



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Masterarbeit

Jan Martin

Spatial Auditory Displays



Wahrnehmungstheoretische Eignungsuntersuchung
akustischer Benutzeroberflächen

Jan Martin

Spatial Auditory Displays

Wahrnehmungstheoretische Eignungsuntersuchung
akustischer Benutzeroberflächen

Masterarbeit

im Studiengang Next Media
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Erstprüfer: Prof. Dr. Kai von Luck
Zweitprüfer: Prof. Dr. Wolfgang Fohl

Abgegeben am 21.05.2015

Jan Martin

Thema der Arbeit

Spatial Auditory Displays – Wahrnehmungstheoretische Eignungsuntersuchung akustischer Benutzeroberflächen

Stichworte

Spatial Auditory Displays, Audio-Augmented Reality, Sonification, Klangumgebung, akustische Holografie, Wahrnehmung

Kurzzusammenfassung

Der menschliche Wahrnehmungsapparat verfügt grundsätzlich über fünf Sinneskanäle durch die Informationen aus der Umwelt gewonnen werden können. Unsere heutige digitale Welt bedient, bis auf wenige Ausnahmen, aber nur einen dieser potentiellen Kommunikationswege – die Vision. Gleichzeitig böte die Einbeziehung weiterer Sinne die Möglichkeit, die visuelle Wahrnehmung zu entlasten und entsprechend holistischer Wahrnehmungsprinzipien angemessen zu komplementieren. Durch ihre wahrnehmungstheoretische Eignung erscheinen akustische Repräsentationsformen prädestiniert diese Aufgabe zu erfüllen. Akustische Signale sind einerseits in der Lage hoch komplexe Information mit großer Dichte zu transportieren, andererseits ist der akustische Wahrnehmungsapparat mittels einer Kaskade kognitiver Mechanismen fähig, diese Informationen adäquat zu decodieren und zu interpretieren. Als Disziplin der akustischen Datenpräsentation wird in der vorliegenden Arbeit die Domäne der auditorischen Displays identifiziert. Durch diese Displaytechnologie ist es möglich, Nutzern handlungsweisende Informationen verständlich zu präsentieren, eine spielerische Interaktion zu provozieren und kontrolliert Emotionen zu induzieren. Der Einsatz auditorischer Displays und Benutzeroberflächen erscheint deshalb aus wahrnehmungstheoretischer Perspektive *sinnvoll*.

Zusätzlich kann die Sicherheit im Umgang mit akustischen Informationen durch eine Spatialisierung der akustischen Signale weiter unterstützt werden. Aus wahrnehmungstheoretischer Sicht leitet sich dieser Zusammenhang aus der engen Verknüpfung zwischen Audition und räumlicher Wahrnehmung ab. Technologien und Konzepte zur Implementierung spatial-auditorischer Anwendungen existieren bereits heute in Form sogenannten spatialer, auditorischer Displays. Im Rahmen der vorliegenden Arbeit wird die wahrnehmungstheoretische Eignung dieser Systeme auf dem Hintergrund klassischer und auditionsspezifischer Wahrnehmungsmodelle untersucht.

Jan Martin

Title of the paper

Spatial Auditory Displays – Wahrnehmungstheoretische Eignungsuntersuchung akustischer Benutzeroberflächen

Keywords

Spatial Auditory Displays, Audio-Augmented Reality, Sonification, Soundscapes, acoustical Holography, Perception

Abstract

The human perceptory system does essentially comprise five senses through which information from the environment can be gleaned. However, in the contemporary digital age, save for a few exceptions, merely one of these potential cognitive pathways – vision – is attended to. Concurrently, inclusion of further senses would provide an opportunity to relieve and to adequately complement visual perception in accordance with holistic perception and cognition principles. Due to their inherent suitability in terms of perception theory, acoustic forms of representation seem predestined to fulfil this task. On the one hand, acoustic signals are able to convey highly complex information at high density, on the other hand, the acoustic perceptory system, via a deluge of cognitive mechanisms, is able to decode this information adequately and to interpret it. Recognition of the domain of auditory displays as a discipline of acoustic data presentation is explored and substantiated in this thesis. Such display technology allows the possibility of presenting instructional information comprehensibly to its users, to provoke playful interaction and to induce emotion in a controlled manner. Accordingly, utilisation of auditory displays and user interfaces does, from the perspective of perception theory, seem sensible. Additionally, acoustic information handling skills can further be supported by spatialisation of acoustic signals. Under aspects of the theory of perception, this correlation is derived from the close interrelation between audition and spatial perception.

Technologies and concepts aimed at implementation of spatial auditory applications do exist already in the shape of spatial auditory displays. Within the scope of this thesis, suitability of these systems in terms of perception theory is researched on the background of classical and audition-specific models of perception.

Inhaltsverzeichnis

1	Einleitung	7
1.1	Primat der visuellen Sphäre	7
1.2	Ansatz und Argumentationsstruktur	8
1.3	Aufbau	9
2	Vorangegangene Projekte.....	11
2.1	Automated Tour Guide	11
2.2	Hear & There	12
2.3	Audio Aura.....	13
2.4	Resumée.....	14
3	Auditorische Information.....	16
3.1	Grundlagen der Akustik und Pseudoakustik	17
3.2	Hören – Codierung und Terminologie.....	21
3.2.1	Codierung des Hörens.....	21
3.2.2	Terminologie des Hörens	22
3.3	Auditorische Ordnungsmechanismen.....	23
3.4	Kognitive Modelle	25
3.4.1	Modularität (nach Fodor).....	26
3.4.2	Adaptive Resonance Theory (nach Grossberg)	26
3.4.3	Congruence-Associationist Framework (CAF).....	29
3.5	Informationeller Gehalt akustischer Ereignisse.....	31
3.6	Auditorische Displays und Benutzeroberflächen.....	35
3.6.1	Ausgestaltung und Terminologie	36
3.6.2	Gestaltungskriterien auditorischer Displays (Subtasks).....	39
3.6.3	Psychoakustische Dimensionen auditorischer Displays.....	39
3.6.4	Anwendung und Limitationen.....	42
3.7	Lerneffekte	45
3.8	Auditorische Emotionsinduktion.....	46
3.9	Kontext	51
3.9.1	Begrifflichkeit	52
3.9.2	Soundscapes (nach Truax) und akustische Ambiguität.....	55
3.9.3	Mehrdimensionalität des spatialen Kontexts	56
3.10	Konklusion.....	57

4	Audition und Spatialität.....	60
4.1	Schizophonie (nach Truax)	60
4.2	Kognitive Relation von Klang und Raum	61
4.2.1	Spatiale Kognition	61
4.2.2	Spatialität als Vehikel der Kognition	63
4.2.3	Aural Fluency Theorie (nach Droumewa & Wakkary).....	64
4.3	Bedeutung von Spatialität für Hören und Zuhören	66
4.3.1	Hören – Spatialität als Wiedergabeparameter	66
4.3.2	Zuhören – Spatialität als narrativer Kontext	71
4.4	Konklusion	73
5	Spatial-auditorische Systeme	74
5.1	Defintion und Eignungsuntersuchung von Spatial Auditory Displays.....	74
5.2	Audio-Augmented Reality	76
5.3	Praktische Ausgestaltung	80
5.3.1	Landscape Music	81
5.3.2	Auditory Attitude Indicator	82
5.3.3	Navigation durch Acoustic Landmarks	83
5.3.4	VR-Computerspiele	85
5.3.5	Hörbücher als akustisches Hologramm.....	85
5.3.6	SADs für sehbehinderte Nutzer.....	86
5.4	Ausblick	86
6	Konklusion	88
	Literaturverzeichnis	90
	Abbildungsverzeichnis	95

1 Einleitung

1.1 Primat der visuellen Sphäre

Wahrnehmung ist ein Prozess, der fast immer automatisch und mühelos funktioniert. Alle sinnlichen Einflüsse der Umgebungen werden scheinbar selbstverständlich in eine komplexe Kaskade neuraler Impulse umgewandelt, deren Interpretation die Grundlage der subjektiven Erfahrung unserer Umwelt bildet. Diese *Transformation physikalischer Energie in Bedeutung* findet innerhalb eines Bruchteils einer Sekunde statt. Die Leichtigkeit und Geschwindigkeit mit der der menschliche Wahrnehmungsapparat diese Herkulesaufgabe bewältigt, sollte aber keinesfalls über ihre Komplexität hinwegtäuschen. Das Studium der fundamentalen Mechanismen und Prozesse der menschlichen Wahrnehmung, i.e. die Wahrnehmungstheorie, ist deshalb für jede Form der informationellen Interaktion zwischen einem Nutzer und einem System von zentraler Bedeutung; die Domäne der spatial-auditorischen Datenpräsentation bildet hierzu keine Ausnahme (Neuhoff 2011, S.63).

Die Ausgestaltung und Struktur des menschlichen Wahrnehmungsapparates hat sich evolutionär den wahrnehmungspraktischen Gegebenheiten unserer Umwelt angepasst. Während sich natürliche Wahrnehmungskontexte in der Regel aber durch eine ausbalancierte Sinnesansprache auszeichnen, hat sich die Informationstechnologie der vergangenen Jahrzehnte mehr und mehr auf die visuelle Präsentation von Informationen konzentriert (Truax 2000 et al.). Ein holistischer Ansatz, wie er dem menschlichen Wahrnehmungsmodell entspräche, wird in der Regel nicht verfolgt.

„The design of effective auditory displays that support the human operators of complex systems has received substantially less attention in the human factors literature than the design of visual displays. [...] The presentation of data to other sensors (auditory, vestibular, haptic and olfactory) has not been addressed. The relative neglect of other sensory modalities in EID [Ecological Interface Design] is puzzling in view of the emphasis in EID on representing the world in a way that matches human perceptual processes.“ (Sanderson et al. 2000, S.259)

Gleichzeitig durchsetzen artifizielle Signale gerade in urbanen Umgebungen heute einen Großteil aller Wahrnehmungskontexte. Eine wahrnehmungstheoretische Untersuchung unserer digitalisierten Umwelt scheint deshalb durchaus notwendig. Die vorliegende Untersuchung widmet sich dem akustischen Spektrum dieser Umwelt.

Die Effizienz und Effektivität der visuellen Datenpräsentation steht dabei offenkundig außer Frage. Gerade in Anwendungsszenarien, in denen eine interpretationsrobuste und konkrete Information unabdingbar ist, werden visuelle Systeme aber auch in Zukunft unverzichtbar blei-

ben. Die Ergebnisse der folgenden, wahrnehmungstheoretischen Eignungsuntersuchung scheinen dennoch ein deutliches Ungleichgewicht zwischen dem Potential auditorischer Displays bzw. Benutzeroberflächen und ihrer praktischen Würdigung zu signalisieren (Shinn-Cunningham et al. 2005, Shinn-Cunningham 2002, Walker & Kramer 2004, Sanderson et al. 2000 et al.). Die eingangs erwähnte, unausgewogene Fokussierung auf visuelle Kommunikationsmittel folgt nicht aus einer wahrnehmungstheoretischen Notwendigkeit: Auditorische Displays und Benutzeroberflächen im Allgemeinen und spatial-auditorische Anwendungen im Speziellen sind durchaus in der Lage, komplexe Informationen mit hoher Dichte in verständlicher Form an einen Adressaten zu übermitteln. Gleichzeitig erscheint die einseitige Beanspruchung des visuellen Wahrnehmungskanals schon per se nicht ratsam. Einerseits existieren Anwendungsfälle in denen die akustische Datenrepräsentation nachweislich bessere Ergebnisse erzielt, andererseits könnte eine Umverteilung helfen, den visuellen Wahrnehmungsapparat zu entlasten (Sanderson et al. 2000). Die vorliegende Arbeit untersucht deshalb die Eignung von *Spatial Auditory Displays* als Komplement in der digitalen Datenrepräsentation auf dem Hintergrund des menschlichen Wahrnehmungsmodells.

1.2 Ansatz und Argumentationsstruktur

Grundsätzlich lässt sich die Argumentationsstruktur dieser Arbeit vergleichsweise einfach zusammenfassen. Ausgangspunkt ist die bereits beschriebene Dominanz visueller Benutzeroberflächen und Displays, die nicht auf eine wahrnehmungstheoretische Notwendigkeit zurückgeführt werden kann. Gleichzeitig bieten sich auditorische Anwendungen im Sinne der zentralen Hypothese H1 als potentielle Komplementärtechnologie an.

H1 Der Einsatz kontextabhängiger, auditorischer Displays und Benutzeroberflächen ist aus wahrnehmungstheoretischer Sicht *sinnvoll*.

Die Kontextsensitivität akustischer Signale stellt nach wie vor eine der größten Herausforderungen für den erfolgreichen Einsatz auditorischer Benutzeroberflächen dar. Im Hinblick auf eine Vielzahl wahrnehmungstheoretischer Kriterien lässt sich Spatialität als bedeutendste Kontextgröße akustischer Information identifizieren. Gerade spatialisierte Audio-Anwendungen zeichnen sich entsprechend der Hypothese H1.1 deshalb im Hinblick auf die Beschaffenheit des menschlichen Wahrnehmungsapparates durch eine besondere Eignung aus.

H1.1 Die Spatialisierung auditorischer Displays und Benutzeroberflächen ist aus wahrnehmungstheoretischer Sicht *sinnvoll*.

Auf Basis der beiden Hypothesen dieser Arbeit scheint der Verweis auf die Technologie der *Audio-Augmented Reality (AAR)* geradezu zwingend. Die inhärente, spatiale Dimension von AAR-Systemen prädestiniert diese als Kandidat für einen Einsatz spatial-auditorischer Benutzeroberflächen nach dem Vorbild der wahrnehmungstheoretischen Argumentation der vorlie-

genden Arbeit. Eine abschließende Betrachtung dieser Technologie erfolgt deshalb im letzten Kapitel.

Eine Spezialfallbetrachtung soll zum Zweck einer möglichst allgemeinen wahrnehmungstheoretischen Herleitung bewusst ausbleiben. Bezug auf sprachgesteuerte Systeme oder den Einsatz spatial-auditorischer Anwendungen für sehbehinderte Nutzer wird deshalb nur am Rande genommen.

1.3 Aufbau

Kapitel 2 dient als praktischer Einstieg in die folgende wahrnehmungstheoretische Untersuchung auditorischer Benutzeroberflächen. Die vorgestellten Projekte illustrieren zum einen die Vielfalt spatial-auditorischer Anwendungen und werden zum anderen im Laufe der Arbeit immer wieder als sinnstiftende Beispiele herangezogen.

In Kapitel 3 wird zunächst die allgemeine Eignung auditorischer Displays und Benutzeroberflächen nach wahrnehmungstheoretischen Kriterien untersucht. Zu diesem Zweck werden zu Beginn die notwendigen Grundlagen der Akustik zusammengefasst und in die Domäne der Psychoakustik überführt. Anschließend befasst sich die Ausarbeitung mit der sinnlichen Wahrnehmung akustischer Signale und der mechanoelektrischen Signaltransduktion. Die resultierenden Wahrnehmungsreize lösen eine Kaskade kognitiver Prozesse aus, die auf dem Hintergrund klassischer Kognitionsmodelle der menschlichen Wahrnehmung diskutiert werden. Aus dieser Betrachtung lässt sich die Qualifikation des Adressaten, i.e. des auditorischen Wahrnehmungsapparates, ableiten. Die informationelle Güte akustischer Signale per se wird im folgenden Kapitel betrachtet. Als korrespondierende Technologie im Licht der angestellten Überlegungen wird in Kapitel 3.6 das Forschungsfeld der auditorischen Displays und Benutzeroberflächen vorgestellt. Zusätzlich findet eine Untersuchung potentieller Lerneffekte im Umgang mit derartigen Systemen statt. Zum Zweck einer abschließenden Eignungsbeurteilung im Sinne der zentralen Hypothese dieser Arbeit (H1) wird in Kapitel 3.8 darüber hinaus die Theorie der musikalischen Emotionsinduktion herangezogen. In Überleitung zum nächsten Hauptteil werden außerdem die grundlegenden, kontextuellen Einflussgrößen der akustischen Informationsübertragung besprochen. Im letzten Unterkapitel folgen dann die Überprüfung der Hypothese H1 und eine Konklusion hinsichtlich der wahrnehmungstheoretischen Eignung auditorischer Displays und Benutzeroberflächen.

Nachdem bereits in Kapitel 3.9 Spatialität als bedeutendste Einflussgröße eines Klangkontextes herausgearbeitet wurde, widmet sich das 4. Kapitel dem Verhältnis von Audition und Räumlichkeit. Auf diesem Hintergrund wird Spatialität als Vehikel der auditorischen Kognition, als Wiedergabeparameter spatial-auditorischer Anwendungen und als narrativer Kontext für akustische Erzählungen untersucht. Abschließend folgt eine zusammenfassende Konklusion der Bedeutung von Spatialität für die Audition.

Kapitel 5 befasst sich im Folgenden mit der praktischen Implementierung der abgeleiteten wahrnehmungstheoretischen Prinzipien in Form von Spatial Auditory Displays. Auf dem Hintergrund dieser Technologie findet im Anschluss auch die explizite Überprüfung der Hypothese H1.1 statt. Im Rahmen der folgenden Anwendungsbeispiele wird außerdem noch einmal gesondert auf die Technologie der Audio-Augmented Reality hingewiesen. Zum Abschluss des Kapitels folgt ein Ausblick, der in die Konklusion dieser Arbeit im letzten Kapitel überleitet.

2 Vorangegangene Projekte

Im Zentrum dieser Arbeit steht die Annahme, dass der eingangs erwähnte Primat der visuellen Informationsübertragung, wahrnehmungstheoretisch keine Zwangsläufigkeit ist. Dass auch akustische Benutzeroberflächen effektiv für verschiedene Anwendungskontexte eingesetzt werden können und dabei ohne die Unterstützung der visuellen Sphäre auskommen, ist in der Forschung längst bekannt. Einige Projekte aus dem Bereich der Audio-Augmented Reality (AAR) sollen deshalb im Folgenden dargestellt werden, um einen praktischen Einstieg in die Welt spatial-auditorischer Displays und Benutzeroberflächen zu bieten.

2.1 Automated Tour Guide

Im Mai 1995 wurde der *Automated Tour Guide* von Benjamn B. Bederson erstmals im Rahmen seines Kurzartikels „Audio Augmented Reality: A Prototype Automated Tour Guide“ (1995) für die „CHI ’95 Mosaic of Creativity“ Konferenz vorgestellt. Ziel Bedersons Bemühungen war es eine Museumsumgebung akustisch so zu augmentieren, dass sich die Besucher zu jedem Kunstwerk zuvor aufgenommene Audio-Kommentare und Beschreibungen anhören konnten:

„We have built a prototype audio augmented reality-based tour guide. This system replaces analog audio tapes with random access digital audio. [...] The idea is to allow visitors to hear descriptions of pieces just by walking up to them.“ (Bederson 1995, S. 210)

Als Grundlage für dieses Projekt dienten frühere, rudimentäre AAR-Systeme in denen die Besucher über Kopfhörer eine einfache, fortlaufende Audio-Aufnahme anhören konnten. Solange die Museumsgäste einer vorgegebenen Route in der vorgesehenen Geschwindigkeit folgten, wurde so eine akustische Museumsführung imitiert. Leider war es bei diesen älteren *Tour Guides* für den Nutzer nicht möglich, individuell über die Verweildauer oder die Reihenfolge mit der die Ausstellungsstücke betrachtet wurden, zu entscheiden. Eine freie und natürliche Erschließung der Museumsfläche wurde somit ausgeschlossen.

Um diese prinzipielle Einschränkung auszuräumen, sieht der Automated Tour Guide eine Ortung¹ der Museumsbesucher mittels Infrarot-Transmittern (angebracht über den Exponaten) und Empfängern (getragen von allen Museumsbesuchern) vor. Betritt ein Nutzer die *Zielzone*² eines Exponates, übermittelt der Transmitter eine eindeutige Kennung, die einem Audi-Kommentar zugeordnet ist, welches anschließend für den Besucher ausgegeben wird. Verlässt der Benutzer diese Zone, endet die Wiedergabe. Durch diese simple Erweiterung ist es den Be-

¹ Streng genommen handelt es sich nicht um eine Ortung der Besucher; die Transmitter, die über den Exponaten angebracht sind, emittieren kontinuierlich ein Infrarotsignal welches durch die Empfänger der Besucher erkannt und an das Wiedergabesystem weitergegeben wird. Eine Berechnung der Position des Besuchers auf der Museumsfläche ist nicht vorgesehen.

² Der Radius dieser Zone ist in Bederson’s Ausführung nicht angegeben.

suchen möglich sich frei im Museum zu bewegen, sich mit anderen Personen auszutauschen oder ein Ausstellungsstück näher zu untersuchen, ohne den Limitationen einer klassischen Führung unterworfen zu sein.

Der *kontextuelle Schlüssel* für diese deutliche Verbesserung der Nutzerfreundlichkeit ist die Ortung der Besucher – die Nutzung der spatialen Dimension der Museums Umgebung. Erst diese Erweiterung ermöglicht, so Bederson, eine sinnvolle, akustische Augmentation: „A more technologically sophisticated tour guide, on the other hand, can offer the benefits of automation without the social conflicts caused by the taped tour guide.” (1995, S. 210).

2.2 Hear & There

Anders als das vorangegangene Projekt konzentriert sich Joseph Michael Rozier’s „Hear & There“ (2000) nicht auf eine kontrollierbare, überwiegend statische Museums Umgebung, sondern öffnet sich für alle Umgebungstypen. Grundgedanke der Anwendung ist es, den Nutzern die Möglichkeit zu bieten, sogenannten *akustische Abdrücke* (Audio Imprints) in einer digital angereicherten Umgebung zu hinterlassen, die durch andere Nutzer gefunden und angehört oder sogar erweitert werden können. Diese akustischen Spuren könnten beispielsweise private Gedanken eines Autors zu einem historischen Monument oder eine kurze Geschichte über die persönlichen Erlebnisse an einem bestimmten Ort beinhalten. Langfristig soll so eine zusätzliche Dimension akustischer Hinterlassenschaften entstehen, die die Realität anreichert und jederzeit abgerufen werden kann. Das gesamte System ist dabei darauf ausgelegt, vollständig als AAR-Anwendung ausgeführt zu werden:

“Hear & There is a system that provides many important additions to the field of augmented reality. These include the focus on audio (including spatialized audio), the focus on authoring in the space, the notion of linking components in an augmented space together, and the focus on highly accurate location information.” (Rozier 2000, S. 48)

Auch dieses Projekt stellt kontextuell vornehmlich auf eine *direkte Verquickung von Raum und akustischer Information* ab. Im Kontrast zum Automated Tour Guide geschieht dies bei Rozier gleich auf drei unterschiedlichen Ebenen. Zum einen wird jede Audio-Aufnahme direkt an eine Geoposition gekoppelt, an der sie durch andere Nutzer wiedergegeben werden kann. Zum anderen beziehen sich die Aufnahmen in der Regel, wie im Beispiel des Kommentars zu einem historischen Monument, auch inhaltlich auf ein spatiales Objekt (i.e. das Monument) oder eine reale Umgebung. Darüber hinaus tritt die räumliche Dimension auch als Wiedergabeparameter in einer dreidimensionalen (spatialen) Klangumgebung in Erscheinung:

„Without visual cues, the system must rely entirely on audio to indicate the presence of SoundSpots and to direct a user into the center of a SoundSpot to hear the primary audio of the SoundSpot [...]. To alleviate this problem, we chose to investigate the more

complex system of spatialized sound. Using spatialized, or 3D sound, a sound can “appear” to be in any location in 3D space.” (Rozier 2000, S. 39)

Auch dieses Projekt wurde, mit geringeren Einschränkungen, prototypisch umgesetzt. Details zur technischen Umsetzung und eine Diskussion der klassischen Herausforderungen vergleichbarer Systeme finden sich in Rozier’s Ausarbeitung „Hear & There: An Augmented Reality System of Linked Audio“ (2000).

2.3 Audio Aura

Während sich die ersten beiden Projekte, wenn auch nicht ausschließlich, auf Augmentation durch die Vertonung von Sprache fokussieren, widmet sich das Projekt „Audio Aura“ der digitalen Anreicherung einer klassischen Büroumgebung durch non-verbale akustische Hinweise. Entwickelt wurde dieses Konzept von Elizabeth D. Mynatt, Maribeth Back, Roy Want und Ron Frederick 1997 im Xerox Palo Alto Research Center. Durch die Verwendung akustischer Signale aus einem natürlichen Kontext soll eine *leichtgewichtige Interaktion* entstehen, die akustische Hinweise nicht als Alarmsignale, sondern innerhalb eines intuitiven, beiläufigen Informationskonzeptes einsetzt. Beispielszenarien für die Implementation wären:

- Ein Mitarbeiter bewegt sich frei in der Büroumgebung und trägt mit Audio Aura verknüpfte Kopfhörer. Während der Abwesenheit von seinem Arbeitsplatz sammeln sich langsam wichtige Mails in seinem Posteingang, auf die er hingewiesen werden soll. Im Hintergrund werden deshalb, zunächst ganz leise und in weiter Ferne, die Schreie einiger Möwen für ihn hörbar. Relativ zur verstrichenen Zeit und der Menge ungelesener Mails, verdichtet sich diese Geräuschkulisse, um den Mitarbeiter zu veranlassen, zu seinem Arbeitsplatz zurückzukehren.
- Mehrere Mitarbeiter arbeiten parallel an der Konzeption eines neuen Projektes. Mitarbeiter A sieht Gesprächsbedarf mit Mitarbeiter B hinsichtlich bestimmter Details; beide Mitarbeiter nutzen Audio Aura. Mitarbeiter A aktiviert die entsprechende Programmoption. Wenn sich Mitarbeiter B nun das nächste Mal in der Nähe des Arbeitsplatzes von Mitarbeiter A aufhält, wird er durch ein Hintergrundsignal (e.g. Schwingen einer Klangschale, Knistern eines Feuers) auf diesen Umstand aufmerksam gemacht und kann entscheiden, ob er das Gespräch jetzt oder zu einem anderen Zeitpunkt suchen möchte.

Wie in den Beispielszenarien abgebildet, kommen in Audio Aura primär abstrakte Signale zum Einsatz, die verschiedenen, natürlichen Klangumgebungen entlehnt sein können.

Durch diese elektroakustische Erweiterung der Arbeitsumgebung können verschiedene Effekte erzielt werden. Die Augmentation der gesamten Bürofläche macht eine beiläufige Interaktion möglich, die nicht mehr strikt an den Schreibtisch gebunden ist. Mitarbeiter werden dadurch animiert, den Arbeitsplatz häufiger zu verlassen, um sich beispielsweise mit Kollegen auszutauschen. Ob dieser Effekt erstrebenswert ist, hängt letztlich vom Tätigkeitsfeld und der strukturel-

len Ausrichtung des Unternehmens ab. Darüber hinaus sind die Benutzer nicht mehr gezwungen, wiederholt bewusst auf die üblichen Kommunikationsinterfaces wie Pager, Handy oder Desktop-PC zurückzugreifen, sondern können gezielt beiläufig durch Audio Aura informiert werden. Zusätzlich ermöglicht die Anwendung, dass der häufig überlastete, visuelle Wahrnehmungskanal entlastet und eine sinnvolle Ergänzung durch akustische Einflüsse erzeugt werden kann (Mynatt et al. 1997). Die Ungewohntheit dieses Interfacedesigns (*Strange Interface*³) führt bei Nutzern außerdem häufig zu einer neugierigen, fast spielerischen Interaktion mit dem System, innerhalb derer die Funktionen der Anwendung schnell selbstständig entdeckt und ausprobiert werden (Vazquez-Alvarez et al. 2011, Droumeva & Wakkary 2008 et al.).

In diesem Projekt wird weniger auf die direkte Ortung einzelner Nutzer abgestellt, als auf die *feste Verortung digitaler Informationen in der augmentierten Benutzeroberfläche* – der Büroumgebung.

„Rather than attempt to determine the precise location of an individual, Audio Aura focuses on information that can be associated with pre-determined points in space. Tags are placed at those locations, providing an easy mechanism of determining where the user is in the augmented space.” (Rozier 2000, S.12)

Durch diese *Verquickung* einer zusätzlichen, elektroakustischen Dimension mit einer realen Umgebung wird die räumliche Eigenschaft des Büros (Spatialität⁴), als Interface nutzbar gemacht. Prototypisch wurde auch Audio Aura umgesetzt, zu einer dauerhaften Installation auf einer Bürofläche kam es allerdings nicht.

2.4 Resumée

Bereits diese Auswahl wohlkonzipierter Projekte zeigt, wie durch den gezielten Einsatz akustischer Signale beiläufig oder unmittelbar Informationen übertragen und Geschichten erzählt werden können, während der visuelle Wahrnehmungskanal entlastet wird. Durch die spatiale Verortung der Klangelemente, in diesem Fall durch ihre Anwendung in einem AAR-System, kann zusätzlich die *ordnende Struktur des Raumes* instrumentalisiert werden, um eine intuitive Interaktion zwischen Nutzer und System zu unterstützen (Droumeva & Wakkary 2008, Sonnenschein 2011 et al.). Dass keines dieser Projekte nach 2000 entworfen wurde, unterstreicht, wie lange das Potential spatialer, akustischer Benutzeroberflächen in der Forschung bereits bekannt ist. Natürlich existieren auch bei der Umsetzung solcher Projekte die klassischen Herausforderungen hinsichtlich adäquater Ortung, dem Echtzeit-Rendering oder einer belastbaren Datenübertragung. Sie unterscheiden sich aber von ihren visuellen Pendanten meist dadurch, dass eine

³ Als „Strange Interface“ werden üblicherweise Benutzeroberflächen bezeichnet, deren Funktionsweise neuartig und dem Nutzer deshalb unbekannt sind. Der Vorteil dieser Interface-Klasse besteht darin, dass häufig ein überdurchschnittliches Maß an Eigenmotivation für die Erschließung der Bedienstruktur, bei gleichzeitig hoher Frustrationstoleranz für Fehlfunktionen beobachtet werden kann.

⁴ Gemeint ist mit „Spatialität“ hier und im Folgenden die Eigenschaft der Räumlichkeit gemäß der direkten Übersetzung aus dem Englischen („Spatiality“), in Abgrenzung zu der gängigen, literaturwissenschaftlichen Verwendung des Wortes.

praktische Umsetzung bereits heute technisch möglich wäre (Rozier 2000, Mynatt et al. 1997, Lokki & Järveläinen 2001 et al.).

Warum sich akustische Konzepte in bestimmten Kontexten häufig sogar besser eignen, als klassische visuelle Lösungen, lässt sich aus einer bloßen Projektpräsentation allerdings nicht schließen. Der wahrnehmungstheoretische Hintergrund spatial-auditorischer Anwendungen soll deshalb in den folgenden beiden Kapiteln umfassend beleuchtet werden.

3 Auditorische Information

Mit einem Anteil von 11% fungiert die auditorische Wahrnehmung als zweitbedeutendste Sinnesmodalität heutiger Menschen.⁵ Die in Abbildung 1 dargestellte Übersicht verdeutlicht die Dominanz visueller Stimuli im Rahmen des menschlichen Wahrnehmungsprozesses. Dass diese Veranlagung auch durch digitale Anwendungen bedient wird, erscheint zunächst sinnvoll und natürlich. Ob visuelle Technologien diese Verteilung weiter zugunsten der Vision verschoben haben und dies in Zukunft weiter tun werden, bleibt Spekulation.

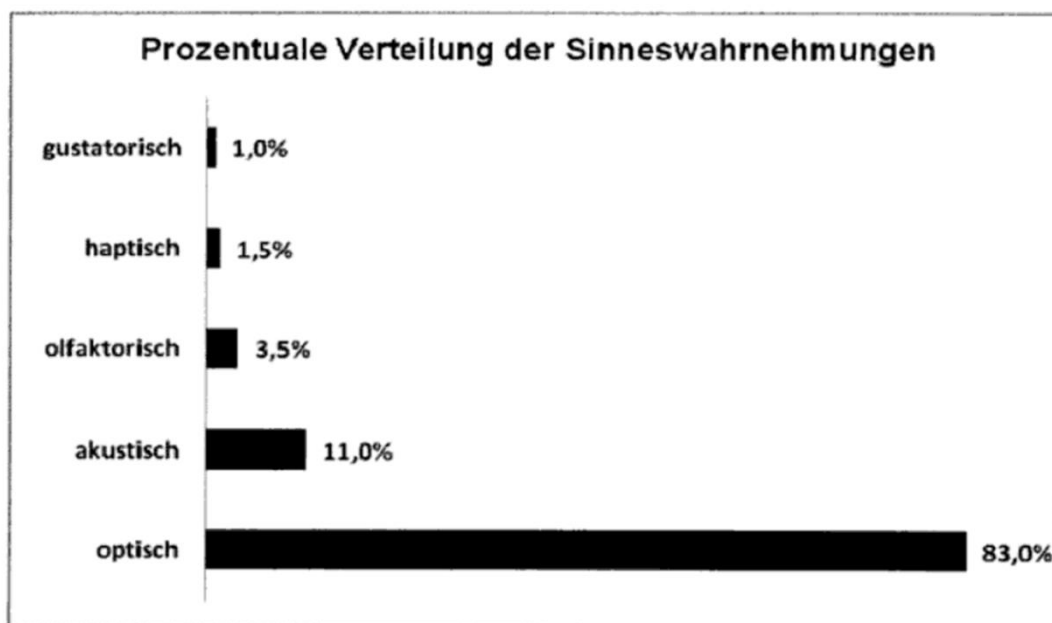


Abbildung 1: Prozentuale Verteilung der Sinneswahrnehmung (Steiner 2003, S.29)

Akustische Signale stellen in unserer digitalen Welt meist allenfalls den Rahmen oder das Beiwerk einer ansonsten primär visuellen Sinnesansprache dar. Häufig soll Klang dabei die Wahrnehmung des visuellen Inhaltes unterstützen (e.g. Filmmusik) oder in Form eines Alarms auf ihn aufmerksam machen (e.g. Wecker) (Cohen 2000, Walker & Kramer 2004 et al.). Dass und wie Klang auch eigenständig in der Lage ist komplexe Informationen mit hoher Dichte zu transportieren, Emotionen zu induzieren oder Neugier und Aufmerksamkeit im Adressaten zu erzeugen, soll im Laufe des 3. Kapitels verdeutlicht werden. Um auf eine wahrnehmungstheoretische Eignung auditorischer Displays und Benutzeroberflächen zu schließen, wie es Ziel der vorliegenden Arbeit ist, müssen zunächst zwei Komponenten betrachtet werden. Einerseits müssen akustische Signale in der Lage sein, ein gewisses Maß an Information mit einer angemessenen Dichte transportieren zu können. Andererseits muss der auditorische Wahrnehmungsapparat über die

⁵ Dieser Wert ist lediglich als Mittelwert zu verstehen. Speziell wahrnehmungsbehinderte Menschen, e.g. sehbehinderte Menschen, werden offensichtlich andere Werte aufweisen.

notwendigen Mechanismen verfügen, um diese Informationen auszulesen und angemessen zu interpretieren. Sowohl der Adressat als auch das Trägermedium der Information sind insofern für die vorliegende Arbeit von Bedeutung und sollen in diesem Kapitel umfassend Würdigung finden.

3.1 Grundlagen der Akustik und Pseudoakustik

Unsere akustische Umwelt besitzt einige phänomenologische Eigenschaften die zentral für die menschliche Wahrnehmung sind. Zur Aufnahme dieser akustischen Reize⁶, dient das Gehör. Im Kontrast zum visuellen Wahrnehmungsapparat, liefert das Gehör dauerhaft und im Normalfall ununterbrochen Informationen, die durch die unmittelbare Fernwirkung des Schalls übertragen werden. Das menschliche Ohr ist zwar weniger beweglich, allerdings wegen der physikalischen Eigenschaften des Schalls und der physiologischen Konstruktion des Ohres in der Lage, Klangeinflüsse einer 360-Grad-Umgebung zu erfassen. Eine bewusste Fokussierung auf Umgebungsausschnitte ist per se nicht notwendig, weshalb akustische Stimuli oft beiläufig und weniger intendiert wahrgenommen werden. Erzeugen akustische Reize keine Signalwirkung oder werden nicht gezielt *erhört* kommt es deshalb häufig zu einer unterbewussten oder gänzlich unbewussten Wahrnehmung (Raffaseder 2000, Hug 2008).

Damit ein akustischer Reiz durch das menschliche Gehör wahrgenommen werden kann, es also zu einem sogenannten *Hörereignis* (Empfindung) kommt, muss zunächst ein akustisches Ereignis (Reiz) vorliegen. Da sich dieser Begriff auf die physikalische Schallentwicklung bezieht, wird häufig auch von einem *Schallereignis* gesprochen (Blauert 1969, Blauert & Braasch 2007, et al.). Grundlage jedes akustischen Ereignisses ist nach Hannes Raffaseder der Schall, definiert als mechanische Schwingung eines elastischen Mediums im Frequenzbereich des menschlichen Hörens (ca. 20 – 20.000 Hz⁷) (Raffaseder 2000). Voraussetzungen hierfür sind ein Anregungsmechanismus (e.g. das Schlagen einer Glocke), ein schwingendes Objekt (e.g. eine Glocke) und ein Ausbreitungs- oder Trägermedium (e.g. Luft). Neben dem Frequenzbereich ist der erzeugte Schalldruckpegel eines akustischen Ereignisses, auch als Lautstärke bezeichnet, entscheidend für die menschliche Wahrnehmung. Damit ein Schallreiz überhaupt durch das Gehör erfasst werden kann, muss zunächst die Hörschwelle, die untere Lautstärkenbeschränkung des menschlichen Hörens, erreicht werden. Überschreitet der Schalldruckpegel allerdings die Schmerzschwelle des Gehörs, die das erträgliche Lautstärkenmaximum markiert, kann das akustische Signal nicht mehr schmerzfrei wahrgenommen werden. Die jeweiligen Schwellenwerte sind zum einen durch die individuelle Konstitution des Empfängers (e.g. Alter, Gesundheit), zum anderen aber durch

⁶ Ausgenommen seien an dieser Stelle Impulse, die bei ausreichender Lautstärke als eigene Körperschwingung wahrgenommen werden können.

⁷ Im Hinblick auf die Behandlung von Auditory Displays in späteren Kapiteln soll darauf hingewiesen sein, dass dieses Intervall nur eine Näherung darstellt. Zum einen sind nicht alle Menschen in der Lage Schwingungen am Rande dieses Spektrums wahrzunehmen, zum anderen wird beim Design akustischer Displays nur auf den mittleren Teil des Intervalls abgestellt, um das Erhören von Signalen für den Nutzer zu erleichtern.

eine Wechselwirkung untereinander bestimmt⁸. Als allgemeiner Richtwert für die Schmerzschwelle wird häufig der Wert von 130 dB angenommen. Die Menge aller Wertepaare, die innerhalb dieser beiden Intervalle liegt, wird üblicherweise als *Hörfläche* bezeichnet.

Akustische Ereignisse beinhalten als Folge ihrer physikalischen Umgebung deren Eigenschaften und Charakteristika und sind deshalb Träger von Information (Truax 2000, Raffaseder 2000, Hug 2008). Eine Vielzahl zusammengehöriger, akustischer Einzelereignisse bildet eine akustische Szene, die in Barry Truax' „Acoustic Communication“ auch als „Acoustic Community“ bezeichnet wird (2000). Der Zuhörer kann sich selektiv einer Klangquelle zuwenden („Cocktail-party-Effekt“) oder die akustische Szene als Ganzes wahrnehmen. In jedem Fall werden die Einzelereignisse aber zueinander und zu allen weiteren Sinneseinflüssen in Kontext gesetzt. Zusätzlich werden Erfahrungswerte, Erwartungswerte und Erinnerungen herangezogen, um ein Hörereignis angemessen zu interpretieren. So kann erklärt werden, weshalb beispielsweise ein Berufsfischer das Rauschen des Meeres mit einem gewöhnlichen Arbeitstag assoziiert, während etwa ein Bürokaufmann diese akustische Szene in der Regel intuitiv mit einem Strandurlaub in Verbindung bringt (Raffaseder 2000).

Wie sich ein akustisches Ereignis „anhört“, wird darüber hinaus durch verschiedene Parameter bestimmt, die durch die Psychoakustik untersucht werden. Nach Jens Blauert und Jonas Braasch setzt sich ein Hörereignis durch die untergeordneten Parameter der Überkategorien Raum, Zeit und Eigenschaft bzw. *Eigenschaftlichkeit* zusammen:

„Das, was wir hören ist räumlich zeitlich und eigenschaftlich bestimmt. Unsere Hörereignisse (... Hörobjekte, Laute) existieren jeweils zu ihrer Zeit an ihrem Ort und sind jeweils mit spezifischen Eigenschaften ausgerüstet.“ (Blauert & Braasch 2007, S.1)

Die Überkategorie der Klangeigenschaft umfasst hierbei Faktoren wie Tonhöhe, Klangfarbe, Lautstärke, Schärfe, Tonheit, Rauigkeit, Tonhaltigkeit, Impulshaltigkeit, Schwankungsstärke, etc. Abhängig vom Untersuchungshintergrund kann diese Reihe erweitert oder reduziert werden und ist letztlich an das subjektive Empfinden des Hörenden gebunden.

Während es sinnvoll erscheint, diese psychoakustischen Parameter physikalischen Schalleigenschaften zuzuordnen, konnte ein konkretes Modell bis dato leider nicht erstellt werden. Zwar ordnet man klassischerweise die empfundene Tonhöhe der Frequenz, die Lautstärke der Amplitude und die Klangfarbe dem Spektrum zu, eine formale Äquivalenz kann allerdings nicht abgeleitet werden. Zusätzlich scheint ein Großteil der oben angeführten Klangeigenschaften durch Wechselwirkungen mehrerer physikalischer Schalleigenschaften bedingt zu sein (Raffaseder 2000 et al.).

Der Informationsgehalt eines Schallsignals kann nach Raffaseder in Anlehnung an Blauert deshalb wie folgt zusammengefasst werden:

⁸ Bei einer Frequenz von etwa 1.000 Hz wird beispielsweise in der Regel schon bei geringerem Schalldruckpegel die individuelle Schmerzschwelle erreicht, als bei einer Frequenz von ca. 10.000 Hz.

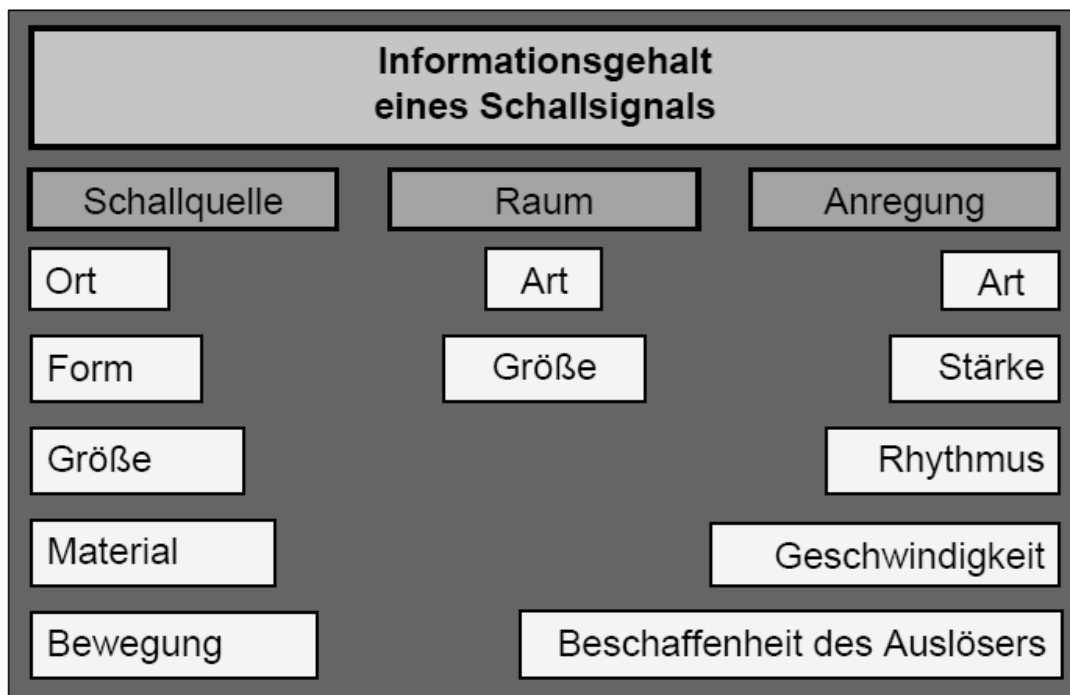


Abbildung 2: Informationsgehalt eines Schallsignals (Raffaseder 2003, F.14)

Ausgehend von diesen *physikalischen und informationellen Eigenschaften* erzeugt unsere akustische Umwelt eine facettenreiche Phänomenologie, die auf dem Hintergrund der vorliegenden Arbeit zusammenfassend auf sieben Kernbereiche heruntergebrochen werden kann. In seinem Artikel „Towards a Hermeneutics and Typology of Sound for Interactive Commodities“ (2008) identifiziert Daniel Hug in Anlehnung an Wolfgang Welsch (1996) und Don Ihde (1976) diese Kernelemente als:

Pervasiveness (Omnipräsenz)

Das menschliche Gehörn ist fast immer aktiv und nimmt Geräusche wahr. Selbst wenn wir schlafen werden akustische Ereignisse registriert und interpretiert. Das menschliche Ohr kann nicht wie das menschliche Auge geschlossen oder abgewendet werden. „Weghören“ ist insofern praktisch unmöglich.

Temporality (Vergänglichkeit)

Akustische Ereignisse sind vergänglich und immer an einen dynamischen Prozess gebunden. Weil sie nicht, dem visuellen Paradigma entsprechend, in einem statischen Zustand analysiert werden können, entziehen sie sich häufig unserer intendierten Wahrnehmung.

Multidimensionality (Mehrdimensionalität)

Unser Gehör ist ein mehrdimensionaler Wahrnehmungskanal, durch den mehrere simultane Schallereignisse in kürzester Zeit separiert und verarbeitet werden können, um alle relevanten Informationen aus einem akustischen Reiz zu extrahieren.

Spatiality (Spatialität oder Räumlichkeit)

Diese Eigenschaft akustischer Ereignisse ist direkt mit dem oben genannten Aspekt der Mehrdimensionalität verbunden. Das wahrgenommene, akustische Abbild unserer Umwelt ist holistisch und durch simultane Einflüsse aus allen Umgebungsrichtungen geformt. Faktoren wie Wind oder die spatialen Eigenschaften unserer räumlichen Umgebung drücken sich in bestimmten Klangeigenschaften aus.

Sociality (Soziale Kollektivität)

Akustische Ereignisse sind in der Lage Menschen zu verbinden. Die *gemeinsame Wahrnehmung* erzeugt häufig ein starkes Zusammengehörigkeitsgefühl, das sich in sozialen Phänomenen von Stammestänzen bis hin zu Rockkonzerten widerspiegelt. Selbst Ereignisse mit eingeschränkter Sichtbarkeit können durch ihr akustisches Abbild von hunderten Zuhörern erlebt werden.

Emotionality (Emotionalität)

Ein weiteres Charakteristikum des Klangs ist seine enorme, emotionale Wirkkraft. Der emotionale Kontext einer Filmszene beispielsweise – der einen zentralen Aspekt der gesamtheitlichen Bedeutung des Filmes ausmacht – kann mithilfe der entsprechenden Filmmusik maßgeblich verändert werden. Auch alltägliche Geräusche erzeugen Emotionen, wie i.a. durch die umfassende Untersuchung der Disstressinduktion⁹ durch Störgeräusche belegt werden konnte.

Physicality (Körperlichkeit)

Jedes Hörereignis ist an physikalische Bewegungen - nämlich die Ausbreitung von Schallwellen im Raum und die Impulskraft der Klangquelle - geknüpft. Gleichzeitig ist der Klang eines akustischen Ereignisses immer an die physikalischen Eigenschaften des erzeugenden Artefaktes gebunden. Viele Studien haben sich bereits mit der (eingeschränkten) Fähigkeit des menschlichen Gehörs auseinandergesetzt, diese Eigenschaften aus einem Hörereignis zu extrahieren. Konkrete Aussagen über die zugrundeliegenden Analysemechanismen sind allerdings kaum möglich.

Während sich manche dieser phänomenologischen Charakteristika direkt aus den physikalischen Eigenschaften des Schalls selbst ableiten lassen, bedürfen andere einer tiefgreifenderen Begründung.

Mit Einsetzen des 20. Jahrhunderts ist es notwendig geworden entsprechende Überlegungen auch im Hinblick auf die Elektroakustik anzustellen, die sich mit der Übersetzung, Speicherung, Wiedergabe und dem Transfer akustischer Impulse beschäftigt:

„Similarly, the entire field of electroacoustics, the result of the application of 20th-century technology to sound [...] is concerned with the energy transfer from acoustic to electrical forms, a process called transduction, as well as the subsequent processing and/or storage of the resultant audio signal.” (Truax 2000, S.9)

⁹ Disstress bezeichnet nach Hans Selye's Modell, im Kontrast zum positiven Eustress, alle negativen Stresseinwirkungen auf den Menschen (Selye 1976).

Darüber hinaus sind auch die Bearbeitung und Synthese von Klang als ihre Teildisziplin definiert: „[T]he application of electrical and electronic technology to the processing or synthesizing of sound, the latter being the electronic generation of sound without an original acoustic source (Truax 2000, S.9).“ Grundsätzlich kann die Elektroakustik in den theoretischen Rahmen der natürlichen Akustik eingebettet werden; einige kommunikationstheoretische Prinzipien müssen auf diesem Hintergrund allerdings reinterpretiert werden. Nie zuvor war es möglich, ein akustisches Ereignis außerhalb seiner natürlichen Klangumgebung zu hören, Klänge synthetisch zu erzeugen oder diese aufzunehmen und beliebig wiederzugeben. Die resultierenden Manipulationsmöglichkeiten bergen das Potential den informationellen Gehalt eines akustischen Ereignisses erheblich zu verändern.

„Before audio technology, no sound had ever been heard twice, exactly the same, nor had any sound ever been heard outside its original context. The communicational significance of any sound can only be judged within its complete context in the broadest environmental, social, and cultural sense. In fact, it is through context that we understand how a sound functions.“ (Truax 2000, S.12)

Diese *elektroakustische Revolution* unserer Umwelt hat Projekte ermöglicht, wie sie beispielhaft bereits im 2. Kapitel vorgestellt wurden und soll im Folgenden als Bewertungshintergrund für eine wahrnehmungstheoretische Analyse dienen.

3.2 Hören – Codierung und Terminologie

Um die Auswirkung dieser elektroakustischen Revolution besser begreifbar zu machen, stellt Truax in seinem Buch ein umfassendes, kommunikationstheoretisches Modell vor. Dieses Modell setzt sich aus dem Zuhörer (Adressat), dem akustischen Ereignis (Information) und der Umgebung (Medium/Kontext) zusammen. Mittelpunkt dieser Kommunikationsstruktur und der folgenden wahrnehmungstheoretischen Untersuchung ist ihr Adressat (Truax 2000).

3.2.1 Codierung des Hörens

Die menschliche Wahrnehmung von Klang beginnt mit dem Eintritt von Schallwellen in das Außenohr bzw. die Ohrmuschel (Pinna bzw. Auricula auris). Nachdem der Schall in das Mittelohr vorgedrungen ist, versetzt dieser die feine Membran des Trommelfells (Membrana tympani) in Schwingung. Der mechanische Impuls wird von dort über die Gehörknöchelchen (Ossicula auditūs) an die mit Flüssigkeit gefüllte Gehörschnecke im menschlichen Innenohr (Cochlea) weitergeleitet, wo ein Wellenimpuls initiiert wird, der die Basilarmembran (Membrana basilaris) in Bewegung versetzt. Die Bewegung der Membran gleicht dabei einer Wellenbewegung (Wanderwelle), die ihre maximale Auslenkung an dem Punkt der Basilarmembran hat, dessen Spannung und Dimension der Frequenz des eingehenden Tons entspricht. Ähnlich wie eine Schwimmblosse ist die Basilarmembran an ihrem Apex weit und beweglich und wird zunehmend schmaler und steifer in Richtung ihrer Basis. Handelt es sich um eine hochfrequente Klangein-

wirkung, ist diese stärker an der Basis der Basilarmembran verortet und vice versa (Lotto & Holt 2010, Truax 2000 et al.).

Diese Eigenschaft der Basilarmembran ermöglicht eine Spektralanalyse des eingehenden Reizes, der dadurch in seine einzelnen Frequenzkomponenten aufgeteilt werden kann. Die auf der Basilarmembran befindlichen auditorischen Rezeptoren, die inneren Haarsinneszellen, werden durch die Wellenbewegung der Membran stimuliert und erzeugen in Folge eine Kaskade chemischer Prozesse. Der ursprüngliche, mechanische Impuls wird so in den Axonen des Hörnervs in Form von Aktionspotentialen an das Gehirn weitergeleitet (mechanoelektrische Signaltransduktion). Die äußeren Haarsinneszellen, die sich auf der Basilarmembran befinden, unterstützen dabei, frequenzabhängig, die Bewegungsreaktion der Membran und verbessern so die Transduktion. Der resultierende neuronale Code enthält außerdem zusätzliche Informationen über die Reizfrequenz, die sich im zeitlichen Muster der neuronalen Aktivierungen äußert (Phase Locking). Gleichzeitig schlägt sich die Reizintensität in der Anzahl der aktiven Nervenstränge und deren Aktivitätsrate nieder. Zusätzlich werden im oberen Olivenkernkomplex des Ohres parallel Laufzeitunterschiede bzw. Laufzeitdifferenzen (ITD) und Intensitätsunterschiede bzw. Intensitätsdifferenzen (ILD) zwischen beiden Ohren analysiert, um eine Ortung oder Richtungszuweisung der Schallquelle zu ermöglichen. Das skizzierte Schema stellt natürlich nur ein rudimentäres Ablaufmodell der menschlichen Klangwahrnehmung dar. Streng genommen lösen gerade komplexere Klangeinwirkungen eine Reihe von neuronalen Prozessen und Wechselwirkungen zwischen Reizkomponenten aus, die im Rahmen dieser Arbeit nicht abschließend betrachtet werden können:

„Thus, for complex signals, there is quite a bit of neural processing and stimulus component interaction occurring even at the level of the auditory nerve. Substantial processing and feature extraction occurs all along the auditory pathway, which is more complex than the visual pathway with significant subcortical processing and interactions between the signal from both ears.” (Lotto & Holt 2010, S.5)

Die hohe Informationsdichte, die deshalb durch akustische Signale erreicht werden kann, soll in Kapitel 3.4 genauere Betrachtung finden. Sie bildet die Grundlage für die effektive Nutzung solcher Signale in auditorischen Displays.

3.2.2 Terminologie des Hörens

Abseits des physiologischen Hörvorgangs ist für die individuelle Wahrnehmung eines akustischen Ereignisses oder einer akustischen Szene entscheidend, wie sich der Adressat verhält – hört dieser nur unbewusst, hört er „hin“ oder hört er vielleicht sogar einem bestimmten akustischen Ereignis, e.g. einer sprechenden Person, „zu“. Michel Chion unterscheidet in seinem für die Filmwissenschaften grundlegenden Werk „Audio-Vision: Sound On Screen“ (1994) drei Modi des Hörens bzw. Zuhörens:

- **Reduced (Reduziert):** Ein Modus des Zuhörens, der sich allein auf das Erhören der Klangeigenschaften fokussiert, unabhängig von der Bedeutung des Klangs oder der Ortung seiner Quelle.
- **Semantic (Semantisch):** Ein Modus des Zuhörens, der sich auf die Decodierung der Nachricht bzw. Bedeutung eines Klangs, ob verbal oder non-verbal, konzentriert.
- **Causal (Kausal):** Ein Modus des Zuhörens, der versucht, den Ursprung (räumlich und kausal) eines akustischen Ereignisses zu ermitteln, ob kategorisch oder konkret.

Während diese Betrachtung zwar die einzelnen informationellen Dimensionen, die ein akustisches Ereignis konstituieren, angemessen herausstellt, bleibt das *Aktivitätsniveau des Zuhörers* weitestgehend außer Acht. Für die menschliche Wahrnehmung und die Auswirkungen eines Hörereignisses auf das Bewusstsein ist allerdings entscheidend, ob Klang passiv aufgenommen oder aktiv verfolgt wird. Die Definition nach Truax (2000) erlaubt eine solche Unterscheidung:

- **Listening-in-Search:** Ein Zustand des aktiven Zuhörens zum Zwecke der Extraktion von Informationen aus einem Schallereignis, e.g. die Ortung einer Schallquelle.
- **Listening-in-Readiness:** Ein Zustand beiläufiger Beobachtung der akustischen Umgebung. Dieser Modus ist an die Ausbildung von Klangassoziationen gebunden und erzeugt einen Aktionsreiz, wenn ein akustisches Schlüsselereignis wahrgenommen wird. Dieser Aktionsreiz versetzt Zuhörer in aller Regel in den Zustand Listening-in-Search, e.g. die Melodie eines Eiswagens.
- **Background Listening:** Ein Zustand der passiven Klangwahrnehmung. Akustische Reize werden unbewusst oder unterbewusst wahrgenommen, erzeugen im Empfänger aber eine (meist passive) Reaktion, e.g. das unbewusste Hören eines Eiswagens in der Ferne erzeugt ein Gefühl kindlicher Unbefangenheit.

Die beiden vorgestellten Perspektiven sind in der Lage, sich angemessen hinsichtlich der Begriffe des Hörens (englisch „Hearing“, beschreibt eine tendenziell passive Wahrnehmung) und des Zuhörens (englisch „Listening“, beschreibt eine aktive Wahrnehmung) zu ergänzen. Welche Kategorisierung in einer Untersuchung Anwendung findet, ist letztlich von ihrem Betrachtungsgegenstand und Kontext abhängig.

3.3 Auditorische Ordnungsmechanismen

Wurde Schall durch das Gehör aufgenommen, wird dieser, wie in Kapitel 3.2.1 beschrieben, in Nervenimpulse übersetzt (mechanoelektrische Signaltransduktion) und via die auditiven Nervenfasern in das Gehirn geleitet. Dort setzt eine Kaskade kognitiver Verarbeitungsschritte ein, die den Informationsgehalt des eingegangenen Signals decodieren.

Durch *auditorische Gruppierung (Auditory Grouping)* werden zunächst gleichzeitige akustische Ereignisse voneinander getrennt (Segregation) oder als ein Hörereignis zusammengefasst (Fusion). Darauf aufbauend erfolgt auf dieser Ebene die Trennung oder Integration aufeinanderfolgender

akustischer Ereignisse in einen auditorischen Klangstrom (Auditory Stream). Dadurch ist, wie etwa bei melodischen Klangereignissen, die Wahrnehmung mehrerer aufeinanderfolgender Elemente als zusammengehörende Einheit möglich. Neben einer Vielzahl von Einflussfaktoren hängt die Zuordnung zu einem kontinuierlichen Klangstrom vor allem von der Präsentationsrate der Signale, i.e. wie schnell nacheinander die einzelnen akustischen Ereignisse wahrgenommen werden, und der Ähnlichkeit der Klangeigenschaften, i.a. Tonfarbe, Tonlage, ab. Je nach Beurteilung durch den auditorischen Wahrnehmungsapparat kann so eine Struktur einzelner Klangsignale auf verschiedene Arten und Weisen mental repräsentiert werden. Eine bildliche Verdeutlichung dieses Effektes findet sich in Abbildung 3.

Während die durchgezogenen Linien einzelne, akustische Ereignisse symbolisiert, verdeutlicht die gepunktete Linie den Zusammenhang, den das auditorische System zwischen ihnen aufbaut. Ob, wie im oberen Beispiel, die Klangereignisse als zusammengehöriger Klangstrom aufgefasst werden oder nicht, wie es im unter Beispiel der Fall ist, hängt von ihren Eigenschaften und dem auditorischen Kontext der Signale ab. Wie bereits angedeutet, ist diese Zuordnung in aller Regel nicht zwingend, weshalb es durchaus zu unterschiedlichen Interpretationen hinsichtlich der Gruppierung der einzelnen, akustischen Ereignisse kommen kann.



Abbildung 3: Klangstromanalyse (Neuhoff 2011, S.76)

Diese Wahrnehmungsebene ist eng mit der *Auditory Scene Analysis (ASA)* verbunden, durch die akustische Ereignisse in einen szenischen Bedeutungskontext gebracht und interpretiert werden können.

Nachdem eine Repräsentation der einzelnen Klangelemente im auditorischen System erzeugt wurde, kann diese auf dem Hintergrund der evozierten, *abstrakten Wissensstrukturen (Abstract Knowledge Structures)* interpretiert werden. Die auditorischen Repräsentationen werden erkannt oder identifiziert, indem sie mit dem bisherigen Erfahrungsschatz des Zuhörers (top-down) und dem Reizkontext abgeglichen werden. Inwiefern ein eingehender Reiz als relevant einstuft und welche Bedeutung ihm beigemessen wird, ist maßgeblich von diesem Abgleich abhängig. Wurde einem bestimmten Geräusch in der Vergangenheit beispielsweise ein warnender Signalcharakter zugeordnet (e.g. eine Sirene), erhält dieser Reiz durch den Abgleich mit dem für die Wahrnehmung relevanten Teil der abstrakten Wissensbasis eine prioritäre Beachtung. Zusätzlich ermöglicht dieser Prozess die schnelle und sichere Identifikation wohlbekannter, wenn auch ansonsten

bedeutungsarmer Geräusche wie beispielsweise den Klang der Schritte des Nachbarn, mit denen er den Hausflur entlang zu seiner Tür schreitet. Eine detailliertere Betrachtung dieser Ebene folgt in Kapitel 3.5.3.

Die subjektiv wahrgenommene Beziehung einzelner akustischer Ereignisse erzeugt anschließend einen Interpretationsrahmen (Kontext), der die Bedeutung aller nachfolgenden Schallereignisse maßgeblich beeinflusst oder ihnen überhaupt Bedeutung beimisst. Dieser Rahmen bildet gleichzeitig die Grundlage für umfassendere, strukturelle Beziehungen zwischen Hörereignissen untereinander (hierarchischer und assoziativer Natur). Die so konstituierte Wahrnehmungsebene wird als *auditorische Ereignisstrukturierung (Event Structure Processing)* bezeichnet. Im Kontext eines Konzertes können somit beispielsweise die ansonsten alarmierenden, lauten Schallereignisse in einen angemessenen Zusammenhang gebracht und entsprechend verstanden werden. Eine entscheidende Implikation dieses Strukturierungsprozesses ist deshalb seine Bedeutung für zukünftige Hörereignisse.

Sobald die eingehenden akustischen Ereignisse zugeordnet, identifiziert und strukturiert wurden, entsteht eine *mentale Repräsentation der akustischen Umgebung (mental representation of the structure of the sound)* des Zuhörers. Speziell im Hinblick auf die mangelnde Persistenz akustischer Ereignisse ist diese mentale Repräsentation von zentraler Bedeutung für die kognitive Psychologie. Akustische Ereignisse können nur mithilfe dieser über die Zeit akkumulierten Informationen korrekt verstanden werden. Der dynamische Abgleich neuer Reize mit den bereits gesammelten Informationen ist somit grundlegend für das Verständnis der akustischen Wahrnehmung. (McAdams & Bigand 1993a, 1993b).

Mithilfe dieser Kaskade kognitiver Prozesse ist der auditorische Wahrnehmungsapparat allein auf Basis akustischer Signale in der Lage, eine sich nähernde Schallquelle oder einen komplexen, akustischen Kontext korrekt zu identifizieren. Schlussendlich entsteht so die Schnittstelle zwischen einer physikalischen Welt, in der sich Schallwellen von einer Quelle weg im Raum ausbreiten und einer mentalen Welt individueller Hörereignisse, in der akustische Szenen repräsentiert oder Gefühle induziert werden können.

3.4 Kognitive Modelle

Es muss wohl eingeräumt werden, dass die beschriebenen Verarbeitungsschritte akustischer Stimuli von jenen kognitiven Prozessen, die mit abstrakteren, intellektuellen Aktivitäten (e.g. Denken, Logik, Imagination, Entscheidungsfindung) einhergehen, unterschieden werden können. Dabei sollte allerdings bedacht werden, dass erstens, diese elementaren Prozesse bereits nur auf dem Hintergrund höherer Niveaus der menschlichen Kognition (e.g. Evokation abstrakter Wissensstrukturen) möglich und in vielerlei Hinsicht mit diesen verflochten sind. Zweitens darf nicht vernachlässigt werden, dass die *Begreifbarmachung der physikalischen Umwelt* - die sinnliche

Wahrnehmung - einerseits die Grundlage aller Kognition bildet¹⁰ und andererseits, wie eingangs angemerkt, nicht allein deshalb als weniger komplex eingeschätzt werden sollte, weil sie weitestgehend *automatisch* abläuft. Die kognitiven Modelle, mit deren Hilfe die Grundlagen dieser Begreifbarmachung erklärt werden können, sollen im folgenden Kapitel in Kürze vorgestellt werden.

3.4.1 Modularität (nach Fodor)

Die menschliche Wahrnehmung ist grundsätzlich *modular*, i.e. alle Sinne nehmen ihre Umgebung zunächst autonom wahr (Fodor 1983, Cohen 2000). Diese Modularität beginnt mit den physiologischen Grundlagen der Wahrnehmung. So decodiert die Fähigkeit des peripheren Sehens beispielsweise Aktivitäten auf der Netzhaut, nicht aber der Basilarmembran. Seit Jerry Fodors berühmter Ausarbeitung über die Modularität des menschlichen Geistes (1983) ist bekannt, dass zwischen visuellen und akustischen Stimuli insofern per se keine kognitive Abhängigkeit besteht. Zum Zweck einer multimodalen Umgebungsrepräsentation werden die Informationen der einzelnen Sinneskanäle allerdings integriert, um eine holistische Wahrnehmungserfahrung zu ermöglichen. Dabei werden die Ergebnisse der ursprünglich unabhängigen Sinneskanäle auf Konsistenz überprüft und im Konfliktfall entsprechend reinterpretiert. Die resultierende, holistische Wahrnehmungserfahrung ist folglich synthetisch – eine Illusion. Die Wahrnehmung der Umgebungsrealität ist so weit ein synthetisches Konstrukt, wie die multimodale Integration es erforderlich macht (Cohen 2000). Wegen dieser *Autonomie des Wahrnehmungsapparates* erfolgt die Integration der Sinneskanäle noch vor der kognitiven Verarbeitung ihrer ursprünglichen Informationen. Nicht alle dieser Informationen können auf Basis der holistischen Umgebungsrepräsentation durch bewusste Reflektion wiedererlangt werden.

3.4.2 Adaptive Resonance Theory (nach Grossberg)

Erreicht diese holistische Umgebungsrepräsentation die Ebene der Kognition (es entsteht eine Wahrnehmung), entscheidet sich welche Aufmerksamkeit den einzelnen Wahrnehmungsreizen zugeordnet wird. Im Zweifel bedeutet dies, dass ein Großteil der Informationen gar keine bewusste Beachtung findet (Cohen 2000). Während alle Umwelteinflüsse gleichermaßen um Aufmerksamkeit buhlen, erhalten nach Stephen Grossberg's „Adaptive Resonance Theory“ (ART) (1995) nur jene Beachtung, die unseren typischen Erwartungsstrukturen (Abstract Knowledge Structures) entsprechen. Grundlage dieser Konzeption bildet das Zusammenspiel zwischen *Top-down* und *Bottom-up* als Arbeitsrichtung der menschlichen Wahrnehmung (Sarter et al. 2001, Buschman & Miller 2007 et al.).

Top-down beschreibt dabei den Vorgang der Relevanzbewertung von Informationen auf Basis der abstrakten Wissensstrukturen und wirkt vom Abstrakten, Übergeordneten schrittweise hin zum Konkreten und Speziellen. Martin Sarter et al. definieren die Top-down-Verarbeitung in

¹⁰ Wegen der naturwissenschaftlichen Ausrichtung dieser Arbeit können die theoretischen Implikationen des Kantschen Begriffs a priori synthetischen Wissens (und alle ihm zugeordneten epistemischen Konzeptionen) an dieser Stelle vernachlässigt werden.

ihrer Abhandlung „The cognitive neuroscience of sustained attention: where top-down meets bottom-up“ wie folgt:

„‘Top-down’ processes describe knowledge-driven mechanisms designed to enhance the neuronal processing of relevant sensory input, to facilitate the discrimination between signal and ‘noise’ or distractors, and to bias the subject toward particular locations in which signals may appear.“ (Sarter et al. 2001, S.148)

Durch den ständigen Abgleich mit Erfahrungswerten entwickelt der Wahrnehmungsapparat bestimmte Erwartungshaltungen. Diese Erwartungen bilden entsprechend ein *Realitätskriterium*, das an alle eingehenden Reize herangetragen wird. Stimulipaarungen, wie etwa das Rauschen des Meeres beim Anblick eines Strandes, können so beispielsweise als adäquat erkannt werden. Der Organismus erwartet diese Paarung vielmehr schon bevor sie sinnlich wahrgenommen wird. Im Rahmen der ART nimmt Grossberg an, dass nur solche Umgebungsreize auch Aufmerksamkeit erhalten, die dieses Realitätskriterium erfüllen. Dem der ART zugrundeliegenden Wirkungszusammenhang zufolge, gelangen die Informationen im Kurzzeitgedächtnis nur dann ins Langzeitgedächtnis, wenn sie den *vorhergesagten Realitäten des Langzeitgedächtnisses* entsprechen (Grossberg 1995). Welche Vorhersagen vom Langzeitgedächtnis getroffen werden, hängt allein von der abstrakten Wissensbasis des Individuums ab, welches sich über die Zeit verändert. In aller Regel wird diese langfristige Anpassung als Lernprozess verstanden.

Schenkt man den Ausführungen von Eviatar Zerubavel zu „Social Mindscapes“ und „Thought Communities“ Glauben, durchdringen diese Erwartungen unsere Wahrnehmung so weit, dass inkonsistente Reize nicht nur ausgeblendet, sondern in Extremfällen sogar angepasst werden können (1999). Abweichende Stimulipaarungen¹¹ können so beispielsweise in Kongruenz gebracht werden, auch wenn die resultierende Umgebungsrepräsentation anschließend nicht mehr der Umgebungsrealität entspricht. Natürlich darf nicht vergessen werden, dass die Wahrnehmungswirkung der Top-down-Verarbeitung durch die bewusste Konzentration auf bestimmte Umgebungsreize durchbrochen werden kann. Gleiches gilt für das Realitätskriterium der ART (Cohen 2000, Grossberg 1995).

Der Wirkweise von Top-down steht die Bottom-up-Verarbeitung entgegen, die Wahrnehmungsreize auf dem Hintergrund ihrer inhärenten Charakteristika bewertet:

„Such a ‘top-down’ biasing of attentional performance contrasts with ‘bottom-up’ perspectives that describe attentional functions as driven mainly by the characteristics of the target stimulus and its sensory context. ‘Bottom-up’ perspectives attempt to explain a subject’s ability to detect targets and target-triggered attentional processing largely by

¹¹ In der Regel handelt es sich natürlich um ungewöhnliche und deshalb nur scheinbar abweichende Stimulipaarungen. Wird beispielsweise das Bild des Meeres mit dem Klang einer stark befahrenen Autobahn kombiniert, mag dies nicht erwartungsgemäß sein; es muss sich deshalb aber nicht zwangsläufig um einen Wahrnehmungsfehler oder eine Illusion handeln, die der Korrektur durch das Top-down-Verfahren bedarf.

the sensory salience of the targets, and their ability to trigger attentional processing by recruiting ‘higher’ cortical areas in a bottom-up manner.“ (Sarter et al. 2001, S.148)

Weisen Stimuli bestimmte Eigenschaften (i.a. Lautstärke, Informationsdichte) auf, kann ihnen auch ohne Erfahrungswertabgleich Aufmerksamkeit zugeteilt werden. Die beiden Wirkungsmechanismen können, abhängig vom Kontext, in Einklang oder Widerspruch stehen, in jedem Fall ergänzen sie aber einander, um eine maximal effiziente Aufmerksamkeitsallokation zu erreichen (Sarter et al. 2001). Eine vereinfachte Modelldarstellung dieses bidirektionalen Wirkungszusammenhangs findet sich in Abbildung 4.

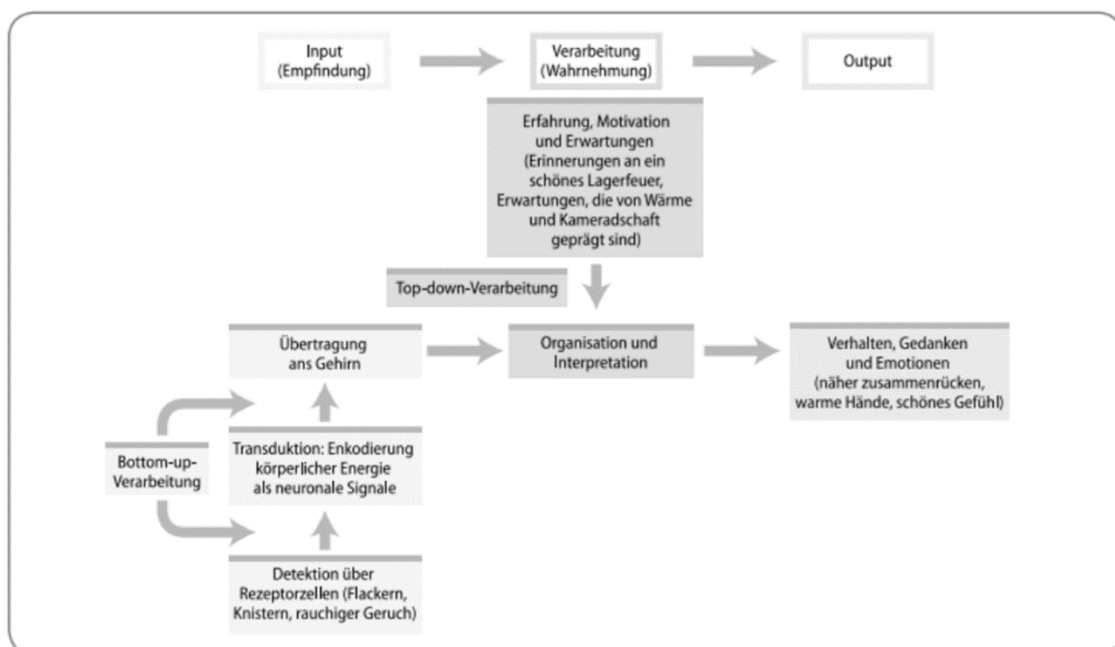


Abbildung 4: Top-down und Bottom-Up (Myers 2008, S.215)

Die von Grossberg vorgestellte Theorie (ART) stellt speziell auf den Einfluss der abstrakten Wissensbasis (Top-down) ab. Durch den stetigen Konsistenzabgleich mit ihr werden die Ergebnisse der Bottom-up-Verarbeitung reevaluiert.

Auf dem Hintergrund der auditorischen Wahrnehmung müssen für die genauere Betrachtung zwei Komponenten akustischer Signale unterschieden werden: die strukturelle oder akustische Komponente (Eigenschaften und Art) und die affektive Komponente (Bedeutung) (Cohen 2000). Ein klassisches Musikstück setzt sich insofern aus seiner strukturellen Komponente, i.e. einer komplexen melodischen Komposition verschiedener Klangströme die durch ein Orchester oder einzelne Instrumente erzeugt werden, und seiner affektiven Komponente, i.e. der spannende oder euphorische Impetus des Stückes, zusammen. In ihrer Ausarbeitung mit dem Titel „Film Music – Perspectives from Cognitive Psychology“ wählt Annabel J. Cohen (2000) das Beispiel einer Filmszene, in der ein solches, klassisches Musikstück einen Ausschnitt eines Baseballspiels untermalt. Eine Paarung der strukturellen Komponente mit dem visuellen Kontext entspricht nicht der Erwartungshaltung des Langzeitgedächtnisses; in der Realität werden Base-

ballspiele selten durch ein klassisches Orchester begleitet. Die affektive Komponente allerdings (e.g. Spannung), deckt sich mit dem emotionalen Kontext des Baseballspiels in der Filmszene. Während deshalb der Affekt des akustischen Signals wahrgenommen wird, wird die strukturelle Komponente, ohne bewusste Evokation, ausgeblendet. Werden Zuschauer nach dem Film befragt, welche Musik diese Szene unterstützt hat, können sie in der Regel zwar Aussagen über die emotionale Situation, nicht aber über die Art oder die Eigenschaften der Filmmusik machen (Cohen 2000).¹²

Im Rahmen von Grossbergs ART kann argumentiert werden, dass die inkonsistente Stimulipaarung nicht das Realitätskriterium erfüllt und deshalb nicht bewusst wahrgenommen wird. Sie nimmt eine Funktion ein, die der Schriftart dieser Arbeit gleicht – sie bildet die *Struktur des Trägermediums*, wird aber nicht bewusst wahrgenommen, solange die Aufmerksamkeit des Lesers nicht mit Intention auf sie gelenkt wird.

3.4.3 Congruence-Associationist Framework (CAF)

Auf Basis dieser formalen Grundlagen entwickelt Cohen für den Kontext der Filmmusik ihre Theorie des „Congruence-Associationist Framework“ (CAF), mit der sie einen *wahrnehmungstheoretischen Rahmen für die Domäne synthetischer Kontexte und akustischer Stimuli* legt. Wurde ein akustisches Signal als relevant eingeschätzt und der Wahrnehmung zugeführt, kann aus diesem Kontext ein Interpretationsrahmen für alle weiteren, sinnlichen Einflüsse entstehen – eine Fähigkeit, die in dieser Form nur der Audition zugesprochen werden kann.

„For example, Michel Chion emphasizes that what we see is always altered by what we hear. He refers to the “added value” of sound that “engages the very structuring of visions ... by rigorously framing it.”“ (Cohen 2000, S.360)

Diese Einschätzung lässt sich ad hoc auf dem Hintergrund der Filmmusik validieren. Viele Studien konnten nachweisen, dass subtile Spannungen im Erzählkontext eines Filmes ohne die entsprechende, melodische Unterstützung häufig gar nicht erst erzeugt, jedenfalls aber durch den intendierten Einsatz harmonischer Musik vollständig entspannt werden können. Wird eine fröhliche Sommermelodie über die Szenen des Kinoklassikers „Der weiße Hai“ gelegt, in denen der Hai selbst nicht zu sehen ist, entsteht schnell der Eindruck einer belebenden Urlaubsgeschichte an einem Strand in Neuengland.

Wie weit die *Dominanz des akustischen Interpretationsrahmens* reicht, hängt von der Bedeutungsschwere anderer kontextueller Einflussgrößen ab. Während ambigüe Informationen durch akustische Manipulation dominiert werden können, lassen sich eindeutige Szenen nicht durch bloßen Klang relativieren. Cohen selbst nutzt hier das Beispiel einer Gewaltszene, die auch durch beruhigende Musik nicht in einen harmonischen Kontext überführt werden kann (Cohen 2000).

¹² Überschreiten die Signalintensität (Lautstärke) oder Informationsdichte (Konzentration verschiedener Klangströme) beispielsweise ein bestimmtes Niveau, kann die Bottom-up-Priorisierung auch einem kontextlosen, akustischen Stimulus priorisiert Aufmerksamkeit gewähren.

„Thus, all [...] studies reviewed above found that music helps define the meaning of a scene and that the effect of music is most pronounced when the situation depicted is ambiguous.“ (Cohen 2000, S.364)

Sind (synthetische) Hörereignisse kongruent mit den Assoziationen des Langzeitgedächtnisses erfüllt der akustische Interpretationsrahmen in Cohen's CAF zusammengefasst drei Aufgaben:

- ***Suspension of Disbelief:*** Durch eine passende Paarung mit einem akustischen Stimulus (strukturell oder affektiv) kann das Realitätskriterium für einen Umgebungsreiz leichter erfüllt werden (ART).
- ***Meaning:*** Umgebungsreize erhalten innerhalb ihres akustischen Kontextes eine Bedeutung. Je ambiger die Situation, desto größer die Dominanz des akustischen Kontextes.
- ***Memory:*** Erinnerungen an diese Umgebungsreize und ihre Bedeutung werden in Assoziation zum entsprechenden akustischen Kontext abgespeichert und können anschließend sowohl durch das Erinnern an den Reiz selbst, als auch an den akustischen Kontext abgerufen werden. In der Regel wird durch diesen Zusammenhang gleichzeitig die Gedächtnisleistung gefördert.

Gerade auf dem Hintergrund synthetischer Wahrnehmungskontexte, i.a. Film, Augmented Reality, Virtual Reality, Digital Displays, wird die Bedeutung der wahrnehmungstheoretischen Eigenschaften akustischer Reize offensichtlich. Bei der Gestaltung solcher Wahrnehmungskontexte muss deshalb zum einen auf die Modularität der Wahrnehmung geachtet werden, i.e. physikalistische Wahrnehmungskonflikte¹³ einzelner Sinneskanäle müssen vermieden werden. Zum anderen muss dem mit der ART etablierten Realitätskriterium der menschlichen Wahrnehmung Rechnung getragen werden, i.e. Reize erhalten nur dann entsprechende Aufmerksamkeit wenn sie einem Abgleich mit der abstrakten Wissensbasis des Adressaten standhalten. Sind diese Bedingungen erfüllt, kann ein akustischer Interpretationsrahmen geschaffen werden, durch den nachfolgende Impulse bewusst manipuliert oder in einen intendierten Kontext gerückt werden können (Cohen 2000, Truax 2000, et al.).

Diese Wahrnehmungsstrukturen stellen eine angemessene Verarbeitung akustischer Stimuli seitens des auditorischen Wahrnehmungsapparates (Empfänger) sicher. Um eine Aussage hinsichtlich der Eignung akustischer Signale für eine Interaktion mit einem menschlichen Benutzer abzuleiten, müssen zusätzlich die Qualitäten akustischer Signale als Trägermedium geprüft werden. Das folgende Kapitel widmet sich deshalb der Untersuchung des Informationsgehaltes von Klangereignissen.

¹³ Als physikalistische Wahrnehmungskonflikte werden hier Konflikte der Ergebnisse einzelner Sinneskanäle auf elementarster Ebene verstanden, wie sie in Kapitel 3.4.1 beschrieben wurden.

3.5 Informationeller Gehalt akustischer Ereignisse

Obwohl seine Leistungen in einer visuell dominierten Welt nur selten Anerkennung finden, sind die Fähigkeiten des akustischen Wahrnehmungsapparates beachtlich. In ihrem Artikel „Psychology of auditory perception“ beschreiben Andrew Lotto und Lori Holt (2010) anekdotisch das Potential der menschlichen Audition:

„Imagine sitting on a pier with a friend over a swimming beach where the swimmers are out of your direct view. Your friend offers you this challenge: from the movement of the water waves that you see below, can you tell how many swimmers there are? Where are these swimmers relative to each other? And what kind of stroke is each one doing? The offer of such a seemingly impossible challenge would probably lead you to question whether you needed a better class of friends. However, our auditory system performs similarly improbable feats every day.” (Lotto & Holt 2010, S.1)

Damit akustische Signale als Trägermedium für Displays oder Benutzeroberflächen in Frage kommen können, müssen sie in der Lage sein, eine hohe Informationsdichte zu transportieren. Dass akustische Signale dieses Kriterium erfüllen, stellen Lotto & Holt anhand der Beispiele der Segmentation, Lokalisation und Kategorisierung heraus.

Spätestens seit E. Colin Cherrys Betrachtung des „Cocktail Party Problem“¹⁴ (1953), beschäftigt sich die Forschung mit der systematischen *Segmentation* kontinuierlicher, simultaner Geräuschquellen, im Speziellen auf dem Hintergrund der menschlichen Sprache. Nachdem ein akustisches Signal das menschliche Innenohr erreicht hat, wird es, wie in Kapitel 3.2.1 beschrieben, in seine Frequenzbereiche zerlegt. Der auditorische Wahrnehmungsapparat muss nun einerseits unabhängige Komponenten voneinander trennen und andererseits zusammengehörige Komponenten erkennen und integrieren. Albert S. Bregman bezeichnet diesen Vorgang als Auditory Scene Analysis (ASA)¹⁵: Akustische Komponenten derselben Quellen verhalten sich tendenziell gleichmäßig relativ zur Zeit, weisen eine harmonische Beziehung auf (Frequenzen A ist ein ganzzahliges Vielfaches von Frequenz B), beginnen und enden zusammen und weisen keine Diskontinuitäten auf. Bregman fasst dieses heuristische Konzept in den Prinzipien der *similarity*, *harmonicity*, *contemporaneity* und *good continuation* zusammen (1953). Unsere Fähigkeit die Stimme unseres Gesprächspartners von dem Hintergrund der Geräuschkulisse einer Cocktailparty zu lösen, rührt aus der Befriedigung eines Großteils dieser Prinzipien her. Der grundlegende Teil der dafür notwendigen Informationen wird durch das akustische Signal selbst bereitgestellt. Darüber hinaus ist eine erfolgreiche auditorisch-szenische Analyse aber auch vom akustischen Kontext, sowie von der Aufmerksamkeit des Zuhörers und seiner abstrakten Wissensbasis abhängig. Bregman weist außerdem darauf hin, dass die auditorische Analyse zur Orientierung

¹⁴ Allgemeiner als „Cocktail Party Effect“ bekannt, beschreibt dieser Begriff das Phänomen der auditorischen Isolation einer Geräuschquelle innerhalb einer komplexen akustischen Szene, die es Gesprächspartnern auf einer Cocktailparty ermöglicht, einander zu verstehen.

¹⁵ Siehe auch Kapitel 3.3.

schemenhaft auf bereits erfolgreich organisierten Klangmustern aufbaut. Als Folge ist der akustische Wahrnehmungsapparat deshalb beispielsweise in der Lage, eine bereits erlernte Zielmelodie leichter von ihrem akustischen Hintergrund zu trennen als unbekannte Klangströme (Lotto & Holt 2010).

Die *inhärenten Informationen auditorischer Signale* ermöglichen neben der Isolation und Integration auch eine Lokalisation akustischer Ereignisse. In Umgebungen, die eine hohe akustische Komplexität oder eine schwache Akustik aufweisen, e.g. stark verwinkelte Räume, werden Zielobjekte im Normalfall visuell lokalisiert. Obwohl das Auge prädestiniert für das Orten von Objekten im Raum scheint, verfügt der auditorische Wahrnehmungsapparat über eine wesentlich höhere, temporale Umgebungsauflösung. Wegen seiner omnidirektionalen Ausrichtung ist das Gehör außerdem nicht auf einen Umgebungsausschnitt beschränkt und erfasst so simultan alle akustischen Ereignisse in seiner Umwelt unabhängig von deren zeitlicher Begrenztheit. Ist die visuelle Wahrnehmung kompromittiert, e.g. bei Nacht oder schlechter Sicht, steigt der Einfluss der Akustik auf den Vorgang der *Lokalisierung*.

Zur Bestimmung der räumlichen Position einer Klangquelle werden hauptsächlich zwei integrale Bestandteile jedes akustischen Ereignisses herangezogen: die Differenz mit der der Klang beide Ohren erreicht (Interaural Time Difference (ITD) oder Interaural Phase Difference (IPD)) und die Differenz der Klangintensität mit der beide Ohren erreicht werden (Interaural Level Difference (ILD)). Ein relativ zum Empfänger rechtsseitiger Klang wird das rechte Ohr früher erreichen, als das linke Ohr. Gleichzeitig wird dieser Klang wegen des Impulsverlustes durch die höhere Distanz (i.e. Abstandsgesetz oder Entfernungsgesetz) und dem Klangschatten des Kopfes das linke Ohr mit geringerer Intensität erreichen. Eine bildliche Darstellung findet sich in Abbildung 5.

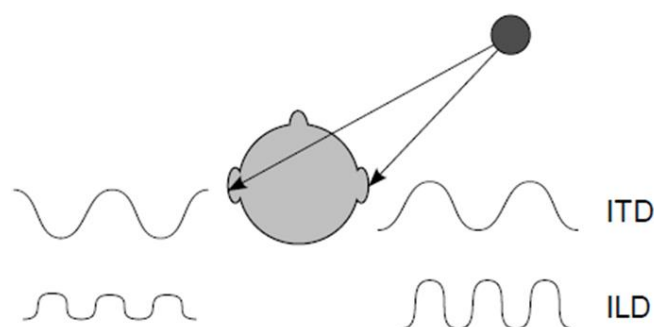


Abbildung 5: Lokalisierung durch ITD und ILD (Stewart 2010, S.24)

In Alltagssituationen erfolgt eine akustische Ortung primär auf Basis interauraler Phasendifferenzen, speziell weil Phasendifferenzen robust im Hinblick auf Umgebungsgeräusche sind. Räume oder Umgebungen mit sehr hohem Lärmpegel erschweren hingegen beispielsweise die

Ortung mittels ILDs¹⁶. Durch ihre Ausrichtung im dreidimensionalen Raum können die einzelnen Klangquellen sowohl vertikal (bzw. frontal), als auch horizontal verschoben sein (siehe Abbildung 6). Verschiebungen in der frontalen Ebene (e.g. Elevation) können durch dieses Verfahren allerdings prinzipiell nicht unterschieden werden, da weder Laufzeit- noch Intensitätsunterschiede zwischen den Ohren entstehen (Lotto & Holt 2010).

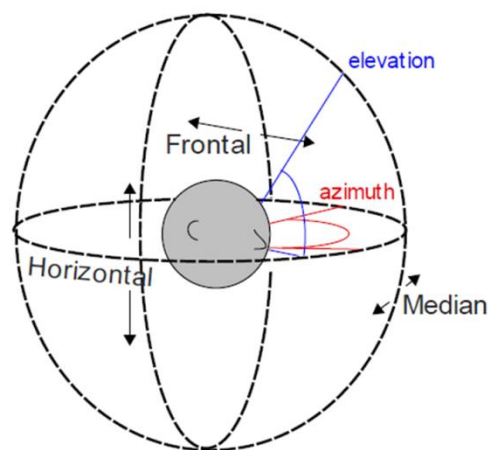


Abbildung 6: 3D-Modell (Stewart 2010, S.22)

Dennoch ist die Ortung in der horizontalen Ebene grundsätzlich möglich. Hierzu dient das individuelle Verhältnis von Kopf, Außenohr (Pinna) und Rumpf des Empfängers – die *Head-Related Transfer Function (HRTF)*.¹⁷ Die HRTF wertet die Amplitudenunterschiede eingehender Schallereignisse aus. Diese nehmen mit der Frequenz zu, weil Beugungseffekte an Kopf und Rumpf mit kürzerer Wellenlänge stärker werden. Winkelabhängige Resonanzerscheinungen an

¹⁶ An dieser Stelle scheint erwähnenswert, dass das klassische Stereo-Panning, das heute noch den größten Teil aller akustischen Simulationen ausmacht, dennoch vollständig ohne ITDs/IPDs auskommt und eine subjektive Ortsverschiebung einer Klangquelle lediglich durch Lautstärkenunterschiede erzeugt. Obwohl die Position der Klangquelle leicht aufgefächert anmutet, kann die virtuelle Position der Klangquelle, gerade bei dynamischen bzw. bewegten Klangbildern in der Regel sicher und genau bestimmt werden. Das Hörereignis, das sich bei einer echten Klangquelle oder einer virtuellen Verortung mittels Wellenfeldsynthese einstellt, ist dennoch ein anderes. Das Geräusch scheint subjektiv „direkter“ oder „unmittelbarer“ von einer bestimmten Position auszugehen, was auch erklärt, weshalb moderne Surround-Anlage häufig über einen zentralen Lautsprecher verfügen, der beispielsweise Sprecherstimmen wiedergibt. Der auditorische Wahrnehmungsapparat scheint insofern alle zur Verfügung stehenden Informationen heranzuziehen, um anschließend eine maximal kohärente, spatiale Repräsentation der Klangumgebung zu generieren. Obwohl der fehlende Laufzeitunterschied suggerieren müsste, dass sich die Quelle beispielsweise unmöglich links vom Zuhörer befinden kann, wird auf Basis des Intensitätsunterschiedes auf eine entsprechende Anordnung geschlossen.

¹⁷ Auch im Deutschen wird in der Regel der englische Begriff der HRTF herangezogen. Gelegentlich wird diese Funktion aber auch als „kopfbezogene Übertragungsfunktion“ oder „Außenohrübertragungsfunktion“ bezeichnet.

der Pinna ermöglichen auch bei weiter steigenden Frequenzen die Lokalisation in der Elevationssebene (vertikale Ebene).¹⁸

Die theoretische Grundlage der Wahrnehmung einzelner Schallereignisse mit geringer Komplexität erscheint heute recht fundiert. In welchem wahrnehmungstheoretischen Verhältnis allerdings die akustische Lokalisation und Segmentation stehen oder welche zusätzlichen Mechanismen bei der Lokalisierung komplexer Klangströme einsetzen, wirft für die Forschung weiterhin Fragen auf:

„While our understanding of localization of single sound sources (especially for simple sounds such as tones) is fairly well developed, there is still much to be learned about how spatial location is determined for multiple simultaneous sounds. One chicken-or-egg question is whether we segregate sound events prior to localization or whether we use spatial location cues to segregate sound events.“ (Lotto & Holt 2010, S.7)

Als Produkt seiner realen Umgebung transportiert ein akustisches Ereignis außerdem immer auch Hinweise über die physikalischen Charakteristika und Eigenschaften seines Ursprungs. Diese Informationen werden vom menschlichen Gehör zum Zwecke der *Kategorisierung* herangezogen. Werden beispielsweise zwei Klangstäbe aneinander geschlagen, ist ein Zuhörer für gewöhnlich in der Lage das Material, die Größe, die Form etc. der Klangstäbe kategorisch zu bestimmen. Typischerweise lässt die Genauigkeit dieser Bestimmung allerdings Raum für Verbesserungen (Lotto & Holt 2010, Raffaseder 2003 et al.). Während eine konkrete *Identifikation* des Ursprungs ohne spezifische Vorkenntnisse meist nicht möglich ist, sind Zuhörer in der Regel fähig, robuste Kategorisierungen vorzunehmen. Diese Einschätzung deckt sich mit der *Kernthese der kategorialen Wahrnehmungstheorie* (Massaro 1987, Schouten et al. 2003 et al.) und ist gleichzeitig der Grund, weshalb es dem auditorischen Wahrnehmungsapparat so leicht fällt, beispielsweise verbale von non-verbale Stimuli zu unterscheiden¹⁹. Welche Eigenschaftsdimensionen für die Zuordnung in eine bestimmte Kategorie entscheidend sind, konnte bis jetzt nicht abschließend ermittelt werden. Fest steht aber, dass häufig speziell die Tonhöhe bzw. Lage eines akustischen Ereignisses für seine kategorische Einordnung relevant ist. Grundsätzlich erfolgt zum Zweck der Kategorisierung aber eine Gewichtung der einzelnen Eigenschaftsparameter (e.g. Tonhöhe, Tonfarbe). Das Gewicht eines Parameters verringert sich dabei prinzipiell mit steigender Varianz, i.e. mit wachsender Unregelmäßigkeit konzentriert sich das Gehör zunehmende auf andere

¹⁸ Diese Beschreibung des Vorganges der akustischen Lokalisation geht von singulären Schallereignissen aus. Handelt es sich um Klangströme oder wiederholte Schallereignisse einer gleichbleibenden Quelle, können Bewegungen des Kopfes beispielsweise zum Zwecke der akustischen Triangulation genutzt werden. Da sich dieses Kapitel grundsätzlich mit der Informationsdichte akustischer Signale als Hinleitung zur Thematik auditorischer Displays und Benutzeroberflächen beschäftigt, wird auf eine weiterführende Erklärung an dieser Stelle verzichtet.

¹⁹ Die kategoriale Wahrnehmung kann heute bis hin zur Aktivität unterschiedlicher Hirnareale bei der Wahrnehmung kategorisch unterschiedlicher Informationen zurückverfolgt werden. Als verbal erkannte, akustische Stimuli werden so, anders als non-verbale Stimuli, beispielsweise dem „Wernicke Areal“ zugeführt. Diese Kategorisierung und anschließende, angepasste Verarbeitung von Informationen entscheidet sich offensichtlich auf dem Hintergrund ihrer Funktion (Dingwall 1993 et al.).

akustische Hinweise, die aus einem Schallereignis für eine Kategorisierung herangezogen werden können (Lotto & Holt 2010). Wie aus Kapitel 3.1 hervorgeht, verfügen auditorische Reize über eine ganze Palette von Eigenschaften, die sich für ihre Einordnung eignen.

Die von Lotto und Holt ins Feld geführten Beispiele der Segmentation, Lokalisation und Kategorisierung verdeutlichen die Eignung akustischer Signale als Träger komplexer Information. Gleichzeitig ist das menschliche Gehör in der Lage, den inhärenten Informationsgehalt eines Schallereignisses in Echtzeit zu extrahieren und auszuwerten. Diesen Umstand machen sich auditorische Displays oder Benutzeroberfläche zunutze und kommunizieren mit dem Anwender durch den gezielten Einsatz akustischer Stimuli.

3.6 Auditorische Displays und Benutzeroberflächen

Bruce N. Walker und Gregory Kramer definieren auditorische Displays in ihrer grundlegenden Arbeit „Ecological Psychoacoustics and Auditory Displays: Hearing, Grouping, and Meaning Making“ eher allgemein:

„Auditory display is a broad term referring to the use of any type of sound to present information to a listener. This may include, but is certainly not limited to, warnings, alarms, status indicators, and data sonification.“ (Walker & Kramer 2004, S.3)

Enger gefasst, bezeichnet der Begriff auditorischer Displays die Informationsübermittlung mittels non-verbaler Signale (Walker & Brewtser 2001, Walker 2002, Walker & Kramer 2004 et al.). Für alle folgenden Untersuchungen wird der weiter gefasste Begriff als Referenz dienen.

Wird das *Kriterium des intendierten Präsentierens* in Walker und Kramers Definition vernachlässigt, wird deutlich, dass bereits die alltägliche Umwelt über eine Vielzahl auditorischer Displays verfügt:

Weltweit setzt sich heute bei fast jedem Automobil das Zentrum der Fahrerkonsole aus einem Geschwindigkeitsmesser und einem Drehzahlmesser zusammen, mit dessen Hilfe die Umdrehungen pro Minute angezeigt werden. Fahranfänger werden darauf verwiesen, diesem Drehzahlmesser entsprechend ihr Schaltverhalten auszurichten. Erreicht ein durchschnittliches Dieselfahrzeug etwa 2000 Umdrehungen pro Minute, kann der Gang nach oben gewechselt werden. Fahranfänger orientieren sich kontinuierlich an diesem visuellen Display, müssen ihm dafür allerdings einen beträchtlichen Teil ihrer kognitiven Aufmerksamkeit widmen. Fortgeschrittene Fahrer gehen deshalb in der Regel intuitiv dazu über, die Umdrehung approximativ am Motorengeräusch des Fahrzeugs zu erkennen (auditorisches Display). Dadurch entsteht eine scheinbar natürlichere und beiläufige Interaktion mit dem Fahrzeug, welche dem Fahrer ermöglicht, seine visuelle Aufmerksamkeit anderen Reizen zu widmen. So kann zum einen die Fahrsicherheit enorm erhöht werden, andererseits steigert sich spürbar der Bedienkomfort für den Fahrer, der nicht mehr für jedes Schaltmanöver bewusst den Drehzahlmesser beobachten muss. Dieses Beispiel verdeutlicht, dass durchaus Kontexte existieren, in denen (sogar unintendierte) auditori-

sche Displays eine wahrnehmungstheoretisch effizientere Information des Benutzers ermöglichen, als andere Displayklassen.

Obwohl die reale Umwelt häufig eine Vielzahl akustischer Informationsplattformen bietet, konzentriert sich das Wissenschaftsfeld auditorischer Displays auf die intendierte Informationspräsentation im Allgemeinen und die Information mittels digitaler Technologie im Speziellen (Walker & Brewtser 2001, Walker 2002, Walker & Kramer 2004 et al.). Akustische Benutzeroberflächen (im Kontrast zu Displays) bieten neben der unidirektionalen Präsentation von Informationen zusätzlich Bedienmöglichkeiten, die eine Interaktion dem Nutzer und dem System zulassen. Während die Präsentation der resultierenden Informationen auf Basis akustischer Signale erfolgt, können die Bedienmodalitäten sowohl akustisch, als auch manuell etc. ausgestaltet sein.

3.6.1 Ausgestaltung und Terminologie

Die einfachste und bis heute noch immer häufigste Klasse auditorischer Displays wird durch *klassische Alarmer und akustische Notifikationen* konstituiert (Sorkin 1987, Walker & Kramer 2004). Hierbei werden, wie etwa beim Klingeln eines Telefons, auffällige akustische Ereignisse genutzt, um den Zuhörer darauf hinzuweisen, dass etwas geschehen ist, geschehen wird oder eine Aktion des Zuhörers erforderlich ist. Häufig werden diese akustischen Hinweise heute dafür genutzt, den Nutzer auf eine Informationspräsentation durch ein visuelles Display aufmerksam zu machen oder vorzubereiten. Während diese Struktur schon lange i.a. durch Wecker genutzt wird, wird im digitalen Zeitalter durch Notifikationen beispielsweise auf Nachrichten mobiler Applikationen auf dem persönlichen Smartphone hingewiesen.

Akustische Warnsignale bilden strenggenommen eine Unterkategorie dieser Klasse, die ausschließlich der Warnung dient und in aller Regel auf Signale zurückgreift, die entweder evolutionsbiologisch (e.g. bloße Lautstärke eines Warngeräusches) oder durch soziale Prägung (e.g. Wiedererkennung des Klanges einer Feuerwehrsirene) als Indikatoren für Gefahrenpotentiale erkannt werden (Walker & Kramer 2004). Alarmer und Notifikationen sind binäre auditorische Displays: ein Feueralarm zeigt insofern nur an, dass oder ob ein Brand vorliegt, nicht etwa wo oder wie gravierend dieser ist.

Auditory Icons bilden das akustische Pendant zu *Visual Icons* und bezeichnen die symbolische Repräsentation von Objekten oder Prozessen. Heute werden Icons genutzt, um die Funktionen eines Computersystems, Systemobjekte, den Systemstatus, Befehle, den Nutzerfokus, Menüpunkte, Abbildungsfenster, Auswahlmöglichkeiten, Dienstleistungsprogramme, Prozesse, Programme und andere Informationen abzubilden (Uzilevsky & Andreev 1993). Icons werden eingesetzt um mittels der Wiedererkennung durch den Nutzer kompakt und vereinfacht Informationen darzustellen. Sie werden schneller durch den Nutzer verarbeitet und sind häufig in der Lage linguistische und kulturelle Barrieren zu umgehen (Walker & Kramer 2004). In der Domäne akustischer Signale wäre so als Icon für einen Druckbefehl beispielsweise ein prägnanter Ausschnitt charakteristischer Druckgeräusche anstelle eines Bildes vorstellbar. Eine direkte Verknüpfung von realen Klangeigenschaften eines Objektes und dem entsprechenden auditorischen Icon ist nicht zwingend notwendig.

Mit wachsendem Abstraktionsgrad, speziell aber für die symbolische Repräsentation von Objekten oder Prozessen die nicht im engeren Sinne über ein eigenes Klangbild in der Realität verfügen, werden sogenannte *Earcons*²⁰ eingesetzt. Da Earcons ohnehin synthetisch generiert werden müssen, werden diese häufig in hierarchischer Struktur integriert, um eine kontrollierte Informationsübermittlung zu gewährleisten. Earcons haben häufig melodischen Charakter und versuchen oft eine metaphorische Assoziation zu bedienen. Dennoch sind sie deutlich beliebiger als Auditory Icons und müssen deshalb vom Nutzer zunächst gewissermaßen erlernt werden. Ein bestimmter Ordner könnte so durch eine bestimmte Melodie repräsentiert werden. Wird der Ordner gelöscht, könnte dies beispielsweise durch eine dreistufige Reduzierung der Wiedergabelautstärke bis zum Erlöschen des Klanges symbolisiert werden (Walker & Kramer 2004, Vazquez-Alvarez et al. 2011).

Während sich auditorische Icons und Earcons auf eine symbolische oder metaphorische Repräsentation stützen, konzentriert sich die *Audification* auf die direkte Übersetzung von Daten in hörbare Schallwellen (Walker & Kramer 2004). Um die relevanten Daten „hörbar“ zu machen ist häufig eine Frequenzverschiebung in den Raum der menschlichen Hörfläche²¹ notwendig. Das wohl eingängigste Beispiel ist die Audification seismischer Datensätze. Die Wellenstruktur der zugrundeliegenden Seismogramme konnte in einem prototypischen Projekt so akustisch wiedergegeben werden, dass eine abschließende Analyse der Daten allein auf Basis der auditorischen Repräsentation möglich war. Beeindruckend ist dabei aber vor allem, dass durch diese Darstellung eine zuvor nicht zu bewältigende Datenmenge offenbar wahrnehmungspraktisch vereinfacht und anschließend für die Forscher transparent gemacht werden konnte (Speeth 1961, Walker & Kramer 2004 et al.).

Die heute wohl fortschrittlichsten auditorischen Displays allerdings, verzichten schlicht auf eine Einschränkung der Ausgestaltungsform und machen sich im Rahmen der *Sonification* kontextsensitiv alle der vorgestellten Konzepte zunutze:

„Sonification is the use of non-speech audio to convey information such as that used in the interpretation of scientific results. Specifically, data sonification is “the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation”. That is, scientific data, of any sort, is used to change the parameters of a synthesized tone, usually in a largely metaphorical or symbolic manner.“ (Walker & Kramer 2004, S.6)

Dieses Verfahren verdeutlicht sich gut am Beispiel des Geigerzählers, dessen Klickrate die Höhe der Strahlungswerte einer Umgebung repräsentiert und gleichzeitig eine Warnfunktion erfüllt. Häufig ist es hilfreich sich Sonification als die Darstellung von Informationen durch akustische

²⁰ Auch dieser Terminus bildet ein akustisches Pendant zu dem aus der visuellen Repräsentation stammenden Begriff des Icons, der Aussprache nach auch als „eye-con“ lesbar. Der Begriff des „ear-con“ wurde 1985 von Denise A. Sumikawa geprägt. Gelegentlich findet sich außerdem der Begriff des „spear-con“ (speech based earcon), der speziell auf den Einsatz verbaler Signale abstellt.

²¹ Siehe Kapitel 3.1.

Graphen vorzustellen. Dabei darf allerdings nicht vergessen werden, dass in diesem Zusammenhang akustische Mittel für die Übersetzung originär visueller (i.e. graphischer) Datenrepräsentationen eingesetzt werden, um diese auditorisch zugänglich zu machen. Die Wertepaarabbildung in einem zweidimensionalen Koordinatenraum beispielsweise, wie sie für die Integration von Daten in einem Graph eingesetzt wird, muss aus historischer Perspektive als Vehikel der visuellen Informationsübermittlung verstanden werden - ein akustischer Graph ist insofern immer ein Transfer einer visuellen Repräsentationsform in die Sphäre der Audition. Dieser Umstand erklärt auch, weshalb sich einige graphische Darstellungen zwar angemessen für die Audition aufbereiten lassen, andere aber nicht. Abbildung 7 fungiert als eingängige Illustration dieser Übersetzungsproblematik. Während die erste Zeile der Abbildung einen Fall darstellt, in dem durch eine geschickte binaurale Wiedergabe zweier verschiedener Klangströme eine verständliche Repräsentation möglich ist (in diesem Beispiel wird durch zwei verschiedene Tonlagen eine schiefe Ebene beschrieben), zeigt die zweite Zeile einen Fall, in dem eine solche Repräsentation unter Bemühung desselben Ansatzes ausgeschlossen ist.

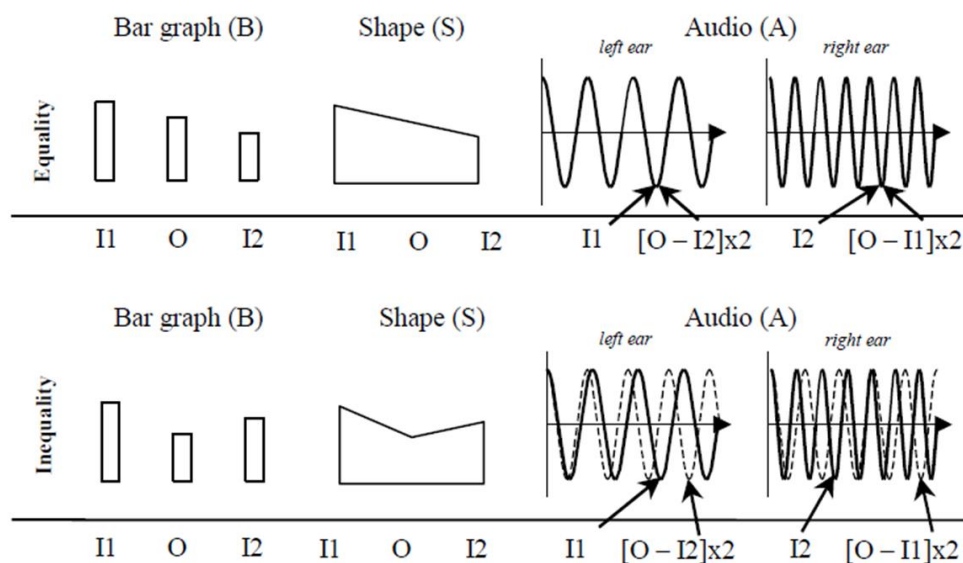


Abbildung 7: Akustische Datenrepräsentation (Sanderson et al. 2000, S.264)

Die Erforschung originär akustischer Repräsentationen könnte in Zukunft dabei helfen, die Leistungsfähigkeit auditorischer Displays und Benutzeroberflächen deutlich zu verbessern.

Natürlich verfolgt dieser Überblick keinen Anspruch auf Vollständigkeit. Stattdessen soll verdeutlicht werden, welche Dimensionen akustischer Signale schon heute erfolgreich zum Zwecke der Informationsübertragung eingesetzt werden.

3.6.2 Gestaltungskriterien auditorischer Displays (Subtasks)

Wie und ob ein Nutzer diese akustischen Informationen allerdings versteht, hängt sowohl von einer Reihe akustischer und wahrnehmungstheoretischer Parameter, als auch den Eigenschaften seiner Umgebung ab. Grundlegend für die Gestaltung eines auditorischen Displays sind deshalb nach Walker und Kramer (2004) drei Kernkriterien:

- **Perception (Wahrnehmung):** Dass präsentierte Inhalte durch den Adressaten überhaupt wahrgenommen werden, bildet die Grundbedingung jeder sinnvollen Ausgestaltung eines Displays. Theoretische Leitlinien für eine solche Grundlage lassen sich aus den vorangegangenen Kapiteln ableiten. Es ist entscheidend, dass der Zuhörer Veränderungen und Nuancen in den Klangsignalen erkennen kann, da über diese der Großteil der Informationen transportiert wird. Natürlich muss deshalb auch abgeschätzt werden, welche Umgebungsbedingungen für den Gebrauch zu erwarten sind (e.g. Lärmpegel der Umgebung).
- **Stream Analysis (Klangstromerkennung):** Es muss dem Nutzer möglich sein, Klangströme zu unterscheiden und relevante Klangströme zuzuordnen.²² Welcher Kanal welche Information transportiert, kann ansonsten nicht decodiert werden. Diese Überlegung ist vor allem dann wesentlich, wenn bekannt ist, ob sich der Nutzer in einer Umgebung mit vielen Geräuschquellen befindet oder Klangströme absichtlich spatial verortet werden, um Informationen einfacher zugänglich zu machen. Signaleigenschaften, die die Klangstromerkennung erleichtern, lassen sich aus der Wahrnehmungstheorie ableiten.
- **Conception and Meaning-making (Konzeption und Bedeutung):** Wurden die Signale des auditorischen Displays erfolgreich wahrgenommen und erkannt, kann der Nutzer ihre Bedeutung und Information auslesen. Zentral für die Bedeutung eines Signals ist seine Wahrnehmungswirkung. Während bestimmte Klangeigenschaft über fundamentale Effekte entscheiden, e.g. ein lauter, schriller Ton wird häufig schnell als Warnsignal erkannt, führen andere Eigenschaften zu komplexen Assoziationen, die, wie beschrieben, von der abstrakten Wissensstruktur des Langzeitgedächtnisses abhängig sind. Ob eine bestimmte Klangfarbe deshalb eher ein Gasleck als einen Brand suggeriert, muss fallspezifisch Betrachtung finden. Sicher ist aber, dass das Interpretationsorgan des menschlichen Wahrnehmungsapparates letztlich darüber entscheidet, was ein bestimmtes Signal für den Nutzer bedeutet.

3.6.3 Psychoakustische Dimensionen auditorischer Displays

Neben diesen allgemeinen Gestaltungskriterien lassen sich, anschließend an die wahrnehmungstheoretischen Überlegungen dieser Arbeit, verschiedene psychoakustischer Dimensionen auditorischer Displays identifizieren (Walker & Kramer 2004 et al.). Diese entscheiden über das indi-

²² Siehe Kapitel 3.3, relevant ist hier vor allem der Abschnitt zu auditorischer Gruppierung (Auditory Grouping).

viduelle Hörereignis eines Nutzers und verdeutlichen die praktische Relevanz der wahrnehmungstheoretischen Grundlagen, die in den vorangegangenen Kapiteln erarbeitet wurden:

- **Lautstärke:** Lautstärke gilt als primäre Klangeigenschaft akustischer Ereignisse. Bereits allein durch diese Klangeigenschaft können Warnungen von Notifikationen unterschieden und Lautstärkeniveaus für bestimmte Umgebungen ermittelt werden.
- **Verdeckung:** Wie laut ein Ton absolut sein muss, um gehört zu werden, ist in der Regel weniger relevant als die Fragen, wie laut ein Ton relativ zur Klangumgebung sein muss, um Aufmerksamkeit zu erhalten. Diese Frage kann überhaupt erst auf der Basis kognitiver Wahrnehmungsmodelle vollständig verstanden werden.
- **Tonlage/Tonhöhe:** Obwohl diese Eigenschaft üblicherweise der Frequenz zugeordnet wird, kann, wie bereits beschrieben, keine psychoakustische Äquivalenz abgeleitet werden. Da die Tonhöhe weniger von der Klangumgebung abhängig ist als beispielsweise die Lautstärke einer Klangquelle, wird dieser Parameter sehr häufig eingesetzt um verschiedene Dimensionen, Trends oder Intensitäten zu verdeutlichen. Die Hörfläche des menschlichen Gehörs dient zwar als sinnvolle Orientierung, in der Praxis wird allerdings fast ausschließlich der Frequenzbereich 200 – 5.000 Hz mit dem Sensitivitätsmaximum bei etwa 3.000 Hz für auditorische Benutzeroberflächen bemüht.
- **Wechselwirkungen:** Während einige Klangeigenschaften unabhängig betrachtet werden können, sind andere durch psychoakustische Wechselwirkungen miteinander verknüpft. Eine der Hörfläche angepasste Parameterpaarung von Lautstärke und Tonhöhe kann beispielsweise die Warnwirkung von Alarmsignalen verbessern oder die Genauigkeit bei der Erkennung von Tonhöhenveränderungen unterstützen.
- **Tempo und Rhythmus:** Das menschliche Gehör reagiert sehr sensibel auf Tempoveränderung und ist in der Lage aperiodische Elemente in einem Klangstrom zu erkennen oder periodische (rhythmische) Klangströme zu identifizieren. Die Sicherheit mit der das auditorische System Muster auch auf dem Hintergrund schwerer Störgeräusche erkennen kann, übersteigt die Fähigkeiten des visuellen Wahrnehmungsapparates bei Weitem. Dieser Umstand kann beispielsweise für die Erkennung von Mustern mittels Audification oder Sonification großer Datensätze ausgenutzt werden, die sich durch ein schlechtes Signal-Stör-Verhältnis auszeichnen (e.g. Seismogramme).
- **Klangfarbe/Timbre:** Die Klangfarbe fungiert gewissermaßen als Sammelbegriff für alle Klangeigenschaften, die nicht mit Tempo, Tonhöhe oder Lautstärke zusammenhängen. Trotz der wissenschaftlichen Ambiguität des Begriffes, ist die Klangfarbe grundlegend für die Zuordnung zwischen Daten und Signal. Erst sie macht beispielweise eine Klarinette als solche erkennbar und ermöglicht ein Wiedererkennen bestimmter Klänge, mit deren Hilfe Daten akustisch repräsentiert werden können
- **Klangströme:** Die Erkennung von Klangströmen kann durch das Design der akustischen Signale sowohl unterstützt, als auch erschwert werden. Klangströme werden in auditorischen Displays häufig zur Aufgliederung paralleler Informationskanäle oder als Referenzrahmen verwendet. Ein konstantes Piepen mit einer bestimmten Geschwin-

digkeit kann beispielsweise ein Maximum eines Datensatzes markieren. Je mehr sich das Tempo anderer Klangströme dieser Markierung nähert, desto näher liegt das jeweilige Datum am Maximum. Ein Klangstrom kann so als Referenz anderen Klangströmen Kontext geben und zum Beispiel als ein Pendant zu einer graphischen Achse eingesetzt werden. Gleichzeitig bietet die Aufgliederung paralleler Informationsströme die Möglichkeit, mehrdimensionale auditorische Displays zu erstellen.

- **Spatialität:** Die wahrgenommene räumliche Position einer Klangquelle wird, abgeleitet aus natürlichen Hörkontexten²³, durch den menschlichen Wahrnehmungsapparat als maßgeblich für ihre Bedeutung angesehen. Fortschrittlichere und effektivere Displays konnten durch die Einbeziehung der spatialen Dimension nicht nur die Klangstromerkennung verbessern, sondern durch den räumlichen Bezug auch eine vollständige, neue Referenzebene entwickeln. Die Thematik der Spatialität als wichtigste Kontextebene auditorischer Displays wird in Kapitel 4 genauere Betrachtung finden.

Jede dieser Dimensionen bietet hinsichtlich ihres praktischen Einsatzes Vor- und Nachteile. Die Klangeigenschaften der Lautstärke und der Tonlage sind nach wie vor die am häufigsten instrumentalisierten Dimensionen heutiger auditorischer Displays (Neuhoff 2011 et al.).

Die Tonhöhe eines akustischen Signals lässt sich leicht manipulieren und kann deshalb inhaltlich sehr direkt bestimmten Datenstrukturen zugeordnet werden. Außerdem ist das menschliche Gehör in der Lage schon Veränderungen von weniger als 1 Hz bei einer Frequenz von 100 Hz wahrzunehmen. Eine nuancierte Repräsentation innerhalb dieser Dimension ist deshalb durchaus möglich. Zusätzlich sind musikalisch talentiert oder geübte Menschen in der Lage sogar konkrete Niveauunterschiede in Form von Noten zu unterscheiden. Eine der häufigsten Repräsentationsmechanismen durch Manipulation der Tonlage, ist das Steigen und Fallen der Tonhöhe im Verhältnis zur Veränderung des Wertes einer zugeordneten Variable. Eine solche Repräsentationsform bezeichnet man als *Polarität* (Neuhoff 2011 et al.). Zusätzlich wird zwischen positiver und negativer Polarität unterschieden. So kann beispielweise bei der Sonification von Wetterinformationen eine steigende Temperatur durch eine steigende Tonhöhe in Form einer positiven Polarität abgebildet werden. In anderen Anwendungskontexten empfiehlt sich hingegen eine negative Polarität um einen inhaltlichen Zusammenhang zu verdeutlichen. Diese Repräsentation eignet sich beispielsweise für Lagerbestände, bei denen eine geringe Restmenge durch eine sehr hohe Tonlage verdeutlicht wird. Die Palette möglicher Repräsentationsformen ist letztlich nur durch das Kriterium der Nachvollziehbarkeit einerseits und der Kreativität des Audio-Designers andererseits eingeschränkt. Untersuchungen weisen allerdings darauf hin, dass Zuhörer durchaus unterschiedliche Fähigkeiten hinsichtlich der Tonhöhenenerkennung besitzen. Zusätzlich ist der umgebende Klangkontext bzw. Frequenzbereich der Klangumgebung von erheblicher Relevanz im Hinblick auf die subjektive Beurteilung einer bestimmten Tonlage.²⁴

Obwohl es sich bei der Lautstärke immer noch um eine der am häufigsten genutzten psychoakustischen Parameter auditorischer Displays handelt, ist ihre Eignung für die Datenrepräsentation

²³ „Natürlich“ wird hier im Kontrast zum synthetischen Kontext auditorischer Displays auf dem Hintergrund der speziellen Definition auditorischer Displays aus Kapitel 3.6 verwendet.

²⁴ Eine weiterführende Beschreibung findet sich in Neuhoff 2011, S.65.

tion eher limitiert. Dass diese Einflussgröße so häufig Anwendung findet liegt vor allem daran, dass auditorische Displays, wie eingangs beschrieben, häufig in Form von Alarmsignalen oder Notifikationen genutzt werden. Die Lautstärke eines akustischen Ereignisses fungiert auf natürliche Weise als *Mediator für Aufmerksamkeit*. Während laute Geräusche schon aus evolutionsbiologischen Gründen mehr Beachtung durch den menschlichen Wahrnehmungsapparat erhalten, finden Signale nahe der Wahrnehmungsschwelle zunächst nur wenig Aufmerksamkeit. Die Genauigkeit der Lautstärkenunterscheidung ist allerdings in keiner Weise mit der Auflösung zu vergleichen, die durch Manipulationen der Tonhöhe erreicht werden können. Gleichzeitig ist das menschliche Gehirn, im Kontrast zur Tonlage, kaum in der Lage eine konkrete Lautstärke zu speichern und später widerzuerkennen. Zusätzlich wird die Lautstärkenwahrnehmung unmittelbar durch den Schallpegel der Klangumgebung beeinflusst. Ein objektiv lautes Signal kann so in einer lauten Umgebung kaum mehr als ein solches erkannt werden. Mit steigender Komplexität der Ausgestaltung eines auditorischen Displays sinkt deshalb der Nutzen der Lautstärke als Repräsentationsdimension (Neuhoff 2011).

Neben dieser Auswahl psychoakustischer Dimensionen auditorischer Displays im Kontext dieser Arbeit kann, je nach Arbeitshypothese, fraglos eine Vielzahl anderer Dimensionen als relevant identifiziert werden (Walker & Kramer 2004).

3.6.4 Anwendung und Limitationen

Auditorische Displays funktionieren heute vor allem dann gut, wenn Trends, Wendepunkte (Minima und Maxima) oder Abweichungen von einer Basislinie erkannt werden müssen. Mithilfe einer sinnvollen Sonification kann beispielsweise die Entwicklung der Verkaufszahlen eines Einzelhandels über ein Jahr eingängig dargestellt werden (Proctor & Reeve 1990 et al.). Frequenz oder Klickzahl könnten in Verhältnis zu einer Grundlinie (e.g. Monatsziel) gebracht und parallel ausgegeben werden (e.g. ist die Frequenz des Datensatzes höher als die der Basislinie, wurde das Monatsziel überschritten). So entsteht eine akustische Relation der beiden Größen, die sich den meisten Nutzern intuitiv erschließt. Gleichzeitig können auditorische Displays Annäherungen an ein Maximum in bestimmten Anwendungsszenarien gleitend in ein Warnsignal umwandeln. Im Fall des Geigerzählers fungiert die Verdichtung der Klicks bei der Detektion hoher Strahlenwerte so parallel als Alarm zum Schutz des Nutzers.

Der auditorische Wahrnehmungsapparat erweist sich außerdem auch dann als besonders leistungsstark, wenn komplexe Muster²⁵ (e.g. Rhythmus) in Datensätzen erkannt werden müssen. Im Fall der Audification seismischer Aufzeichnungen war es den Forschern durch Mustererkennung möglich, Erdbeben von den Erschütterungen atomarer Explosionen mit einer Wahrscheinlichkeit von mehr als 90% korrekt zu unterscheiden, während eine Unterscheidung auf Basis graphischer Darstellungen wegen der großen Datenlast zuvor praktisch ausgeschlossen war (Speeth 1961). Die Sonification und Audification von Datensätzen bietet sich speziell dann an, wenn Datensätze besonders groß sind oder ein schlechtes Signal-Rausch-Verhältnis aufwei-

²⁵ In der Regel existieren konkrete Muster in Datensätzen natürlich nur näherungsweise. Die auditorische Erkennung ist insofern robust, als dass sie in der Lage ist unregelmäßige Abweichungen auszugleichen oder als solche zu identifizieren.

sen. Sollen solche Datensätze zunächst auf Relevanz oder Auffälligkeiten untersucht werden, sind akustische Repräsentationen ihren visuellen Pendanten häufig klar überlegen. Allein die Limitationen die durch die Eingeschränktheit des menschlichen Sichtfeldes entstehen, führen häufig dazu, dass bei großen Datenmengen ohnehin nicht alle Daten gleichzeitig betrachtet werden können. Frederick L. Scarf (1979) hat das Verfahren der Audification beispielsweise bei der Untersuchung der Messdaten herangezogen, die durch die Voyager-2-Raumsonde bei der Durchquerung der Ringe des Saturns gesammelt werden konnten. Graphisch konnte der gleiche Datensatz wegen des schlechten Signal-Rausch-Verhältnisses nicht wissenschaftlich interpretiert oder gedeutet werden. Die akustische Repräsentation hingegen ermöglichte eine qualifizierte Analyse der Messdaten.

Generell existieren Kontexte in denen sich visuelle Repräsentationen weder technisch, noch inhaltlich anbieten. Teilnehmer einer Telefonkonferenz etwa durch visuelle Hologramme darzustellen, liefert heute eher unbefriedigende Ergebnisse. Werden die Sprecher aber, wie in der Konzeption von Andrew Jay Singer et al.²⁶, durch ein *Spatial Auditory Display (SAD)* virtuell im Raum verteilt und durch separierte Klangströme repräsentiert, kann ein Niveau kommunikativer Interaktion erzielt werden, dass sich einer realen Konferenz wenigstens erkennbar nähert (Singer et al. 1996, Walker & Kramer 2004 et al.). Zusätzlich kann eine räumliche Verortung (i.e. Spatial Auditory Display) immer dann herangezogen werden, wenn eine weitere Dimension unterstützend bei der Datenpräsentation wirkt. Im Beispiel der Sonification von Umsatzzahlen eines Einzelhandels wurde beispielsweise erfolgreich eine Konzeption für den Vergleich zweier Standorte untereinander entworfen. In diesem Fall wurden die Klangströme der beiden Standorte spatial fixiert, um dem Nutzer eine Identifikation zu ermöglichen. Während links des Nutzers zum Beispiel die Zahlen aus Los Angeles präsentiert wurden, wurde auf der rechten Seite das Geschäftsjahr in Detroit dargestellt (Proctor & Reeve 1990).

Bereits diese Anwendungsbeispiele verdeutlichen, wie vielseitig und effektiv der Einsatz auditorischer Displays grundsätzlich ist. Ganz andere informationelle Effekte könnten aber durch eine Einbeziehung von verbalen Ein- und Ausgabemodi erreicht werden. Während es non-verbalen Signalen außerhalb eines bestimmten Kontextes häufig an Eindeutigkeit mangelt, ist die menschliche Sprache in der Lage, sehr konkret Informationen zu transportieren. Verbale Systeme sind so im Stande die Vorteile der akustischen Informationsübermittlung mit der Bestimmtheit des gesprochenen Wortes zu vereinen. Es ist deshalb nicht überraschend, dass entsprechende Investitionen in derartige Technologien auch stattfinden (Neuhoff 2011 et al.). Dennoch sollen verbale, auditorische Displays, wie eingangs erwähnt, als Spezialfall in dem ansonsten allgemeinen, wahrnehmungstheoretischen Ansatz dieser Arbeit keine gesonderte Beachtung finden.

Penelope M. Sanderson et al. (2000) identifizieren in ihrem Artikel „Extending Ecological Interface Design to Auditory Displays“ zusammenfassend drei Kernbereiche für den Einsatz auditorischer Displays: “There are three areas where auditory displays should be considered and may offer an advantage over visual displays [...]“ (S.259)

²⁶ Andrew Jay Singer et al. haben am 30. März 1999 das U.S.-amerikanische Patent für ihr „Audio Conference System“ erhalten.

- **Überwachungsaufgaben (*Vigilance Tasks*):** Überwachungsaufgaben zeichnen sich durch eine kognitive Belastung über einen langen Zeitraum aus. Zu diesem Bereich zählen i.a. die Überwachung von Kraftwerken, die Flugzeugsteuerung auf Reise Flughöhe oder die Pflege auf Intensivstationen. Die Beiläufigkeit auditorischer Anwendungen erzeugt weniger Aufwand für den Anwender und unterstützt so eine mühelose Bedienung.
- **Aufgaben mit hoher kognitiver Belastung (*High Cognitive Load Tasks*):** Derartige Aufgaben werden durch ein extrem hohes Maß an Informationslast charakterisiert, das für den Nutzer in der Regel zu einer starken kognitiven Belastung führt. Da häufig allerdings lediglich Normabweichungen erkannt werden müssen, stellt die Repräsentation durch kontinuierliche Klangströme eine effiziente Lösung für entsprechende Anwendungskontexte dar. Als Beispiele für High Cognitive Load Tasks können Warnsysteme in Kampfjets oder Signalschleifen im NASA Kontrollzentrum angeführt werden.
- **Prädestinierte Datenstrukturen (*Constraints on visual presentation*):** Neben diesen Anwendungsfeldern können auditorische Displays eingesetzt werden, um Datenstrukturen zu repräsentieren, die visuell schlicht und ergreifend nicht angemessen abgebildet werden können. Als Beispiel ist hier die eingangs beschriebene Audification seismologischer Datensätze zu nennen.

In der Tat zeichnet sich auf dem Hintergrund dieser Betrachtung ein Ungleichgewicht zwischen der wahrnehmungstheoretischen Eignung auditorischer Displays und ihrer praktischen Anwendung ab (Sanderson et al. 2000). Trotzdem sollte an dieser Stelle darauf verwiesen werden, dass der Forschungsbereich heute fraglos immer noch vor einigen Herausforderungen steht. Dass viele dieser Probleme noch nicht gelöst werden konnten, hängt nicht zuletzt damit zusammen, dass sich klinische Testergebnisse in der stark symbol- und interpretationsabhängigen Domäne auditorischer Displays häufig kaum auf reale Nutzungsszenarien übertragen lassen. Zusätzlich findet die psychoakustische Grundlagenforschung in der Regel in akustisch regulierten Umgebungen, häufig sogar in vollständig schallisolierten Räumen statt. Die resultierenden Umgebungsbedingungen sind offensichtlich nicht mit denen verwandt, die ein alltägliches Anwendungsszenario mit sich brächte. Erkenntnisse im Hinblick auf individuelle Hörereignisse sind deshalb häufig kaum verwertbar (Walker & Kramer 2004).

Aus konzeptioneller Sicht stellt sich außerdem die Frage, welche Klangeigenschaft (psychoakustisch oder physikalisch) in welcher Form bemüht werden sollten, um Informationen möglichst eindeutig und effizient zu transportieren. Während einige Forschungsergebnisse zu suggerieren scheinen, dass zusätzliche Dimensionen, i.e. Veränderungen mehrerer Klangeigenschaften, zu einer Verbesserung der Informationsübertragung beitragen, muss eine schlussendliche Überforderung des Nutzers in jedem Fall ausgeschlossen werden. Gleichzeitig mangelt es an kreativen Codierungen von Klang für das Transportieren von Information. Ein Warnsignal zu erhören oder das Tempo eines Klangstroms zu ermitteln ist für die meisten Nutzer zwar durchaus möglich, das Maß an Information, das auf diesem Weg übermittelt werden kann, ist allerdings zügig erschöpft. Auch deshalb stellen auditorische Displays häufig auf Symbolik und Metaphorik ab. Kritisch ist bei diesem Ansatz leider vor allem die Abhängigkeit von der individuellen Wahr-

nehmungsprägung des Nutzers. In einer Studie von Bruce N. Walker und David M. Lane (2001) konnte beispielsweise festgestellt werden, dass Probanden mit teilweise oder vollständig eingeschränkter Sehfähigkeit gegebenenfalls andere mentale Modelle zur Interpretation akustischer Reize heranziehen, als Probanden mit uneingeschränkter Sehkraft. Im Rahmen einer Sonification-Sequenz wurden Geldwerte durch entsprechende Frequenzveränderungen symbolisiert, i.e. eine hohe Frequenz entsprach einem hohen Geldwert und vice versa (positive Polarität). Überraschenderweise wurde seitens der Probandengruppe mit eingeschränkter Sehkraft einstimmig angemerkt, dass ein größerer Geldwert sinnvollerweise durch eine niedrige Frequenz dargestellt werden sollte und vice versa (negative Polarität). Diese Erkenntnis ist insofern problematisch, als dass eine adäquate Ausgestaltung auditorischer Displays immer auch davon abhängig ist, dass eine Ambiguität metaphorischer oder symbolischer Zuordnungen im besten Fall vollständig ausgeschlossen werden kann; mehrdimensional ambigüe Displays sind nicht in der Lage die intendierte Bedeutung einer Nachricht zu garantieren.

Einen Großteil dieser Herausforderungen adressiert die Disziplin der Wahrnehmungstheorie, andere Probleme werden nur durch eine Verbesserung der Testverfahren gelöst werden können. Welcher Faktor allerdings häufig unbeachtet bleibt, ist die Auswirkung von Lerneffekten auf die erfolgreiche Bedienung akustischer Benutzeroberflächen. Das folgende Kapitel widmet sich deshalb diesem Aspekt.

3.7 Lerneffekte

Es steht außer Frage, dass die heutige digitale Welt eine visuell dominierte Sphäre ist. Vor allem komplexere auditorische Displays, die das Niveau bloßer Notifikationen und Warnsignale hinter sich lassen, finden sich nur selten. Es ist deshalb nicht verwunderlich, dass der größte Teil der Nutzer mit ihrer Bedienung nicht vertraut ist. Probanden für Tests zur Untersuchung auditorischer Displays müssen häufig genau eingewiesen werden und bedürfen einer gewissen Eingewöhnungsphase (Shinn-Cunningham et al. 2005, Droumewa & Wakkary 2008, Aggarwal et al. 2014, Sundareswaran et al. 2003). Andererseits ist offenkundig bekannt, dass Trainingseinheiten im Umgang mit auditorischen Benutzeroberflächen sehr schnell zu messbaren Erfolgen führen. Personen die beispielsweise im Rahmen ihrer Anstellung auditorische Benutzeroberflächen bedienen, zeichnen sich häufig durch bemerkenswertes Geschick im Umgang mit den entsprechenden Bedienungsmodalitäten aus:

„The special abilities of skilled sonar operators is another example that shows how learning can significantly enhance the efficiency with which auditory patterns can be discerned.“ (Walker & Kramer 2004, S.17)

Dass in der Regel keine Kompetenzen im Umgang mit akustischen Interfaces erworben werden, ist letztlich auf eine historische Designentscheidung zurückzuführen. Sanchit Aggarwal et al. (2014) stellen in ihrer Studie fest, dass sich durch das Training des auditorischen Gedächtnisses mittels adäquater Lernstrukturen überzeugende Erfolge erzielen lassen. Auf diesem

Hintergrund scheint die umfassende Dominanz visueller Displays weniger auf einer wahrnehmungstheoretischen Notwendigkeit, als vielmehr auf einer zumindest in Teilen willkürlichen technologischen Entwicklung zu fußen. Stetige Lernerfolge erlauben langfristig sogar eine Einflussnahme auf die abstrakte Wissensbasis und die damit verbundenen Erwartungshaltungen der Probanden:

„With long-term training, subjects can localize accurately even when the acoustic localization cues are inconsistent with previous experience.“ (Shinn-Cunningham et al. 2005, S.418)

Diese Einschätzung unterstreicht die Relevanz der in Kapitel 3.4 diskutierten kognitiven Modelle für die Konzeption auditorischer Displays. Auch Walker und Kramer sprechen sich nicht zuletzt deshalb explizit für die kognitions-psychologische Untersuchung akustischer Benutzeroberflächen aus: „Since learning results in basic perception and higher-level cognitive processes becoming streamlined, a great amount of the cognitive psychology literature can also be useful for auditory display designers.“ (Walker & Kramer 2004, S.17)

Aus den Untersuchungen von Milena Droumewa und Ron Wakkary (2008) im Rahmen ihres Artikels „Understanding aural fluency in auditory display design for ambient intelligent environments“ geht darüber hinaus hervor, dass speziell die Bewegung in einer akustischen Umgebung und das räumliche Erleben von Klang überdurchschnittliche Lernerfolge ermöglicht. Welchen Einfluss die spatiale Dimension akustischer Signale auf die menschliche Wahrnehmung hat, soll ausführlich, auch auf dem Hintergrund Droumewa und Wakkary's *Aural Fluency Theory*, in Kapitel 4 untersucht werden.

3.8 Auditorische Emotionsinduktion

Die Vielfalt gestalterischer Möglichkeiten für akustische Displays verdeutlicht, wie komplex die kognitiven Strukturen sind, die für die akustische Informationsübermittlung genutzt werden können. Neben dem eher grundlegenden Einsatz als einfaches Warnsignal bis hin zur Verwendung symbolisch und metaphorisch aufgeladener Repräsentationen, eignet sich die Domäne akustischer Stimuli besonders für eine weitere Disziplin – *die Induktion und Kommunikation von Emotionen* (Truax 2000, Narmour 1991, Juslin & Västfjäll 2008, Juslin et al. 2008). Juslin und Västfjäll (2008) adressieren diesen Aspekt in ihrer Grundlagenarbeit zum Verhältnis von Emotion und Musik wie folgt:

„With an increasing number of studies devoted to exploring emotional responses to music, we are in a good position to answer more definitively the long-standing question of whether music really can induce emotions. [...] Music appears to induce a wide range of both basic and complex emotions.“ (Juslin & Västfjäll 2008, S.561-562)

Studien hinsichtlich der Induktion von Emotionen beziehen sich bis jetzt, aus naheliegenden Gründen, hauptsächlich auf das Feld der Musikologie. Eine der wenigen Ausnahmen dazu bilden beispielsweise die psychologischen Untersuchungen zum Verhältnis von Stressempfinden und Störgeräuschen (Truax 2000). Da komplexe, melodische Kompositionen von harmonischen Klangströmen, i.e. Musik, fraglos der Akustik zugeordnet werden können, stellt dieser Umstand keine eigentliche Einschränkung für die weitere Argumentation oder die gewählten Beispiele aus Kapitel 3.4 dar. Eine Unterscheidung von Musik und Klang erweist sich entsprechend weder als sinnvoll, noch als notwendig. Im Rahmen des minimalistischen Modells von Truax (2000) (Abbildung 8) lässt sich Musik als melodische Klangart mit extrem ausgeprägten Struktureigenschaften verstehen. Gerade wegen dieser strukturellen Eigenschaften ist Musik in der Lage emotionale Bedeutung zu entwickeln.²⁷

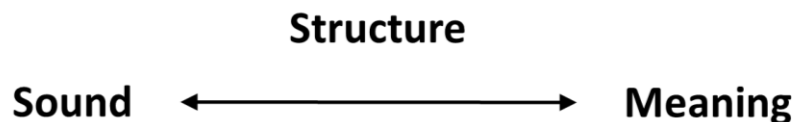


Abbildung 8: Klang, Struktur und Bedeutung (Truax 2000, S.55)

Ein Nebenschauplatz der Diskussion um die Emotionsinduktion durch Musik oder Klang ist die Suche nach einem Konsens hinsichtlich einer gelungenen Definition von Emotion. Eine ausführliche Diskussion dieses Gesichtspunktes findet sich in Juslin & Västfjäll 2008. Entsprechend der Konzeption von Patrik N. Juslin et al. (Juslin & Västfjäll 2008, Juslin et al. 2008) sollen Emotionen im Folgenden als „relativ kurze, aber intensive, affektive Reaktionen auf potentiell wichtige Ereignisse oder Veränderungen der externen oder internen Umgebung“²⁸ (übersetzt aus Juslin & Västfjäll 2008, S.561-562²⁹) definiert werden. Dass Musik in der Lage ist, Emotionen im Sinne dieser Definition zu evozieren, steht heute faktisch außer Frage.

Eine affektive Reaktion des Zuhörers, i.e. eine Emotion, kann durch verschiedene, kognitive Mechanismen provoziert werden. Die folgende Aufzählung soll einen angemessenen Überblick über diese Mechanismen bieten, um den Vorgang der Emotionsinduktion zu plausibilisieren. Die Liste ist dabei zum besseren Verständnis so konzipiert, dass zumindest grundsätzlich von einem zunehmenden Abstraktions- und Komplexitätsgrad der Mechanismen mit steigender Listenposition gesprochen werden kann.

²⁷ Eine ausführliche Diskussion des Modells findet sich in Truax 2000, Kapitel „Speech-Music Soundscapes“ (S.50 ff).

²⁸ Diese relativ weite Definition wurde bewusst gewählt um Gemütszustände, Emotionen im engeren Sinne und affektive Zustände oder Reaktionen gleichermaßen zu inkludieren. Eine Diskussion der offensichtlichen Definitionsproblematik kann an dieser Stelle nicht geleistet werden. Dass der Begriff der Emotion, speziell im individualpsychologischen Kontext, häufig enger definiert wird, ist bekannt.

²⁹ Originalzitat: „[...]relatively brief, though intense, affective reactions to potentially important events or changes in the external or internal environment [...].“

Brain Stem Response / Reflexes (Stammhirnresonanz)

Stammhirnresonanzen erzeugen immer dann im Zuhörer Affekt, wenn eine oder mehrere Klangeigenschaften des Signals dem Stammhirn suggerieren, dass dieses ein potentiell wichtiges Ereignis einleitet. Speziell Geräusche, die unvermittelt auftreten, besonders laut oder dissonant sind, eignen sich für diesen Mechanismus. Diese Beschreibung verdeutlicht auch, dass Stammhirnresonanzen stärker von rudimentären, physikalischen Klangeigenschaften beeinflusst werden, denen sich vor allem die Bottom-up-Verarbeitung widmet, als von Verknüpfungen mit der abstrakten Wissensstruktur des Zuhörers. So erscheint auch nachvollziehbar, dass dieser Mechanismus von einigen Musikologen als weniger relevant für komplexe, musikalische Klangergebnisse eingeschätzt wird (Juslin & Västfjäll 2008, Juslin et al. 2008).

Evaluative Conditioning (Emotionale Konditionierung)

Dieser Begriff bezeichnet einen kognitiven Mechanismus der auf einem Zustand der auditorischen Konditionierung aufsetzt. Bestimmte Emotionen werden deshalb durch bestimmte Klangmuster oder Melodien ausgelöst, weil sie in der Vergangenheit wiederholt mit dem akustischen Reiz kombiniert wurden. Dies setzt eine Paarung eines ursprünglich neutralen bedingten Reizes, i.e. ein akustischer Reiz, mit einem unbedingten, affektiven Stimulus, i.e. einer Emotion, voraus. Der Klang des Schulgongs aus der Kindheit, der die häufig ersehnten Pausen einläutete, kann so auch noch im Erwachsenenalter ein Gefühl der Erleichterung oder Befreiung induzieren. Ebenso kann beispielsweise das bloße Geräusch eines Zahnarztbohrers innere Ängste auslösen. Im Bereich auditorischer Displays wird dieser Mechanismus beispielsweise für Warnsignale herangezogen, die den Zuhörer spontan in einen alarmierten Zustand versetzen sollen.

Eine derartige Stimulipaarung bildet sich häufig mit der Unterstützung des bereits angesprochenen Phänomens des „Background-Listening“ aus (Truax 2000). Der bedingte Reiz, im Kontext des vorherigen Beispiels der Schulgong, wird dabei nur unterbewusst oder gänzlich unbewusst wahrgenommen. Einzelne Studien suggerieren, dass dieser direkte Zugang zum Unterbewusstsein des Zuhörers dem Reiz die Etablierung einer bestimmten, emotionalen Bedeutung sogar erleichtert. Aufmerksamkeit kann eine Konditionierung in diesem Zusammen deshalb gegebenenfalls sogar beeinträchtigen (De Houwer et al. 2005).

Konditionierungsprozesse können dennoch durchaus intendiert stattfinden. Als Unterdisziplin des Marketings versucht sich beispielsweise die akustische Markenbildung mehr oder weniger erfolgreich daran, den Klang eines Produktslogans oder Werbejingles mit einer positiven Empfindung zu kombinieren, um so potentielle Abnehmer zum Kauf zu motivieren (Juslin & Västfjäll 2008, Juslin et al. 2008).

Emotional Contagion (Emotionale Ansteckung)

Eine *emotionale Ansteckung* findet dann statt, wenn ein Klangbild eine bestimmte Emotion transportiert, die durch den Zuhörer „imitiert“ wird. So ist es möglich, dass sich die dramatische Traurigkeit eines langsamen Klavierstückes auf alle Konzertbesucher überträgt, ohne dass im Publikum zunächst eine Disposition für eine negative Stimmungslage vorzuherrschen schien. Dieser Mechanismus ist im Kontrast zur emotionalen Konditionierung direkter von der Aufmerksamkeit und emotionalen Offenheit des Zuhörers abhängig.

Befindet sich dieser in einem akuten, emotionalen Zustand oder hört er nicht „zu“, ist eine Beeinflussung des Gemütszustandes eher unwahrscheinlich. Gleichen sich die emotionale Botschaft und der Gemütszustand des Zuhörers schon vor Einsatz des akustischen Signals, wird die präexistente Gemütslage in aller Regel einfach unterstützt.

Trotz gelegentlicher Abweichungen belegen zahlreiche Studien, dass viele Klangmuster durchaus interindividuell robust Emotionen induzieren können. Bereits Kinder sind in der Lage die emotionalen Botschaften musikalischer Kompositionen „korrekt“ zu interpretieren. Schenkt man Juslins *Super-expressive Voice Theory* Glauben, entsteht der emotionale Ausdruck eines Musikstückes ähnlich wie der einer Stimme. Laute, schnelle Tonfolgen mit bestimmter Klangfarbe werden so beispielsweise sowohl in musikalischer Form, als auch in der menschlichen Sprache tendenziell als aggressiv oder fordernd wahrgenommen. Dieser Aspekt der emotionalen Interpretation musikalischer Stimuli gleicht insofern der emphatischen Analyse, mit der auch verbale Reize beurteilt werden (Juslin & Västfjäll 2008, Juslin et al. 2008).

Visual Imagery (Bildliche Assoziation)

Erzeugt ein akustischer Reiz beim Zuhörer durch eine konkrete Assoziation ein bestimmtes Bild, e.g. eine schöne Landschaft, vor dessen geistigem Auge, liegt eine bildliche Assoziation vor. Trotz der fehlenden visuellen Stimuli, ähneln die resultierenden kognitiven Prozesse durchaus denen, die bei der Betrachtung eines realen Landschaftspanoramas einsetzen würden. Ob es sich dabei um ein rein fiktives Bild handelt, ist für die emotionale Wirkung, die durch diesen Prozess hervorgerufen wird, nicht von Relevanz.³⁰ Aus dieser Kombination von akustischem Signal und der inneren Visualisierung können schlussendlich entsprechende Emotionen entstehen. Obwohl einige Details dieses Mechanismus noch nicht restlos decodiert werden konnten, kann heute angenommen werden, dass der Zuhörer mittels der bildlichen Assoziation die wahrgenommene, musikalische Struktur metaphorisch in sogenannten Bildschemata konzeptualisiert. In der Regel entstehen diese metaphorischen Bildschemata zunächst spontan und unwillkürlich. Dennoch ist es dem Zuhörer häufig möglich durch bewusste Konzentration zu beeinflussen, welche Inhalte visualisiert werden und damit letztlich auch, welche Emotionen induziert werden können. Der Prozess der bildlichen Assoziation ist eng mit dem Vorgang der episodischen Assoziation verwandt (Juslin & Västfjäll 2008, Juslin et al. 2008).

Episodic Memory (Episodische Assoziation)

Nicht selten wird ein bestimmtes Ereignis oder eine Episode aus der Vergangenheit des Zuhörers mit einem bestimmten Lied oder einem akustischen Reiz assoziiert. Wird ein entsprechendes, akustisches Signal wahrgenommen, werden gleichzeitig die assoziierten Erinnerungen vergegenwärtigt. Die mit diesen Ereignissen oder Episoden verbundenen Emotionen werden so häufig „wiedererlebbar“. Statt der prävalenten Emotionen der vergegenwärtigten Erlebnisse wird in einigen Fällen allerdings auch schlicht ein Gefühl von Nostalgie erzeugt. In beiden Fällen geht die Induktion von Emotionen unmittelbar vom akustischen Stimulus aus. Der be-

³⁰ Handelt es sich beispielsweise um einen Ort, der emotional mit einem bestimmten, vergangenen Ereignis zusammenhängt, e.g. das Grundstück der verstorbenen Großmutter, entsteht in der Regel auch eine Kopplung mit dem Vorgang der episodischen Assoziation.

schriebene Vorgang wird, aus gegebenem Anlass, i.a. auch als „Darling, they are playing our tune“-Phänomen bezeichnet (Davies 1978). Die episodische Assoziation wird häufig sogar bewusst eingesetzt, um bestimmte Erinnerungsvorgänge oder eine emotionale Immersion zu unterstützen (Juslin & Västfjäll 2008, Juslin et al. 2008).

Musical Expectancy (Musikalisch-syntaktische Verarbeitung)

Das Phänomen der musikalisch-syntaktischen Verarbeitung von melodischen Stimuli ist größtenteils auf musikalische Reize limitiert. Durch diesen Vorgang können Emotionen induziert werden, falls die spezifischen melodischen Strukturen eines Musikstückes den musikalischen Erwartungen des Zuhörers widersprechen, diese hinauszögern oder erfüllen. Die Tonsteigerung von E auf F suggeriert dem Zuhörer beispielsweise, dass nun ein G folgen wird. Wird diese Erwartung nicht erfüllt, könnte dies den Zuhörer zum Beispiel überraschen. Durch die gezielte Verzögerung oder Erfüllung bestimmter Erwartungen können entsprechend andere affektive Reaktionen provoziert werden (Juslin & Västfjäll 2008, Juslin et al. 2008). Die Aggregation dieser affektiven Reaktionen und deren Ausgestaltung ist letztlich entscheidend dafür, ob Musik als gut oder schlecht empfunden wird und deshalb essentiell für jede musikalische Komposition. Während eine weiterführende Betrachtung dieser Disziplin für die vorliegende Arbeit nicht zielführend wäre, soll an dieser Stelle dennoch auf die wegweisende Arbeit „Music Perception: An Interdisciplinary Journal“ von Eugene Narmour (1991) hingewiesen sein, in der alle grundlegenden musikologischen Erwartungskonzepte auf einem fundierten, kognitionswissenschaftlichen Hintergrund aufgearbeitet werden.

Das Phänomen musikalisch-syntaktischer Erwartungshaltungen ist vor allem auch deshalb interessant, weil sich die besagten Erwartungen speziell mit dem Niveau der musikalischen Bildung verändern können. Insofern kann von einer direkten Abhängigkeit von den abstrakten Wissensstrukturen des Zuhörers gesprochen werden (Narmour 1991).

Cognitive Appraisal (kognitive Bewertung)

Obwohl der Vorgang der kognitiven Bewertung auf durchaus abstraktem Niveau stattfindet, geht ein Großteil der für gewöhnlich durch Musik induzierten Emotionen von ihm aus. Affekt wird hier durch die subjektive Bewertung der Musik auf dem Hintergrund abstrakter Wünsche oder Lebensziele des Zuhörers induziert. Für diese Bewertung werden neben musikalischen Kriterien auch Erwartungen und Dimensionen außerhalb der strikt musikalischen Sphäre herangezogen. Besucht ein Zuhörer beispielsweise ein Konzert und ist über dessen hohe Qualität erstaunt, kann dies beispielsweise ein Gefühl von Überraschung und Befriedigung erzeugen. Verläuft ein Konzertbesuch eher enttäuschend, kann durch ihn negativer Affekt entstehen. Genauso kann Musik dem Zuhörer helfen, bestimmte, negative Erlebnisse besser zu verarbeiten und auf diesem Weg positiv konnotiert Emotionen zu generieren.

Es sollte allerdings angemerkt werden, dass diese Form der Emotionsinduktion weniger vom musikalischen Stimulus selbst als von der Natur des Gesamterlebnisses (e.g. des Konzertbesuches) abhängig ist. Ob und inwiefern die Form der kognitiven Bewertung deshalb als spezifisch für die Verarbeitung akustischer Reize angesehen werden kann, bleibt strittig. Eine ausführliche Diskussion dieser Problematik findet sich in Juslin & Västfjäll 2008 und Juslin et al. 2008.

Die Mechanismen der Emotionsinduktion, wenn auch verwandt, können durch verschiedene Kriterien, e.g. Induktionsgeschwindigkeit, Steuerbarkeit, Modularität, Abhängigkeit von musikalischer Struktur, deutlich voneinander unterschieden werden. Gleichzeitig schließen sich die Vorgänge nicht gegenseitig aus, sondern wirken häufig ergänzend oder treten parallel auf. Die Palette der resultierenden affektiven Reaktionen kann sich über das gesamte menschliche Emotionspektrum erstrecken (Juslin & Västfjäll 2008, Juslin et al. 2008).

3.9 Kontext

Im Rahmen der vorliegenden Eignungsuntersuchung erschließt sich, dass und weshalb akustische Signale in der Lage sind komplexe Information mit hoher Dichte zu transportieren. Anders als visuelle Einflüsse müssen sie nicht fokussiert werden und eignen sich entsprechend gut für Begleittechnologien, die den Nutzer beiläufig informieren. Außerdem verfügen akustische Impulse über eine sehr große Reichweite und breiten sich mit Schallgeschwindigkeit³¹ von ihrem Ursprung her aus. Gleichzeitig ist der (menschliche) auditorische Wahrnehmungsapparat in der Lage, sowohl die direkten Klangeigenschaften, als auch die transportierte Bedeutung des Schallsignals in Echtzeit zu decodieren. Durch die Elektroakustik und in ihrer Folge die Entwicklung moderner auditorischer Displays und Benutzeroberflächen, könnten die Potentiale des akustischen Wahrnehmungsapparates gezielt ausgeschöpft werden. Zusätzlich befindet sich die akustische Holografie mit Technologien wie der Wellenfeldsynthese und 3D-Spatial-Audio-Anwendungen heute auf einem Niveau, dass seinem visuellen Pendant weit voraus ist (Vazquez-Alvarez et al. 2011).

Dennoch gestaltet sich die akustische Informationsübertragung häufig schwieriger als diese praktischen Grundlagen vermuten ließen. Wie schon in Kapitel 3.6 skizziert, neigen akustische Signale deutlich stärker zur Ambiguität, als visuelle Stimuli³². Kontextuelle Faktoren können die wahrgenommene Bedeutung eines Hörereignisses völlig verändern. Andererseits kann intendierter Kontext gezielt eingesetzt werden, um Signale zu disambiguieren und die *Robustheit ihres interindividuellen Bedeutungswertes* zu unterstützen, wie es das Zitat von Walker und Kramer prägnant zusammenfasst:

„Recent studies have pointed out that successful interpretation of the sonified data requires more than just the sonified data—context is also necessary.“ (Walker & Kramer 2004, S.6)

³¹ Entspricht in trockener Luft von 20° C etwa 343 m/s.

³² Obwohl eine breitere Betrachtung inklusive haptischer, olfaktorischer und gustatorischer Stimuli andernorts sicherlich sinnvoll erscheinen könnte, konzentriert sich die vorliegende Arbeit vornehmlich auf die visuelle Informationsübertragung. Auf dem Hintergrund ihrer dominanten Stellung in der digitalen Sphäre, erscheint ein Vergleich mit der Technologie der visuellen Datenpräsentation am ehesten tauglich um eine potentielle Eignung auditorischer Anwendungen abzuleiten.

3.9.1 Begrifflichkeit

Damit die Meriten von Kontext nutzbar gemacht werden können, muss zunächst seine Terminologie bestimmt werden, um abzuleiten, wie Kontext zum Zweck der gezielten Informationsübermittlung instrumentalisiert werden kann (Bradley & Dunlop 2005, Dey & Abowd 2000, Niessen et al. 2008, Wigelius & Väättäjä 2009). Obwohl die meisten Menschen ein implizites Verständnis davon haben, was Kontext bedeutet, sind sie in der Regel nicht in der Lage eine Begriffserklärung zu explizieren. Kontext wird deshalb häufig nur beispielhaft oder durch Synonyme erklärt. Sonstige Versuche Kontext adäquat zu definieren, stellen speziell auf die spatiale und geolokale Umgebung ab, lassen aber beispielsweise den Tätigkeits- oder sozialen Horizont regelmäßig außeracht (Dey & Abowd 2000 et al.).

Bill Schilit et al. versuchen diesen Impetus in ihrem Artikel zum ersten internationalen Workshop für Mobile Computing Systems und Applications (1994) zu öffnen und etablieren im Folgenden drei grundlegende Fragen zur Bestimmung eines Kontextes: wo befindet sich die Entität, deren Kontext bestimmt werden soll, mit „wem“ befindet sich die Entität dort und welche Ressourcen befinden sich in der unmittelbaren Umgebung. Kontext wird in diesem Sinne als *eine sich kontinuierlich verändernde Umsetzungsumgebung* verstanden. Auf Basis dieser Überlegungen können 3 Umgebungsdimensionen identifiziert werden (Dey & Abowd 2000 et al.):

- **Computing Environment:** Welche Einrichtungen zur Eingabe von Input durch den Nutzer und der Wiedergabe von Output zum Zwecke der Informationsübermittlung stehen zu Verfügung.
- **User Environment:** Wo und in welcher sozialen Situation befindet sich der Nutzer und welche Personen befinden sich in seiner unmittelbaren Umgebung.
- **Physical Environment:** Welche physischen Gegebenheiten (e.g. Belichtung, Geräuschpegel) liegen vor.

Annind K. Dey und Gregory D. Abowd (2000) greifen diese Perspektive sechs Jahre später auf und formulieren einen Kontextbegriff, der im Folgenden auch als Arbeitsdefinition dieser Untersuchung dienen soll:

„Jede Information, die eingesetzt werden kann, um die Situation bzw. den Zustand einer Entität zu charakterisieren, ist Teil ihres Kontextes. Eine Entität ist eine Person, ein Ort oder ein Objekt, das als relevant für die Interaktion zwischen einem Nutzer und einer Anwendung angesehen werden kann, inklusive dem Nutzer und der Anwendung selbst.“ (übersetzt aus Dey & Abowd 2000, S.3-4³³)

Dass diese Definition sehr weit gefasst ist, steht zwar außer Frage, dass dies der Fall ist liegt aber nicht an ihrer sprachlichen Ungenauigkeit, sondern an der *komplexen und allumfassenden Tragweite*

³³ Originalzitat: „Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.“

des Kontexts. Pauschale Einschränkungen zu treffen hieße wissentlich Informationen zu ignorieren, die als Kontext die Bedeutung einer Nachricht beeinflussen und im unkontrollierten Zustand deshalb verzerren könnten. Dennoch verweisen auch Dey und Abowd (2000) im Anschluss an diese Definition auf eine kategorische Unterscheidung kontextueller Einflussgrößen. Eine Situation ist demnach zunächst durch seine übergeordneten oder primären Kontextfaktoren Standort, Identität, Zeit und Aktivität bestimmt. Alle weiteren Informationen unterstützen eine differenziertere Charakterisierung der Situation, werden aber als sekundäre Einflussgrößen betrachtet. Die Studie von Heli Wigelius und Heli Väättäjä (2009) zur Untersuchung mobiler Anwendungen legt eine ähnliche Unterscheidung zugrunde. Der primäre Kontext ergibt sich bei Wigelius und Väättäjä aus den sozialen, spatialen, temporalen, infrastrukturellen und tätigkeits-spezifischen Eigenschaften einer Situation. Veranschaulichende Beispiele zu den einzelnen Eigenschaften finden sich in Abbildung 9, die der Darstellung von Wigelius und Väättäjä (2009) entlehnt ist.

Sozial	Spatial	Temporal	Infrastrukturell	Tätigkeit
Arbeitsklima	Standort	Termine	Netzwerkverbindung	Arbeitsziel
Kulturelle Prägung	Temperatur	Uhrzeit und Datum	Verfügbare Technologien	Interaktionsniveau
Anwesende Personen	Lautstärke	Arbeitsgeschwindigkeit		
	Beleuchtung	Irregularitäten		

Abbildung 9: Kontextbeispiele (Wigelius & Väättäjä 2009, S.10)

Um kontextbezogen bzw. kontextspezifisch zu agieren, müssen die zur Verfügung stehenden Kontextinformationen vom System *zum Zwecke der Adaption* genutzt werden. Auf dem Hintergrund der Eignungsuntersuchung dieser Arbeit erscheint es sinnvoll, den Kontextbegriff angemessen in die Systemkonzeption auditorischer Displays und Benutzeroberflächen einfließen zu lassen. Abgeleitet aus ihrer Kontextdefinition beschreiben Dey & Abowd (2000) kontextbezogene Systeme wie folgt:

„Ein System arbeitet kontextbezogen, wenn es Kontextinformationen verwendet, um dem Nutzer relevante Inhalte und/oder Dienste in Abhängigkeit seiner aktuellen Tätigkeit (Nutzungsintention) anzubieten“ (übersetzt aus Dey & Abowd 2000, S.6³⁴)

Obwohl möglichst neutrale Definitionen als Grundlage für diese Arbeit herangezogen wurden, schlägt sich schon an der Wahl der kontextuellen Primärfaktoren nieder, dass sie der Disziplin

³⁴ Originalzitat: „A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user’s task.“

der Informationstechnologie (IT) entstammen und entsprechend speziell auf Kontextfaktoren im Umgang mit informationstechnologischen Anwendungen fokussieren. Nicholas A. Bradley und Mark D. Dunlop (2005) stellen in ihrer Ausarbeitung „Towards a Multidisciplinary Model of Context to Support Context-Aware Computing“ einen interdisziplinären Ansatz vor, der auf ein breiteres Spektrum an Primärfaktoren rekurriert. Erklärtes Ziel dieses Ansatzes ist es, die Kontextbegriffe der Linguistik, der Psychologie und der Informationstechnologie in einer umfassenden Konzeption zusammenzuführen. Auch Bradley und Dunlop erkennen, dass Kontext zwar zunächst alle effektiven und potentiellen Einflussfaktoren einer Umgebung einschließt, gleichzeitig stellen sie aber heraus, dass nicht alle Kontextinformationen unmittelbar relevant sind. Kontextfaktoren werden in ihrem Ansatz deshalb in beiläufig („incidental“) und bedeutend („meaningful“) unterteilt. Darüber hinaus findet eine Abgrenzung interner und externer Einflussgrößen statt. Während die Menge der externen Einflussgrößen alle Umgebungsfaktoren eines klassischen Kontextbegriffes umfasst, konstituiert sich der interne Kontext (in der Betrachtung des Nutzerkontextes, nicht des Kontextes der Applikation) durch den kognitiven Zustand des Nutzers. Diesen Zustand definieren Bradley und Dunlop als „[a] user’s cognitive processing abilities, short- and long-term memory abilities, dislikes and preferences, opinions and beliefs, cultural interpretations, perceptual sensing abilities, cognitive mapping strategies for encoding spatial information, and so on.“ (Bradley & Dunlop 2005, S.426) Die wahrnehmungstheoretischen Grundlagen der vorangegangenen Kapitel werden in dieser Konzeption also als interne Kontextfaktoren integriert.

Streng genommen expliziert die Konzeption von Bradley und Dunlop damit, was in der Definition von Dey und Abowd implizit bleibt. Die zusätzlichen Untergliederungen wirken zugegeben strukturgebend auf die Begrifflichkeit des Kontextes, zum Zweck dieser Arbeit ist eine detailliertere Betrachtung allerdings nicht notwendig.

Das zugrundeliegende Verständnis von Umgebung setzt sich dabei aus zwei Komponenten zusammen. Einerseits verfügt die Umgebung über direkte, spatiale Eigenschaften, die kontextgegeben sind, e.g. Offenheit des Raumes, geolokale Position des Nutzers. Andererseits fungiert der Begriff der Umgebung als *räumliche Bestimmung des Geltungsbereichs eines Kontextes* im Hinblick auf den Großteil aller anderen kontextuellen Einflussgrößen. Der soziale Kontextfaktor „Arbeitsklima“ ist beispielsweise nur im Hinblick auf das den Nutzer unmittelbar umgebende Arbeitsklima relevant; gleiches gilt etwa für infrastrukturelle Faktoren.

Auch wenn dieser Zusammenhang auf den ersten Blick trivial erscheint, unterstreicht er dennoch die Dominanz der spatialen Dimension des Kontextbegriffes.

Formal umfasst die zugrunde gelegte Kontextdefinition von Dey und Abowd wie beschrieben alle potentiellen Einflussgrößen. Dennoch erscheint es im Rahmen dieser Arbeit sinnvoll, besonderes Augenmerk auf den für die akustische Welt so entscheidenden *Klang- und Bedeutungskontext* zu legen. Eine detaillierte Betrachtung dieses Aspektes folgt im nächsten Kapitel.

3.9.2 Soundscapes (nach Truax) und akustische Ambiguität

Der Begriff „Soundscape“ (deutsch „Klanglandschaft“) kann als Bezeichnung für jede mögliche Allokation von Klangquellen in einer beliebigen Umgebung eingesetzt werden. Sowohl das akustische Innenleben eines Cafés an einem Sonntagnachmittag, als auch die charakteristische Geräuschkulisse eines Urwaldes können als Klanglandschaft bzw. Soundscape verstanden werden (Truax 2000). Natürliche (nicht artifizielle, i.e. vom Menschen kaum oder nicht veränderte) Klangumgebungen, so Truax, verfügen in aller Regel über eine *akustische Balance*, welche den Informationsaustausch durch akustische Signale ermöglicht.

„When the system is well balanced (what we have loosely referred to as a “hi-fi” environment), there is a high degree of information exchange between its elements and the listener is involved in an interactive relationship with the environment.“ (Truax 2000, S.65)

Einzelne Schallereignisse können auf dem Hintergrund ihrer Soundscape disambiguiert und deshalb angemessen interpretiert werden – sie werden einem Klangkontext zugeordnet. Im Falle des Urwalds könnte beispielsweise die Klangkulisse (unabhängig von anderen Sinneswahrnehmungen) ausreichend sein, um eine schrille Klangwiederholung als das Zwitschern eines Vogels zu identifizieren. Befindet sich der Zuhörer stattdessen in einem Café, dürfte eine derartige Zuordnung trotz der relativen Eindeutigkeit dieses Beispiels deutlich schwerer fallen. Maria E. Niessen et al. (2008) verdeutlichen diesen Zusammenhang anhand einer Studie für die Arbeit „Disambiguating Sound through Context“. Grundlage der Arbeit ist die Annahme, dass alltägliche Hörereignisse immer sowohl anhand der Informationen des Signals selbst (Bottom-up), als auch dem kontextuellen Vorwissen des Zuhörers beurteilt werden (Top-down). In der Studie von Niessen et al. wurde 42 Probanden zunächst ein homonymes³⁵ Geräusch präsentiert. Schon definitorisch war eine eindeutige Zuordnung durch die Probanden in diesem Fall ausgeschlossen. Im zweiten Durchgang folgte das homonyme Geräusch auf ein zusätzliches, kontextverwandtes Geräusch. Das Prasseln von Regen konnte isoliert beispielsweise von einer Dusche mit entsprechend angepasstem Wasserdruck nicht unterschieden werden. Folgte das Geräusch allerdings auf einen Donnerschlag, wurde das akustische Ereignis sicher als Regen erkannt.

Die skizzierte Studie konnte erfolgreich nachweisen, dass die Probanden im zweiten Fall in der Lage waren die intendierte Bedeutung nachzuvollziehen bzw. das akustische Ereignisse korrekt zu identifizieren (Niessen et al. 2008). Der synthetisch erzeugte Klangkontext war damit erfolgreich in der Lage, homonyme Klänge zu disambiguieren.

„We have shown that context can have a facilitatory effect in identification, and used this asset to automatically disambiguate sounds that may have multiple causes.“ (Niessen et al. 2008, S.13)

³⁵ Ein homonymes Geräusch ist als nicht eindeutiges Geräusch definiert, dass zwei oder mehr potentiellen Klangquellen zugeordnet werden könnte (Niessen et al. 2008).

Natürlich verdeutlicht die bloße Existenz ambiguer Schallereignisse ein gewisses Maß an Manipulationspotential. Auf dem Hintergrund der Eignungsuntersuchung dieser Arbeit ist der Klangkontext allerdings vor allem deshalb interessant, weil er herangezogen werden kann, um geeignete auditorische Displays zu konzipieren, die die Dynamiken akustischer Landschaften nicht nur beachten sondern für ihre Zwecke instrumentalisieren.

Truax stellt explizit darauf ab, dass äußerlich beeinflusste Soundscapes (wie etwa die Klanglandschaft einer Stadt) nicht mehr angemessen in der Lage sind, Geräuschen innerhalb ihrer akustischen Kulisse eine eindeutige Bedeutung zuzuordnen. Dies liegt zum einen daran, dass häufig eine zu hohe Konzentration akustischer Reize vorliegt, die eine wahrnehmungstechnische Verarbeitung erschweren; zum anderen aber ist dieser Effekt auf die elektroakustische Wende zurückzuführen, die es möglich gemacht hat, synthetische Klangquellen zu erzeugen und in praktisch jeden Umgebungskontext zu überführen. Gerade eine symbolische Aufladung akustischer Signale, wie auditorische Displays sie häufig einsetzen, wird dadurch deutlich verkompliziert (Truax 2000). Eine weiterführende Diskussion dieser Thematik findet sich in Kapitel 4.

3.9.3 Mehrdimensionalität des spatialen Kontexts

In Anlehnung an die Unterscheidung nach Christian Freksa (2013) ist Spatialität gleich in mehrerlei Hinsicht relevant für die akustische Informationsübermittlung:

Erstens dient *Räumlichkeit als Vehikel für die auditorische Kognition* per se - Schall breitet sich im Raum aus. Entsprechend transportiert ein Schallereignis immer auch Informationen über die räumliche Umgebungsstruktur (e.g. offene, weite Fläche versus enger Korridor). Eine entsprechende Allokation von Schallquellen im Raum ist deshalb in der Lage, Spatialität zumindest konzeptionell zu repräsentieren. Durch die Fähigkeiten der Lokalisation, der Identifikation bestimmter Hallstrukturen und anderer Erkennungsmechanismen des auditorischen Wahrnehmungsapparates, ist die Audition in der Lage eine mentale Repräsentation der räumlichen Umgebung zu erzeugen. Mittels der spatialen Audiowiedergabe ist es heute allerdings ebenso möglich, ein entsprechendes, akustisches Hologramm einer solchen Umgebung synthetisch zu erzeugen und beispielsweise mittels Kopfhörer an den Zuhörer zu übermitteln (Duraismami et al. 2005, Härmä et al. 2003 et al.).

Zweitens existiert eine enge, *kognitive Relation zwischen Spatialität und Audition*, die nicht nur entscheidend für die Sicherheit im Umgang mit akustischen Eindrücken ist, sondern auch maßgeblich Lernerfolge im Umgang mit auditorischen Benutzeroberflächen beeinflusst (Sonnenschein 2011, Droumewa & Wakkary 2008 et al.).

Drittens fungiert *Spatialität als Grundlage für den Kontextbegriff der akustischen Informationsübertragung*. Wie in Kapitel 3.9 dargestellt, ist die räumliche Umgebung gleich in mehreren Dimensionen fundamental für Kontext im Allgemeinen und den akustischen Kontext im Speziellen. Truax' (2000) Auffassung der *spatialen Identität akustischer Signale* innerhalb einer Klanglandschaft verdeutlicht diese wahrnehmungstheoretische Abhängigkeit (Truax 2000, Dey & Abowd 2000, Niessen et al. 2008 et al.)

Angesichts dieser zusammengefassten Aspekte soll die Bedeutung der Spatialität für die Audition und die Ausgestaltung auditorischer Benutzeroberflächen in Kapitel 4 genauere Betrachtung finden.

3.10 Konklusion

Zunächst soll allerdings eine Überprüfung der eigentlichen Arbeitshypothese (H1) angestellt werden.

H1 Der Einsatz kontextabhängiger, auditorischer Displays und Benutzeroberflächen ist aus wahrnehmungstheoretischer Sicht *sinnvoll*.

Um H1 zu konkretisieren, ist es notwendig, den Begriff der Sinnhaftigkeit inhaltlich auszugestalten. „Sinnvoll“ bedeutet gemäß der Definition des Dudens „durchdacht und zweckmäßig“ oder „für jemanden einen Sinn habend, eine Befriedigung bedeutend“. Im Kontext auditorischer Benutzeroberflächen bezieht sich der hintere Teil dieser Definition unmittelbar auf den Nutzer. Auch wenn eine inhaltlich eindeutige Ausgestaltung durch formale Kriterien offenkundig ausgeschlossen ist, sollte eine angemessene Näherung für die Überprüfung der Hypothese zweifelsfrei ausreichend sein. Auf diesem Hintergrund kann angenommen werden, dass eine Anwendung zum Zweck der akustischen Informationsübertragung immer dann *sinnvoll* ist, wenn:

- (i) sie *handlungsermöglichend* oder *handlungsweisend* ist, i.e. dem Nutzer bestimmte Handlungen durch adäquate Information ermöglicht.
- (ii) sie *intendiert Affekt erzeugt*, i.e. den Nutzer emotional beeinflusst.
- (iii) sie der *Unterhaltung* dient, i.e. Neugier im Nutzer erzeugt oder ihn in eine spielerische Interaktion verwickelt.

Alle drei Kriterien zielen auf eine beabsichtigte Einflussnahme im Sinne einer Zustandsveränderung auf den Nutzer durch akustische Information ab.

Die Kapitel 3.4 bis 3.6 dieser Arbeit beschäftigen sich mit der Leistungsfähigkeit des menschlichen, auditorischen Wahrnehmungsapparates, der Informationsdichte akustischer Signale und der Nutzung dieser Gegebenheiten durch den Einsatz auditorischer Benutzeroberflächen. Angesichts dieser Ausführungen steht die Eignung auditorischer Displays als klassisches Informationsmedium außer Frage. Ob eine Sirene den Nutzer vor einem Feuer warnt und ihm so eine rechtzeitige Evakuierung ermöglicht oder die Audification seismischer Aufzeichnungen die Interpretation von Seismogrammen erlaubt, ist letztlich lediglich eine Frage des Komplexitätsgrades der Anwendung; (i) kann insofern im Rahmen dieser Untersuchung als erfüllt angesehen werden.

Dass und wie auditorische Displays, speziell durch den Einsatz von Musik, durchaus in der Lage sind, affektive Reaktion zu erzeugen, wurde ausführlich in Kapitel 3.8 aufgearbeitet. Allein auf Basis der Beispiele und Ausführungen wird deutlich, dass auditorische Displays und Benutzeroberflächen dem Kriterium (ii) entsprechen. Wie etwa im Fall der Filmmusik dominiert häufig

sogar der akustische Reiz den Inhalt einer emotionalen Botschaft. Dass die Bedeutung ambiguer, visueller Inhalte durch den geschickten Einsatz akustischer Stimuli manipuliert werden kann, unterstreicht den Einfluss der Audition auf die menschliche Gefühlswelt. Wenn auf diesem Hintergrund auch nicht von einer Übererfüllung von (ii) gesprochen werden muss, darf dennoch eine gewisse *Prädestination der Akustik für die Emotionsinduktion* angenommen werden.

Befragungen der Testpersonen aus Projekten wie Audio Aura (Mynatt et al. 1997), Hear & There (Rozier 2000) oder Soundgarden (Vazquez-Alvarez et al. 2011) zeigen außerdem, wie groß das Unterhaltungspotenzial ungewohnter, auditorischer Benutzeroberflächen ist. Sind Benutzer noch nicht mit der Bedienung auditorischer Systeme vertraut, verfügen sie meist über eine überdurchschnittlich hohe Frustrationstoleranz im Rahmen der „Entdeckung“ des Interfaces. Bereits nach kurzer Zeit stellen sich aber in aller Regel beachtliche Lernerfolge ein (siehe Kapitel 3.7), die eine spielerische und befriedigende Interaktion mit dem System weiter fördern. Auch das Kriterium (iii) kann insofern als erfüllt betrachtet werden.

Diese Auswertung scheint zu implizieren, dass die eklatante Konzentration auf visuelle Displays und Benutzeroberflächen, zumindest in Teilen, als sich *perpetuierende Designentscheidung* aufgefasst werden kann. Die wahrnehmungstheoretische Eignung auditorischer Systeme für die Übertragung von Affekt oder Information kann jedenfalls als unstrittig betrachtet werden. Wie in der Einleitung vorweggenommen, stellt diese Untersuchung absichtlich nicht auf die Nutzung auditorischer Benutzeroberflächen durch sehbehinderte Nutzer oder die Informationsübertragung mittels verbaler Signale ab. Fraglos erscheinen Nutzen und Sinnhaftigkeit akustischer Systeme noch klarer unter Einschluss dieser Signale und Benutzergruppe.

Interessant ist, dass sich bereits mit Blick auf die zugrunde gelegten Sinnhaftigkeitskriterien die Relevanz von Spatialität für die akustische Informationsübertragung ableiten lässt. Während eine spatiale Verortung akustischer Signale in der Lage ist, die Übertragungseffizienz auditorischer Displays messbar zu erhöhen (Walker & Kramer 2004, Singer et al. 1996 et al.) und die Verbindung akustischer Benutzeroberflächen mit einer räumlichen Erfahrung größere Lernerfolge verspricht (Droumewa & Wakkary 2008 et al.), nimmt eine spatiale Ausgestaltung auch nachhaltigen Einfluss auf den Unterhaltungswert auditorischer Anwendungen:

„Based on the user feedback, the extra time spent stationary and the extra distance covered when audio spatialization was used did not lead to frustration, rather it appears to be related to the enjoyment and sense of discovery of the participants. In contrast, for the conditions lacking audio spatialization, participants behaved more like in a navigation environment setting themselves the task of finding all the landmarks by systematically walking through the park. This behaviour emerged despite participants in all conditions being given the same set of instructions before starting the exploration of the garden.“ (Vazquez-Alvarez et al. 2011, S.7)

Im Folgenden konzentriert sich diese Ausarbeitung deshalb, ganz im Sinne des Arbeitstitels, auf die Untersuchung der wahrnehmungstheoretischen Beziehung zwischen Räumlichkeit, i.e. Spatialität, und auditorischer Information.

4 Audition und Spatialität

4.1 Schizophonie (nach Truax)

Die Funktionsweisen und Wahrnehmungskonzepte des menschlichen, auditorischen Systems sind originär auf die Signalverarbeitung natürlicher Klanglandschaften ausgerichtet. Wie Truax (2000) es in seinen *kommunikationalen Ansatz* beschreibt, verfügen diese Soundscapes häufig über eine akustische Balance, durch die ein Klangkontext etabliert wird, der bedeutungsgebend für alle ihn konstituierenden Klangsignale ist. Erst die Disziplin der Elektroakustik hat eine Synthese akustischer Ereignisse und Klangkontexte ermöglicht:

„Before audio technology, no sound had ever been heard twice, exactly the same, nor had any sound ever been heard outside its original context. The communicational significance of any sound can only be judged within its complete context in the broadest environmental, social, and cultural sense. In fact, it is through context that we understand how a sound functions.“ (Truax 2000, S.12)

Als Resultat dieser elektroakustischen Wende sind teilsynthetische Klanglandschaften entstanden, deren akustische Versatzstücke in keinem kommunikativen Verhältnis mehr zueinander stehen. Die Klangtreue (high-fidelity oder hi-fi) heutiger Wiedergabesysteme garantiert zwar ein im Sinne des Signaltransfermodells praktisch identisches Schallereignis, der kommunikative Kontext des reproduzierten Signals findet in diesem System allerdings keine Beachtung. Das akustische Signal kann deshalb häufig nicht mehr mit Inhalt oder Bedeutung gefüllt werden – es kommt zur *Schizophonie*³⁶ (*Schizophonia*).

„A simpler way to define the problem is to refer to the split between an original sound and its electroacoustic reproduction as “schizophonia”[.]“ (Truax 2000, S.134)

Die Herausforderung einer schizophonen Situation für den Zuhörer besteht darin, die Parallelität zweier verschiedener Klangkontexte angemessen zu interpretieren. Verbale Signale aus der Richtung einer massiven Betonwand auf einem Flughafen nicht nur korrekt zu interpretieren, sondern sogar zu antizipieren, setzt ein hohes Maß an sozial-technologischer Prägung voraus. Genauso verändert ein klassisches Stück, gespielt von einem hundertköpfigen Orchester, das aus den eigenen Kopfhörern ertönt, um die morgendliche Laufleistung zu fördern, sicherlich die Umgebungswahrnehmung. Das Signal kann nur deshalb angemessen eingeordnet werden, weil der veränderte Kontext bewusst generiert wurde. Eine Vielzahl synthetischer Klangkontexte

³⁶ Der Begriff leitet sich aus der psychologischen Bezeichnung der „Schizophrenie“ ab und wurde 1969 von R. Murray Schafer etabliert.

erzeugen aber, gerade bei Zuhörern ohne entsprechende sozial-technologische Prägung³⁷, auditorische Verwirrung und Verunsicherung (Truax 2000).

Eine sinnvolle, spatiale Ausgestaltung teilsynthetischer Klanglandschaften gemäß dem natürlichen Vorbild unterstützt stattdessen eine sichere und intuitive Wahrnehmung (Vazquez-Alvarez et al. 2011, Sonnenschein 2011 et al.). Diese Effekte müssen bei der Konzeption zukünftiger auditorischer Displays und Benutzeroberflächen Beachtung finden. Der Einsatz elektroakustischer Technologien nach wahrnehmungstheoretischen Prinzipien könnte so in der Lage sein, die Symptome der entstandene Schizophonie wieder zu lindern. Insbesondere ermöglicht die Nutzung der räumlichen Dimension einer Klanglandschaft eine Verknüpfung synthetischer Signale mit der realen Umgebungswelt. Das daraus resultierende, *spatiale Audio-Hologramm* imitiert einen natürlichen Wahrnehmungskontext, in den die synthetischen Audiosignale *sinnvoll* reintegriert werden könnten (Truax 2000).

Um eine solche, adäquate Verknüpfung zwischen Räumlichkeit und synthetischen Audiosignalen zu gewährleisten, müssen die *wahrnehmungstheoretischen Logiken der Audition und der räumlichen Wahrnehmung* zusammengeführt werden.

4.2 Kognitive Relation von Klang und Raum

Aus wahrnehmungstheoretischer Perspektive tritt Spatialität in ihrer Relation zur Audition in zweierlei Form in Erscheinung. Einerseits ist Spatialität eine Eigenschaft, ein kognitives Vehikel, das die Wahrnehmung anderer sinnlicher Eindrücke beeinflusst. Diese eher kontextuelle oder rahmengebende Funktion von Räumlichkeit wurde in den vorherigen Kapiteln immer wieder beiläufig hervorgehoben. Andererseits kann Räumlichkeit auch selbst Objekt der Kognition sein, i.e. mentale Konzeptionen von Räumlichkeit werden entworfen, um sich zu orientieren oder die spatiale Umgebung manövrierbar zu machen (Freksa 2013). In jedem Fall ist Spatialität aber von zentraler Bedeutung für die menschliche Kognition und Wahrnehmung:

„Space and our dealings with space provide for many of the skills essential for intelligent behaviour.“ (Bateman et al. 2014, S.13)

Dieses Unterkapitel widmet sich deshalb der Untersuchung beider Funktionen von Spatialität.

4.2.1 Spatiale Kognition

Anders als etwa die auditorische Wahrnehmung, ist die räumliche Wahrnehmung keinem Sinnesorgan direkt zugeordnet. Sie setzt sich kohärent aus allen relevanten, sinnlichen Informationen, vor allem aber den visuellen und akustischen Einflüssen der Umgebung zusammen. Während das dreidimensionale Sehen beispielsweise das Bestimmen von Entfernung oder Position einer visuellen Erscheinung ermöglicht, transportiert die charakteristische Hallstruktur eines Raumes Informationen über seine räumliche Beschaffenheit.

³⁷ E.g. Kleinkinder oder Mitglieder nativer Völker.

Nach der Ausführung von Freksa (2013) in seinem Artikel „Understanding Cognition through Synthesis and Analysis“ definiert sich spatiale Kognition wie folgt:

„Spatial cognition is concerned with cognitive agents [i.e. Menschen, Tiere oder Roboter] in space with the nature of that space (spatial environment), and with the knowledge of the agents about their spatial environment (knowledge representation).“
(Freksa 2013, S.3)

Von zentraler Bedeutung für die Orientierung und Manövrierbarkeit der räumlichen Umgebung ist ihre mentale Repräsentation. Wie stark spatiale Informationen abstrahiert werden können hängt davon ab, für welchen Zweck sie eingesetzt werden sollen. Während ein höherer Abstraktionsgrad tendenziell die Komplexität einer Aufgabe verringert und weniger Speicherkapazitäten beansprucht, wird gleichzeitig erforderlich, dass bestimmte konkrete Informationen ausgespart werden. Eine in diesem Sinn ausbalancierte, mentale Repräsentation ist entscheidend für den erfolgreichen Umgang mit spatialen Informationen (Freksa 2013). Bei der Beschreibung einer Route beispielsweise, werden häufig die Gegebenheiten der spatialen Umgebung selbst als Grundlage einer abstrakten Repräsentation herangezogen. Wird ein Weg in einer urbanen Umgebung so als Abfolge von Abbiegungen beschrieben (e.g. rechts, geradeaus, links, links, rechts), wird die Beschaffenheit der Umgebung selbst, nämlich ihre durch Kreuzungen geordnete Struktur, bereits als Abstraktionsebene eingesetzt. Soll hingegen eine Karte einer realen Umgebung entworfen werden, ist das notwendige Informationsniveau deutlich höher und orientiert sich am gewünschten Abstraktionsgrad der geplanten Karte. Genauso wie in der Audition werden deshalb die kognitive Verarbeitung und Repräsentation von Spatialität selbst durch die räumliche Umgebung und ihre Eigenschaften maßgeblich beeinflusst (Freksa 2013 et al.).

Im Kontrast zu derartigen, abstrakten mentalen Modellen, lassen sich konkrete Repräsentationen dagegen häufig auf unmittelbare Erlebnisse oder Erinnerungen einer Person zurückführen und sind deshalb eng mit dem episodischen Gedächtnis verbunden, das sie aufzeichnet. Kann eine bereits bekannte Route beispielsweise vor dem inneren Auge vollständig visualisiert werden, findet praktisch keine Abstraktion statt; die Repräsentation ist konkret.

Zusätzlich sind die Orientierung und mentale Repräsentation der räumlichen Umgebung von sogenannten „Landmarks“ (deutsch Marksteine oder Orientierungspunkte) abhängig. Erst die Kombination dieser Orientierungspunkte und deren spatiale Relation, e.g. „gehe rechts bei X und gehe von Y aus in Richtung Norden“, erlauben in der Regel einen gekonnten Umgang mit der räumlichen Umgebung. X und Y stellen in diesem Beispiel exemplarische Landmarks dar. Diese können sowohl abstrakter, e.g. ein Knotenpunkt, der auf einer Landkarte identifiziert wurde, als auch konkreterer Natur sein, e.g. eine Kirche, die der jeweiligen Person wohl bekannt ist. Gengen He et al. (2014) stellen in ihrer Untersuchung „Where are you now? Dynamics of Simultaneous, Collaborative Navigation“ heraus, dass die ontologischen Eigenschaften dieser Orientierungspunkte prinzipiell willkürlich gewählt werden können. Gleichzeitig weisen sie aber darauf hin, dass sich manche Landmarks beispielsweise besser für eine Navigationsaufgabe eignen, als andere. Aus den Studienergebnissen geht so etwa hervor, dass Probanden mit schlechte-

rem Orientierungssinn vermehrt dazu neigen, ephemere³⁸ Landmarks (e.g. andere Personen oder Fahrzeuge) zur Navigation zu nutzen. Einige der in dieser Arbeit bereits erwähnten Projekte setzen, wenn auch nicht als Objekt der Untersuchung, eine akustische Form dieser Orientierungspunkte ein – *Acoustic Landmarks* (Rozier 2000, Sundaeswaran et al. 2003, Vazquez-Alvarez et al. 2011 et al.). Aus wissenschaftlicher Sicht gibt es durchaus Grund zur Annahme, dass eine akustische oder akustisch unterstützte Form der Navigation sowohl praktisch, als auch wahrnehmungstheoretisch als sinnvoll erachtet werden kann. Eine genauere Betrachtung dieses Ansatzes folgt in Kapitel 5.

4.2.2 Spatialität als Vehikel der Kognition

Wie einleitend beschrieben, tritt die Räumlichkeit nicht nur als Objekt der Kognition selbst in Erscheinung, sie wirkt auch strukturgebend auf andere Wahrnehmungsprozesse. In ihrer Funktion als Vehikel der Kognition, fungiert die Spatialität zwangsläufig als Rahmen oder Medium aller sinnlichen Erfahrungen. Ebenso wie akustische Signale, breiten sich alle Wahrnehmungsreize im Raum aus und werden so durch seine Struktur und Ordnung *geprägt*. Allein diese permanente Immersion in spatiale Strukturen, so Freksa (2013), führt dazu, dass kognitive Prozesse maßgeblich durch die Spatialität beeinflusst werden. Diese wahrnehmungstheoretische Perspektive greift Freksa auch in seiner Untersuchung auf:

„The [...] perspective is concerned with implications of the fact that cognitive agents and their internal processes are immersed in spatial structures and are themselves spatially structured. As a consequence of the immersion of cognitive agents and their information processing structures in space, spatial structures in the environment must be in some way reflected in the spatial structures of the cognitive agents - otherwise they could have no knowledge about their environment.” (Freksa 2013, S.6)

Nach Freksas Einschätzung geht dieser Einfluss so weit, dass häufig selbst logische Prozesse von der spatialen Struktur der wahrnehmbaren Welt geprägt sind.

Auf dem Hintergrund dieser Einschätzung scheint nicht überraschend, dass auch der auditorische Wahrnehmungsapparat auf räumliche Strukturen rekurriert. In seiner Arbeit „Sound Spheres: A Model of Psychoacoustic Space in Cinema“ beleuchtet David Sonnenschein (2011) speziell diesen Aspekt der Audition. Während hinsichtlich auditorischer Ortungsprozesse bereits umfassende Studien existieren, ist wenig darüber bekannt, welchen Einfluss die Verortung einer Klangquelle auf ihre Bedeutung hat:

„We react to not only the content of the sound, but also where it exists in our physical space. If it is coming toward us, it can have a very different meaning than if it is going away. If we can see it, it may have a very different impact than if we cannot see it. These conditions are inherent in our perceptual construct of the physical world [...]“ (Sonnenschein 2011, S.14)

³⁸ I.e. nicht permanent, vergänglich.

Aus der kognitiven Psychologie ist, im Rahmen der Auditory Scene Analysis, inzwischen bekannt, dass durchaus verschiedene auditorische Wahrnehmungsniveaus existieren, die in Abhängigkeit zu den spatialen Informationen stehen, die über eine Klangquelle bekannt sind (Sonnenschein 2011, Bregman 1990 et al.). Wenn ein Signal beispielsweise als separat erkannt wird, es dem Zuhörer aber nicht möglich ist, seine Position oder Herkunft zu bestimmen, liegt eine bloße *Diskriminierung (Discrimination)* vor. Kann das Geräusch jedenfalls einer bestimmten Klangart oder Klasse zugeordnet werden, ist das Niveau der *Klassifizierung (Classification)* erreicht. Kann das Geräusch wiedererkannt oder einem bekannten, akustischen Ereignis zugeordnet werden, handelt es sich bereits um eine *Erkennung (Recognition)*. Wenn der Zuhörer in der Lage ist, das Signal einer eingeschränkten Menge an potentiellen Quellen in der Umgebung zuzuordnen, wird von einer *Zuordnung (Matching)* gesprochen. Die höchsten beiden Wahrnehmungsniveaus bilden die *Identifikation (Identification)* und *Korrespondenz (Correspondence)*. Eine Identifikation setzt ein, sobald der Zuhörer ein Signal sicher einer Quelle zuordnen kann. Befindet sich diese Quelle zusätzlich in der sichtbaren Umgebung des Zuhörers und kann vollständig verortet werden, ist die Ebene der Korrespondenz erreicht. Je höher die erreichte Wahrnehmungsebene, desto sicherer auch der Umgang mit dem entsprechenden akustischen Ereignis. Sonnenschein beschreibt diesen Zusammenhang anhand eines eigenen Erlebnisses:

„For example, when I couldn't figure out what the source of a sound was, that caused me anxiety. Upon seeing where it came from I would be able to relax or go into action depending on my reaction to the true nature of the sound source.“ (Sonnenschein 2011, S.13)

Akustische Informationen, die in der räumlichen Umgebung angemessen verortet sind, werden intuitiv als verlässlicher wahrgenommen. Entscheidungen auf Basis dieser subjektiv verlässlicheren Informationen zu treffen (e.g. das Einschlagen einer Richtung), fällt dem Zuhörer deshalb in der Regel leichter. Dieser Aspekt könnte beispielsweise dafür eingesetzt werden, die Unsicherheit unerfahrener Benutzer bei der Bedienung auditorischer Displays zu kompensieren.

Im Rahmen von Kapitel 3.4.2 wurde auf dem Hintergrund Grossbergs Adaptive Resonance Theory (ART) das Realitätskriterium für Wahrnehmungsreize erläutert. Im Fall der Spatialisierung von Audiosignalen könnte der räumliche Wahrnehmungsimpuls in diesem Sinne eine geeignete Paarung zum akustischen Ereignis erzeugen, mit deren Hilfe das Realitätskriterium erfüllt werden könnte. Dass eine räumliche Aufteilung akustischer Signale dem Wahrnehmungsapparat mit anderen Worten „realer“ erscheint, als die monotone Wiedergabe aus einem Lautsprecher, könnte entsprechend erklären, weshalb spatialisierte Audiosignale ein höheres Wahrnehmungsniveau erreichen (Grossberg 1995, Sonnenschein 2011).

4.2.3 Aural Fluency Theorie (nach Droumewa & Wakkary)

Neben diesen unmittelbaren Wahrnehmungseffekten scheint Spatialität auch einen eher indirekten Einfluss auszuüben, der das Bedienerlebnis von akustischen Benutzeroberflächen prägt und über potentielle Lernerfolge entscheidet. Wie in Kapitel 3.7 bereits angedeutet wurde, bilden Droumewa und Wakkary in ihrer „Aural Fluency“-Theorie (Droumewa & Wakkary 2008) ab,

dass speziell die Bewegung in einer akustischen Umgebung und das räumliche Erleben von Klang überdurchschnittliche Lernerfolge unterstützen.

Grundlage ihrer Studie ist ein „Ambient Intelligence“-Spiel mit dem Titel „socio-ec(h)o“. Socio-ec(h)o wird durch die unmittelbare körperliche Interaktion eines vierköpfigen Teams innerhalb der Ambient Intelligence Umgebung gespielt. Informationen werden in erster Linie durch akustische Signale transportiert, einige Lichteffekte unterstützen die Informationsübertragung. Die auditorische Benutzeroberfläche setzt sich aus insgesamt drei Komponenten zusammen: einer Echtzeit-Sonification verschiedener Klangumgebungen für jedes Level, einem antizipatorischen Feedbacksignal, das die Zusammenarbeit der Gruppe in Richtung eines Ziels unterstützt und einem Bestätigungssignal, das die erfolgreiche Beendung einer Aufgabe anzeigt. Während des Spiels müssen Worträtsel in sechs verschiedenen Schwierigkeitsgraden bewältigt werden. Gesteuert wird das Spiel durch verschiedene Bewegungen oder Körperpositionen der Spieler. Ziel der Studie war die Untersuchung der spielerischen Interaktion mit dem beschriebenen System. Dabei wurde den Testgruppen genügend Zeit eingeräumt, den Umgang mit den spatialsierten Klanginformationen der interaktiven Umgebung zu erlernen.

Das Nutzerfeedback und die im Rahmen der Studie gesammelten Beobachtungen bilden die Grundlage der Theorie von Droumewa und Wakkary, die die Kompetenzzaneignung im Umgang mit Spatial Auditory Displays als *auditorische Literalität* (i.e. die Fähigkeit komplexe, akustische Informationen nach dem Vorbild der Lese- und Schreibfähigkeit kognitiv zu verarbeiten) beschreiben (Droumewa & Wakkary 2008). Auch Aggarwal et al. (2014) berichten in ihrer Untersuchung von einer grundsätzlichen Kompetenz der Probanden auditorische Benutzeroberflächen erfolgreich und sicher zu bedienen. Die Aural Fluency Theorie ist insofern interessant für die vorliegende Arbeit, als dass sie eine Abhängigkeit dieser auditorischen Literalität von einer physikalischen Verortung akustischer Ereignisse unterstellt:

„[T]he type of auditory literacy and fluency that is needed and seems to develop in more physical, situated technological environments such as ambient intelligent spaces seems more akin to everyday listening [...]. It is situated in a social, contextual and shared physical environment and takes on characteristics of listening in everyday life, including dynamic shifts of attention, listening comprehension and aural expertise.“ (Droumewa & Wakkary 2008, S.5)

Ähnlich wie die Lese- und Schreibfähigkeit, die zunächst in einem langwierigen Prozess erlernt werden muss, muss auch der Umgang mit auditorischen Displays geübt werden. Durch die Dominanz der visuellen Informationspräsentation in der digitalen Welt bleiben entsprechende Lernprozesse allerdings aus. Wie der wahrnehmungstheoretische Hintergrund dieser Arbeit und die von Droumewa und Wakkary präsentierte Theorie nahelegen, sind sowohl die Sicherheit im Umgang mit akustischer Information, als auch der auditorische Lernprozess maßgeblich von der spatialen Repräsentation akustischer Ereignisse abhängig.

4.3 Bedeutung von Spatialität für Hören und Zuhören

Jedes akustische Signal setzt sich aus seiner physikalischen Struktur (akustische Syntax) einerseits und seiner Bedeutung als Hörereignis (auditorische Semantik) andererseits zusammen. Auf dem Hintergrund der Unterscheidungen von Truax und Chion wurde bereits in Kapitel 3.2.2 darauf verwiesen, dass in entsprechender Art und Weise auch zwischen dem reinen Hören und einem bewussten Zuhören unterschieden werden kann. Während sich Hören mit der mentalen Repräsentation der physikalischen Klangumgebung befasst, konzentriert sich der Prozess des Zuhörens auf die Interpretation der individuellen Hörereignisse. Beide Prozesse werden allerdings gleichermaßen durch ihre spatiale Umgebung beeinflusst.

4.3.1 Hören – Spatialität als Wiedergabeparameter

Für das auditorische System besteht hinsichtlich der mentalen Umgebungsrepräsentation ein deutlicher Unterschied zwischen der monotonen Wiedergabe eines Klangstroms oder der spatial aufgefächerten Struktur eines akustischen Hologramms. Elektroakustische Artefakte, die der räumlichen Anordnung innerhalb einer natürlichen Klanglandschaft zu widersprechen scheinen, werden in aller Regel als künstlich erkannt und lassen deshalb wenig Hoffnung auf eine natürliche Interaktion mit dem Nutzer. Erzeugt beispielsweise die Wiedergabe mittels Kopfhörer einen intrakraniellen Klangstrom, dessen Quelle scheinbar im Kopf des Hörers liegt, steht für das auditorische System außer Frage, dass es sich um einen artifiziellen Reiz handelt.

Folgt man den Ausführungen von Vazquez-Alvarez et al. (2011) hängt das Nutzererlebnis bei der Interaktion mit auditorischen Benutzeroberflächen häufig unmittelbar von der Spatialität als Wiedergabeparameter ab.

Der von Vazquez-Alvarez et al. für ihre Studie entworfene Soundgarden ist, ganz seinem Namen entsprechend, als auditorische Erlebnisumgebung in Form eines frei begehbaren Gartens angelegt³⁹, der durch verschiedene Wiedergabemodi akustisch (e.g. Vogelgezwitscher) ausgestaltet wird. Die Übertragung der akustischen Signale findet über ein mobiles Kopfhörersystem statt. Alle Testpersonen erhalten die Aufgabe den Soundgarden zu erkunden, während ihre Bewegungsmuster über ein Trackingsystem erfasst werden.

³⁹ Natürlich handelt es sich bei diesem Projekt um eine Audio-Augmented Reality (AAR) Umgebung im klassischen Sinne. Da eine formale Zuordnung der AAR zur Gruppe der Spatial Auditory Displays erst in Kapitel 5.2 erfolgt, wird auf eine Spezifizierung an dieser Stelle noch verzichtet, um unnötige Unklarheiten zu vermeiden.



Abbildung 10: Soundgarden (Vazquez-Alvarez et al. 2011, S.4)

Auf der linken Seite von Abbildung 10 ist auf dem höchsten Punkt der Kopfhörerschiene der JAKE-Sensor (1) zu erkennen, mit dessen Hilfe die Ausrichtung des Kopfes erkannt und via Bluetooth übermittelt wird. Außerdem befindet sich linksseitig am Kopfhörergestell auch der GPS-Receiver (2), der in Echtzeit die Position des Nutzers bestimmt und ebenfalls via Bluetooth mit dem Mobiltelefon (3) verbunden ist. Auf der rechten Seite der Abbildung findet sich eine schematische Darstellung des Soundgardens. Insgesamt existieren auf dem Gelände fünf Landmarken, an die akustische Signale geheftet wurden. Die räumlichen Landmarken wurden so zu Acoustic Landmarks erweitert, wie sie bereits in Kapitel 4.2.1 Erwähnung gefunden haben. Jede akustische Landmarke verfügt über einen Nahbereich (Proximity Zone) mit einem Radius von 25 m und einen Aktivierungsbereich (Activation Zone) mit einem Radius von 10 m, die genutzt werden können, um die Ausgabe des Audiofeedbacks zu steuern.

Das Gesamtexperiment teilt sich auf 4 Durchgänge auf, in denen die spatiale Ausgestaltung der synthetischen Klanglandschaft sukzessiv erhöht wird. Im ersten Modus (*Baseline*) wird der zugeordnete, akustische Inhalt einmalig ausgegeben, wenn der Nutzer den Aktivierungsbereich betritt. Der Nahbereich wird nicht verwendet. Im zweiten Modus (*Earcons*) werden zusätzlich Earcons eingesetzt. Im dritten Modus (*Spatial*) findet eine räumliche Eingliederung durch die Nutzung des Nahbereichs statt. Befindet sich der Nutzer im Nahbereich eines Acoustic Landmarks, wird die Lautstärke des Audiofeedbacks entfernungsgerecht angepasst. Im letzten Modus (*Spatial 3D*) ist die synthetische Klanglandschaft vollständig spatialisiert. Alle Klangquellen sind räumlich so fixiert, dass sie sich mittels Lautstärkenabweichung und Richtungsveränderung eindeutig lokalisieren lassen - sie verhalten sich annähernd wie ihr natürliches Pendant.

Im Anschluss an jeden Durchgang wird eine umfassende Nutzerbefragung durchgeführt. Anhand der Aussagen der Testpersonen zeichnet sich ab, wie entscheiden die spatiale Ausgestaltung und physikalische Begehbarkeit für das Nutzererlebnis ist:

„The users highlighted that ‘the experience of moving around to get the information was good’ and the ‘novelty of coming across the message like stumbling across something on your way. Serendipity and wonder’. [...] During the Spatial 3D condition, participants reported a sense of ‘discovery’ and that the sound garden was ‘quite immer-

sive'. The participants in this condition liked the experience because 'you rely only on your hearing' and often closed their eyes in order to listen to the Earcons. They found the system curious because 'you know sounds come from headphones but it sounds like it is coming from the outside'.“ (Vazquez-Alvarez et al. 2011, S.8-9)

Die räumliche Komposition der synthetischen Klangumgebung nach ihrem natürlichen Vorbild unterstützt in diesem Beispiel eine Immersion des Nutzers in den Erlebnisraum des Soundgardens. Die Ergebnisse von Vazquez-Alvarez et al. verdeutlichen die Diskrepanz zwischen auditivischen Systemen, die eine spatiale Dimension integrieren und solchen, die darauf verzichten. Gleichzeitig verweisen Vazquez-Alvarez et al. auf die technische Umsetzbarkeit spatialisierter Systeme:

„In conclusion, the combination of 3D spatial audio techniques together with Earcons was the most effective auditory display. This work suggests that the location and orientation sensing technologies now present in commercially available smartphones can be used to create rich and compelling outdoor audio-augmented environments.“ (Vazquez-Alvarez et al. 2011, S.12)

Die Grundlage jeder Form der Spatialisierung akustischer Signale ist die Reproduktion oder Imitation des natürlichen Wellenfeldmusters des physikalischen Schalls (Duraiswami et al. 2005, Faller & Baumgarte 2001, Zotkin et al. 2004, Härmä et al. 2003). Dieser Effekt kann sowohl durch Rauminstallationen, als auch durch mobile Wiedergabesysteme erzielt werden.

Im Falle des Soundgardens wäre es beispielweise auch möglich gewesen, einen Lautsprecher an jeder akustischen Landmarke zu installieren. Anstelle einer Simulation könnte so eine Klangumgebung räumlich nachgebaut werden, in der die natürlichen Klangquellen (e.g. Vogelgezwitscher) durch elektroakustische Klangquellen (i.e. Lautsprecher) ersetzt würden. Die resultierende Klanglandschaft entspräche so natürlich (näherungsweise) der zu simulierenden Gartenumgebung. Eine derartige Installation ist aber vor allem deshalb nicht zu empfehlen, weil entsprechend jede einzelne Klangquelle durch einen eigenen Lautsprecher repräsentiert werden müsste. Sollte die Konzeption der virtuellen Klanglandschaft jemals geändert werden, würde dies umfassende Umbaumaßnahmen nach sich ziehen. Heutige *Wellenfeldsynthese-Systeme (WFS-Systeme)* bieten eine sinnvolle Alternative zu diesem Ansatz, die eine physikalische Repräsentation obsolet werden lässt. WFS-Anwendungen erzeugen durch eine Lautsprecherreihung akustische Hologramme, deren Wellenfeld synthetisiert wird. Derartige Einrichtungen sind zwar häufig kostspielig, können aber prinzipiell beliebige Klangquellenkonstellationen (näherungsweise) simulieren und eignen sich insofern speziell für die Disziplin der *Virtual Reality (VR)*.

Die physikalisch-theoretische Grundlage synthetisierter Wellenfelder findet auch in mobilen Systemen Anwendung. Mobile Systeme zeichnen sich vor allem durch ihre Portabilität, Flexibilität und zumeist auch relative Kosteneffizienz aus. Da sie frei durch den Raum bewegt werden können, sind sie speziell für den Bereich der *Augmented Reality (AR)* geeignet, wie er beispielsweise im Rahmen des Soundgarten-Projektes vorgestellt wurde. In der Regel werden akustische

Reize hier über Kopfhörer, i.e. binaural, wiedergegeben. Grundsätzlich lässt sich das schematische Wirkprinzip dieser binauralen Wiedergabe wie in Abbildung 10 zusammenfassen:

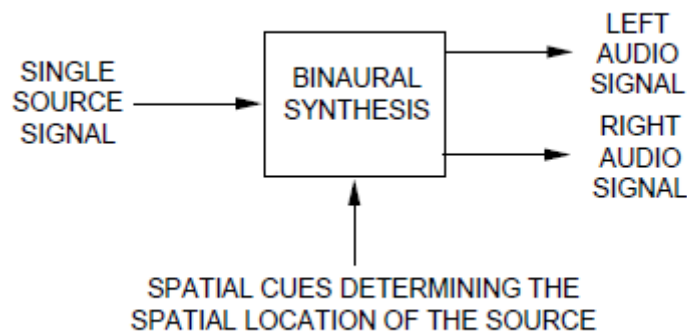


Abbildung 11: Modell der binauralen Synthese (Faller & Baumgarte 2001, S.1)

Zunächst liegt eine einfache Klangquelle ohne spatiale Zuordnung vor. Diese wird im Raum virtuell einer bestimmten Position zugeordnet, deren spatiale Parameter über die spätere Wiedergabe der Klangquelle entscheiden. Dieser Vorgang wird für alle Quellen wiederholt, die im Raum verteilt werden sollen. Auf Basis der Theorie der Wellenfeldsynthese wird anschließend mit Hilfe der spatialen Parameter virtuell ein Wellenfeld synthetisiert, welches als spatiale Aggregation der einzelnen Klangströme bzw. Klangquellen verstanden werden kann. Durch die Anpassung der interauralen Phasendifferenz (ITD/IPD), der interauralen Niveaudifferenz (ILD) und gegebenenfalls auch der (näherungsweise)⁴⁰ individuellen Head-Related Transfer Function (HRTF) in der binauralen Wiedergabe, können die Klangquellen virtuell jeder Position im Raum zugeordnet werden (Duraiwami et al. 2005, Faller & Baumgarte 2001). Das resultierende Signal wird anschließend über zwei Kanäle dem linken und rechten Ohr zugeordnet. Im Hinblick auf die Grenzen und den Fokus der vorliegenden Arbeit soll an dieser Stelle lediglich auf die mathematische Ausarbeitung von Ramani Duraiswami et al. (2005) und Christof Faller & Frank Baumgarte (2001) verwiesen sein. Auf eine vollständige Darstellung der mathematischen Grundlagen muss leider verzichtet werden.

Die Position und Lage des Kopfes des Nutzers im Wellenfeld sind speziell wegen der Verteilung auf beide Gehörgänge und der Anwendung der spatialen Parameter von zentraler Bedeutung für das Verfahren. Da die akustischen Eigenschaften, die der auditorische Wahrnehmungsapparat zur Lokalisierung von Klangquellen heranzieht, imitiert werden, entsteht für den Benutzer ein natürliches, räumliches Hörereignis. Im besten Fall ist die synthetisierte Klanglandschaft nicht mehr von ihrem natürlichen Pendant zu unterscheiden:

⁴⁰ Da sich die Berechnung der individuellen HRTF als äußerst komplex erweist, werden häufig bestehende Näherungswerte aus Datenbanken herangezogen. Eine umfassendere Beschreibung findet sich in Zotkin et al. (2004).

„The listening test results show that indeed pseudo-acoustic sound can be almost indistinguishably imitated. The test subjects were able to recognize or guess the source of the sound correctly only in 68% of the cases (pure guessing gives 50 %).“ (Härmä et al. 2003, S.6)

Ob der Nutzer in der Lage ist, die artifizielle von einer natürlichen Klangumgebung zu unterscheiden, fungiert gleichzeitig als Kriterium für den von Aki Härmä et al. (2003) beschriebenen *Turing-Test* für virtuelle Audio-Umgebungen. Härmä et al. präsentieren in ihrem Artikel „Techniques and applications of wearable augmented reality audio“ (2003) die vollständige Architektur für ein Spatial Auditory Display im Rahmen einer AAR-Anwendung mit dem Titel *WARA*⁴¹.

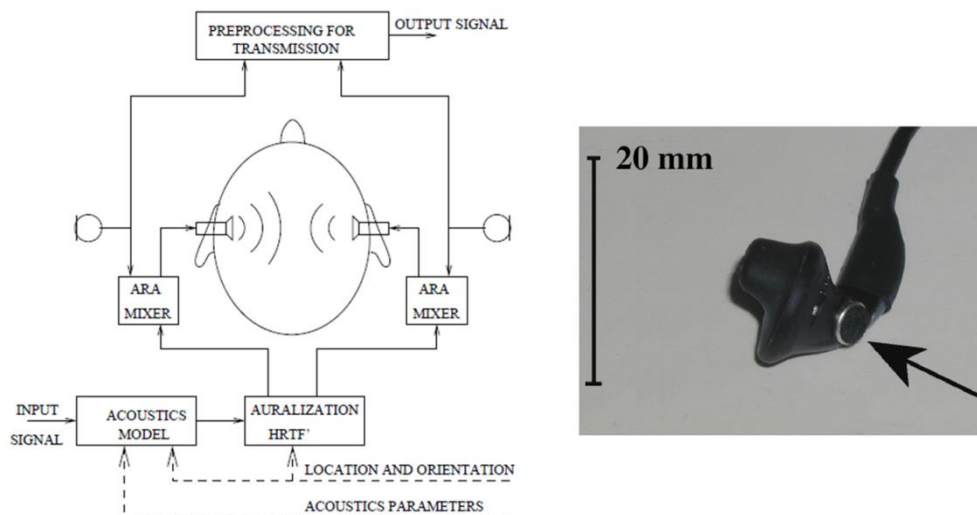


Abbildung 12: Wearable Augmented Reality Audio (WARA) (Härmä et al. 2003, S.5 ff)

Wie sich aus Abbildung 12 erschließt, findet die Übertragung akustischer Reize an den Nutzer mittels Kopfhörer statt. Um eine Integration synthetischer Klangquellen in die reale, akustische Umgebung des Zuhörers zu ermöglichen, verfügen die Kopfhörer über Mikrophone, die eingehende Schallereignisse separat für den linken und rechten Gehörgang aufnehmen. Die Impulse der realen Umgebung werden anschließend nach dem spatialen Audio-Modell des Systems zusammen mit den artifiziiellen Klangquellen der AAR-Anwendung in ein gemeinsames, synthetisiertes Wellenfeld integriert und wiedergegeben. Ziel dieser Integration ist die Ausgestaltung der elektroakustischen Klanglandschaft nach holistischem Prinzip. Obwohl die akustische Umgebungsrealität des Nutzers (annähernd) originalgetreu wiedergegeben wird⁴², ist es für diesen im

⁴¹ Wearable Augmented Reality Audio (WARA).

⁴² Auch bei diesem System kann es sich natürlich nur um eine gute Näherung handeln. Auf dem Hintergrund erfolgreicher Turingtests für diese Architektur kann aber jedenfalls von einer annähernden Transparenz der Technologie gesprochen werden. Härmä et al. bezeichnen die reproduzierten, realen Klangumgebungen deshalb als pseudo-akustische Umgebungen (Pseudo-Acoustic Environments).

besten Fall⁴³ nicht mehr möglich, reale von synthetischen Klangquellen zu unterscheiden. In ihrem Artikel erwägen Härmä et al. (2003), dass ein AAR-System nach diesem Vorbild genutzt werden könnte, um alle existierenden, auditorischen Anwendungen (e.g. Telefonie, verbale Navigationsunterstützung, Musikwiedergabe, AAR-Effekte) für den Nutzer zusammenzuführen. Gleichzeitig könnte die Integration in die reale Klangumgebung eine *nahtlose Wahrnehmungserfahrung* unterstützen. Hinsichtlich der technischen Umsetzbarkeit schließen Härmä et al.: „Mechanical and electrical integration is already feasible.“ (2003, S.3)

Natürlich bringt auch diese Architektur einige Herausforderungen mit sich. Zunächst ist eine störungsfreie Anwendung überhaupt nur dann denkbar, wenn das System keine wahrnehmbaren Latenzen erzeugt. Obwohl die spatiale Auffächerung akustischer Signale zunächst einen großen Rendering-Aufwand suggeriert, hält sich das Datenvolumen im Vergleich zu visuellen Anwendungen durchaus in bewältigbaren Grenzen. Zusätzlich zeigen Arbeiten wie die von Faller & Baumgarte (2001) vielversprechende Ansätze auf, durch die der Rendering-Aufwand akustischer Systeme weiter reduziert werden könnte. Wie in praktisch jedem Augmented Reality-System (AR-Systeme) stellen außerdem auch in dieser Architektur das erfolgreiche Indoor- und Outdoor-Tracking und die ausreichende Akkulaufzeit der mobilen Geräte kritische Systemgrößen dar. Auf dem Hintergrund dieser Arbeit muss aber darauf hingewiesen werden, dass diese Probleme heute in mindestens gleichem Maße bei der Entwicklung visueller AR-Systeme auftreten. Gleichzeitig liegt das Niveau räumlicher Audio-Hologramme deutlich über dem ihrer visuellen Pendanten. Studien wie die von Dmitry N. Zotkin et al. wiesen schon 2004 auf das Potential akustischer Displays für die glaubwürdige Integration synthetischer Signale in bestehende Klangumgebungen hin. Im Zentrum dieser Technologie, steht die Spatialität als Wiedergabeparameter.

4.3.2 Zuhören – Spatialität als narrativer Kontext

Die räumliche Erlebbarkeit von Klang ist entscheidend für die *Art und Weise* in der akustische Ereignisse wahrgenommen werden. Gleichzeitig fungiert die spatiale Dimension als Bedeutungskontext für akustische Informationen. Maribeth Beck und D. Des von Xerox Parc (1996) stellen deshalb in ihrem Konzept für „Micro-Narratives in Sound Design“ die *Mikrostruktur des Raums* (*The Microstructure of Space*) als Kontextebene der Narration vor. Die räumliche Struktur einer akustischen Erzählung und die mit ihr einhergehenden Klangcharakteristika können entscheidend für die Vorstellungskraft des Nutzers sein. Erst die Eingliederung in eine sinnvolle Klangumgebung weist einem akustischen Signal die gewünschte Bedeutung zu. Diese Einschätzung deckt sich mit Truax' Darstellung des kommunikativen Ansatzes (Truax 2000). Gemessen an der Vehemenz mit der sich die individuelle Hallstruktur und erlebbare Klangumgebung einer Eishöhle von den Eigenschaften eines Waldstückes unterscheiden, ist nicht verwunderliche, dass diese Faktoren starken Einfluss auf die Dynamik einer Geschichte nehmen. Die Mög-

⁴³ Gerade komplexe Hallmuster bestimmter augmentierter Umgebungen erschweren eine überzeugende Replikation.

lichkeit beispielsweise Hörbücher durch akustische Hologramme zu erzählen soll deshalb in Kapitel 5.3 genauer betrachtet werden.

Ebenso wie die fiktive Umgebung einer Erzählung kann auch ein realer Ort die Bedeutung akustischer Reize beeinflussen oder unterstützen. Wie in Kapitel 3.8 beschrieben, stellt sich gerade bei musikalischen Einflüssen häufig das Phänomen der bildlichen Assoziation (Visual Imagery) ein. So erzeugt ein bestimmtes Stück beispielsweise die Impression einer belebten Stadt oder einer hügeligen Landschaft. Folgt man dieser kognitiven Stoßrichtung, liegt die Eingebung nicht weit, Musik speziell für bestimmte Orte zu komponieren, um so eine harmonische Gesamtkomposition aus räumlicher Umgebung und akustischer Untermalung zu erschaffen. Ryan Holladay hat eben diese Idee im Oktober 2013 in San Francisco im Rahmen eines TED-Talks präsentiert (Holladay 2013).

„This is Central Park, one of the most beautifully designed public spaces in America. But to anyone who hasn't visited, these images can't really fully convey. To really understand Central Park, you have to physically be there. Well, the same is true of the music, which my brother and I composed and mapped specifically for Central Park.“ (Holladay 2013, S.1)

Das Album, das Holladay in dieser Rede vorstellt, ist nicht im Handel erhältlich. Die einzige Art und Weise seine Musik zu hören, ist durch einen Besuch im Central Park. Die landschaftliche Umgebung wird insofern als intrinsisch für die Komposition betrachtet:

„We're sometimes contacted by people in other parts of the world who can't travel to the United States, but would like to hear this record. Well, unlike a normal album, we haven't been able to accommodate this request. When they ask for a C.D. or an MP3 version, we just can't make that happen, and the reason is because this isn't a promotional app or a game to promote or accompany the release of a traditional record. In this case, the app is the work itself, and the architecture of the landscape is intrinsic to the listening experience.“ (Holladay 2013, S.2)

Die einzelnen Stücke wurden intendiert auf bestimmte Teile des Parks abgestimmt und entsprechend komponiert. Das Resultat bezeichnet Holladay als „location-aware music“. Ziel ist es, der Musik im unmittelbar wörtlichen Sinne *Raum* zu geben und so das Hörerlebnis eklatant zu bereichern. Durch diese spatiale und landschaftliche Erweiterung wird eine Kaskade kognitiver Prozesse ausgelöst, wie sie in Kapitel 3.8 beschrieben wurden. Es ist so für den Künstler möglich viel kontrollierter und umfassender Gefühle und Emotionen zu induzieren oder eine bestimmte Geschichte zu erzählen. Entsprechend unbescheiden fällt auch Holladay's abschließende Einschätzung aus:

„[I]t speaks to a larger vision for a music industry that's sometimes struggled to find its footing in this digital age, that they begin to see these new technologies not simply as

ways of adding bells and whistles to an existing model, but to dream up entirely new ways for people to interact with and experience music.“ (Holladay 2013, S.2-3)

Eine solche *Symbiose zwischen Räumlichkeit und akustischer Narration* könnte in Zukunft häufiger bemüht werden, um gezielt die Erzählstrukturen zu generieren, die für eine Geschichte notwendig sind. Auch die in Kapitel 4.2.3 im Rahmen der Aural Fluency Theorie vorgestellten Ergebnisse von Droumewa und Wakkary (2008) weisen auf einen konkreten Zusammenhang zwischen Narration, räumlicher Umgebung und der Verkörperung akustischer Stimuli hin: Die holistische Erfahrung der auditorischen Literalität, so Droumewa und Wakkary, erwächst letztlich aus der Interaktion mit einer bestimmten Umgebung auf dem Hintergrund einer entsprechenden Narration. Ihre Beobachtungen unterstützen demnach die Annahme, dass Nutzer einer spatialisierten Audio-Anwendung eine *narrative Assoziation* zu ihrer akustischen Umgebung entwickeln (Droumewa & Wakkary 2008, S.5-6).

4.4 Konklusion

Ob die Zukunft der digitalen Akustik langfristig in ihrer Verwurzelung in der spatialen Dimension liegt, bleibt abzuwarten. Dass der Spatialität aber als entscheidende, wahrnehmungstheoretische Einflussgröße ein entsprechendes Gewicht bei der Gestaltung auditorischer Anwendungen eingeräumt werden sollte, steht jedenfalls theoretisch außer Frage.

Die gezielte Verortung synthetischer Klangquellen im Raum erlaubt es der Elektroakustik, die Ordnung natürlicher Klanglandschaften zu imitieren, statt schizophone Hörereignisse billigend in Kauf zu nehmen. Diese für das auditorische System gewohnten Strukturen erzeugen ein höheres Maß an subjektiver Wahrnehmungssicherheit und unterstützen so eine benutzerfreundliche und intuitive Bedienung auditorischer Benutzeroberflächen (Grossberg 1995, Sonnenschein 2011 et al.). Zusätzlich können die narrativen Strukturen einer akustischen Erzählung durch eine Verquickung mit der spatialen Umgebung des Nutzers stärker herausgearbeitet oder geschärft werden. Der resultierende Bedeutungszuwachs ermöglicht die Erfüllung der originären Funktion akustischer Signale innerhalb der Narration – die Übertragung von Information (Beck & Des 1996, Holladay 2013 et al.).

Um diese Vorzüge zugänglich zu machen, muss die Spatialität als Parameter für die Wiedergabe akustischer Inhalte bemüht werden. Anders als in der Disziplin der visuellen Informationsübertragung stehen die dafür notwendigen Technologien und Architekturen bereits auf angemessenem Niveau zur Verfügung und sind verhältnismäßig frei zugänglich (Duraiswami et al. 2005, Faller & Baumgarte 2001, Zotkin et al. 2004, Härmä et al. 2003 et al.). Es stellt sich deshalb in der Tat die Frage, weshalb die Technologie spatialer Audio-Benutzeroberflächen bis dato vergleichsweise wenig praktische Anerkennung gefunden hat. Das nächste Kapitel beschäftigt sich deshalb mit dem praktischen Einsatz von Spatial Auditory Displays und deren wahrnehmungstheoretischer Eignung.

5 Spatial-auditorische Systeme

5.1 Definition und Eignungsuntersuchung von Spatial Auditory Displays

Bereits in Kapitel 3.6 wurde die Domäne auditorischer Displays anhand der Definition von Walker und Kramer eingegrenzt:

„Auditory display is a broad term referring to the use of any type of sound to present information to a listener. This may include, but is certainly not limited to, warnings, alarms, status indicators, and data sonification.“ (Walker & Kramer 2004, S.3)

Grundsätzlich ist nur eine Erweiterung hinsichtlich der spatialen Ausgestaltung der auditorischen Displays notwendig, um diese Definition auch auf Spatial Auditory Displays anwenden zu können. Insofern kann ein Spatial Auditory Display (SAD) verstanden werden als „Vorrichtung, die Informationen mittels räumlich repräsentierter, akustischer Signale, i.e. verbal oder non-verbal, an den Nutzer übermittelt“. Ob diese Eingliederung in die spatiale Dimension dabei durch eine Integration in eine reale Klangumgebung erzielt (e.g. Audio-Augmented Reality) oder der Wiedergabemodus schlicht durch räumliche Parameter ergänzt wird (e.g. Virtual Reality mittels Wellenfeldsynthese), ist dabei zunächst irrelevant. Wie im Fall non-spatialer Anwendungen kann die Palette akustischen Stimuli von einfachen Alarmsignalen bis hin zu komplexen Sonifications reichen. Beispiele wie socio-ec(h)o (Droumewa & Wakkary 2008) oder das von Singer et al. (1996) entwickelte Audio-Konferenz-System, die im Laufe dieser Arbeit vorgestellt wurden, unterstreichen die Vielseitigkeit spatialisierter Audio-Displays.

Aus den Kapiteln 3 und 4 lassen sich alle notwendigen, wahrnehmungstheoretischen Grundlagen für den Einsatz von SADs ableiten. Speziell im 4. Kapitel wird die kognitive Relation von Klang und Raum thematisiert, die als elementar für alle spatialen Audio-Anwendungen verstanden werden kann. Ziel dieser Arbeit war es, die wahrnehmungstheoretische Eignung auditorischer Displays und Benutzeroberflächen im Allgemeinen und spatialisierter Anwendungen im Speziellen zu überprüfen. Nachdem diese Evaluation für auditorische Displays und Benutzeroberflächen bereits in Kapitel 3.10 durchgeführt wurde, liegt es nahe, dieselben Sinnkriterien auch für die wahrnehmungstheoretische Bewertung von SADs heranzuziehen. Zum Zweck dieser Bewertung wird die Hypothese H1.1 als Unterhypothese zu H1 zugrunde gelegt:

H1.1 Die Spatialisierung auditorischer Displays und Benutzeroberflächen ist aus wahrnehmungstheoretischer Sicht sinnvoll.

In Anlehnung an die Herleitung aus Kapitel 3.10, ist eine solche Spatialisierung akustischer Signale innerhalb eines SADs immer dann sinnvoll, wenn:

- (i) sie *handlungsermöglichend* oder *handlungsweisend* ist, i.e. dem Nutzer bestimmte Handlungen durch adäquate Information ermöglicht.
- (ii) sie *intendiert Affekt erzeugt*, i.e. den Nutzer emotional beeinflusst.
- (iii) sie der *Unterhaltung* dient, i.e. Neugier im Nutzer erzeugt oder ihn in eine spielerische Interaktion verwickelt.

Nachdem H1 („Der Einsatz kontextabhängiger, auditorischer Displays und Benutzeroberflächen ist aus wahrnehmungstheoretischer Sicht *sinnvoll*“) bereits in Kapitel 3.10 bestätigt werden konnte und sich die Klasse der SADs formal als Unterkategorie auditorischer Displays definiert, ist von einer wahrnehmungstheoretischen Eignung natürlich auch im Falle der SADs auszugehen. Dennoch lohnt sich eine detailliertere Betrachtung.

Ausarbeitungen wie die von Sonnenschein (2011) oder Singer et al. (1996) zeigen, wie durch die Verknüpfung mit der natürlichen Umgebung des Nutzers ein sicherer Umgang mit auditorischen Displays gefördert werden kann. Gleichzeitig erhöht die Einbeziehung der spatialen Dimension innerhalb eines auditorischen Displays die Informationsdichte der akustischen Signale. Im Beispiel von Robert W. Proctor und T. Gilmour Reeve (1990) symbolisiert die räumliche Position der übertragenen Schallereignisse beispielsweise die Zugehörigkeit zu einem bestimmten Unternehmensstandort. Wie bereits in Kapitel 3.6.4 dargestellt, können dem Nutzer so parallel die Umsatzzahlen verschiedener Standorte präsentiert werden, ohne ihn zu verwirren. Die Ordnungsstrukturen räumlicher Umgebungen (i.e. Höhe, Breite, Tiefe) können so als zusätzliche Referenz für das auditorische Displays instrumentalisiert werden. Das Informationspotential auditorischer Displays und Benutzeroberflächen wird insofern durch eine Spatialisierung erhöht. Nachdem das Kriterium (i) bereits in der ersten Untersuchung erfüllt werden konnte, steht außer Frage, dass auch SADs diesem Anspruch gerecht werden.

Im Hinblick auf Kriterium (ii) kann zunächst geschlossen werden, dass eine spatialisierte Anwendung zumindest im selben Maße in der Lage ist Emotionen zu induzieren, wie ihr nicht-spatiales Pendant. Allerdings ergeben sich gegebenenfalls Unterschiede hinsichtlich der Zuordnung von Aufmerksamkeit im Wahrnehmungsprozess. Als „Suspension of Disbelief“ wird von Cohen im Rahmen des Congruence-Associationist Framework (CAF) auf den Effekt von Stimulipaarungen im Hinblick auf das Realitätskriterium der Wahrnehmung hingewiesen⁴⁴ (Cohen 2000). Ob der Wahrnehmungsapparat einen (synthetischen) Umgebungsreiz als „real“ anerkennt hängt davon ab, in welchem Wahrnehmungskontext der Reiz steht. Durch eine adäquate Paarung mit der räumlichen Umgebung des Nutzers könnte diesem Kriterium leichter entsprochen

⁴⁴ Siehe Kapitel 3.4.3.

werden. Dies liegt zum einen daran, dass akustische Reize gemäß ihrem natürlichen Vorbild immer von einer räumlichen Quelle ausgehen. Zum anderen erleichtert bereits das erhöhte Informationsvolumen, wie beschrieben, die Wahrscheinlichkeit, dass der Reiz als real eingestuft wird (Cohen 2000, Grossberg 1995). Da nur jene Reize auch Affekt induzieren können, die durch den Wahrnehmungsapparat Beachtung finden, liegt der Schluss nahe, dass eine Spatialisierung auch auf die Emotionsinduktion unterstützend wirken könnte. Obwohl empirische Untersuchungen zu diesem Zusammenhang bis jetzt nicht angestellt wurden, kann das Kriterium (ii) allein im Licht der Argumentation in Kapitel 3.10 bereits als erfüllt angesehen werden.

Dass auditorische Displays im Sinne von (iii) Neugier im Nutzer erzeugen oder ihn in eine spielerische Interaktion verwickeln können steht angesichts der Beispiele von Mynatt et al. (1997), Rozier (2000) oder Vazquez-Alvarez et al. (2011) offenkundig außer Frage. Genau genommen handelt es sich allerdings bereits bei den in Kapitel 3.10 angeführten Projekten um spatialisierte Anwendungen. Da die Argumentation aus dem vorangegangenen Kapitel insofern ohnehin auf die Unterkategorie der SADs rekurriert, kann auch dem Kriterium (iii) entsprochen werden.

Gemäß dieser Auswertung kann durchaus von einer wahrnehmungstheoretischen Sinnhaftigkeit spatialisierter Audio-Anwendungen gesprochen werden. Natürlich sorgt jeder Anwendungskontext für Anforderungen, denen individuell begegnet werden muss. Dass visuelle Benutzeroberflächen akustischen Lösungen deshalb je nach Kontext in vielerlei Hinsicht überlegen sind und dies auch in Zukunft bleiben werden, steht außer Frage. Dennoch weist diese Untersuchung gleichzeitig darauf hin, dass spatiale Audio-Displays aus wahrnehmungstheoretischer und praktischer Sicht durchaus in der Lage sind, für eine adäquate Information des Nutzers zu sorgen. In Anwendungsfällen, die den Einsatz spatialer Audio-Displays begünstigen, könnten so zumindest in Teilen visuelle durch akustische Interfaces ersetzt werden. Eine solche Komplementierung würde zusätzlich zu einer Entlastung des visuellen Wahrnehmungsapparates führen, die auf dem Hintergrund der *Omnipräsenz visueller Stimuli in der digitalen Welt* dringend notwendig erscheint. Welche praktischen Anwendungskontexte einen sinnvollen Einsatz von SADs möglich machen, soll in den folgenden Kapiteln genauere Betrachtung finden.

5.2 Audio-Augmented Reality

Eine Option für die praktische Implementierung spatialer Audio-Benutzeroberflächen bildet die bereits erwähnte Audio-Augmented Reality (AAR). AAR-Anwendungen können allgemein der Klasse der SADs zugeordnet werden. Sie unterscheiden sich insofern von der klassischen Definition von SADs, als dass sie eine physikalisch begehbbare, reale Umgebung durch elektroakustische Signale und Wiedergaben augmentieren. Wie so häufig ist eine scharfe Unterscheidung dieser Unterkategorie allerdings kaum möglich. Auch aus diesem Grund definieren Oliver Bimber und Ramesh Raskar in ihrem Buch „Spatial Augmented Reality: Merging Real and Virtual

Worlds“ (2005) Augmented Reality (AR) nicht positiv⁴⁵, sondern in Abgrenzung zu Virtual Reality (VR):

„The fact is that in contrast to traditional VR, in AR the real environment is not completely suppressed; instead it plays a dominant role. Rather than immersing a person into a completely synthetic world, AR attempts to embed synthetic supplements into the real environment.“ (Bimler & Raskar 2005, S.2)

Im Kontext der AAR sind die im Zitat erwähnten “synthetic supplements” als synthetisch-akustische Signale zu verstehen. Abbildung 13 fasst die Komponenten einer AR-Anwendung übersichtlich zusammen. Die Präsentationstechnologie im Rahmen der AAR, die in der Abbildung als „Display Technology“ bezeichnet wird, wurde in den vorherigen Kapiteln ausführlich unter dem Titel der SADs vorgestellt. Während sich die gewählte Definition durch eine wahrnehmungsneutrale Beschreibung auszeichnet, ist die Disziplin der AR für viele Ausarbeitungen mit der Betrachtung der visuellen Sphäre scheinbar selbstverständlich abgeschlossen. Diese wohl pathologische Konzentration auf visuelle Anwendungen muss als symptomatisch für das aktuelle wahrnehmungstheoretische Ungleichgewicht verstanden werden, das als Motivation für die vorliegende Arbeit dient. Seit dem 2. Kapitel sollte ersichtlich sein, dass nicht nur eine Fülle unterschiedlicher AAR-Anwendungen existiert, sondern diese gerade durch ihre Vielseitigkeit und ihr kreatives Potential für Audio-Designer interessant sein dürfte.

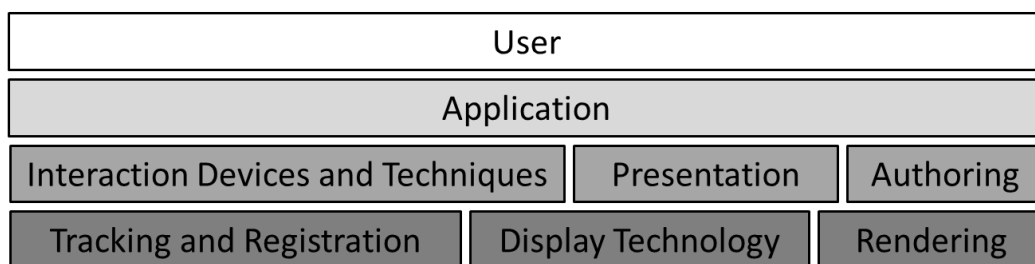


Abbildung 13: Komponenten von AR (Bimler & Raskar 2005)

Auf dem Hintergrund der bisherigen Arbeit ist vor allem das wahrnehmungstheoretische Profil von AAR bemerkenswert. Schon definitorisch verfügen AAR-Anwendungen über eine natürliche, spatiale Komponente – die Realität. Wahrnehmungstheoretisch ergeben sich daraus speziell zwei Vorteile, die für eine Eignungsuntersuchung von Relevanz sein sollten:

Zum einen lässt sich nur schwer ein Anwendungskontext erwägen, der sich besser dazu eignen könnte, die von Truax proklamierte Schizophonie zu überwinden. Wie bereits beschrieben hat

⁴⁵ Der Begriff der „positiven“ Definition wird an dieser Stelle als Definition der Struktur „X ist a, b und c“ im Kontrast zu einer negativen Definition der Struktur „X ist nicht d,e und f“ verwendet. Eine normative Bewertung der Definition soll nicht erreicht werden.

die elektroakustische Wende, so Truax, zu teilsynthetischen Klangumgebungen geführt, die nicht mehr in der Lage sind den Schallereignissen die sie konstituieren, angemessen Bedeutung zuzuordnen. Zu Beginn seiner Arbeit stellt Truax deshalb die Frage: „Is the split between a sound and its source, or what we will call “schizophonia”, a sign of disintegration for the listening process and human communication? “ (Truax 2000, S.XXII)

Im Kontrast zu natürlichen, akustisch ausbalancierten Klanglandschaften erzeugen teilsynthetische Umgebungen in der Regel ambig und undurchsichtige Wahrnehmungskontexte, die Truax als schizophon beschreibt. Der Ursprung dieses Phänomens findet sich in der Trennung von Klangquelle und akustischem Signal. Durch ihre zwangsläufige Einbettung in reale Umgebungen erscheinen AAR-Anwendungen geradezu prädestiniert um entlehnte Klangsignale in einen adäquaten Wahrnehmungskontext zu reintegrieren. Klangquellen werden durch die akustische Augmentierung von Realität nicht nur im virtuellen Raum verteilt, sondern in der Regel konkret physikalischen Entitäten zugeordnet. Wie im Beispiel des Soundgarden (Vazquez-Alvarez et al. 2011) kann etwa das Zwitschern eines Vogels real einem Baum zugeordnet werden, um so eine sinnvolle Klanglandschaft zu modellieren. Natürlich beschränken sich die Möglichkeiten der AAR nicht auf die scheinbare Imitation natürlicher Klangumgebungen. Dennoch gibt es auch für kreative oder ungewohnte Ausgestaltungen akustischer Räume überzeugende und weniger überzeugende digitale Arrangements. Die Domäne der AAR eröffnet so Audio-Designern und Informatikern die Chance ihre akustische Realität neu auszugestalten und sich dennoch innerhalb der Grenzen der (pseudo-)natürlichen Wahrnehmung zu bewegen. Gerade weil die elektroakustische Schizophonie in der digitalen Welt unumgänglich scheint, lohnt es sich, einen bewussteren Umgang mit ihr zu üben.

„Schizophonia is an inevitable fact of audio technology, but our concern with it will be to understand its implications, not to condemn its existence altogether.“ (Truax 2000, S.134)

Zum anderen erfüllen AR-Anwendungen ein allgemeines Technologiekriterium, das sich leider nach wie vor noch immer keiner besonderen Popularität erfreut. Während moderne Technologien wie im Zitat von Bimber und Raskar (2005) üblicherweise die Umgebungsrealität unterdrücken oder den Nutzer an seinen Schreibtisch fesseln, handelt es sich bei AR um eine Technologie, die der Umgebungsrealität zugewandt ist. In seinem berühmten Aufsatz „The Computer of the 21st Century“ bricht Mark Weiser offen mit Technologien, die versuchen, Umgebungseinflüsse zu eliminieren, statt sich *nahtlos* in sie zu integrieren:

„The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it. [...] We are therefore trying to conceive a new way of thinking about computers, one that takes into account the human world and allows the computers themselves to vanish into the background. Such a disappearance is a fundamental consequence not of technology but of human psychology. [...] The idea of integrating computers seamlessly into the world at large runs counter to a number of present-day trends.“ (Mark Weiser 1991, S.3)

Systeme, wie die vorgestellte Architektur für WARA (Härmä et al. 2003), könnten eine solche nahtlose Integration potentiell ermöglichen und damit allen in dieser Arbeit vorgestellten wahrnehmungstheoretischen Kriterien entsprechen. Wie im Fall natürlicher Wahrnehmungsprozesse nutzen derartige Systeme die reale Umwelt als Benutzeroberfläche; anders als visuelle Anwendungen sind sie dabei nicht mehr auf die fraglichen technologischen Entwicklungen der Zukunft angewiesen.

So vielversprechend eine Öffnung für die reale Umgebung deshalb anmuten mag, so viele Probleme und Herausforderungen bringt sie auch mit sich, um das eingangs erwähnte Zitat von Bimber und Raskar zu einem Ende zu bringen: „This leads to a fundamental problem: a real environment is much more difficult to control than a completely synthetic one.“ (Bimber & Raskar 2005, S.2)

Um eine Umgebung zu einer Benutzeroberfläche oder einem Display umzufunktionieren, ist zunächst einmal entscheidend, wo sich der Benutzer befindet, in welche Richtung er sich ausrichtet und welche Entitäten entsprechende Komponenten der Benutzeroberfläche oder des Displays repräsentieren. Eine robuste *Registrierung* und die konstante *Ortung* des Nutzers sind deshalb gerade für AR-Anwendungen von essentieller Bedeutung. Als Registrierung wird in diesem Zusammenhang die konkrete Zuordnung synthetischer Supplemente, im Falle der AAR akustischer Signale, zu realen Entitäten (e.g. das Zwitschern eines Vogels wird beim Betreten des Nahbereiches eines Baumes wiedergegeben) bezeichnet. Diese Zuordnung muss nicht nur robust sein, damit der Nutzer nicht verwirrt wird, sondern sollte, wie hergeleitet, auch als *sinnvoll* wahrgenommen werden können. Während es heute eine ganze Reihe verschiedener Ortungsmechanismen gibt (i.a. Global Positioning System (GPS), mechanisches Tracking, optisches Tracking) aus denen gewählt werden kann, stellen vor allem Übergänge zwischen Indoor- und Outdoor-Tracking, sowie hohe Anforderungen an die Ortungsgenauigkeit aktuell die größte Herausforderung für heutige Trackingsysteme dar (Bimber & Raskar 2005 et al.). Diese Problematik ergibt sich in AAR-Systemen wie in visuellen Anwendungen gleichermaßen, auch wenn die Anforderungen an die Registrierungsgenauigkeit im visuellen Kontext in der Regel deutlich höher sind, um eine nahtlose Integration zu gewährleisten.

Zusätzlich muss dem Nutzer ein vernünftiges Maß an Mobilität gewährt werden, damit er sich angemessen in der augmentierten Umgebung bewegen kann. Das heißt zum einen, dass die eingesetzte Apparatur nicht zu sperrig sein darf, als dass sie die Bewegungsfreiheit des Benutzers massiv einschränkt (Vazquez-Alvarez et al. 2011). Zum anderen muss ein Minimum an potentieller Akkulaufzeit sichergestellt werden, damit es überhaupt zu einer ungestörten Interaktion zwischen Nutzer und System kommen kann, ohne dass die Anwendung regelmäßig für Ladevorgänge oder einen Austausch unterbrochen werden muss. Der von Härmä et al. (2003) vorgestellte Prototyp wird diesen Anforderungen grundsätzlich gerecht. Die aufwendige Displaytechnologie erfolgsversprechender, visueller Systeme, deren holographische Leistung ausreicht um einen überzeugenden Wahrnehmungskontext erzeugen zu können, führt leider bis dato zu ei-

nem Konflikt mit diesem Mobilitätskriterium⁴⁶. Dies liegt nicht zuletzt an der großen Datenlast die es im Rahmen visueller Anwendungen zu bewältigen gilt. Das visuelle AR zweifelsfrei visionäres Potential bietet, soll mit dieser Einschätzung nicht in Frage gestellt werden. Wann allerdings mit einer praktischen Ausschöpfung dieses Potentials durch voll einsatzfähige Technologien gerechnet werden kann, bleibt, im Kontrast zu AAR, weiterhin abzuwarten (Bimber & Raskar 2005 et al.).

Dennoch stellt die besagte Datenlast auch für akustische Anwendungen immer noch eine zentrale Herausforderung dar. Wie bereits in Kapitel 3.4.1 festgestellt wurde, müssen wahrnehmbare Latenzen vollständig ausgeräumt werden, um vom auditorischen Wahrnehmungsapparat nicht automatisch als künstlich verworfen zu werden (Grossberg 1995, Cohen 2000). Gleichzeitig erzeugen resultierende Positionsverschiebungen synthetischer Supplemente immer auch Fehlregistrierungen, die vom Nutzer in aller Regel schnell bemerkt werden und das Bedienerlebnis deutlich beeinträchtigen. Diesem Erfordernis kann letztlich nur durch ein effektives Echtzeit-Rendering begegnet werden (Bimber & Raskar 2005 et al.). Erprobte Methoden zur Reduktion dieser Datenlast für AAR-Anwendungen wurden bereits entwickelt (Faller & Baumgarte 2001 et al.).

Aus wahrnehmungstheoretischer Perspektive bleibt festzuhalten, dass schon heute akustische Systeme existieren, deren holographisches Leistungsniveau in der Lage ist einem Turing-Test standzuhalten, wie ihn Härmä et al. vorschlagen (2003). Auf dem Hintergrund einer realistischen Prognose kann von vergleichbaren Standards für fotorealistentes Rendering im visuellen AR-Bereich in näherer Zukunft nicht gesprochen werden.

5.3 Praktische Ausgestaltung

Wie im eingangs beschriebenen Beispiel des Drehzahlmessers⁴⁷ finden sich immer wieder Anwendungskontexte, in denen eine Information durch auditorische Displays superiore Ergebnisse erzeugt. Häufig ist die akustische Ausgestaltung einer Anwendung (e.g. Simulation) aber auch schlicht erforderlich und unterstützt ein Gesamtsystem. Im Laufe dieser Arbeit wurden zum Zweck der Illustration immer wieder Anwendungsbeispiele angeführt, die in Teilen der AAR zugeordnet werden können, häufig aber auch der klassischen Definition eines SADs entsprechen. Zu diesen Beispielen gehören:

⁴⁶ An dieser Stelle muss darauf hingewiesen werden, dass natürlich nicht alle AR-Anwendungen den Anspruch verfolgen, die Grenze zwischen artifiziellen und realen Wahrnehmungsinhalten zu verwischen. Viele Systeme konzentrieren sich stattdessen auf die reine informationelle Augmentierung realer Umgebungen. Gerade im industriellen Bereich werden AR-Systeme häufig eingesetzt, um bestimmte Arbeitsschritte zu vereinfachen oder effizient zu gestalten, nicht aber um dem Nutzer ein besonders angenehmes oder immersives Bedienerlebnis zu ermöglichen. Gerade dieser Anspruch erscheint aber zentral für eine wahrnehmungstheoretische Eignungsuntersuchung. Aus dieser Einschätzung ergibt sich auch die bewusste Betonung bestimmter Aspekte unter Vernachlässigung anderer Untersuchungsgegenstände.

⁴⁷ Siehe Kapitel 3.6.

- Automated Tour Guide (AAR), Kapitel 2.1 (Bederson 1995)
- Hear & There (AAR), Kapitel 2.2 (Rozier 2000)
- Audio Aura (AAR), Kapitel 2.3 (Mynatt et al. 1997)
- Telefonkonferenz (SAD/AAR), Kapitel 3.6.4 (Singer et al. 1996)
- Umsatzzahlenbeispiel (SAD), Kapitel 3.6.4 (Proctor & Reeve 1990)

Zudem wurde in Kapitel 3.6.1 die Audification seismologischer Datensätze für ein auditorisches Display beschrieben (Speeth 1961). In der Grundauführung verfügt das eingesetzte Display nicht über eine Spatialisierung. Auch in diesem Fall wäre aber vorstellbar, zwei oder mehr Seismogramme angemessen im Raum zu verteilen, um so einen Echtzeitvergleich zu ermöglichen. Eine solche Modifikation würde die Anwendung folglich in die Sphäre der SADs verschieben. Da die in der vorliegenden Arbeit angestellte, wahrnehmungstheoretische Untersuchung die Eignung von SADs und spatial-auditorischen Benutzeroberflächen unterstützt, liegt der Schluss nahe, dass sich der sporadische Einsatz dieser Systeme auf einen gewissen Mangel kreativer Technologiekonzepte zurückführen lässt. Die folgende Auswahl unterschiedlicher Konzeptideen soll deshalb zum Abschluss einen Einblick in den Facettenreichtum spatialer, auditorischer Benutzeroberflächen und Displays bieten.

5.3.1 Landscape Music

In Kapitel 4.3.2 wurde im Rahmen Holladay's TED-Talk (2013) bereits die Idee vorgestellt, Musik an bestimmte Umgebungen zu binden oder im Zweifel direkt für eine bestimmte Landschaft zu komponieren. Dieses narrative Konzept könnte in ähnlicher Form fraglos auch in andere Anwendungskontexte überführen werden.

So wäre es beispielsweise vorstellbar jedes Viertel einer Stadt mit einem individuellen Soundtrack zu versehen, den jeder Nutzer durch eine mobile Anwendung verfolgen kann. Wird beispielsweise ein japanischer Garten betreten, könnten die beruhigenden Melodien von Ryuichi Sakamoto dieses Erlebnis sinnlich komplettieren. Folgt man diesem Konzept, scheint es vorstellbar einem gewöhnlichen Stadtrundgang durch eine interessante, melodische Ausgestaltung eine neue Dimension zu verleihen. Gleichzeitig ist natürlich nicht davon auszugehen, dass schon zu Beginn alle Umgebungen über eine melodische Untermalung verfügen. Das Finden solcher Orte könnte einen zusätzlichen Reiz für Besucher der Stadt ausmachen, aber auch Bewohner motivieren, ihre heimische Umgebung genauer zu erkunden.

Zusätzlich wären Mischformen zwischen konventioneller Musikwiedergabe und einem umgebungsgebundenen Modus vorstellbar. Die Premiere eines neuen Albums könnte so beispielsweise für einen eingeschränkten Zeitraum an einen bestimmten Ort gebunden werden, bevor das Produkt frei im Handel erhältlich wird. Dadurch könnte jedenfalls vorübergehend dem Konzept von Holladay Rechnung getragen werden, ohne den Zuhörern anschließend zu verwehren, die Musik auch in anderen Kontexten zu genießen. Je nach öffentlichem Interesse könnten so ein Waldstück, ein verlassenes Stadtviertel oder ein freies Feld vorübergehend in eine Pilgerstätte musikalisch Interessierter verwandelt werden.

5.3.2 Auditory Attitude Indicator

Im Kontrast zu diesem unterhaltungsorientierten Konzept stellen Douglas S. Brungart und Brian D. Simpson in ihrem Artikel „Design, Validation, and In-Flight Evaluation of an Auditory Attitude Indicator Based on Pilot-Selected Music“ (2008) ein SAD als Begleittechnologie für die Flugzeugsteuerung vor. Ziel der Benutzeroberfläche ist es, den visuellen Kognitionsapparat des Piloten zu entlasten und eine begleitende Parameterinformation zu ermöglichen, um Ermüdungs- und Frustrationserscheinungen des Nutzers vorzubeugen.

In ihrem Konzept verwenden Brungart und Simpson spatialisierte Klangsignale, um die Ausrichtung des Flugzeugs relativ zum Horizont abzubilden. Während seitliche Neigungen des Flugzeugs durch das Verschieben der Klangquelle nach links oder rechts im virtuellen Raum signalisiert werden, führen Steig- oder Sinkneigungen zu einer spatial diffusen Veränderung der Tonhöhe. Neigt sich das Flugzeug nach links, wird die Klangquelle in entsprechendem Maß nach rechts verschoben, um dem Piloten eine Seitenlage anzuzeigen und vice versa. Befindet sich das Flugzeug im Steigflug wird das Klangsignal in einer tieferen Tonlage wiedergegeben. Befindet sich das Flugzeug im Sinkflug, wird die Tonlage gesteigert. Durch diese Zuordnung wird gleichzeitig erreicht, dass das Signal bei extremen Frontneigungen durch die schrille Tonlage als Alarm wahrgenommen wird und den Piloten veranlasst, das Flugzeug unverzüglich in eine ausbalancierte Lage zurückzusetzen. Wie beschrieben wurde das auditorische Display prinzipiell als Begleittechnologie entworfen, die planmäßig vor allem für das Beibehalten der Reiseflughöhe eingesetzt werden sollte. Natürlich ist es dem Nutzer bei Start und Landung deshalb möglich, das Display abzuschalten, um unnötige Schallbelastungen zu vermeiden. Da das Interface nicht vom Signalinhalt (sondern nur der Signalposition bzw. Tonlage) abhängig ist, ist dem Piloten zusätzlich freigestellt, welches Signal eingesetzt werden soll. Musik oder kompatible Umgebungsgeräusche haben sich in den Untersuchungen als Favorit bewiesen. Auch dieser Faktor führt zu einer zusätzlichen Vermeidung unnötiger Stresseinflüsse. Die Architektur des Konzeptes findet sich in Abbildung 14.

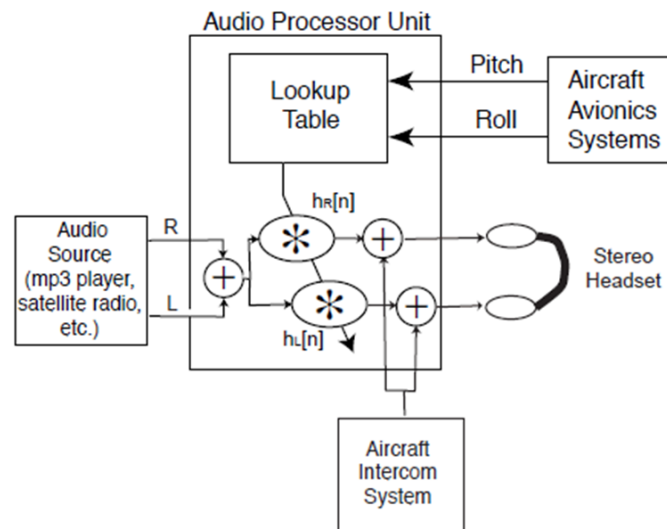


Abbildung 14: Auditory Attitude Indicator (Brungart & Simpson 2008, S.4)

Die Nutzung eines solchen auditorischen Kontrolldisplays ermöglicht es dem Piloten außerdem seine visuelle Aufmerksamkeit ununterbrochen auf den ihn umgebenden Luftraum zu richten. Neigungsmesser und andere Anzeigen müssen nicht mehr dauerhaft unter Beobachtung stehen, was häufig zu einer spürbaren, kognitiven Entlastung des Piloten führt. Brungart und Simpson's Einschätzung hinsichtlich der Eignung ihres Displays fällt deshalb insgesamt entsprechend positiv aus:

„[W]e think the results of this experiment show very strong potential for the operational utility of audio attitude indicators of the type described here. These displays clearly can provide supplemental information about aircraft attitude, and they are likely to achieve a high degree of acceptability among pilots. We also believe that they could be implemented in a very affordable way.“ (Brungart & Simpson 2008, S.8)

Auch auf dem Hintergrund dieses Beispiels bestätigt sich die These, dass auditorische Displays in bestimmten Kontexten bessere Leistungen erzielen können, als ihre visuellen Pendanten, sofern sie einer durchdachten Konzeption folgen.

Vergleichbare Anwendungen für den Automobilbereich existieren bereits (Neuhoff 2011).

5.3.3 Navigation durch Acoustic Landmarks

Der Großteil der Weltbevölkerung ist heute im Besitz von Smartphones, mit denen er die meisten ihrer navigatorischen Alltagsherausforderungen meistert. Dennoch treten bei der Navigation mit dem Handy immer wieder Probleme auf, die häufig zur Frustration des Nutzers führen. Den

meisten Nutzern ist es oft ad hoc nicht möglich sich in der meist stark abstrahierten und vogelperspektivischen Karte zu orientieren; ein einziger Blick auf das Handy ist in der Regel nicht ausreichend. Der permanente Kontrollblick auf das Smartphone ist notwendig, um eine verlässliche Navigation zu gewährleisten und prägt schon heute die Art und Weise, in der Städte erkundet werden. Leider erweist sich diese Herangehensweise in aller Regel nicht als besonders weltgewandt – die Navigation lässt eine freie Interaktion mit der unbekanntenen Umgebung kaum noch zu. Angesichts der in dieser Arbeit untersuchten Veranlagungen des menschlichen Wahrnehmungsapparates, wäre vorstellbar, derartige Navigationsaufgaben in Zukunft durch spatialisierte Audiohinweise zu lösen:

„Spatial audio is an important feature in virtual environments as it helps users orient themselves, and can provide 360 aural awareness, independently of sight.“ (Larsen et al. 2014, S.70)

Acoustic Landmarks, wie sie bereits in Kapitel 4.2.1 vorgestellt werden, würden sich für eine solche Anwendung besonders in urbanen Umgebungen anbieten (Rozier 2000, Sundareswaran et al. 2003, Vazquez-Alvarez et al. 2011). Einzelnen räumlichen Landmarks (e.g. Kirchen, Plätze) könnten Klangquellen zugeordnet werden, die eine Navigation sogar dann erlaubten, wenn die eigentlichen Orientierungspunkte verdeckt sind. Die Forschungsergebnisse von Camilla H. Larsen et al. (2014) geben Hinweise darauf, dass derartige akustische Orientierungspunkte wegen der engen Verknüpfung des auditorischen Systems mit der räumlichen Wahrnehmung sogar dann unterstützend in Anspruch genommen werden, wenn visuelle Landmarks für die Navigation zur Verfügung stehen:

„Search latencies were significantly decreased [...] even with the visual stimuli being within the field of view, suggesting that auditory stimuli is used as an aid to visual search even within the field of view.“ (Larsen et al. 2014, S.73)

Wegbeschreibungen wie “nach dem Wasserfall biegen Sie rechts ab und folgen von dort an der Jahrmarktsmelodie” wären grundsätzlich nicht ausgeschlossen. Stattdessen wäre aber auch vorstellbar eine einzelne Klangquelle virtuell auf dem Zielort zu fixieren. Die spatialen Informationen der Klangquelle würden so, ähnlich der Funktionsweise eines klassischen Kompasses, kontinuierlich die Himmelsrichtung angeben, in der sich das Ziel befindet. Wie im Beispiel von Brungart und Simpson könnte schon das persönliche Lieblingslied als Klangquelle ausreichen, um eine rudimentäre Orientierung zu gewährleisten. Die Beiläufigkeit solcher Systeme könnte dem Anwender ermöglichen, sich wieder mehr auf seine reale Umgebung zu konzentrieren und so eine natürlichere Interaktion zwischen Nutzer und Benutzeroberfläche fördern. Es steht sicherlich außer Frage, dass vergleichbare Anwendungen durch Elemente der visuellen Navigation komplementiert werden müssen. Dies ist vor allem dann erforderlich, wenn der Anwendungsfall eine besonders genaue Navigation erforderlich macht. Ein Großteil der herkömmlichen Navigation zwischen Start und Zielort könnte so dennoch deutlich angenehmer gestaltet werden, um eine kognitive Entlastung des Nutzers zu erzielen.

5.3.4 VR-Computerspiele

Der Weltmarkt für Computerspiele ist nach wie vor eine der treibenden Kräfte informationstechnologischen Fortschritts. Es ist kaum verwunderlich, dass innovationsbegierige Endnutzer auch im Hinblick auf die akustische Ausgestaltung der immer realistischer werdenden Spielwelten höhere Ansprüche entwickeln:

„One of the largest and most rapidly growing application areas for spatial auditory displays is in the entertainment industry. [...] Computer games increasingly rely on stereo speakers to simulate sources at various locations. For these applications, subjective realism is the most important goal, but the accuracy and resolution achieved by the display is not as important.“ (Shinn-Cunningham 2002, S.3)

Heutige SADs für binaurale Kopfhörersysteme bergen durchaus das Potential diese Nachfrage in Zukunft zu stillen. Die resultierenden, auditorischen Hologramme wären in der Lage, virtuelle Welten jedenfalls akustisch „begehbar“ zu machen. Das solche Systeme zu einem immersiveren Spielerlebnis beitragen könnten, scheint offensichtlich. Wie lange es aber dauern wird, bis diese Technologie als Standard Einzug in die Computerspielwelt hält, bleibt weiter abzuwarten.

5.3.5 Hörbücher als akustisches Hologramm

Auch für Hörbücher könnte ein solcher Fortschritt in Richtung virtuelle Realität in Zukunft neue narrative Strukturen und immersivere Erzählweisen eröffnen. Die ohnehin oft fantastischen Gedankenwelten des Zuhörers würden so angemessen akustisch ausgekleidet. Ein Besuch des Geisterschlusses mit den „Drei Fragezeichen“ könnte dann bedeuten, dass der Zuhörer rechts von sich einer Unterhaltung von Justus Jonas und Peter Shaw lauschen kann, während sich die Gruppe über den knarrenden Dielenboden bewegt, als plötzlich hinter ihnen die schwere Eisentür ins Schloss fällt und sie von der Außenwelt abschneidet. Abgesehen von einem spatial-auditorischen Softwaresystem wäre lediglich ein Satz Kopfhörer mit einem entsprechenden Lagemesser notwendig, um dem Nutzer zusätzlich zu ermöglichen, sich nach Belieben in der Erzählwelt „umzuhören“. Ein solcher Lagemesser befindet sich heute in jedem handelsüblichen Smartphone, wenn auch zwecks einer robusten Simulation gegebenenfalls auf höherwertige Geräte zurückgegriffen werden sollte⁴⁸. Dem Zuhörer wäre es in dieser Konzeption zwar nicht möglich, sich frei in der Erzählwelt zu bewegen, er könnte durch Bewegungen seines Kopfes aber auf intuitive Weise mit ihr interagieren. Würde der Nutzer also als Reaktion auf die hinter ihm zugefallene Eisentür (i.e. auf 180°) seine Kopfposition um 90° nach links rotieren, befände sich das Signal entsprechend des virtuellen Klangraumes genau in der Richtung seines linken Ohres. Das Ergebnis wäre eine manövrierbare, virtuelle Klangumgebung in der der Zuhörer die Erzählgeschichte auf eine völlig neue Art und Weise erleben könnte.

⁴⁸ Systeme, die über Beschleunigungssensoren ohne eine absolute Fixierung die Ausrichtung in einer räumlichen Umgebung berechnen, erzeugen in der Regel sogenannte „Drifts“. Die Verwendung eines einfachen Handysensors ist deshalb gegebenenfalls nicht zu empfehlen.

Um ein möglichst realitätsnahes, akustisches Hologramm zu erzeugen, ist es notwendig, dass schon die einzelnen Klangquellen des Hörbuches separat aufgenommen und anschließend einer konkreten, spatialen Position zugeordnet werden. Die klassische Erzählerstimme könnte sinnvollerweise als intrakraniale Wiedergabe, direkt in den Kopf des Zuhörers gelegt werden. Unabhängig von der letztlichen Ausgestaltung eines solchen Hörbuchs ist zu erwarten, dass ein System nach diesem Vorbild die Vorstellungskraft des Nutzers entscheidend unterstützen könnte.

5.3.6 SADs für sehbehinderte Nutzer

Wie schon zu Beginn verdeutlicht, verzichtet diese Arbeit bewusst darauf, die Nutzung von SADs durch sehbehinderte oder blinde Nutzer zu untersuchen. Die wahrnehmungstheoretische Eignungsuntersuchung soll in der Lage sein, ein generelles Argument für den Einsatz spatialisierte Audio-Systeme zu liefern, ohne dabei die Nutzergruppe einschränken oder spezifizieren zu müssen.⁴⁹ Gleichzeitig steht außer Frage, dass die Technologie der SADs oder Benutzeroberflächen gerade für die Gruppe der sehbehinderten Nutzer von besonderem Wert ist.

„Blind users are increasingly making use of spatial auditory displays. For instance, spatial auditory displays are used in navigational aids or as a substitute for graphical computer interfaces. As with command and control applications, realism of the spatial auditory display is not important by itself with these applications; what matters is the amount of spatial information conveyed to the user. [...] [H]eadphone-based systems are generally preferred for use as sensory aids.“ (Shinn-Cunningham 2002, S.3)

Natürlich ließen sich alle bisher vorgestellten Anwendungen auch für sehbehinderte Nutzer einsetzen. Von der Navigation durch akustische Landmarken, über die wissenschaftliche Datenanalyse, bis hin zur Steuerung von Fortbewegungsmitteln könnten SADs und auditorische Benutzeroberflächen ein breiteres Spektrum der Nutzerinteraktion abdecken und so für sehbehinderte Personen eine vielversprechende Schnittstelle zur digitalen Welt darstellen. Da sich der Einsatz von SADs getreu dem Ergebnis der vorliegenden Untersuchung auch für uneingeschränkte Nutzer anbietet, könnten beide Gruppen in der Nutzung auditorischer Systeme zusammengeführt werden. Gerade im Hinblick auf den *inkluisiven Ansatz der Barrierefreiheit* empfiehlt sich deshalb die praktische Integration der vorgestellten Systeme. Schon seit einiger Zeit halten auditorische Interfaces Einzug in die *Assistive Technologie (AT)*. Allein die wahrnehmungstheoretische Untersuchung dieses Einsatzgebietes böte ausreichend Potentiale für eine eigenständige Arbeit.

5.4 Ausblick

Auffallend ist, dass heutige spatial-auditorische Displays zum größten Teil entweder der Unterhaltung dienen oder für die Interaktion mit technischen bzw. wissenschaftlichen Anlagen einge-

⁴⁹ An dieser und allen anderen Stellen liegt implizit die Einschränkung vor, dass die Untersuchung sich nur auf den Wahrnehmungsapparat menschlicher Nutzer fokussiert.

setzt werden (Shinn-Cunningham 2002 et al.). Es ist wahrscheinlich, dass dieser Umstand kein Zufallsprodukt ist, sondern sich von der Beschaffenheit derartiger Systeme ableiten lässt. Ein Einsatz für Unterhaltungszwecke ist in aller Regel verhältnismäßig frustrationsresistent. Die ungewohnte Handhabung des unbekanntem Displays kann in derartigen Anwendungskontexten spielerisch erlernt werden. Treten Bedienfehler oder Fehlinterpretationen auf, können diese eher toleriert werden, weil die Anwendung per se kein konkretes Ziel verfolgt. Die Anwendung in stark technischen oder wissenschaftlichen Kontexten bezeichnet das andere Extrem des Spektrums an Anwendungsszenarien. Ziel und Aufgabe sind hier in der Regel klar festgelegt; die Fehlertoleranz fällt entsprechend gering aus. Stattdessen treten die Benutzerfreundlichkeit und das Bedienerlebnis des Systems in den Hintergrund. Gleichzeitig werden die Nutzer im Normalfall für die Bedienung der Benutzeroberfläche ausgebildet, wodurch eine erfolgreiche Interaktion gewährleistet werden kann. In beiden Kontexten kann den Anforderungen an das System deshalb zwar auf unterschiedliche Art und Weise, aber im gleichen Maße entsprochen werden.

Zwischen diesen beiden Endpunkten liegt ein Kontinuum alltäglicher Anwendungskontexte. Der Einsatz eines SADs für die Fußgängernavigation in einer unbekanntem Stadt beispielsweise, muss den Nutzer zum einen in die Lage versetzen, sicher einen Zielpunkt zu erreichen, zum anderen sollte dabei aber ein möglichst hohes Maß an Bedienkomfort erzeugen werden. Da der größte Teil der potentiellen Nutzer aber nicht mit der Bedienung (spatialer) auditorischer Benutzeroberflächen vertraut ist, wird es in vergleichbaren Anwendungsszenarien zunächst schwierig sein, überhaupt nur eine der beiden Anforderungen zu erfüllen. Ähnlich wie bei der Lesefähigkeit, wird der Nutzer die erfolgreiche und gekonnte Bedienung von SADs erst erlernen müssen. Wie bereits in Kapitel 3.7 beschrieben, muten die Lernkurven bisheriger Untersuchungen für den Umgang mit spatialen, akustischen Interfaces allerdings durchaus vielversprechend an. Eine sukzessive Integration spatial-auditorischer Benutzeroberflächen in den digitalen Alltag scheint deshalb besonders sinnvoll. Benutzer könnten sich so allmählich an die Bedienung entsprechender Systeme gewöhnen, ohne eine vollständige Abhängigkeit einzugehen. Die dadurch induzierten Lernprozesse würden in Zukunft die Nutzung immer komplexerer und eigenständigerer, spatial-auditorischer Anwendungen zulassen, um insgesamt zu einer Entlastung des visuellen Wahrnehmungsapparates und einer ausgewogenen, wahrnehmungspraktischen Kommunikation beizutragen. Erst durch eine solche, langfristige Integration spatial-auditorischer Systeme in die digitale Umwelt könnte eine abschließende Beurteilung der praktischen Eignung von SADs überhaupt ermöglicht werden. Während an dieser Stelle für gewöhnlich auf neue Forschungsgebiete und die dringende Notwendigkeit zusätzlicher Untersuchungen verwiesen wird, muss auf dem Hintergrund der Ergebnisse dieser Arbeit stattdessen die vermehrte, praktische Integration spatial-auditorischer Benutzeroberflächen empfohlen werden. Obwohl auf eine grundsätzliche wahrnehmungstheoretische Eignung geschlossen werden konnte und vereinzelte Untersuchungen zum Einsatz solcher Systeme bereits vorliegen, hat es den Anschein, dass nur eine umfassende Implementierung konkrete Schlüsse hinsichtlich der Kompatibilität mit der menschlichen Wahrnehmung zulassen könnte.

6 Konklusion

Auch in Zukunft bleibt abzuwarten, welche Rolle (spatial-) auditorische Architekturen in der digitalen Welt einnehmen werden. Mit dem Paradigma der visuellen Datenpräsentation könnte im Rahmen der Wahrnehmungstheorie jedenfalls gebrochen werden: Der Primat der visuellen Informationsübertragung leitet sich nicht aus der Beschaffenheit des menschlichen Wahrnehmungsapparates ab, sondern ist vielmehr als sich perpetuierende Designentscheidung der Informationstechnologie zu verstehen. Beide Hypothesen der vorliegenden Arbeit konnten auf dem Hintergrund der wahrnehmungstheoretischen Eignungsuntersuchung bestätigt werden.

Spätestens seit der Entwicklung der Schrift setzt die Informationstechnologie auf den Einsatz visueller Repräsentationen zum Zweck der Kommunikation. Dokumentations- und Darstellungsformen haben sich über den Verlauf mehrerer Jahrhunderte auf die Vision als Vehikel der Information konzentriert. Es ist deshalb schlicht und ergreifend nicht zu ermesen, welche auditorischen Repräsentationsmechanismen aus einem Fokus auf die Audition hätten erwachsen können. Dass der Einsatz eines Graphen zur Darstellung eines Datensatzes als gängiges Instrument herangezogen werden kann, fußt bereits auf der Prämisse einer visuell dominierten Sphäre. Es könnte deshalb möglicherweise irreführend sein, auditorische Displays anhand ihrer Fähigkeit zu beurteilen, originär visuelle Darstellungsformen akustisch zu übersetzen. Die Eignung akustischer Signale als Informationsträger und die Leistungsfähigkeit des auditorischen Wahrnehmungsapparates stehen wissenschaftlich weitestgehend außer Frage. Eine ausbalancierte Sinnesansprache offeriert außerdem eine Reihe von Vorzügen:

„Human responsiveness is likely to remain dominated by the visual domain but we should not neglect other modalities when designing new systems. A well designed system that supports the operator using multiple modalities is likely to support lower levels of cognitive control which are quicker, more effective and reduce effort for the human operator.“ (Sanderson et al. 2000, S.265)

Die Integration auditorischer Displays und Benutzeroberflächen in unseren informationstechnologischen Alltag könnte nicht nur die dringende Entlastung des visuellen Wahrnehmungskanals herbeiführen, sondern auch eine Ausgestaltung der digitalen Welt nach holistischen Wahrnehmungsprinzipien ermöglichen. Parallel würde der vermehrte Einsatz auditorischer Anwendungen Lerneffekte induzieren, die im Umkehrschluss die praktische Eignung solcher Systeme weiter befördern würde. Eine perpetuierende Weiterentwicklung nach dem Vorbild der visuellen Domäne könnte so, zumindest in Zukunft, nachgeholt werden. Wie im vorangegangenen Kapitel bereits festgestellt wurde, ist die praktische Würdigung auditorischer Systeme deshalb unmittelbar von ihrer nachhaltigen Integration in unsere Umwelt abhängig.

Zusätzlich könnten Technologien wie die AAR eingesetzt werden, um der von Truax (2000) proklamierten Schizophonie angemessen zu begegnen. Mit den Mitteln der modernen Elektroakustik werden heute spatial-akustische Hologramme generiert, die eine Unterscheidung zwischen Realität und Virtualität kaum mehr zulassen. Die zugrundeliegende Technologie eröffnet Audio-Designern die Möglichkeit ihre akustische Umwelt neu auszugestalten und einen weltgewandten Umgang mit Informationstechnologien zu fördern. Als Resultat könnten spatial-auditorische Displays und Benutzeroberflächen in einen breiteren Wahrnehmungskontext treten und so eine natürlichere Interaktion unterstützen. Am Ende dieser Entwicklung sollte die nahtlose, nicht mehr wahrnehmbare Integration von Technologie in den natürlichen Wahrnehmungskontext nach Weisers Vision für den Computer des 21. Jahrhunderts stehen:

„All say, in essence, that only when things disappear in this way are we freed to use them without thinking and so to focus beyond them on new goals. [...] [I]nvisible computing is so strong that some of us use the term "embodied virtuality" to refer to the process of drawing computers out of their electronic shells. The "virtuality" of computer-readable data - all the different ways in which they can be altered, processed and analyzed - is brought into the physical world.“ (Weiser 1991, S.3)

Das oberste Prinzip für einen solchen Einzug in die physikalische Welt muss die strikte Verhinderung einer Klangverschmutzung unserer akustischen Umwelt sein. Ob sich eine potentielle Lärmbelastigung allerdings vollständig ausschließen ließe, ist schwer zu beurteilen. Einerseits könnten Kopfhörersysteme wie WARA (Härmä et al. 2003) dazu beitragen, dass jeder Nutzer individuell über die Menge an akustischen Signalen entscheiden kann, denen er sich aussetzt. Zusätzlich führt die konzeptionelle Weiterentwicklung auditorischer Displays zu einer stetigen Abnahme alarmartiger Hinweise (Sanderson et al. 2000). Andererseits könnte eine Ausweitung der akustischen Informationsübertragung den Nutzer letztlich dazu nötigen, ein höheres Niveau akustischer Aktivität in seiner Umgebung in Kauf zu nehmen.

Aus diesem Anlass sollte abschließend noch einmal darauf hingewiesen werden, dass die vorliegende Arbeit die Eignung der visuellen Informationspräsentation nicht in Frage stellt. Auch über die Qualifikation anderer Sinnesmodalitäten soll keine Aussage getroffen werden. Die Quintessenz dieser Untersuchung ist vielmehr ein wahrnehmungstheoretisches Plädoyer für den Einsatz spatial-auditorischer Displays und einer ausgewogenen Gestaltung synthetischer Wahrnehmungskontexte. Während fraglich ist, inwiefern sich olfaktorische oder gustatorische Systeme für eine kontrollierte Informationsübermittlung instrumentalisieren lassen, steht die kontextspezifische Eignung auditorischer Anwendungen nach dem Urteil dieser Arbeit außer Frage. Ob auditorische Systeme deshalb in Zukunft Einzug in unseren digitalen Alltag halten werden, wird letztlich von einer konzeptionellen Entscheidung, nicht aber von einer wahrnehmungstheoretischen Zwangsläufigkeit abhängen.

Literaturverzeichnis

- Aggarwal, S.; Ghosal, K.; Singhal, P.; Srivastava, P. (2014): *Effect of Learning on Audio-Spatial Working Memory*. In: Spatial Cognition 2014: Poster Presentations 8, Universität Bremen, Universität Freiburg, S.1-4.
- Back, M.; Des, D. (1996): *Micro-Narratives in Sound Design: Context, Character, and Caricature in Waveform Manipulation*. In: Third International Conference on Auditory Display.
- Bateman, J.; Kutz, O.; Mossakowski, T.; Sojic, A.; Codescu, M. (2014): *Space for Space - SpacePortal: the 21st Century Home for Spatial Ontologies*. In: Spatial Cognition 2014: Poster Presentations 8, Universität Bremen, Universität Freiburg, S.13-16.
- Bederson, B. (1995): *Audio Augmented Reality: A Prototype Automated Tour Guide*. Conference Companion on Human factors in computing systems 1995, S.210-211.
- Bimber, O.; Raskar, R. (2005): *Spatial augmented reality: Merging real and virtual worlds*. A. K. Peters LTD.
- Blauert, J. (1969): Die Beschreibung von Hörversuchen an Hand eines einfachen systemtheoretischen Modells. *Kybernetik*, Band 5, S.45-49.
- Blauert, J. (1970): *Ein neuartiger Präsenzfilter*. In: Fernseh- und Kinotechnik, 1970, Nr. 3, S.75-78.
- Blauert, J.; Braasch, J. (2007): *Räumliches Hören*. In: Handbuch der Audiotechnik (Kapitel 3, Stefan Weinzierl, Ed.), Springer, Berlin, Heidelberg.
- Bradley, N.A.; Dunlop, D. (2005): *Towards a Multidisciplinary Model of Context to Support Context-Aware Computing*. In: Human-Computer Interaction 20, S.403-446.
- Bregman, A.S. (1990): *Auditory Scene Analysis*. Massachusetts, MIT Press.
- Brungart, D.S.; Simpson, B.D. (2008): *Design, validation, and in-flight evaluation of an auditory attitude indicator based on pilot-selected music*. In: Proceedings of the International Conference on Auditory Display.
- Buschman, T.J.; Miller E.K. (2007): *Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices*. In: Science 315, S.1860-1862.
- Cherry, E.C. (1953): *Some experiments on the recognition of speech, with one and with two ears*. In: The Journal of the Acoustical Society of America 25, S.975-979.
- Chion, M. (1994): *Audio-Vision: Sound On Screen*. New York, Columbia University Press.
- Cohen, A.J. (2000): *Film Music: Perspectives from Cognitive Psychology*. In: Buhler, James, Flinn, Caryl & Neumeyer, David (Eds.) Music and Cinema. Hanover, NH, University Press of New England, S.360-377.
- Davies, J.B. (1978): *The psychology of music*. Hutchinson.
- De Houwer, J.; Baeyens, F.; Field, A.P. (2005): *Associative learning of likes and dislikes: Some current controversies and possible ways forward*. In: Cognition and Emotion 19, S.161-174.

- Dey, A.K.; Abowd, G.D. (2000): *Towards a Better Understanding of Context and Context-Awareness*. CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness.
- Dingwall, W.O. (1993): *The Biological Basis of Human Communicative Behavior*. In: Psycholinguistics, Gleason, J.B. und Ratner, N.B., Ed., New York (Harcourt, Brace, Jovanovic), S.40-48.
- Droumeva, M.; Wakkary, R. (2008): *Understanding aural fluency in auditory display design for ambient intelligent environments*. International Conference on Auditory Displays, 2008.
- Duraiswami, R.; Zotkin, D.N.; Li, Z.; Grassi, E.; Gumerov, N.A., Davis, L.S. (2005): *High order spatial audio capture and its binaural head-tracked playback over headphones with HRTF cues*. In: Proceedings of 119th Audio Engineering Society Convention, New York, NY, USA, S.1-16.
- Faller, C.; Baumgarte, F. (2001): *Efficient representation of spatial audio using perceptual parametrization*. In: Proceedings of IEEE Workshop Applications Signal Processing to Audio and Acoustics.
- Fodor, J. (1983): *Modularity of Mind*. Cambridge, Massachusetts, MIT Press.
- Freksa, C. (2013): *Understanding Cognition through Synthesis and Analysis*. In: International Journal of Software Informatics 7(1), S.3-18.
- Grossberg, S. (1995): *The Attentive Brain*. In: American Scientist 38, S.438-449.
- Härmä, A.; Jakka, J.; Tikander, M.; Karjalainen, M.; Lokki, T.; Nironen, H.; Vesa, S. (2003): *Techniques and applications of wearable augmented reality audio*. In: Proceedings of 114th Audio Engineering Society Convention, Amsterdam, Netherlands, S.1-20.
- He, G.; Ishikawa, T.; Takemiya, M. (2014): *Where are you now? Dynamics of Simultaneous, Collaborative Navigation*. In: Spatial Cognition 2014: Poster Presentations 8, Universität Bremen, Universität Freiburg, S.44-47.
- Holladay, R. (2013): *To hear this music you have to be there. Literally*. Transkript eines TED-Talks (TED@BCG), San Francisco, Oktober 2013, Link Stand 10.10.2014: http://www.ted.com/talks/ryan_holladay_to_hear_this_music_you_have_to_be_there_literally/transcript?language=en.
- Hug, D. (2008): *Towards a hermeneutics and typology of sound for interactive commodities*. In: Proceedings of the CHI 2008 Workshop on Sonic Interaction Design, Firenze, 2008.
- Ihde, D. (1976): *Listening and Voice: A Phenomenology of Sound*. Ohio University Press, Athens, Ohio.
- Juslin, P.N.; Västfjäll, D. (2008): *Emotional responses to music: The need to consider underlying mechanisms*. In: Behavioral and Brain Science 31, S.559-621.
- Juslin, P.N.; Liljeström, S.; Västfjäll, D.; Barradas, G.; Silva, A. (2008): *An Experience Sampling Study of Emotional Reactions to Music: Listener, Music, and Situation*. In: Emotion 8(5), S.668-683.
- Larsen, C.H.; Lauritsen, D.S.; Larsen, J.L.; Pilgaard, M.; Madsen, J.B.; Stenholt, R. (2014): *Aurally Aided Visual Search Performance Comparing Virtual Audio Systems*. In: Spatial Cognition 2014: Poster Presentations 8, Universität Bremen, Universität Freiburg, S.70-73.
- Lokki, T.; Järveläinen, H. (2001): *Subjective evaluation of auralization of physics-based room acoustics modeling*. In: Proceedings International Conference for Auditory Display, (Espoo, Finland), Juli 2001, S.26-31.

- Lotto, A.; Holt, L. (2010): *Psychology of auditory perception*. In: Wiley Interdisciplinary Reviews: Cognitive Science.
- McAdams, S.; Bigand, E. (1993a): *Introduction to auditory cognition*. In: McAdams und Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition*, Oxford, Clarendon, S.146–198.
- McAdams, S.; Bigand, E. (1993b): *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford University Press, Oxford.
- Massaro, D.W. (1987): *Categorical partition: a fuzzy logical model of categorization behavior*. In: Harnad, S., Ed. *Categorical Perception: The Groundwork of Cognition*, Massachusetts: Cambridge University Press, S.254-283.
- Mynatt, E.D.; Back, M.; Want, R.; Frederick, R. (1997): *Audio Aura: Light-Weight Audio Augmented Reality*. In: *Proceedings User Interface Software Technology 1997 (UIST 1997)*, Banff, Kanada, 14.-17. Oktober 1997, S.211-212.
- Myers, D.G. (2008): *Psychologie*. Springer Berlin.
- Narmour, E. (1991): *The Top-down and Bottom-up Systems of Musical Implication: Building on Meyer's Theory of Emotional Syntax*. In: *Music Perception: An Interdisciplinary Journal* 9(1), S.1-26.
- Neuhoff, J.G.; Hermann, T.; Hunt, A. (2011): *The Sonification Handbook*. Logos Verlag, Kapitel 4, S.63-85.
- T. Hermann, A. Hunt, and J. G. Neuhoff., Logos Verlag, 2011.
- Niessen, M.; van Maanen, L.; Andringa, T. (2008): *Disambiguating sounds through context*. In: *Proceedings of Semantic Computing, 2008 IEEE International Conference*, S.88-95.
- Proctor, R.W.; Reeve, T.G. (1990): *Stimulus-response compatibility: An integrated perspective*. Amsterdam: North Holland.
- Raffaseder, H. (2002): *Audiodesign*. Fachbuchverlag Leipzig im Carl-Hanser-Verlag.
- Raffaseder, H. (2003): *Audio-Design*. Foliensatz für den FH-Studiengang Medientechnik und – design, Hagenberg, MMD-3, WS 2003/2004 (www.raffaseder.com).
- Rozier, J.M. (2000): *Hear & There: An Augmented Reality System of Linked Audio*. Massachusetts Institute of Technology (MIT).
- Scarf, F.L. (1979): *Possible Traversals of Jupiter's Distant Magnetic Tail by Voyager and by Saturn*. In: *Journal of Geophysical Research* 84.
- Sanderson, P.M.; Anderson, J.; Watson, M. (2000): *Extending Ecological Interface Design to Auditory Displays*. In: *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OzCHI2000)*, CSIRO: Sydney, S.259-266.
- Sarter, M.; Givens, B.; Bruno, J.P. (2001): *The cognitive neuroscience of sustained attention: where top-down meets bottom-up*. In: *Brain Research Reviews* 35, S.146-160.
- Schilit, B.; Adams, N.; Want, R. (1994): *Context-Aware Computing Applications*. 1st International Workshop on Mobile Computing Systems and Applications, S.85-90.

- Schouten, B.; Gerrits, E.; Van Hessen, A. (2003): *The end of categorical perception as we know it*. Speech Community 41, S.71–80.
- Selye, H. (1976): *Stress in health and disease*. Woburn (MA), Butterworth.
- Shinn-Cunningham, B.G. (2002): *Spatial auditory displays*. In: International Encyclopedia of Ergonomics and Human Factors, 2nd edition, W. Karwowski, Ed. London: Taylor and Francis, Ltd.
- Shinn-Cunningham, B.G.; Streeter, T.; Gyss, J.-F. (2005): *Perceptual plasticity in spatial auditory displays*. In: ACM Transactions on Applied Perception 2, S.418-425.
- Singer, J.A.; White, S.M.; Edens, G.T.; Meike, R.C.; Charnley, D.; Hindus, D.; Burdick, W.; Stifelman, L. (1996): *Methods and Systems for Creating a Spatial Auditory Environment in and Audio Conference System*. Patent Nr. 5,889,843, Patent ID US5889843, filed 1996, granted 1999, United States Patent.
- Sonnenschein, D. (2011): *Sound Spheres: A Model of Psychoacoustic Space in Cinema*. The New Sound-track 1.1: S.13-27, Print.
- Sorkin, R.D. (1987): *Design of auditory and tactile displays*. In: G. Salvendy (Ed.), Handbook of human factors, New York: Wiley & Sons, S.549-576.
- Speeth, S.D. (1961): *Seismometer sounds*. In: Journal of the Acoustical Society of America 33, S.909-916.
- Steiner, P. (2009): *Sound Branding – Grundlagen der akustischen Markteinführung*. Wiesbaden, Gabler Verlag.
- Stewart, R. (2010): *Spatial Auditory Display for Acoustics and Music Collections*. Dissertation, School of Electronic Engineering and Computer Science, Queen Mary, University of London.
- Sundareswaran, V.; Wang, K.; Chen, S.; Behringer, R.; McGee, J.; Tam, C.; Zahorik, P. (2003): *3D audio augmented reality: implementation and experiments*. In: Proceedings of ISMAR '03, ACM Press, S.296-297.
- Truax, B. (2000): *Acoustic Communication*. Ablex, 2. Edition.
- Uzilevsky, G.; Andreev, V. (1993): *Iconic signs and languages in user interface development*. In: L. J. Bass, J. Gomostaev & C. Unger (Eds.), Human-Computer Interaction, Selected Papers from Third International Conference, EWHCI'93, Heidelberg: Springer-Verlag, S.115-124.
- Vazquez-Alvarez, Y.; Oakley, I.; Brewster, S. (2011): *Auditory Display Design for Exploration in Mobile Audio-Augmented Reality*. In: Personal and Ubiquitous Computing (2011), S.1–13.
- Walker, A.; Brewster, S. (2001): *"Sitting too close to the screen can be bad for your ears": A study of audio-visual location discrepancy detection under different visual projections*. In: Proceedings of the 7th International Conference on Auditory Display, Espoo, Finland, S.86-89.
- Walker, B.N. (2002): *Magnitude estimation of conceptual data dimensions for use in sonification*. In: Journal of Experimental Psychology 8(4), S.211-221.
- Walker, B.N.; Kramer, G. (2004): *Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making*. In: Ecological psychoacoustics, Neuhoff, J.G., Ed. New York: Academic Press, S.150-175.

- Walker, B.N.; Lane, D.M. (2001): *Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners*. In: Proceedings of the 7th International Conference on Auditory Display, Espoo, Finland, S.90-94.
- Weiser; M. (1991): *The Computer for the 21st Century*. In: Scientific American.
- Welsch, W. (1996): *Auf dem Weg zu einer Kultur des Hörens?* In: Grenzgänge der Ästhetik, Reclam, Stuttgart.
- Wigelius, H.; Vääätäjä, H. (2009): *Dimensions of Context Affecting User Experience in Mobile Work*. In: Proceedings of INTERACT 2009 (Uppsala, Sweden, Aug 26-28).
- Zerubavel, E. (1999): *Social mindscapes: an invitation to cognitive sociology*, Harvard University Press.
- Zotkin, D.N.; Duraiswami, R.; Davis, L.S. (2004): *Rendering localized spatial audio in a virtual auditory space*. In: IEEE Transactions on Multimedia 6(4), S.553–564.

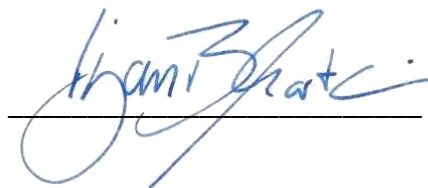
Abbildungsverzeichnis

Abbildung 1: Prozentuale Verteilung der Sinneswahrnehmung	16
Abbildung 2: Informationsgehalt eines Schallsignals	19
Abbildung 3: Klangstromanalyse	24
Abbildung 4: Top-down und Bottom-Up	28
Abbildung 5: Lokalisierung durch ITD und ILD	32
Abbildung 6: 3D-Modell	33
Abbildung 7: Akustische Datenrepräsentation	38
Abbildung 8: Klang, Struktur und Bedeutung	47
Abbildung 9: Kontextbeispiele	53
Abbildung 10: Soundgarden	67
Abbildung 11: Modell der binauralen Synthese	69
Abbildung 12: Wearable Augmented Reality Audio (WARA)	70
Abbildung 13: Komponenten von AR	77
Abbildung 14: Auditory Attitude Indicator	83

Versicherung über Selbständigkeit

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.

Hamburg, den 12.05.2015



A handwritten signature in blue ink, appearing to read 'Khan Zart', is written over a horizontal line.