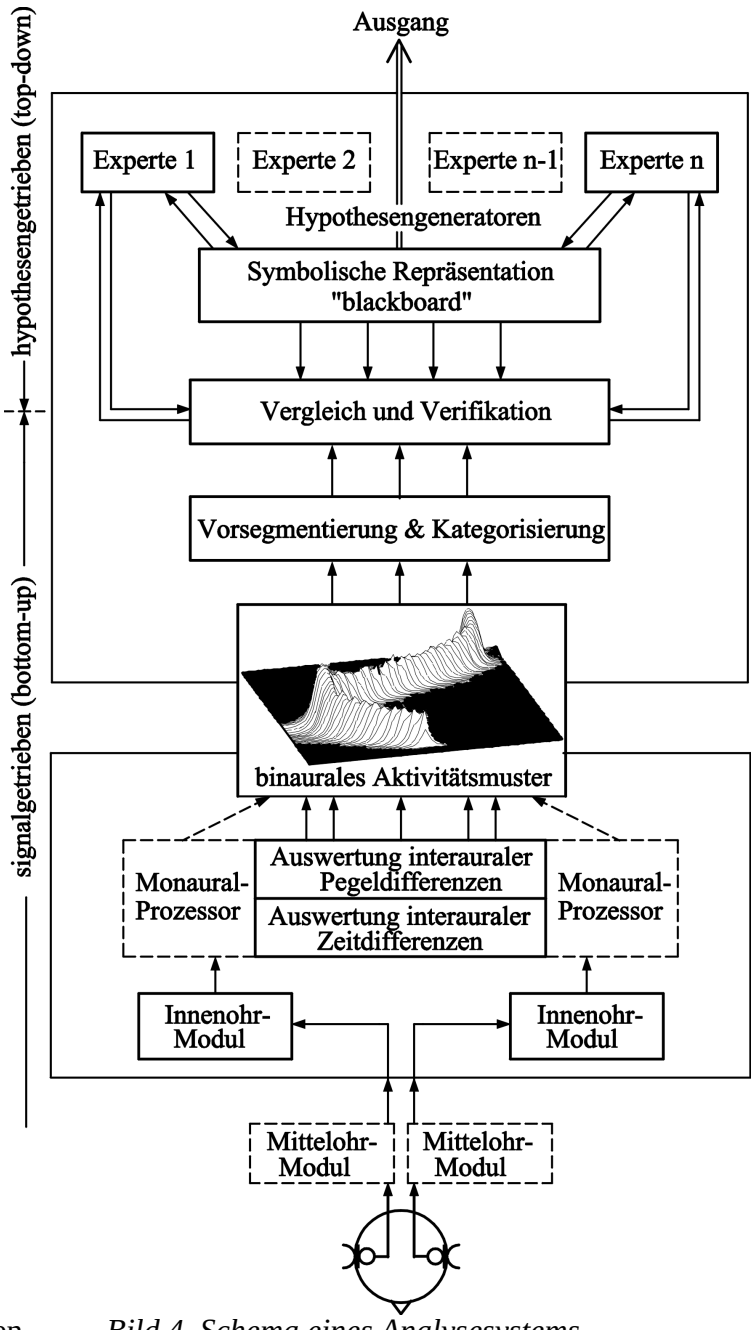

Bild 4 Schema eines Analysesystems für auditive Szenen

Die Expertenmodule erzeugen Hypothesen mit dem Ziel, eine plausible Erklärung für die gleitenden Aktivitätsmuster zu finden, die auf eine sinnvolle Identifikation und Analyse der Szene führt. Diese Hypothesen werden dann schrittweise geprüft, ggf. variiert und schließlich angenommen oder abgelehnt.

Jedes Expertenmodul agiert auf der Grundlage eines bestimmten Wissensbereiches. Das Wissen kann jeweils z.B. in Form expliziter Regeln oder als Datenbasis vorhanden sein. Wissensbereiche können z.B. sein: Wissen über die augenblickliche Kopfmikrophonposition, Vorwissen über die Szene, Information aus anderen Sinnesbereichen (z.B. Sehen,. Tasten), Wissen über Eigenschaften der Schallquellen und ihrer Schallsignale.

Wenn schließlich eine plausible Repräsentation der auditiven Szene ermittelt wurde, ist das weitere Vorgehen - d.h. die Verwendung der ermittelten Daten - aufgabenabhängig.

5. Synthese auditiver Szenen

Die instrumentelle Synthese auditiver Szenen ist gegenwärtig von noch größerer technologischer Relevanz als deren instrumentelle Analyse. Dies gilt insbesondere für die Synthese interaktiver auditiver Szenen, das heißt solcher, in denen die Zuhörer in den generierten Szenen agieren und diese interaktiv beeinflussen können. Man spricht in diesen Zusammenhang auch von „virtuellen auditiven Umgebungen“.

Wegen der Vielzahl der Anwendungen, die für virtuelle auditive Umgebungen diskutiert und implementiert werden, werden hier exemplarisch solche genannt, an denen das Bochumer Institut für Kommunikationsakustik beteiligt war, noch ist oder sein wird. Solche Anwendungen sind:
Auditive Displays für Zivilflugzeugpiloten, für die binaurale Simulation zur raumakustischen Planung und Evaluation von Räumen für Musikdarbietungen, für individuellen und interaktiven Kino-Sound, für Telekonferenzsysteme. Weiterhin virtuelle Tonstudios und Abhörräume, virtuelle Musik-Überäume sowie raumakustische Effektgeräte. Ferner die auditive Repräsentation in multimodalen virtuellen Umgebungen für u.a. folgende Zwecke: Motorrad- und KFZ-Fahrsimulation, Unterstützung von motorischen Rehabilitationsmaßnahmen, Archivierung von Kulturgütern (virtual heritage), Training von Sicherheitskräften in aggressiven Menschenmengen (riots), Internet-Kiosks. Außerdem wurde ein Generator zur Erzeugung von auditiv/taktilen virtuellen Umgebungen für Forschungszwecke erstellt, der z. Zt. zur Erforschung des auditiven Präzedenzeffektes eingesetzt wird.

In Folgenden wird der prinzipielle Aufbau eines Generators für virtuelle auditive Umgebungen kurz geschildert. Um deutlich zu machen, dass solche Systeme zumeist in der Regel in komplexere eingebettet sind, ist in Bild 5 als Beispiel das Schema eines multimodalen sog. Virtual-Reality Generators dargestellt. Die folgenden Beschreibung bezieht sich jedoch zunächst allein auf dessen auditive Komponente.

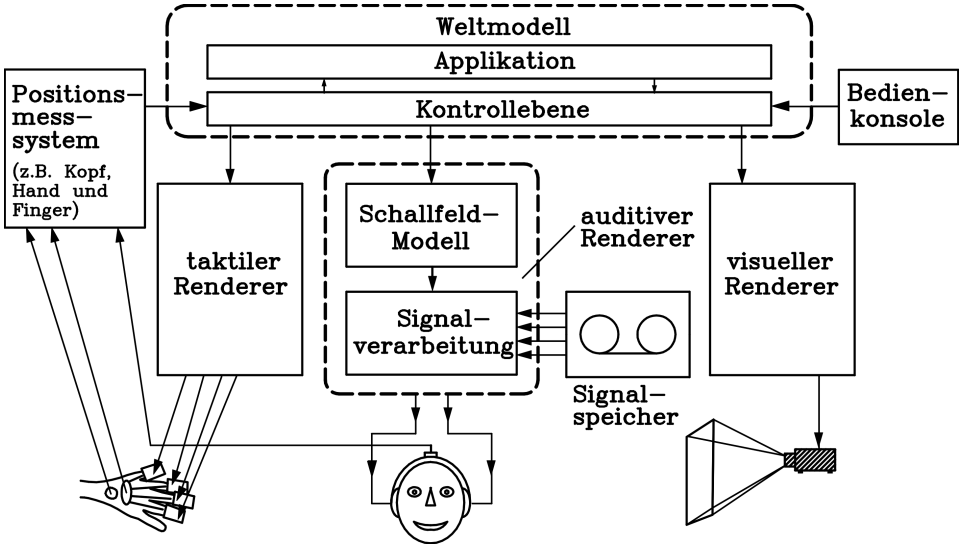

Bild 5: Schema eines Virtual-Reality Generators mit auditivem, taktilem und visuellem Display

Im Bild ist ein System mit Kopfhörerwiedergabe gezeichnet. Prinzipiell sind virtuelle auditive Umgebungen allerdings auch mit Lautsprecherwiedergabe möglich. Die heute in der Konsumelektronik (Kino, TV, CD, DVD, Rundfunk) schon häufig verwendeten Mehrkanal-Wieder-

gabesystem können als Vorstufe virtueller auditiver Umgebungen angesehen werden – ohne allerdings dem Benutzter Möglichkeiten zur Interaktion zu bieten.

Das Beispielsystem enthält im Kern ein Weltmodell, nämlich eine Datenbank, in der Beschreibungen aller vorgesehenen virtuellen Objekte abgelegt sind. Je nach spezieller Anwendung werden Regeln festgelegt, die zwischen diesen Objekten in der virtuellen Welt gelten. Ein zentrales Kontrollmodul erfasst die Aktionen der Personen, die das System interaktiv benutzen und leitet die gewünschten Reaktionen bzw. Aktionen des Systems. Im Beispielsystem werden laufend Kopfpositionen, Handpositionen sowie Fingerpositionen der Benutzer erfasst. Die Kopfpositionen sind z.B. wichtig, da die über Kopfhörer dargebotenen Schallsignale ständig angepasst werden müssen, damit die Zuhörer eine räumlich invariante auditive Szene wahrnehmen - die Szene sich also bei Kopfbewegungen nicht mitbewegt. Mit Händen und Fingern können die Benutzer eingreifen - z.B. Schallquellen im virtuellen Raum bewegen.

Diejenige Systemkomponente, die Signale für den Kopfhörer generiert, heißt „auditiver Renderer“. Sie enthält als wesentlichen Anteil ein Schallfeldmodell. Dies ist ein Modul, welches die folgende Daten erhält und damit einen Satz sog. binauraler Impulsantworten erzeugt: geometrische Daten des virtuellen Raumes, Absorptionsdaten aller Wände und Einrichtungsgegenstände, Positionen und akustische Richtcharakteristiken der virtuellen Schallsender und der Benutzer. Die akustischen Richtcharakteristiken der Benutzer – besser gesagt, ihrer Köpfe und Ohren - sind als Sätze von kopfbezogenen Impulsantworten für alle möglichen Schalleinfallsrichtungen einzugeben. Sie müssen vorher an den Benutzern direkt individuell ausgemessen werden.

Die binauralen Impulsantworten - die im Übrigen ca. 30mal in der Sekunde neu berechnet werden, werden dann mit elektronisch generierten oder natürlichen Signalen (z.B. Sprache, Musik, Produktgeräusche) „gefaltet“. Die Signale sollten vor der Faltung noch keine Reflexionsanteile enthalten, also reflexionsarm sein. Die nach der Faltung erhaltenen (binauralen) Signale werden dann den Zuhörern über Kopfhörer zugeführt.

Es wird bei vielen Anwendungsfällen virtueller Umgebungen angestrebt, dass die Zuhörer sich perzeptiv in diese versetzt wahrnehmen, sich in der virtuellen Umgebung „präsent“ fühlen. Dies gilt insbesondere dann, wenn man möchte, dass die Benutzer der virtuellen Umgebung sich in dieser intuitiv so verhalten, wie sie es in einer korrespondierenden realen Umgebung täten. Bei Mensch-System-Schnittstellen, die auf Prinzipien virtueller Umgebungen beruhen, kann so z.B. die Bedienung vereinfacht werden. Man denke z.B. an Teleoperations-, Entwurfs- oder Dialogsysteme, oder auch an Computerspiele (!).

Der Aufwand den man treiben, muss um perzeptive Präsenz zu erreichen, hängt stark von dem speziellen Anwendungsfall und von den speziellen Aufgaben der Nutzer ab. Der notwendige technische Aufwand ist z.B. bei der auditiven Repräsentation in einem Fahrzeugsimulator wesentlich geringer als bei einem virtuellen Abhörraum für Audioingenieure und Tonmeister.

Grundsätzlich gilt, dass die virtuelle Umgebung unter Berücksichtigung der in ihr zu leistenden Aufgaben hinreichend plausibel (glaubwürdig) sein muss.

Es sind im Rahmen der binauralen Raumsimulation Schallfeldmodelle entwickelt worden, die es erlauben, den Höreindruck in bestehenden realen Räumen so gut zu simulieren, dass Testpersonen im Direktvergleich nicht mehr sicher feststellen können, welche der beiden Darbietungen aus dem realen und welche aus dem simulierten Raum stammt. Bei diesen „statischen“ Simulationen wird allerdings nicht gefordert, dass die Berechnung des Schallfeldes und damit der binauralen Impulsantworten in Realzeit erfolgt. Selbst für die Faltung ist dies nicht notwendig. Entsprechend aufwändig können die Berechnungsalgorithmen für die Simulation gestaltet werden.

Typische solche Systeme verfolgen mehrere tausend Reflexionen individuell und fügen zusätzlich noch einen simulierten Nachhall hinzu.

Die verwendeten Simulationsalgorithmen sind dabei nicht auf Prinzipien der geometrischen Akustik (Verfolgung von Strahlen und Strahlenbündeln, Spiegelschallquellen) beschränkt, son-

dern es werden auch numerische Methoden zur Lösung der Wellengleichung eingesetzt (z.B. finite elements, boundary elements), insbesondere für die tieffrequenten Anteile. Die Berücksichtigung von Streuung und Brechung ist möglich.

Sobald jedoch Interaktionen zugelassen werden, und dies ist bei Generatoren für virtuelle Umgebungen die Regel, ist Echtzeit-Signalverarbeitung zwingend erforderlich ist. Die Reaktion des Systems auf Aktionen des Benutzers muss innerhalb einer perzeptiv glaubwürdigen Zeitspanne erfolgen (für die auditive Komponente innerhalb von ca. 50 ms). Weiterhin muss die Darbietung hinreichend oft aufgefrischt werden (für die auditive Komponente mehr als 30mal pro Sekunde), damit keine „ruckenden“ oder „auditiv flackernden“ Hörereignisse entstehen. Bei Szenen, die sich schnell ändern, sind die Anforderungen ggf. noch strenger.

Bei auditiven Szenen mit Schallquellen- und/oder Zuhörern, die ihre Position schnell ändern, treten Dopplereffekte auf, deren Simulation grundsätzlich möglich ist, aber sehr aufwändig sein kann.

Die Schallfeldmodelle sind wegen der Realzeitanforderungen bei interaktiven Systemen anders aufgebaut als bei statischen Simulationssystemen. An die Stelle einer physikalisch authentischen Simulation tritt eine perzeptiv plausible. Für den Entwickler sind deshalb detaillierte Kenntnisse über die menschliche Wahrnehmung unverzichtbar, denn es muss entschieden werden, welche Signalmerkmale in jedem Augenblick aufgabenspezifisch perzeptiv relevant sind und deshalb unverzüglich mit hoher Genauigkeit dargeboten werden müssen. Weniger relevante Merkmale können ggf. weniger genau und/oder später dargeboten werden.

Wenn die Benutzer innerhalb der virtuellen Umgebung ihre eigene Stimme benutzen – z.B. in Telekonferenzsystemen – so muss diese vom jeweiligen Sprecher selbst als natürlich klingend wahrgenommen werden, sonst stellt sich keine Präsenz ein. Durch sorgfältige Analyse, Simulation bzw. Kompensation der beteiligten Schallübertragungswege ist dieses Ziel erreichbar.

Die moderne Sprachtechnologie bietet Komponenten an, die sich in virtuelle Umgebungen aufgabenspezifisch integrieren lassen. Beispiel sind Systeme zur maschinellen Sprachsynthese und Spracherkennung. Die Interaktion des Nutzers mit dem System kann dann mittels der Stimme erfolgen, das System kann mit Sprechsignalen reagieren, die für den Nutzer verständlich sind. Systeme für Mensch-System-Dialog bzw. Mensch-System-Interaktion, in denen eine virtuelle Umgebung als Mensch-System-Schnittstelle dient, sind schon mehrfach vorgestellt worden – z. Zt. allerdings noch in recht rudimentärer Form.

Da virtuelle Welten künstlich sind, dass heißt mit Hilfe von Computern generiert werden, beruhen sie auf einer parametrischen Repräsentation dieser Welten. Die Parameter, die eine virtuelle Welt und die Aktionen in ihr beschreiben, können codiert und über Raum und Zeit übertragen werden. Es existieren weiterhin bereits Beschreibungsprachen, mit denen sich virtuelle Welten formal definieren und spezifizieren lassen. Die Beschreibungen beziehen sich auch auf inhaltliche (semantische) Aspekte. Die oben bereits erwähnte MPEG7-Codierung spielt auch in diesem Zusammenhang eine Rolle.

Durch die parametrische Beschreibung virtueller Umgebungen ergibt sich z.B. die Möglichkeit dass Nutzer, die sich realiter an unterschiedlichen Orten befinden, sich virtuell in einem gemeinsamen Raum versetzen, z.B. um dort eine Konferenz miteinander abzuhalten (teleconferencing) oder gar gemeinsame handwerkliche Aufgaben zu erledigen (teleoperation). Weiterhin kann man sich beispielsweise in einen virtuelle Raum begeben, um diesen zu inspizieren und zu besichtigen (z.B. virtuelles Museum, virtueller Tourismus). Da der Zugang zu virtuellen Räumen u.a. über das Internet ermöglicht werden kann, sind die denkbaren Anwendungsmöglichkeiten sehr vielfältig.

Virtuellen Umgebungen können weiterhin realen Umgebungen überlagert werden, z.B. um die Navigation in diesen zu unterstützen oder andere On-Line-Handlungshilfen zu geben (sog. augmented reality).

Einen besondere Rolle spielen virtuelle Umgebungen in der Forschung, denn sie erlauben es, komplexe Versuchsszenarien flexibel und aufwandsgünstig darzustellen. Der Szenarienwechsel

erfordert keinen physischen Aufwand und die Datendarbietung und –gewinnung kann automatisiert werden. Die Forschung z.B. auf den Gebieten Psychophysik, Psychologie, Usabilty, Produktqualitätsbeurteilung, kann hieraus großen Nutzen ziehen und tut es zum Teil auch bereits.

Das Institut für Kommunikationsakustik hat auf allen hier der angesprochenen Feldern eigene Beiträge geleistet und die erzielten Forschungs- und Entwicklungsergebnisse für die kommerzielle Nutzung zur Verfügung bereitgestellt, z.B. über Verbundprojekte mit der Industrie.

6. Diskussion und Schlussbemerkungen

Es ist Ziel dieses Beitrages, deutlich zu machen, dass sich die "nachrichtentechnische" Sparte der Akustik in den letzten Jahrzehnten grundlegend gewandelt hat. Aus der Elektroakustik, die sich aus der Symbiose von Elektrotechnik und Akustik gebildet hatte, ist durch Hinzutreten der digitalen Signalverarbeitung und der Informatik die Kommunikationsakustik geworden. Eine deutliche Software-Orientierung ist unverkennbar.

Die Kommunikationsakustik befasst sich mit den akustischen und auditiven Aspekten der Informationstechnik. Ihr Arbeitsgebiet ist interdisziplinär geprägt. Zur Lösung der gebietsspezifischen ingenieurwissenschaftlichen Fragegestellung muss ergänzend Wissen aus unterschiedlichsten Gebieten importiert werden, z.B. aus Psychologie, Biologie, Neuroinformatik, Medizin sowie Sprach- und Musikwissenschaften.

Am Beispiel der Analyse auditiver Szenen wurde gezeigt, dass ein wichtiger Teil laufender kommunikationsakustischer Forschung darin besteht, Algorithmen zu entwickeln, die die Analyse und Erkenntnisfähigkeiten des Menschen modellieren - oder diese sogar übertreffen. Um dies zu erreichen, sind wissensbasierte Komponenten erforderlich. Über die reine Signalverarbeitung hinaus wird also Symbolverarbeitung und Verarbeitung von semantischen Inhalten erforderlich. In der Sprachtechnologie ist dieser Schritt bei der instrumentellen Spracherkennung z.T. schon vollzogen worden.

Auf der Syntheseseite fällt auf, dass die Systeme zunehmend multimodal werden. Im Verbund mit der auditiven wird z.B. die visuelle und die taktile Repräsentation (inkl. Vibrationen) angeboten. Dies alles geschieht oftmals interaktiv. Die Synthese ist parametergesteuert, wobei content-bezogene Parameter in steigenden Maße wichtig werden. Auch hier spielt die Sprachtechnologie eine Vorreiterrolle (z.B. bei Sprach-Dialogsystemen).

In modernen informationstechnischen bzw. kommunikationstechnischen Anlagen sind die akustisch/auditive Komponenten oftmals nur eingebettet (embedded components). Ihre Funktionen isoliert zu betrachten, wäre wenig sinnvoll. Entsprechendes gilt für die Kommunikationsakustik als ingenieurwissenschaftliche Disziplin. Sie erlangt ihre volle Bedeutung erst im Konzert mit all den anderen Disziplinen, die die moderne Informationstechnik ausmachen.

Allerdings machte es wenig Sinn, die Kommunikationsakustik an Hochschulen ledig als Teilgebiet („Haustier") in einem größeren Fachgebiet unterzubringen (z.B. in der Medientechnik). Dazu ist das Gebiet der Kommunikationsakustik zu umfangreich und zu breit gefächert. Um es auch nur einigermaßen umfassend vertreten zu können, ist nach den langjährigen Erfahrungen des Autors eine „kritische Masse" von wenigstens Lehrstuhlgröße notwendig (Die Einwerbung von Drittmittel bietet auf diesem Gebiet in der Regel keine besonderen Schwierigkeiten). Damit kann die Kommunikationsakustik dann zusätzlich etwas liefern, was in keiner Informationstechnik-Fakultät fehlen darf, die sich bestimmungsgemäß u.a. auch mit Mensch-System-Interaktionen befassen muss - nämlich Wissen über die menschlichen Fähigkeiten bei der Informationsaufnahme, -verarbeitung und -ausgabe. Ohne dieses Wissen verfehlte die Informationstechnik ihr Ziel.

8. Danksagung

Der Autor bedankt sich bei den wissenschaftlichen Mitarbeitern des Institutes, insbesondere bei seinen Doktoranden. Eine Titelliste deren fertiggestellter Dissertationen ist angefügt:

Dissertationsliste

(1) Hudde, Herbert, (1980), Messung der Trommelfellimpedanz des menschlichen Ohres bis 19kHz

(2) Schlichthärle, Dietrich, (1980), Modelle des Hörens - mit Anwendungen auf die Hörbarkeit von Laufzeitverzerrungen

(3) Braas, Jürgen, (1981), Ein digitales Leitungsmodell als Hilfsmittel zur Sprachsynthese

(4) Vogelsang, genannt Buschmann, Ulrich, (1982), Untersuchung der Schallausbreitung im inhomogenen Strömungsfeld mit asymptotischen mathematischen Methoden

(5) Lenz, Heino, (1983), Messung der akustischen Wandadmittanzen hochabsorbierender Wandauskleidungen in reflexionsarmen Räumen

(6) Schröter, Jürgen, (1983), Messung der Schalldämmung von Gehörschützern mit einem physikalischen Verfahren (Kunstkopfmethode)

(7) Rühl Hans-Wilhelm, (1984), Sprachsynthese nach Regeln für unbeschränkten deutschen Text

(8) Lindemann, Werner, (1985), Die Erweiterung eines Kreuzkorrelationsmodells der binauralen Signalverarbeitung durch kontralaterale Inhibitionsmechanismen

(9) Els, Hartmut, (1986), Ein Meßsystem für die akustische Modelltechnik

(10) Pösselt Christoph, (1986), Einfluß von Knochenschall auf die Schalldämmung von Gehörschützern

(11) Henrich, Peter, (1988), Sprachenidentifizierung zur automatischen Graphem-zu-Phonem-Umsetzung von Fremdwörtern in einem deutschsprachigen Vorleseautomaten

(12) Letens, Uwe, (1988), Über die Interpretation von Impedanzmessungen im Gehörgang anhand von Mittelohr-Modellen

(13) Jekosch, Ute, (1989), Maschinelle Phonem-Graphem-Umsetzung für unbegrenzten deutschen Wortschatz (Universität/GH Essen)

(14) Gaik, Werner, (1990), Untersuchungen zur binauralen Verarbeitung kopfbezogener Signale

(15) Kesselheim, Michael, (1990), Computergestützte Konstruktion großer Wortklassensysteme

(16) Col, Jean-Pierre, (1990), Localisation auditiv d'un signal et aspects temporels de l'audition spatiale (Universität Aix/Marseille)

(17) Wolf, Siegbert, (1991), Lokalisation von Schallquellen in geschlossenen Räumen

(18) Xiang, Ning, (1991), Mobile Universal Measuring System for the Binaural Room-Acoustic-Model Technique

(19) Bodden, Markus, (1992), Binaurale Signalverarbeitung: Modellierung der Richtungserkennung und des Cocktail-Party-Effektes

(20) Lehnert, Hilmar, (1992), Binaurale Raumsimulation: Ein Computermodell zur Erzeugung virtueller auditiver Umgebungen

(21) Böhm, Arnd, (1992), Maschinelle Sprachausgabe von deutschem und englischem Text

(22) Slatky, Harald, (1993), Algorithmen zur richtungselektiven Verarbeitung von Schallsignalen eines binauralen Cocktail-Party-Prozessors

(23) Pompetzki, Wulf, (1993), Psychoakustische Verifikation von Computermodellen zur binauralen Raumsimulation

(24) Belhoula, Karim, (1996), Ein regelbasiertes Verfahren zur maschinellen Graphem-nach-Phonem-Umsetzung von Eigennamen in der Sprachsynthese

(25) Kraft, Volker, (1996), Verkettung natürlich sprachlicher Bausteine zur Sprachsynthese: Anforderungen, Techniken und Evaluierung

(26) Knohl, Lars, (1996), Prosodiegesteuerte Sprecher- und Umweltadaption in einer Mehrsprecher-Architektur

(27) Giron, Franck, (1997), Investigations about the Directional Characteristics of Sound Sources

(28) Grabke, Jörn, (1997), Ein Beitrag zum Richtungshören in akustisch ungünstigen Umgebungen

(29) Hartung, Klaus, (1998), Modellalgorithmen zum Richtungshören, basierend auf den Ergebnissen psychoakustischer und neurophysiologischer Experimente mit virtuellen Schallquellen

(30) Hegehofer, Thomas, (1998), Ein Analysemodell zur rechnerbasierten Durchführung von auditiven Sprachqualitätsmeßverfahren und seine Realisiertung

(31) Rateitschek, Klaus, (1998), Ein binauraler Signalverarbeitungsansatz zur robusten maschinellen Spracherkennung in lärmerfüllter Umgebung

(32) Bednarzyk, Michael, (1998), Qualitätsbeurteilung der Geräusche industrieller Produkte: Der Stand der Forschung, abgehandelt am Beispiel der KFZ-Innenraumgeräusche

(33) Möller, Sebastian, (1999), Assessment and Prediction of Speech Quality in Telecommunication

(34) Lehn, Karsten, (2000), Unscharfe zeitliche Clusteranalyse von monauralen und interauralen Merkmalen als Modell der auditivem Szenenanalyse

(35) Mersdorf, Joachim, (2000), Sprecherspezifische Parametrisierung von Sprachgundfrequenzverläufen: Analyse, Synthese und Evaluation

(36) Korany, Noha, (2000), A Model for the Simulation of Sound Fields in Enclosures: Integrating the Geometrical and the Radiant Approaches (University of Alexandria)

(37) Strauss, Holger, (2000), Simulation instationärer Schallfelder in auditiven virtuellen Umgebungen

(38) Pörschmann, Christoph, (2001), Eigenwahrnehmung der Stimme in auditiven virtuellen Umgebungen

Zudem in Kürze:

(39) Hempel, Thomas, (2001), Ein Ansatz zur Klassifikation der Beziehungen zwischen auditiven und instrumentellen Merkmalen von KFZ-Innengeräuschen (TU Berlin)

(40) Pellegrini, Renato, (2001), A Virtual Reference Listening Room as an Application of Auditory Virtual Environments

(41) Brüggen, Marc, (2001), Binaurale Klangentfärbung

(42) Dürrer, Bernd, (2001), Gestaltung von Mensch-Maschine- Schnittstellen mit Hilfe auditiver virtueller Umgebungen

(43) Geravanchizadeh, Masoud, (2001), Spektrale Sprecheranpassung durch stückweise lineare Interpolation