

# Interaktion in auditiven Informationsräumen

Palle Klante, Peter Gorny, Michael Gründler

## Einführung

Im INVITE-Teilprojekt „Zugang zum Internet für Blinde“ wurde ein auditiver Interaktionsraum zur Ergänzung der Fenster von gängigen Webbrowsern, z.B. dem Internet Explorer entwickelt, um der Benutzergruppe der Blinden den Zugang zu Webangeboten zu ermöglichen. Das Layout der Webseite wird dem blinden Benutzer als Virtual-Reality-Klangbild präsentiert, der textliche Teil wird vorgelesen. Es wurden mehrere Lösungen in Form von Gestaltungskonzepten, prototypischen Systemen und Anwendungsszenarien entwickelt und evaluiert. Verwendet wurden spezielle Stereosoundcards, die die Ausgabe von virtuellen Klangwelten ermöglichen.

Das Internet hat für blinde Benutzer eine besondere Funktion, da es Defizite auf Grund der Erblindung in vielen privaten und öffentlichen Bereichen kompensieren kann. Es dient zum Kommunizieren mit anderen Menschen, als Einkaufshilfe oder zum Abrufen von allgemeinen Informationen. Entsprechend ist das Surfen zum reinen Vergnügen nebensächlich, und das Navigieren innerhalb einer Webseite oder einer Website ist eher zielorientiert. Meist möchten sich blinde Benutzer rasch einen Überblick über die Internetseite verschaffen und sich nicht durch unnötig kompliziertes, zeitraubendes Suchen aufhalten lassen. Deshalb machen wir den Blinden die wesentlichen Layoutstrukturen zugänglich, so dass sie schneller als mit Hilfe vorhandener Screenreader und den damit verbundenen sequenziellen Erkunden einer Webseite zu ihrem Ziel kommen.

### *Problemstellung*

Der Zugang zum Informationsangebot des World Wide Web wird blinden und stark sehbehinderten Menschen erschwert, weil die Informationen

nicht nur textuell präsentiert, sondern durch grafische Elemente strukturiert (Layout) und oft durch Bilder ergänzt werden. Webseiten (als Bestandteil des Graphical User Interfaces (GUI) der Computer Systeme) sind am besten durch die Interaktionsform der direkten Manipulation zu erschließen. Für normalsichtige Computerbenutzer bringt diese Interaktionsform viele Vorteile, blinden Menschen jedoch eine zusätzliche Barriere.

Die Informationsdarstellung auf Webseiten ist im Laufe der Jahre immer stärker grafisch orientiert und multimedialer geworden. Wurden im HTML 2.0-Standard im Wesentlichen Überschriften, Texte, Verweise und Grafiken genutzt, sind im Laufe der Zeit Tabellen, Formulare, Applets und dynamische Elemente hinzugekommen. Auch die Anzahl der gleichzeitig dargestellten Elemente ist deutlich höher, wodurch die Komplexität einer Webseite gesteigert wird. Guidelines für die Gestaltung von Webseiten sind darauf ausgerichtet, ein gutes optisches Bild zu schaffen und weniger für eine wohlgeformte Struktur des Dokuments zu sorgen. Nachdem diese Problematik erkannt worden war, haben Institutionen wie das World Wide Web-Consortium Regeln aufgestellt, die die Entwickler letztendlich nur daran erinnern sollen, die Webseiten für eine breitere Benutzergruppe zu entwickeln.

Rein textuelle Umsetzungen von vorhandenen Webseiten („text-only“) präsentieren dem Blinden durch ein Braille-Ausgabegerät oder durch ein Vorlesegerät („text-to-speech-system“) den linearisierten Inhalt. Trotz solcher Hilfsmittel gibt es für Blinde weiterhin Barrieren in grafischen Benutzungsoberflächen:

- die Pixel-Barriere – Die von einer Anwendung generierten Texte werden innerhalb des Systems nur als Pixelsammlung im Grafikspeicher abgelegt. Sollen diese Informationen erhalten bleiben, müssen diese vor der Verarbeitung im Grafikspeicher verarbeitet werden.
- die Maus-Barriere – Die Benutzung eines relativen Zeigegerätes auf einer visuellen Ausgabebläche ist Blinden aufgrund des mangelnden Feedback über die aktuelle Position nicht möglich.
- die Grafik-Barriere – Systemzustände und Informationen werden nicht durch textuelle Beschreibungen ausgedrückt, sondern werden in Form von Piktogrammen oder grafischen Darstellungen dem Benutzer angeboten.

Mit Hilfe von Screenreadern, der Beachtung einiger Regeln und der Verwendung von definierten Schnittstellen können diese Barrieren überwunden, bzw. umgangen werden. Bei der Benutzung von Webseiten stellt sich jedoch durch die gestalterische Freiheit des Designers eine weitere Hürde,

- die Layout-Barriere – Die Platzierung von Objekten auf einer Seite erfolgt in der Regel in Abhängigkeit von anderen Objekten. Die Position eines Objektes ist für die Bearbeitung der Seite von Interesse und enthält Informationen über die Bedeutung des Objekts. Das Layout folgt dabei einer Formensprache, die zum Leidwesen der Informatiker nicht wohldefiniert, aber trotzdem – zumindest von professionellen Grafikern systematisch - zur Steuerung der Aufmerksamkeit des Benutzers eingesetzt wird.

Diese Barriere wird von heutigen Screenreadern nicht überwunden, da die notwendigen Informationen nicht textuell in den Ausgabefluss eingebunden werden können.

#### *Lösungsansatz*

Zur Überwindung der Layout-Barriere werden die Komponenten einer Webseite als non-verbale Hearcons dargestellt, so dass der Benutzer sich die Topografie der Webseite mit Hilfe des Raumklangs erschließen kann. Die Idee ist, die zweidimensionale Darstellung des grafischen Webbrowsers direkt auf einen zweidimensionalen virtuellen Hörschirm in einem dreidimensionalen virtuellen auditiven Interaktionsraum mit Hilfe eines Softwaresystems „AIRclient“ (Auditiver Interaktionsraum-Client) vor dem Benutzer einzublenden. Textinformationen werden im AIRclient über ein Text-to-Speech-System vorgelesen. Im auditiven Modell der Webseite wird modelliert, welche Informationen dem Benutzer in welcher Form bei der Interaktion mit einem Hearcon ausgegeben werden. Über die Sprachausgabe können zusätzlich die wesentlichen Metainformationen eines Hearcons mitgeteilt werden.

### **State of the Art**

Die erste Phase intensiverer Forschung zu auditiven Benutzungsoberflächen begann zum Ende der achtziger und zum Beginn der neunziger Jahre und wurde dokumentiert in (Buxton 89) und (Kramer 94). Die Entwicklungen und Ergebnisse werden seither im Wesentlichen durch die Proceedings der jährlich stattfindenden Konferenz der „International Community of Auditory Displays“ (ICAD) dokumentiert.

Die fundamentale Idee der auditiven Benutzungsoberflächen beruht auf dem Gedanken, dass wir Akustik in unserem täglichen Leben ständig um uns haben und uns der Hörsinn hilfreiche Informationen liefert. Diese Erfahrungen aus dem alltäglichen Leben müssen nur auf die Verwendung im Computer übertragen werden. Entsprechend den Entwicklungszyklen bei

grafischen Benutzungsoberflächen, die durch die zur Verfügung stehende Hardware beeinflusst worden sind, lassen sich auch historisch diverse Stufen des Einsatzes von Akustik im Rechner identifizieren. In den Anfängen wurden nur simple „Pieps“ aus dem Systemlautsprecher ausgegeben. Durch die Einführung von Soundkarten konnten qualitativ hochwertige und flexibel einsetzbare Wave- und MIDI-Files ausgegeben werden. Zunächst war die Ausgabe auf eine Wave-Datei beschränkt, später konnten jedoch auch parallel mehrere Dateien über die Soundcard abgespielt werden. In den letzten fünf Jahren war durch die Einführung einer neuen Generation von relativ preiswerten Soundcards auch die räumliche Positionierung von Geräuschen möglich. Bei den Schallerzeugern verlief die Entwicklung über die Anbindung von zwei Lautsprechern, über Surround-Sound bis hin zu achtkanaligen Soundcards.

Es gilt drei Hauptrichtungen der Forschung zu betrachten: Die Physik liefert Vorgaben zur Herstellung notwendiger Hardware und Modelle zur Generierung der Schallquellen auf physikalischer Basis. Die Psychologen beobachten und bewerten die Wahrnehmung beim Menschen und seine Reaktionen auf unterschiedliche Reize. Die Informatik im Teilgebiet der Mensch-Technik-Interaktion beschäftigt sich mit der Integration dieser Modalität in den Rechner zur Erzeugung intuitiv nutzbarer Benutzungsoberflächen. Unterstützt werden die verschiedenen Bereiche durch Komponisten und Musiker, die die geforderten Eigenschaften der Geräusche umsetzen.

### *Konzepte*

Am Deutlichsten lassen sich die Merkmale von Akustik im direkten Vergleich zur Grafik beschreiben:

- Akustik ist abhängig von der Zeit und unabhängig vom Raum, die Grafik dagegen zeitabhängig und ortsgebunden.
- Akustik eignet sich für die Anzeige von Veränderungen und ist nur für eine kurze Zeit vorhanden, Grafik dagegen ist beharrend und wiederholt wahrnehmbar.
- Akustik wird als Vorgang wahrgenommen, Grafik dagegen als Zustand.
- Eine Ausrichtung des Kopfes zur Schallquelle ist unnötig und es sind nur eine limitierte Anzahl von Schallquellen gleichzeitig unterscheidbar, dagegen muss die Blickrichtung auf die Grafik gelenkt werden und dann können höchst komplexe Darstellungen wahrgenommen werden.
- Akustische Objekte sind unabhängig von ihrer Positionierung im Raum, aber nicht unabhängig von der Position des Zuhörers. Grafische

Objekte haben dagegen einen wohldefinierten Ort im Raum, in dem der Benutzer sich bewegt.

In verschiedenen Entwicklungen wurden akustische Elemente in die Benutzungsoberflächen der Computersoftware einbezogen:

Die Grundidee von Gavers **Auditory Icons** (Kramer 94, S. 417) ist das „Everyday Listening“. Wir nehmen mit Geräuschen unterschiedliche Eigenschaften eines zugehörigen Objekts unserer Umwelt wahr. Das wahrgenommene Geräusch ist demnach abhängig von der Art der Interaktion, dem Material des Objekts und der Umgebung, in der diese Interaktion stattfindet. Die Soundquelle wird durch die physikalischen Eigenschaften aller an der Interaktion beteiligten Objekte beeinflusst.

Eine andere Grundrichtung verfolgt Blattner mit **Earcons** (Buxton 89, S. 11). Sie führt eine organisatorische Teilung in modulare, hierarchische und transformationelle Icons auf der Designebene durch und unterteilt grafische Icons entsprechend in drei Teile: Repräsentierende, abstrakte und semi-abstrakte. Abstrakte Earcons stellen den Schwerpunkt ihrer Arbeit dar und werden durch geeignete Verfahren mit standardisierten Elementen in kleine Einheiten von Rhythmen zu großen Motiven zusammengeführt. Ein Motiv wird von Blattner als rhythmische Sequenz von Tonhöhen bezeichnet. Motive werden durch die Parameter Rhythmus, Tonhöhe, Dynamik und Länge beschrieben.

Die vorgestellten Konzepte für das Sounddesign haben bezogen auf die Kriterien, Erlernbarkeit, Differenzierbarkeit, Erkennbarkeit, etc. einige Vor- und Nachteile. Die Verwendung muss dabei kontext- und aufgabenbezogen betrachtet werden.

Eine Erweiterung der Konzepte der Auditory Icons und der Earcons stellen **Hearcons** (Bölke 97) dar. Diese Objekte müssen kein Pendant in einer grafischen Benutzungsoberflächen haben und es kann direkt mit ihnen in einem virtuellen Raum interagiert werden. Neben einem spezifischen Geräusch wird davon ausgegangen, dass ein auditives Objekt, welches auch im Sinne der Benutzungsoberfläche nutzbar ist, eine Lautstärke, eine Position im Raum und eine Interaktionsfläche haben muss. Die direkte Manipulation wird somit von den grafischen Benutzungsoberflächen auf die auditiven Benutzungsoberflächen übertragen. Die Hearcons greifen bei der Wahl des Sounddesign die Vor- und Nachteile der Auditory Icons und Earcons auf und verwenden jeweils das beste Konzept.

### *Anwendungen*

Sehr einfache auditive Umsetzungen von Elementen in grafischen Benutzungsoberflächen sehen gegenwärtig die Verwendung von akustischen Er-

eignissen als Warnmeldungen, Hinweise oder kritische Fehler zur Unterstützung von Dialogfenstern vor. Die Geräusche lenken die Aufmerksamkeit des Benutzers auf die grafische Ausgabe und die eigentlichen Informationen werden erst dort vermittelt. In weiteren Projekten, wie etwa dem AppleSonic Finder von Gaver (Kramer 94, Seite 421-422) wurden Events in einer grafischen Benutzungsoberfläche umgesetzt. Alle Interaktionen, wie etwa das Ziehen eines Dokumentes über den Desktop, bekommen ein Geräusch. Hier wird für normalsichtige Benutzer eine zusätzliche Feedback-Möglichkeit eingebaut. Dieses redundante Feedback hat in einem Ein-Benutzer-System kaum Vorteile, aber bei der Gestaltung von Teamarbeitsplätzen, wie etwa einem Whiteboard (Müller-Tomfelde et al. 2003), werden die Aktivitäten der jeweils anderen Teammitglieder leichter wahrnehmbar. Ähnliche Systeme wurden auf experimenteller Ebene zur Modellierung der Steuerung von Produktionsabläufen eingesetzt. Einzelne Prozesse und Maschinen bekommen einen bestimmten Klang. Je nach Zustand des Systems wird ein Wohlklang erzeugt, wenn etwa die Maschinen störungsfrei arbeiten.

Akustik wird auch als Erweiterung der grafischen virtuellen Realität eingesetzt, um den Realismus zu erhöhen. Grafische Darstellungen, bei denen nicht sofort erkennbar ist, ob ein Objekt, etwa bei einer Verschiebeoperation, ein anders Objekt berührt, können durch einen Sound aufgewertet werden (Begault 94, Seite 173). Aber auch die Erweiterung der Realität, hin zu einer Augmented Reality ist leicht vorstellbar. Vorhandene reale Objekte bekommen in einem virtuellen Raum, der mit dem realen synchronisiert wird, eine Entsprechung und können dort manipuliert werden.

Massiv eingesetzt wird Akustik jedoch im Bereich der Computerspiele. Sie wird genutzt um zusätzliche Informationen zum Spielverlauf einzubringen. In First-Person-Shootern (Quake<sup>1</sup>) werden Informationen über Vorgänge, die sich hinter dem Benutzer zutragen, übermittelt. Es wird hier jedoch nicht der Anspruch einer exakten Modellierung der Welt erhoben, sondern vielmehr steht der Unterhaltungswert im Vordergrund. Dies bedeutet aber häufig auch die Abkehr vom Realismus.

## **Entwicklung des auditiven Webbrowsers AIRclient**

Die Anforderungen an den AIRclient ergeben sich auch aus den Problemen, die mit bisherigen Screenreadern bestehen. Dazu muss eine Lösung

---

<sup>1</sup> <http://www.idsoftware.com/games/quake/quake/>

zur Überwindung der Layout-Barriere gefunden werden und eine zweidimensionale Komponente in den Screenreader eingefügt werden.

Zur Definition der Anforderungen an den AIRclient wurde eine Usability-Studie mit sieben blinden Experten durchgeführt. Die Studie wurde durch einen Prototypen des AIRclient unterstützt. Aus den Ergebnissen dieser Studie und der anschließenden Fokusgruppendifkussion wurde für die Weiterentwicklung des AIRclients folgende Anforderungen abgeleitet:

- Für den Blinden soll das Informationsangebot nicht neu erzeugt werden, sondern es sollen ihm die gleichen Informationen wie den normalsichtigen Benutzern angeboten werden.
- Blinde und Normalsichtige sollen sich über die Webseite unterhalten können.
- Alle Objekte einer Webseite werden in den akustischen Interaktionsraum abgebildet.
- Der AIRclient soll dem Benutzer die Grobstruktur des Layouts zugänglich machen.
- Der Fokus der akustischen Darstellung liegt auf der Darstellung der Struktur der Seite und nicht auf den Details des Webseitendesigns.
- Die semantischen Beziehungen zwischen Objekten auf einer Webseite werden dem Benutzer vermittelt.
- Zur Verarbeitung der Webseite werden nur Standardkomponenten verwendet.
- Hardware-Anforderungen sind zu minimieren, um die Kosten für den Blindenarbeitsplatz möglichst gering zu halten (keine teure Spezialhardware).

Im Folgenden werden die dafür notwendigen Komponenten und theoretischen Überlegungen erläutert und das System vorgestellt.

### **Architektur des AIRclient**

Der auditive Webbrowser AIRclient verwendet eine mehrstufige **Architektur**, um aus der grafischen Darstellung einer Webseite in einem Browser eine auditive Ausgabe zu erzeugen. Die Verarbeitungspipeline in Abbildung 1 verdeutlicht die notwendigen Schritte:

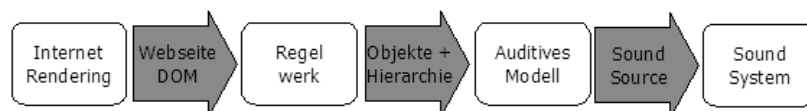


Abbildung 1: Auslesen, Verarbeiten und Ausgeben einer Webseite im auditiven Webbrowser

Um der Anforderung der Kommunikation zwischen Blinden und Normal-sichtigen gerecht zu werden, ist es notwendig, beiden die gleichen Informationen in Quantität und Qualität anzubieten. Zur Erreichung dieses Ziels muss eine Webseite vor der Verarbeitung im AIRclient in einem grafischen Webbrowser gerendert werden.

Die Rendering-Komponente stellt die gerenderte Darstellung aus dem Internet-Explorer zur Verfügung. Dazu wird der Internet Explorer als ActiveX-Modul in die AIRclient-Anwendung integriert und die Webseite über das Document Object Model (DOM) ausgelesen. Diese Struktur modelliert das Layout der Webseite und erleichtert die Generierung von Hierarchien der enthaltenen Objekte. Wesentliche Eigenschaften der Objekte sind deren Ausdehnung und Position auf einer Seite, sowie Informationen, in welchem übergeordneten Objekt sie sich befinden. Der Objektbaum wird in der Regelkomponente auf Prolog-Basis analysiert. Es wurde von uns ein Regelwerk definiert, um anhand der zur Verfügung stehenden Attribute zusammengehörige Objekte zu erkennen, z.B. Überschriften, Absätze, Verweislisten, Navigationsbereiche und Inhaltsbereiche, die nicht explizit in der Webseite durch einen HTML-Tagnamen definiert worden sind, sondern implizit durch das Layout vorliegen. Aus dem Objektbaum wird über eine Metapher die auditive Ausgabe erzeugt.

### Soundsystem

Für die auditive Ausgabe werden spezielle **Soundcards** (Aureal A3D<sup>2</sup>) eingesetzt, die die stereophonische Ausgabe unterstützen.

Zur Darbietung des akustischen Interaktionsraums sind „offene“ **Kopfhörer** verwendet worden, so dass der Benutzer - im Gegensatz zu einem „geschlossenen“ Kopfhörer - kaum von seiner realen akustischen Umgebung abgeschnitten wird. Damit bleibt der Kommunikationskanal aus dem synthetisch erzeugten auditiven Interaktionsraum heraus erhalten. Die Qualität der Kopfhörerausgabe und damit die Lokalisationsfähigkeit ist abhängig von den verwendeten Head-Related-Transfer-Functions (HRTF) den Faltungsfunktionen, die die Kopf- und Ohrenformen des Hörers nach-

<sup>2</sup> <http://www.vortexofsound.com>



bilden. Sie sind für jeden Menschen einzigartig, und die Verwendung von synthetischen HRTFs verschlechtert die Wahrnehmbarkeit.

Wir gingen davon aus, dass die Orientierung wesentlich vereinfacht wird, wenn die virtuellen Objekte eine feste Position im tatsächlich vorhandenen Raum aufweisen. Entsprechend müssten dann die Objekte bei einer Kopfdrehung nachpositioniert werden. Die Realisierung der **Kopferkennung** würde durch einen Headtracker oder eine videobasierte Erkennung erfolgen. Unsere Untersuchungen haben jedoch gezeigt, dass die Wahrnehmungsqualität der Ausgabe und die Lokalisationsfähigkeit kaum steigen und deshalb auf diese Unterstützung verzichtet werden kann. Unterschiede werden erst deutlich, wenn sich der Mensch sehr stark bewegt. Leichte Änderungen seiner Position haben keinen nennenswerten Einfluss auf die akustische Ausgabe.

Der erzeugte virtuelle auditive Interaktionsraum wird durch die ausgegebenen Hearcons aufgespannt. Da bei unserem System keine Sounds hinter dem Benutzer positioniert werden, sind bei der Quadrlautsprecherausgabe alle vier Lautsprecher vertikal vor dem Benutzer platziert, wodurch eine Transformation der Horizontalebene in die Vertikalebene notwendig wird. Damit wird der akustische Interaktionsraum (Surround Mode: um den Benutzer herum) in eine vertikale akustische Interaktionsebene (Front Mode: vor dem Benutzer) überführt. Die Darstellung der auditiven Benutzungsoberfläche bei einer Lautsprecherwiedergabe mit Hilfe von vier Lautsprechern hat den Nachteil, dass der Zuhörer einen streng definierten Hörort einhalten muss, um eine korrekte Platzierung der Hearcons durch Interpolation zu gewährleisten. Der Benutzer ist damit prinzipbedingt auf der Flächennormalen weitestgehend ortsgebunden. Dies verhindert insbesondere, dass mehrere Personen das System gleichzeitig nutzen können.

Es hat sich bei unseren Versuchen als vorteilhaft erwiesen, die Objekte im internen Modell auf die Innenseite einer **Halbkugel** - mit dem Kopf des Benutzers als Mittelpunkt - zu projizieren, um entfernungsbedingte Lautstärkeunterschiede zu eliminieren. Die Lautstärke sollte vielmehr nur von der Entfernung des darzustellenden Objekts von dem Cursor des Zeigegegerätes abhängen. Würden die Hearcons auf eine ebene Interaktionsfläche platziert, ergäben sich sowohl bei Kopfhörer- als auch bei Lautsprecherwiedergabe unerwünschte, ortsabhängige Lautstärkeunterschiede.

### **Akustisches Modell der Webseite**

Das akustische Modell ist die Kernkomponente des AIRclient und führt inhaltlich die visuellen Objekte der Webseite und die auditive Ausgabehardware zusammen, um die Webseite auf eine virtuelle auditive Interak-

tionswand vor dem Benutzer abzubilden. Zusätzlich werden die Objekte und Funktionen zur Steuerung des Webbrowsers eingebunden. Die Extraktion dieser Funktionalitäten ist recht einfach und wird über die Internet-Explorer-Komponente zur Verfügung gestellt.

Die Extraktion der Objekte einer Webseite ist mit größerem Aufwand verbunden. Zur besseren Darstellung und leichteren Verarbeitung werden diese in die Kategorien Meta-Objekte, Container-Objekte und Elementar-Objekte eingeteilt. Die Objekte einer Webseite haben, bezogen auf die auditive Ausgabe, drei wichtige Eigenschaften:

- Sie überlappen sich nicht in der grafischen Darstellung.
- Es wird immer ein rechteckiger Rahmen um die Objekte gelegt.
- Objekte können im Sinne einer Hierarchisierung andere Objekte enthalten.

Objekte, die sich oben links auf einer Webseite befinden, werden auch oben links auf der auditiven Interaktionsebene dargestellt.

### *Fackelmetapher*

Die Anzahl der Elemente übersteigt bei dieser einfachen Umsetzung schnell die Grenze der gleichzeitig wahrnehmbaren Objekte. Es muss also eine Reduktion der Datenmenge erfolgen. Der Mensch ist nur in der Lage, eine begrenzte Anzahl von parallel klingenden Objekten (die Angaben in der Literatur schwanken zwischen 5-8 Objekten) zu unterscheiden. Die „Fackelmetapher“ verändert die akustische Ausgabe einer Webseite, indem sie eine akustische Fackel realisiert („Eine Fackel im nächtlichen Wald lässt nur die nächsten Bäume sichtbar werden.“).



Abbildung 2: Ausschnitt einer Webseite (links) und Auswahl der auszugebenden Objekte durch die Fackelmetapher (rechts)

Durch die akustische Fackel, die mit dem Mauszeiger verknüpft ist (siehe Abbildung 2), wird ein Hearcon auf der auditiven Benutzungsoberfläche lauter, wenn der Benutzer sich ihm mit dem Cursor nähert. Befindet sich

der Mauszeiger auf dem Hearcon, hat es seine Maximallautstärke erreicht. Mit einem Cursor-Hearcon kann der Benutzer nun eines der festpositionierten Hearcons selektieren, analog zu den grafischen Benutzungsoberflächen mit einem Maus-Cursor. Wird der Rand eines Hearcons mit der Maus überfahren, erklingt als Feedback ein Eintritts- bzw. Austrittsgeräusch.

### *Sounddesign*

Sowohl künstliche Generierung von Geräuschen, als auch die Anwendung einer simplen Metaphorik (z.B. Hörbarmachung eines Texteingabefeldes auf einer Webseite durch Schreibmaschinengeräusche) erwies sich als vollkommen ungeeignet. In ersten Vorversuchen stellte sich heraus, dass die Geräusche für den Benutzer möglichst angenehm sein sollten. Die Lästigkeit bzw. Angenehmheit wurden von unseren Versuchspersonen als elementar wichtig für die Funktionsfähigkeit des Systems bezeichnet. Eine Folge der Entscheidung „Angenehmheit vor Metaphorik“ ist, dass die blinden Benutzer die Hearcons nicht metaphorisch erschließen können, sondern die Bedeutung erlernen müssen.

Als Vorgabe zur Erstellung einer Geräuschdatenbank wurde eine Bedarfsliste an relevanten Objekten aus dem akustischen Modell, sowie deren Eigenschaften, erstellt. Insgesamt wurden zwei Geräuschthemenparks (Natur- und Tiergeräusche sowie Musik und Musikinstrumente) für nachfolgende Experimente als Richtung vorgegeben. Die Komposition von Geräuschen mit Minimalmusik hat bei diesem Vorgehen die besten Ergebnisse erzielt und wurde weiter verwendet.

Unter Zuhilfenahme psychoakustischer Kriterien wurden schließlich die Geräusche ausgewählt (Schwankungstärke, spektrale Gestalt, Maskierung von gleichzeitig erklingenden Geräuschen etc.). Anschließend wurden die Signale geschnitten und loop-fähig gemacht. Es stellte sich heraus, dass für gute Resultate die Geräusche eine Dauer von  $\geq 5$  sec. aufweisen müssen, da sonst eine unangenehme, ungewollte und geräuschverfälschende Rhythmik entsteht. Es fand außerdem ein Frequenz- und Dezibelabgleich statt.

Wir haben Klangereignisse verwendet, die dreidimensional leicht zu orten sind und keine Effekte enthalten. Reine Bässe oder sehr hohe Klänge kamen daher nicht in Frage. Des Weiteren wurde besonders darauf geachtet, dass ein Minimum an Obertönen im jeweiligen Klang enthalten ist. Schmalbandige Klänge sind im Wesentlichen nur durch Lautstärkeunterschiede zu orten, was man mit einem Sinuston (z.B. von einem elektrischen Wecker oder Mobiltelefon) sehr leicht nachvollziehen kann.

In der psychoakustischen Forschung sind Signaleigenschaften bekannt, die die Gestaltwahrnehmung von Schallquellen/Hearcons beeinflussen. Im Folgenden werden einige Eigenschaften aufgezählt, die für die Darstellung von Hearcons zu beachten sind:

- Laute Hearcons werden eher in der Nähe des Zuhörers vermutet, dafür rücken leisere Schallquellen subjektiv weiter in den Hintergrund.
- Hearcons, bei denen die tieffrequenten gegenüber den hochfrequenten Signalanteilen dominierend sind, werden als räumlich ausgedehnter wahrgenommen. Im umgekehrten Fall schließt der Zuhörer eher auf ein räumlich kleineres Objekt (brüllender Löwe / zwitschernder Vogel).
- Hearcons mit einer ausgeprägten Zeitstruktur deuten daraufhin, dass es sich um „aktive“ Objekte (wie z. B. Maschinen oder Tätigkeiten/Lautäußerungen von Lebewesen) handelt.

#### *Verbesserung der Fackelmetapher*

Die Grundidee der Fackelmetapher schafft einen verständlichen Rahmen für die Interaktion mit Hearcons im auditiven Interaktionsraum. Für die gebrauchstaugliche Gestaltung der auditiven Benutzungsoberfläche muss diese Metapher jedoch durch Funktionen erweitert werden.

Eine **Abstandsfunktion** berechnet die Lautstärke der darzustellenden Objekte in der Fackelmetapher. Grundlage dieser Berechnung ist jeweils die Distanz zwischen einem Objekt auf der Webseite und dem Cursor des Zeigegerätes. Je weiter ein Objekt vom Cursor entfernt ist, desto leiser wird es ausgegeben. Problematisch erwies sich bei der gleichzeitigen Darbietung von mehreren Klangobjekten, dass diese unterschiedlich stark in den Vordergrund treten. Deshalb muss ein Abgleich der Signale auf gleiche berechnete Lautheitswerte durchgeführt werden. Ferner sollten die Klänge hinsichtlich ihrer temporalen Struktur generiert werden. Hierbei wurde auf einen möglichst gleichbleibenden zeitlicher Verlauf geachtet, denn schnell schwankende Signale treten akustisch eher in den Vordergrund. Demgegenüber können langsam schwankende Signale akustisch in den Hintergrund treten oder gar von anderen verdeckt werden.

Eine **Fisheye-View** zieht die in der Nähe des Cursors liegenden Hearcons auseinander. Demgegenüber werden Hearcons am Rand des Darstellungsbereichs zusammengezogen und erklingen als ein Geräusch. Die Übergänge beim Vergrößern bzw. Verkleinern eines Darstellungsbereichs erscheinen dem Benutzer fließend. Diese „Lupe“ nutzt den zur Verfügung stehenden akustischen Raum effektiver aus. Die Veränderung der **interaktiven Fläche** jedes Hearcons wird genutzt, da grafische Objekte tendenziell wesentlich kleiner gestaltet werden können, als akustische. Um ein

akustisches Objekt zu treffen sollte es jedoch möglichst groß sein. Aus den Untersuchungen der grundlegenden Wahrnehmungs- und Darstellungsmöglichkeiten einer auditiven Benutzungsoberfläche leiten sich die folgenden Anforderungen an die Metapher ab:

- Die kleinsten Layoutobjekte einer Webseite sollten mindestens auf die ermittelte minimale Interaktionsfläche eines Hearcons skaliert werden.
- Als Richtwerte für die optimale Größe eines Hearcons ergaben sich für Kopfhörer eine Ausdehnung in Azimutebene von 9 Grad und in der Elevationsebene von 16 Grad.
- Die Untersuchungen zur Bestimmung des optimalen und minimalen Abstandes zwischen zwei Hearcons führte zu keinen signifikanten Ergebnissen.
- Besonders gute Positionen zur Platzierung von Objekten im akustischen Interaktionsraum liegen in Augenhöhe vor dem Benutzer, bei einer Position von ca. 30 Grad zur linken und zur rechten Seite.

Die **Selektionsfunktion** bestimmt, welche Objekte endgültig ausgegeben werden. Abhängig von den Positionen der jeweiligen Objekte kann es passieren, dass nicht immer die Maximalanzahl der darstellbaren Objekte ausgegeben werden darf, denn die bekannten Forschungsergebnisse zur Lokalisation und Differenzierbarkeit beruhen auf Versuchen, bei denen die Objekte im Raum gleichmäßig verteilt sind. Befinden sich die Objekte jedoch alle im selben Quadranten (etwa unten links in der Ecke des Darstellungsraumes) funktioniert diese Unterscheidung nicht mehr. Ausgehend vom Mittelpunkt dürfen jeweils nur maximal drei Objekte gleichzeitig in einem Quadranten erklingen, da sonst die oben beschriebenen Parameter (Größe und Abstand) nicht mehr eingehalten werden können.

Bei der Anzahl der gleichzeitig darstellbaren Hearcons konnte ein interessanter Effekt beobachtet werden. Das Lokalisationsvermögen verbessert sich, wenn mehr als ein Hearcon gleichzeitig zu hören ist. Durch die Anwesenheit von zwei weiteren Hearcons in der Umgebung des ersten Hearcons lässt sich die Treffsicherheit auf der Azimutachse um 17% und auf der Elevationsachse um 37% erhöhen. Diesen Effekt erklären wir uns durch die Tatsache, dass der Zuhörer das zu suchende Hearcon in Abhängigkeit der zusätzlichen Hearcons suchen kann. Der Darstellungsraum wird durch das zweite Hearcon weiter unterteilt und dadurch der Suchraum verkleinert. Allerdings verschlechtert sich die Lokalisationsfähigkeit ab fünf Hearcons, da die Objekte nicht mehr voneinander differenziert werden können.

Mit der **Geräuschauswahlfunktion** werden den Objekten Geräusche zugeordnet. Die Soundzuordnung sieht drei ähnlich klingende Geräusche für

jeden Objekttyp vor. Die Zuweisung der Geräusche erfolgt in Abhängigkeit der bereits vorhandenen Objekte des gleichen Typs auf der Webseite, die schon akustisch ausgegeben werden. Das Vorgehen wurde gewählt, da dicht nebeneinander liegende Objekte gleichen Typs bei der akustischen Ausgabe nicht mehr zu unterscheiden sind. Sie würden wie ein großes Objekt wahrgenommen werden.

### *Sprachinteraktion*

Allein durch eine non-verbale Ausgabe können nicht alle Attribute eines Objektes wiedergegeben werden. Besonders textuelle Attribute eignen sich sehr gut zur Ausgabe über ein Text-to-Speech-System. Die Anforderung dieser Informationen erfolgt durch eine Spracheingabe. Sie ist gut geeignet für Blinde, sollte jedoch auf ein Minimum an Kommandos reduziert werden, da diese nur schwer erlernt und behalten werden können. Eine gute Spracheingabe, die auch freitextliche Eingaben ermöglicht, ist allerdings hilfreich.

Die wesentlichen Eigenschaften und Inhalte eines Objektes können abgefragt werden. So hat z.B. eine Überschrift grundsätzlich einen Inhalt, ein Verweis hat grundsätzlich ein Ziel und am Zielort liegt ein Dokument eines bestimmten Typs. Während der Sprachausgabe ist die Soundausgabe ausgeschaltet, da beide Ausgaben zusammen nur schwer gleichzeitig wahrnehmbar sind.

### **Regelwerk**

Bei der Wahl der Metapher sind wir von folgenden Forderungen geleitet worden.

- Blinde Benutzer sollen das Layout einer Webseite wahrnehmen und die wichtigsten Layoutstrukturen erkennen.
- Vorhandene semantische Beziehungen, die zwischen Elementen einer Webseite bestehen, sollten erkannt werden.

Diese Forderungen werden in von uns entwickelten Regeln aufgegriffen. Sie bereinigen die Datenbasis und erzeugen auf inhaltlicher Ebene eine Struktur der Objekte und somit eine Hierarchisierung der Objekte auf der Webseite.

Die Bewertung der Qualität der Regeln fällt schwer, da auf Webseiten Bereiche auch von Normalsichtigen nicht eindeutig benannt werden können. Diese Unterschiede wurden durch die visuelle Analyse von zehn Webseiten durch normalsichtige Benutzer ermittelt. Sie markierten dazu auf der Webseite die identifizierten Objekte und Bereiche. Anschließend wurden

unsere Regeln auf die gleichen Webseiten angewendet und mit den Angaben der Testpersonen verglichen.

### *Layoutstrukturen*

Zur Erkennung der vorhandenen Layoutstrukturen in Webseiten wird ein regelbasiertes System auf Prolog-Basis eingesetzt. Es gibt jedoch Einschränkungen durch unterschiedliche

- Modellvorstellungen von Gestalter und den Benutzern der Webseite,
- Inhalte und diverse Verwendungszwecke von Webseiten und
- Bearbeitungsstrategien, mit denen ein Benutzer seine unterschiedlichen Ziele auf einer Webseite erreichen möchte.

Designer nutzen häufig die Layoutmöglichkeiten aus den Styleguides, Regeln der Typografie und Gestaltungsgesetzen anders als die einzelnen Benutzer es später interpretieren. Zusätzlich gibt es häufig Verletzungen der Regeln durch den Designer aufgrund ästhetischer Darstellungswünsche. Ein konkreter Benutzer der Webseite nimmt unter Umständen ein anderes Modell wahr, als es der Designer intendiert hat. Wir haben deshalb beide Modelle bei der Erstellung der Regeln einfließen lassen.

Die Regeln berücksichtigen unterschiedliche **Kategorien** von Webseiten. Kriterium für die Unterscheidung der Kategorien ist das Verhältnis zwischen Inhalts- und Navigationselementen einer Webseite und dem kontinuierlichen Übergang zwischen den beiden Extremen.

Bei der Berücksichtigung der **Bearbeitungsstrategie** des Benutzers wird zwischen Inhaltssuche und dem Absuchen einer Seite nach weiteren Navigationsmöglichkeiten unterschieden. Entsprechend ändern sich auch seine Prioritäten bei der Suche nach bestimmten Elementen auf der Seite. Sieht er die Seite nur als Zwischenstation an, weil er vermutet, von hier aus auf weitere Webseiten zu stoßen, sind für den Benutzer Verweise wichtiger als Absätze.

### *Gruppierung*

Im AIRclient wird zwischen Navigations- und Inhaltsbereich unterschieden. Zunächst werden auf Basis von Ähnlichkeiten zwischen einzelnen Objekten größere Blöcke gebildet. Die Nähe der Objekte zueinander, aber auch die Struktur der Webseite werden dafür genutzt. Die Blöcke werden soweit vergrößert, bis nur noch zwischen 5-8 Gruppen übrigbleiben. Dieser Wert wurde gewählt, da Menschen nur in der Lage sind, bis zu dieser Anzahl noch gleichzeitig dargestellte Objekte zu unterscheiden. Es muss nun analysiert werden, welcher Inhalt in einem Bereich vorherrscht. Dazu wurde ein Index definiert, der das Verhältnis zwischen Anzahl an Verwei-

sen und der belegten Fläche (Pixel) berechnet. Ein Bereich in dem der Index sehr hoch ist, sowohl im Vergleich zum Index der Webseite, als auch untereinander, wird als Navigationsbereich markiert. Bereiche in denen der Index sehr klein ist, werden als Inhaltsbereiche identifiziert.

Innerhalb eines definierten Bereichs wird durch das Regelwerk überprüft, ob es noch weitere Zusammengehörigkeiten gibt. Die Hierarchietiefe darf jedoch nicht über zwei steigen, da sonst der Aufwand zum Navigieren innerhalb der Hierarchie den Vorteil der Struktur aufhebt.

Diese neu erstellte Struktur hat Auswirkungen auf das Sounddesign der Anwendung. Zu den bereits vorgestellten Sounds mussten weitere komponiert werden. Diese definierten zunächst Gruppengeräusche, die eine Unterscheidung zwischen Inhalt- und Navigationsbereich gemacht haben und zudem Geräusche, die Verweislisten etc. hervorhoben.

Unabhängig von den Objekten musste der Übergang zwischen den einzelnen Bereichen, bzw. der Übergang von einer Hierarchiestufe zur nächsten definiert werden. Untersucht werden musste hier die Zeit zum Ein- und Ausblenden und es musste ein Ein- und Ausblendgeräusch festgelegt werden. Der Vorgang des Ein- und Ausblendens liefert zusätzlich Hinweise auf die enthaltenen Objekte in den darunter bzw. darüber befindlichen Hierarchieebenen. Abbildung 3 zeigt die Umwandlung von einer grafischen Webseite in einen auditiven Interaktionsraum unter Verwendung von Layoutinformationen.

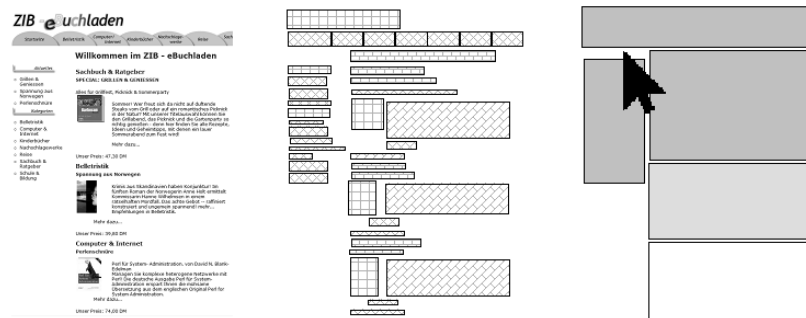


Abbildung 3: Generierung der auditiven Ausgabe mit Layoutinformationen. Von der ursprünglichen Webseite (links) werden Objektinformationen zu Überschrift, Absatz, Verweis und Grafik extrahiert (Mitte), zusammengehörige Objekte zusammengefasst und über die Fackelmetapher ausgegeben (rechts)



### **Eingabegeräte**

Die **Maus** ist für blinde Benutzer nur schwer zu nutzen. Die relative Positionierung des Cursors erfordert eine Hand-Auge-Koordination. Zur Sicherstellung, dass der Benutzer die Maus im akustischen Raum wiederfindet, muss der Cursor ein eigenes Geräusch bekommen. Es hat sich gezeigt, dass dieses Geräusch als störend empfunden wird, da es für die Benutzer eine Doppelcodierung darstellt. Die Nähe zu einem Objekt wird durch dessen Lautstärke dargestellt und es stehen durch ein zusätzliches Hearcon weniger Hearcons für den tatsächlichen Inhalt zur Verfügung.

Mit einem **Grafiktablett** wird diese Problematik umgangen, da es ein absolut positionierbares Eingabegerät ist, bei dem die Grenzen der Eingabefläche mit den Händen ertastet werden können. Die Grenzen des Grafiktablets müssen dafür mit den Grenzen des Darstellungsraumes identisch sein. Tatsächlich hatten die blinden Rechnerbenutzer eine wesentlich sicherere Navigation mit einem Grafiktablett als mit der Maus.

Als weiteres Eingabegerät, welches auf einer Ebene genutzt werden kann, wurde ein **Joystick** getestet. Der Joystick kann auf zwei verschiedene Arten eingesetzt werden. Im „Absolut-Modus“ wird der Ausschlag des Joysticks direkt auf die Ausgabefläche projiziert. Maximalpositionen sind in diesem Modus leicht anzuspringen, aber Positionen zwischen Ursprungspunkt und Maximalposition nicht, da die Positionierung mit dem Joystick sehr unpräzise ist. Im „Relativ-Modus“ wird nicht die Position des Joysticks, sondern die Bewegungsrichtung und Geschwindigkeit des Benutzers im auditiven Interaktionsraum ausgelesen. Die Versuchspersonen verlieren schon nach kurzer Zeit die Orientierung im Raum. Besonders schwer fiel die Wahrnehmung der Richtung. Wurde der Joystick nach rechts ausgelenkt, haben sich die Objekte links weiter entfernt. Dieses Verhalten wurde von den Versuchspersonen anscheinend nicht erwartet.

### **Evaluation auditiver Benutzungsoberflächen**

Zur Unterstützung der Evaluation wurden zahlreiche Erweiterungen in die eigentliche Anwendung integriert. Es wird bei den Methoden darauf geachtet, dass die folgenden Fragestellungen aus den gewonnenen Daten beantwortet werden können:

- Welche Objekte werden von den Benutzern betrachtet?
- Zu welchen Objekten lassen die Benutzer über die Sprachausgabe weitere Informationen ausgeben?
- Welche Strategien werden zur Navigation auf einer Webseite genutzt?
- Welche Kommentare formulieren die Benutzer während der Aufgabebearbeitung?

- Welche Struktur und Layout der Webseite nehmen die Benutzer wahr?
- Wird ein semantischer Zusammenhang zwischen Objekten erkannt?

Um die Versuchsperson während der Evaluation ansprechen zu können, sollte die Lautstärke der Hearconausgabe extern geregelt werden können. Der Evaluationsleiter hört mit und kann bei Problemen gegebenenfalls eingreifen.

Das **Laute Denken** ist im Anwendungsfeld der auditiven Benutzungsoberflächen mit Vorsicht einzusetzen. Das Zuhören schränkt das „Laute Denken“ stark ein. Ein weiteres Problem ist die Verwendung der Sprach-eingabe. Das System muss entscheiden, ob es sich bei dem Gesprochenen um ein Kommando oder um lautes Denken handelt, wodurch es zu vielen Fehlansagen kommen kann. Eine Lösung ist die Simulation der Sprach-eingabe durch den Versuchsleiter: Der Benutzer spricht weiterhin in ein Mikrofon, aber der Versuchsleiter löst an seinem Rechner über Funktionstasten die jeweilige Systemantwort aus.

Die konkrete Darstellung der Webseite durch den grafischen Web-Browser wird vom AIRclient ausgewertet und in eine vereinfachte **schematische Darstellung** überführt. Die tatsächlichen Inhalte treten in den Hintergrund und die Struktur der Seite in den Vordergrund. Jedes Objekt wird durch ein Rechteck mit unterschiedlicher Farbe dargestellt. Um das mentale Modell dokumentieren zu können, haben wir die Testpersonen das wahrgenommene Layout mit einem **Layoutbaukasten** nachbauen lassen, das sich die Versuchsteilnehmer von einer besuchten Webseite gemacht haben. Auf einem Whiteboard werden für die unterschiedlichen Typen von Objekten Magnete mit entsprechenden Markierungen positioniert. Im Anschluss konnte das Ergebnis des Layoutbaukastens mit der schematischen Darstellung verglichen werden. Dieser Vergleich ermöglicht die Überprüfung, ob die wichtigsten Elemente einer Seite von dem Benutzer erkannt worden sind.

Um Usability-Tests mit dem AIRclient durchführen zu können, wurde eine **Interaktionsprotokollierung** konzipiert, entworfen und implementiert. Nach der Evaluation sind die Protokollierungen geeignet, um die Mausspur nachzubilden und sich so Bereiche anzuschauen, die besonders intensiv betrachtet worden sind. Auf diese Weise werden z.B. unnötige Wege mit dem Mauszeiger identifiziert. In der Visualisierung wird deutlich, welche Objekte zu welchem Zeitpunkt zu hören sind. Neben der Mausspur werden auch die Interaktionspfade des Benutzers auf einer Webseite dokumentiert.

Für die Usability-Evaluation muss in einer grafischen Ausgabe der Inhalt visualisiert werden, den die blinden Benutzer des Systems gerade hören.

Die grafische Ausgabe des Internet-Explorers kann leicht aufgezeichnet werden. Dagegen ist die akustische Ausgabe des AIRclients (aufgrund der 3D-Ausgabe) mit herkömmlichen Mitteln nur unzureichend mit einer **Videoaufzeichnung** zu speichern. Deshalb wurde neben der schematischen Ausgabe der Struktur einer kompletten Webseite eine Darstellung implementiert, die die dynamisch berechnete akustische Ausgabe des aktuellen Interaktionszustands visualisiert. Die akustisch dargestellten Objekte werden entsprechend ihrer Lautstärke farblich markiert. Je leiser das Geräusch/Objekt, desto heller die Farbe.

Die subjektive Meinung der Versuchsteilnehmer ist bei der Bewertung der Angenehmheit der verwendeten Sounds wichtig. Deshalb wurden auch die mündlichen Äußerungen protokolliert.

## **Innovationen und Ergebnisse des Projekts**

Mit dem AIRclient wurde ein Software-System geschaffen, das mit preiswerten Hardware-Komponenten auf normalen PCs installiert werden kann, um blinden Benutzern die Nutzung des Internet für die Aus- und Weiterbildung, den Beruf und die Freizeit erheblich zu erleichtern.

Dazu wurden folgende neuartige Methoden entwickelt:

- Methodisches Vorgehen zur Entwicklung von auditiven Benutzungsoberflächen in einem interdisziplinären Entwicklungsteam bestehend aus Informatikern, Psychologen, Physikern und Musikern.
- Das Konzept „Fackelmetapher“ zur Navigation in auditiven Informationsräumen, welches auch auf andere Anwendungen übertragbar ist, sowie Entwicklung von Funktionen zur Verbesserung der Metapher.
- Grundlagenuntersuchungen zum Aufbau von auditiven Informationsräumen, zur Gestaltung von Hearcons und zur Vorgehensweise zum Sounddesign.
- Erweiterung und Entwicklung von Evaluationsmethoden zur Durchführung von Usability-Studien mit blinden Rechnerbenutzern.

Außerdem wurden folgende Komponenten und Werkzeuge entwickelt:

- Entwicklung einer virtuellen Interaktionsfläche in Form einer Halbkugel zum Ausgleich von Lautstärkeunterschieden an unterschiedlichen Positionen.
- Ein Regelwerk zur automatischen Erkennung von Layoutstrukturen und semantischen Beziehungen in Webseiten zur Aufbereitung von Webseiten.

- Toolkit von auditiven Interaktionsobjekten zur Nutzung in mobilen Situationen mit minimalen Ausgabemöglichkeiten.
- Ein Werkzeug zur Erstellung von Prototypen auditiver Benutzungsoberflächen unter Verwendung von vorgefertigten Interaktionsobjekten (Klante 2003).
- Implementierung eines Sound-Frameworks auf Java-Basis zur Verwendung in beliebigen Anwendungen mit unterschiedlichen 3D-Soundcards.
- Anbindung eines Lautsprecherarrays über zwei Achtkanal-Soundcards, zur Ansteuerung von 16 Schallerzeugern.
- Umwandlung eines Quad-Lautsprecheraufbaus von einer Hörebene um den Benutzer zu einer Hörebene vor dem Benutzer.

## **Bewertung und Ausblick**

Der vorgestellte Lösungsansatz hat sich im Laufe des Projekts bewährt. Die Versuche zum Sounddesign, zur minimalen Interaktionsfläche, zum minimalen Abstand zwischen Hearcons und zur Lokalisationsfähigkeit bilden die Grundlagen für die Interaktion mit auditiven Interaktionsräumen. Das von uns entwickelte Konzept der Fackelmetapher wurde im auditiven Webbrowser AIRclient implementiert. Mit diesem Werkzeug gelingt es der Zielgruppe der Blinden sowohl eine Übersicht zu erhalten, als auch eine eigenständige Exploration von Webseiten durchzuführen. Innerhalb der vorgestellten Metapher ist es möglich für detaillierte Informationen Anfragen an die Objekte zu stellen.

Unter Berücksichtigung der Objekte für die auditive Interaktion wegen der schlechten akustischen Auflösung des menschlichen Gehörs aufbereitet werden müssen, wurden Funktionen für die Unterstützung der Fackelmetapher entwickelt. Die interaktive Fläche eines Hearcons wird auf einen Mindestwert erhöht und ermöglicht die Interaktion mit sehr kleinen Objekten. Eine Fish-Eye-View erlaubt die Differenzierung zwischen sehr dicht aneinander positionierten Objekten. Diese Entwicklungen stoßen auf große Akzeptanz bei den potentiellen Benutzern.

Die zum Teil fehlerhafte oder nicht dem HTML-Konzept entsprechende Programmierung von Webseiten erschwert das Auslesen der auf einer Webseite vorhandenen Objekte. Das entwickelte Regelwerk beseitigt diese Mängel und erkennt Strukturen in Webseiten, die nicht explizit im Quellcode vorhanden sind. Es ermöglicht dem Benutzer sich über eine Hierarchisierung und Gruppierung der auf einer Webseite enthaltenen Ob-

jekte einen schnellen Überblick zu verschaffen und Layoutstrukturen zu erkennen. Die Zusammenfassung der Objekte zu Navigations- und Inhaltsbereichen befähigen den Benutzer zu einer zielgerichteten Aufgabebearbeitung. Diese inhaltliche Bewertung der Objekte ist notwendig geworden, da die Auflösung, also die Anzahl gleichzeitig darstellbarer Objekte in einen auditiven Interaktionsraum deutlich geringer ist als in grafischen Benutzungsoberflächen. Die Informationen werden aufbereitet und es findet eine Abstraktion statt.

Blinde Versuchspersonen mit Interneterfahrung verfügen bereits über Methoden für den Zugang. Für sie ist das Erlernen neuer Software ein sehr zeitaufwendiger Prozess und beansprucht einen Zeitraum von mehreren Wochen. In dieser Zeit werden die wichtigsten Funktionen trainiert und auswendiggelernt. Das Erlernen einer zusätzlichen Software ist mit einem nicht unerheblichen Aufwand und deshalb oft mit Widerständen verbunden. In den Laborsituationen haben die Versuchsteilnehmer häufig sehr euphorisch an diesen innovativen Experimenten teilgenommen. Die Grundeinstellung war bei den anschließenden Interviews und Gesprächen entsprechend positiv. Einem Wechsel der bisher genutzten Software wurde jedoch wegen des hohen Aufwandes mit Zurückhaltung begegnet. Eine Langfriststudie muss nun klären, ob das System auch für den täglichen Einsatz geeignet ist und diese Hürde überwunden werden kann. Dazu wird der auditive Webbrowser im Internet<sup>3</sup> zum Download zur Verfügung gestellt. Durch die Java-Implementierung und Anbindung an gängige Soundcards mit Creatives EAX<sup>4</sup> Standard kann eine größere Gruppe, Nutzern angesprochen werden.

Die guten Ergebnisse bestätigen, dass sich auditive Benutzungsoberflächen eignen, um die Layoutstruktur einer Webseite darzustellen. Mit dem AIRclient steht der Benutzergruppe der Blinden ein Hilfsmittel zur Verfügung, das den Mangel vorhandener Screenreader bezüglich des Nichtdargestellens des Layouts einer Webseite, beseitigt. Insbesondere spricht die auditive Benutzungsoberfläche eine Modalität an mit deren Verwendung Blinde vertraut sind.

Das Projekt hat gezeigt, dass auditive Benutzungsoberflächen für eine Interaktion mit Objekten auf Webseiten geeignet sind. Vorhandene Erfahrungen sollten jetzt auf die Bedienung einer Benutzungsoberfläche eines Betriebssystems wie Windows angewendet werden. Die Tatsache, dass Blinde mit dem AIRclient in der Lage sind, Webseiten vollständig zu er-

---

<sup>3</sup> <http://www.offis.de/projekte/mi/zib>

<sup>4</sup> <http://eax.creative.com>

fassen, macht auditive Benutzungsoberflächen auch für Normalsichtige interessant.

In Zukunft ist der Einsatz in vielen anderen Bereichen denkbar. Z.B. wird in parallelen Projekten versucht, auditive Benutzungsoberflächen auch in mobilen Situationen einzusetzen. Eine Standardisierung der verwendeten Interaktionsobjekte ist dazu dringend erforderlich, um die hohen Lernzeiten zu minimieren und die Benutzung und Interaktion mit dem System zu vereinheitlichen. Der Benutzer kann seine Sehkraft als primären Sinn für die Koordination und Kontrolle der realen Bewegungssituation nutzen. Die Akustik kann für die Steuerung eines Rechners genutzt werden. Befindet sich der Benutzer in Umgebungen in denen er aufgrund seiner Arbeitssituation oder aufgrund von Umwelteinflüssen nicht auf die Benutzung seines primären Sinn zurückgreifen kann, ist die Bedienung eines Rechners mit einer auditiven Benutzungsoberfläche eine sinnvolle Alternative.

### *Danksagung*

Dank gilt dem früheren Projektmitarbeiter Hilko Donker und allen studentischen Hilfskräften und Studenten, die durch Diplom- und Studienarbeiten zum Gelingen beigetragen haben.

Darüber hinaus wurde in Zusammenarbeit mit dem Institut für technische und angewandte Physik der Carl von Ossietzky Universität Oldenburg (I-TAP) eine Technik entwickelt, die die Präsentation des akustischen Interaktionsraums mit Hilfe einer Mehrfach-Lautsprecher-Ausgabe statt mit Kopfhörern erlaubt. Das Institut zur Erforschung von Mensch-Umwelt-Beziehungen der Carl von Ossietzky Universität Oldenburg (MUB) führte vergleichende Untersuchungen zu der Eignung von Geräuschen für blinde Benutzer an Prototypen des auditiven Webbrowsers durch.

## **Referenzen**

Begault, R., *3-D Sound for Virtual Reality and Multimedia*, Academic Press Professional, Cambridge, MA. USA, 1994

Bölke, L., *Ein akustischer Interaktionsraum für blinde Rechnerbenutzer*. Dissertation. Berichte aus dem Fachbereich Informatik. 1997

Buxton, W. (ed.), *Journal of Human-Computer Interaction, Special Issue on Non-Speech Audio*, V4; N1. Lawrence Erlbaum Publishers 1989

Klante, P. *Visually Supported Design of auditory User Interfaces*. In Stephanidis, C., Jacko, J. (Ed.): *Human-Computer Interaction: Theory and Practice (Part II)*, Lawrence Erlbaum Publishers (2003) S. 696-700

Kramer, G. (ed.). *Auditory Display – Sonification, Audification and Auditory Interfaces*. A Proceedings volume of the first ICAD conference. Santa Fe Institute, 1994

Müller-Tomfelde, C., Streitz, N.A., Steinmetz, R. *Sounds@Work - Auditory Displays for Interaction in Cooperative and Hybrid Environments*. In Stephanidis, C., Jacko, J. (Ed.): *Human-Computer Interaction: Theory and Practice (Part II)*, Lawrence Erlbaum Publishers (2003) S. 751-755,

### Die Autoren

Dipl.-Inform. Palle Klante



Palle Klante arbeitet seit vier Jahren beim Oldenburger Forschungs- und Entwicklungsinstitut für Informatik-Werkzeuge und -Systeme (OFFIS) im Bereich Multimedia und Internetinformationsdienste. Seit 2000 arbeitet er an seiner Dissertation zur „Gestaltung auditiver Benutzungsoberflächen“ an der Universität Oldenburg. Dort studierte er Informatik und machte 1999 sein Diplom in der Abteilung Computer Graphics und Software-Ergonomie. Sein Arbeitsschwerpunkt ist das Usability Engineering von auditiven Benutzungsoberflächen und die Entwicklung von Methoden, Tools und Interaktionsobjekten.

Prof. Dr-Ing. Peter Gorny



Peter Gorny hat das Diplom im konstruktiven Ingenieurbau (TH Hannover) erworben und in der Angewandten Informatik promoviert (Ruhr-Universität Bochum). Er ist Professor für Angewandte Informatik an der CvO Universität Oldenburg und leitet seit 1986 die Abteilung Computer Graphics und Software-Ergonomie im Fachbereich Informatik. Seine Hauptarbeitsgebiete sind Anwendungen der Virtual-Reality-Technologien, Entwicklungsmethoden für Benutzungsoberflächen und E-Learning.

Dipl.-Inform. Michael Gründler



Michael Gründler studierte an der CvO Universität Oldenburg Informatik. In seiner Diplomarbeit in der Abteilung Computer Graphics und Software-Ergonomie beschäftigte er sich mit Konzepten zur Sprachinteraktion mit Webseiten. Nach dem Diplom ist er seit September 2002 beim Oldenburger Forschungs- und Entwicklungsinstitut für Informatik-Werkzeuge und -Systeme (OFFIS) angestellt und beschäftigt sich im INVITE-Teilprojekt „Zugang zum Internet für Blinde“ mit der Entwicklung eines auditiven Webbrowsers.