

# Ein Gestaltungsansatz für intuitiv verständliches Audiofeedback

**Matthias Peissner**

Fraunhofer IAO  
Nobelstraße 12  
70569 Stuttgart  
matthias.peissner@iao.fraunhofer.de  
www.iao.fraunhofer.de

**Vivien Melcher**

MTO Psychologische Forschung und  
Beratung GmbH  
Schweickhardtstr. 5  
72072 Tübingen  
v.melcher@mto.de  
www.mto.de

## Abstract

In diesem Beitrag wird untersucht, in wie weit paraverbale menschliche Stimmäußerungen geeignet sind, um als Vorbild für die Gestaltung intuitiv verständlicher Feedbacktöne zu dienen.

In einer mehrstufigen Studie werden entsprechende Stimmaufnahmen erfasst und bezüglich ihrer effektiven Anwendbarkeit für grundlegende Interaktionssituationen bewertet.

## Keywords

Auditives Feedback, Sound Design, Sonification, Auditory Icons

## 1.0 Einleitung

Bei der Gestaltung von ergonomischen Benutzungsschnittstellen bietet der Einsatz akustischer Ausgaben erhebliche Vorteile – nicht nur für sehbehinderte Nutzer. Insbesondere in Situationen, in denen man sich nicht permanent auf ein visuelles Display konzentrieren kann, ist eine rein grafische Informationsdarstellung oftmals nicht ausreichend. Beispiele hierfür sind im industriellen Kontext, bei der Führung eines Fahrzeugs oder in mobilen Nutzungskontexten häufig anzutreffen. Akustische Signale helfen, die visuelle Beanspruchung des Nutzers zu reduzieren. Sie eignen sich besonders gut für Alarmer oder Erinnerungen und um die visuelle Aufmerksamkeit auf das grafische Display zu lenken, wo Detailinformationen gegeben werden (vgl. Brewster, 2002). Darüber hinaus gibt es Informationen, die sich akustisch deutlich besser darstellen lassen als visuell, wie z.B. Informationen über interne Prozesse komplexer Objekte (vgl. Gaver, 1989).

Geeignete Kombinationen akustischer und grafischer Darstellungsformen können zu einer gesteigerten Usability beitragen, z.B. durch die Verwendung von akustisch angereicherten Bedienelementen („sonically enhanced wid-

gets“, Brewster, 1998). Schließlich können durch den Einsatz von Audio überfüllte Bildschirmansichten auch auf kleinen Displays mobiler Endgeräte vermieden werden.

Nichtsprachliche akustische Ausgaben sind Sprache in vielen Belangen überlegen, insbesondere wenn ein kurzes Feedback auf Benutzeraktionen gegeben oder über Systemzustände bzw. Zustandsveränderungen informiert wird. Während Sprachausgaben sequenziell erfolgen und damit zeitintensiv sind, können nichtsprachliche Signale auch komplexe Informationen in sehr kurzen Meldungen übermitteln, indem die zur Verfügung stehenden Gestaltungsattribute Tonhöhe, Rhythmus, Lautstärke und Klangfarbe systematisch ausgeschöpft werden (vgl. Blattner et al., 1989).

Buxton et al. (1994) nennen drei wesentliche Einsatzbereiche für nichtsprachliche Audioausgaben: (1) Alarm- und Warnfunktion, (2) Information über den aktuellen Systemzustand und laufende Prozesse und (3) Darstellung von numerischen Daten. Die beiden erstgenannten Funktionen spielen bei fast allen Benutzungsschnittstellen eine zentrale Rolle und stehen auch im Mittelpunkt dieses Beitrags.

## 2.0 Zielsetzung und Fragestellung

Der vorliegende Artikel widmet sich der Frage, wie auditive Hinweise gestaltet werden können, damit sie intuitiv verstanden werden. Dabei konzentrieren wir uns auf einige Situationen und Ereignisse, die unabhängig von der Applikationsdomäne für viele Benutzungsschnittstellen interessant sind. Dem entsprechend abstrakt lassen sich diese wie folgt beschreiben:

1. Positive Bestätigung nachdem der Benutzer erfolgreich eine Funktion aufgerufen hat.
2. Positive Bestätigung nachdem das System einen Prozess erfolgreich abgeschlossen hat.
3. Negative Rückmeldung wenn der Benutzer versucht, eine aktuell nicht verfügbare Funktion aufzurufen.
4. Negative Rückmeldung wenn das System einen Prozess nicht erfolgreich abschließen kann.
5. Rückfrage vor der Ausführung einer irreversiblen Systemaktion
6. Rückfrage zur Klärung unvollständiger und mehrdeutiger Eingaben
7. Hinweis auf einen kritischen Systemzustand mittleren Schweregrads
8. Hinweis auf einen kritischen Systemzustand hoher Priorität
9. Erinnerung

Einige dieser Ereignisse erinnern an Situationen, die häufig in zwischenmenschlichen Dialogen auftreten. In diesen Situationen genügt oft eine kurze paraverbale Äußerung wie etwa »m-mh« oder »mmh?«, um die gewünschte Botschaft auszudrücken. Diese Überlegung führte uns zu der Hypothese, dass paraverbale Äußerungen möglicherweise als eine effektive Designmetapher für die Gestaltung auditiver Hinweise dienen könnten. Dieser Ansatz entspricht Gavers Konzept von »Auditory Icons« (Gaver, 1989). Auditory Icons greifen auf natürliche und alltägliche Klänge zurück, um die Objekte und Aktionen einer Benutzungsschnittstelle zu repräsentieren. Voraussetzung dafür ist eine Analogie zwischen der realen Welt und der »Modellwelt« des Computers (Gaver, 1997).

Die zentrale Fragestellung der durchgeführten Studienreihe lautet demnach: Können nach dem Vorbild paraverbaler menschlicher Äußerungen Signaltöne gestaltet werden, die grundlegende Ereignisse und Situationen von Benutzungsschnittstellen in einer intuitiv verständlichen Weise kommunizieren?

### 3.0 Vorgehensweise

Die Vorgehensweise zur Konkretisierung und Evaluation dieses Gestaltungsansatzes umfasst die folgenden Schritte:

1. Aufnahme paraverbaler menschlicher Äußerungen
2. Bewertung und Selektion der aufgenommenen Klangereignisse
3. Analyse der ausgewählten Klangereignisse und Synthese
4. Evaluation der synthetisch erzeugten Klangereignisse

Im diesem Beitrag können nicht alle Details der durchgeführten Aktivitäten wiedergegeben werden (s. Melcher, 2007 für eine vollständige Darstellung der Methodendurchführung).

### 4.0 Vorstudie: Aufnahme paraverbaler Äußerungen

Als Ausgangspunkt der Studie wurden zunächst paraverbale menschliche Äußerungen aufgenommen, die in einem experimentellen Setting gezielt provoziert wurden.

Hierzu nahmen acht Studenten der Staatlichen Hochschule für Musik und darstellende Kunst in Stuttgart mit einer Sprecherausbildung an einem Rollenspielexperiment teil. In dem Rollenspiel wurden neun alltägliche Kommunikationssituationen nachgestellt, die während einer Autofahrt zwischen Fahrer und Beifahrer auftreten können. Der Proband spielte den Beifahrer. Links neben ihm saß der Versuchsleiter, der den Fahrer spielte. Der Proband wurde aufgefordert, auf Fragen bzw. Anweisungen des Fahrers (Versuchsleiters) oder auf Anweisungen, die er über Kopfhörer erhielt, vokal aber nicht sprachlich zu reagieren. Somit standen dem Probanden nur paraverbale Äußerungen als mögliche Reaktionen zur Verfügung. Es wurden insgesamt neun verschiedene Rollenspiel-Szenarien so ausgewählt, dass eine inhaltliche Entsprechung zu den oben genannten Interaktionssituationen angenommen werden kann. So wurde zum Beispiel entsprechend zu Situation 3 (Negative Rückmeldung bei nicht ausführbarer Systemaktion) der Proband vom Versuchsleiter gebeten: »Gib mir bitte etwas zu trinken!«, während dem Probanden offensichtlich war, dass die Getränkeflasche bereits leer war. Die neun Situationen wurden jedem Probanden in randomisierter Reihenfolge dargeboten. Die Probandenäußerungen wurden mit einem Mikrophon aufgenommen. Nach jedem Szenario wurde die Aufnahme dem Probanden vorgespielt und er wurde gefragt, ob die intendierte Botschaft mit dieser Äußerung verständlich ausgedrückt sei. War dies nicht der Fall,

so wurde das Szenario wiederholt. Damit sollte sichergestellt werden, dass möglichst aussagekräftige Äußerungen für die weiteren Analysen zur Verfügung stehen.

Als Ergebnis dieser Vorstudie lagen insgesamt 72 Aufnahmen vor, jeweils 8 menschliche paraverbale Äußerungen für jede der neun Alltagssituationen<sup>1</sup>.

### 5.0 Bewertung und Selektion der aufgenommenen Äußerungen

Nun wurde in einem Zuordnungsexperiment überprüft, in wieweit die aufgenommenen Stimmäußerungen geeignet sind, einem naiven Hörer eine klare und unmissverständliche Botschaft zu kommunizieren. Neben der Validierung zielte dieser zweite Schritt insbesondere auf die Identifikation der am besten geeigneten Töne, mit denen dann weitergearbeitet werden sollte.

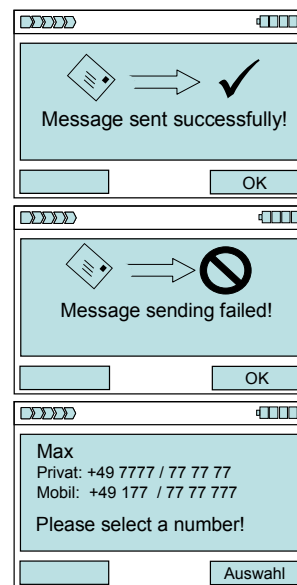


Abb. 1: Schematische Bildschirmansichten der Bedingung »Mobiltelefon« (Beispiele)

<sup>1</sup> Die Aufnahmen sind unter <http://www.hci.iao.fraunhofer.de/de/projekte/sounddesign> öffentlich zugänglich.

Für diese Studie wurden die in Abschnitt 2.0 beschriebenen Interaktionssituationen in konkrete Bildschirmansichten einer Waschmaschine und eines Mobiltelefons »übersetzt« (vgl. Abb. 1). 48 Probanden nahmen an der Studie teil. Sie ordneten jeder der 72 Aufnahmen jeweils die nach ihrem Eindruck am besten passende schematische Bildschirmansicht zu. Die Probanden wurden zufällig und in gleichen Anteilen einer der beiden Versuchsbedingungen »Mobiltelefon« und »Waschmaschine« zugeordnet. Da sich die Zuordnungsergebnisse beider Bedingungen für keine der Aufnahmen signifikant unterscheiden, scheint eine gemeinsame Analyse der Ergebnisse beider Bedingungen angemessen. Darüber hinaus kann dieser Befund als Hinweis dafür gedeutet werden, dass die Ergebnisse auch auf andere ähnliche Anwendungsbereiche übertragen werden können.

In vielen Fällen schienen die Probanden nicht in der Lage zu sein, zwischen zwei semantisch ähnlichen Situationen zu unterscheiden: den beiden Bestätigungen (1 und 2 der oben genannten Liste), den beiden negativen Rückmeldungen (3 und 4), den beiden Rückfragen (5 und 6), sowie den beiden kritischen Zuständen (7 und 8). Aufnahmen, die häufig mit einer Situation eines solchen Paares assoziiert wurden, wiesen auch eine enge Verbindung mit der anderen Situation desselben Paares auf. Daher betrachteten wir in der Folge nur noch die Zuordnungsraten dieser inhaltlichen Paare. Daraus ergaben sich für die einzelnen Aufnahmen Klassifikationsmuster mit Konzentrationen von bis zu 60-80% auf einem der Paare (vgl. Abb. 2). Einzige Ausnahme war die Erinnerung. Keine aufgenommene Äußerung wurde deutlich mit dieser Situation in Verbindung gebracht.

Nun wurde für jede der neun Interaktionssituation die damit am häufigsten und am zweithäufigsten assoziierte pa-

raverbale Äußerung bestimmt. Für Situation 4 (Negative Rückmeldung nach gescheiterter Ausführung) wiesen drei Aufnahmen dieselben hohen Zuordnungsraten auf. Daher gingen insgesamt 19 Audiodateien in die weiteren Analysen ein.

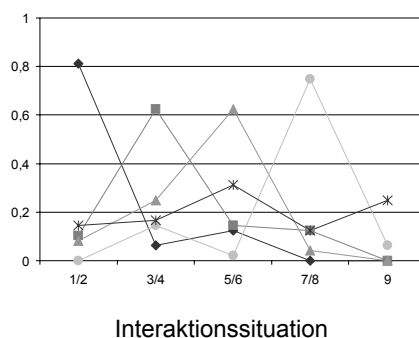


Abb. 2 Zuordnung von Äußerungen zu Interaktionssituationen: empirische Verteilung (relative Häufigkeiten) ausgewählter paraverbaler Äußerungen

Zusammenfassend kann festgestellt werden, dass einige der aufgenommenen paraverbalen Äußerungen zur Vertonung der diskutierten Interaktionssituationen sehr gut geeignet scheinen. Allerdings scheint der gewählte Gestaltungsansatz bezüglich seiner Anwendbarkeit auf bestimmte Situationen beschränkt zu sein.

#### 5.1 Analyse und Synthese der ausgewählten Äußerungen

Trotz der guten Verständlichkeit der ausgewählten Stimmaufnahmen gibt es gute Argumente gegen ihre direkte Verwendung als User Interface-Töne eines interaktiven Produkts. Neben einer gesteigerten Usability zielen akustische Gestaltungselemente insbesondere auch auf den Ausdruck bestimmter Produkteigenschaften, wie z. B. hochwertiger Materialien, und einen attraktiven Gesamteindruck. Darüber hinaus gewinnt die akustische Produktgestaltung auch im Bereich der gezielten Schaffung einer Markeniden-

tität mehr und mehr an Bedeutung. Aus diesem Grund sind wir auf der Suche nach generischen Gestaltungsregeln, die nur die relevanten informationstragenden Klangmerkmale betreffen und die anderen Klangaspekte der Produkt- und Markendifferenzierung überlassen. Durch derart minimale Gestaltungsregeln kann darüber hinaus auch eine wirtschaftliche und einfache Produktion entsprechender Interface-Klänge unterstützt werden.

Nach einer Segmentierung und Analyse<sup>2</sup> der ausgewählten Äußerungen entschieden wir uns in einem ersten Versuch für die folgende Vorgehensweise zur vereinfachten Klangsynthese: (1) Der *Rhythmus* wird unverändert beibehalten, d.h. Anzahl und Dauer der Klangsegmente werden direkt vom Original übernommen. (2) Die *Dynamik* der menschlichen Äußerungen wird in einer vereinfachten Form abgebildet. Dabei wird die durchschnittliche Intensität jedes Klangsegments über die gesamte Länge des entsprechenden synthetischen Klangsegments konstant gehalten. Intensitätsveränderungen innerhalb eines Segments werden damit nicht nachgebildet. (3) Alle *Frequenzmodulationen*, die als unbeabsichtigt angesehen werden können, werden durch die Verwendung der durchschnittlichen Grundfrequenz ( $f_0$ ) des jeweiligen Segments, eliminiert. Neben den zufälligen Frequenzschwankungen, die es in jedem Segment gibt, kommen auch beabsichtigte Frequenzmodulationen vor, die einen wesentlichen Bestandteil der Intonation menschlicher Sprache darstellen. Mit ihrer Hilfe wird beispielsweise zwischen Feststellung, Frage oder Aufforderung unterschieden. Während der

<sup>2</sup> Die Analyse der Aufnahmen wurde mit der Software Praat (www.praat.org, Boersma & Weenink, 2005) und Adobe Audition durchgeführt. Die Klangereignisse wurden in Segmente unterteilt und bezüglich Frequenz, Tonhöhe, Dauer und Intensität parametrisiert (vgl. Melcher, 2007)

Analyse stellten wir fest, dass es sich hierbei meistens um stetige Tonhöhenveränderungen um etwa eine Oktav nach oben oder unten handelt, d.h. etwa eine Verdopplung oder Halbierung der Grundfrequenz von Anfang bis Ende eines Segments. Diese Segmente werden durch einen entsprechenden linearen Oktavverlauf nachgestellt, der dieselbe durchschnittliche Grundfrequenz  $f_0$  besitzt wie das Ausgangsmaterial. (4) Die Klangfarbe der Stimmäußerungen wird vernachlässigt, indem zur Synthese reine Sinustöne verwendet werden.

Als Ergebnis dieses Arbeitsschrittes lagen insgesamt 19 synthetisch erzeugte Klangereignisse vor, die nach der geschilderten Vorgehensweise erstellt worden waren<sup>3</sup>.

## 5.2 Evaluation der synthetisch erzeugten Klangereignisse

Schließlich wurden die synthetisierten Klangereignisse in einer Studie, die dem ersten Zuordnungsexperiment gleich, evaluiert. Dabei stand die Bewertung der entwickelten Reproduktionsmethode im Vordergrund. Wenn die synthetisierten Klangereignisse ähnlich eindeutige Zuordnungsmuster erhalten, wie die menschlichen Äußerungen, dann kann davon ausgegangen werden, dass die Synthesemethode die wesentlichen informationstragenden Klangmerkmale erhält.

An der Untersuchung nahmen 30 Probanden teil. Ein Ausschnitt der Ergebnisse ist in Abb. 3 illustriert. Auch wenn einige synthetische Klangereignisse signifikant unterschiedliche Zuordnungsverteilungen aufweisen als die entsprechenden Originalaufnahmen, deuten die Ergebnisse darauf hin, dass zumindest bezüglich der jeweils am häufigsten assoziierten Situationen die rele-

vanten Informationen durch den Syntheseprozess erhalten bleiben.

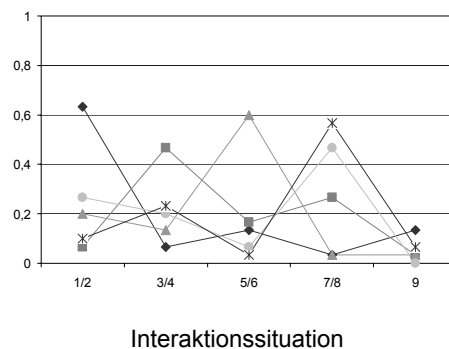


Abb. 3 Zuordnung von Äußerungen zu Interaktionssituationen: empirische Verteilung (relative Häufigkeiten) ausgewählter synthetischer Klangereignisse

## 6.0 Zusammenfassung

Die Ergebnisse der Studie unterstützen die Hypothese, dass paraverbale menschliche Äußerungen ein geeignetes Vorbild für die ergonomische Gestaltung akustischer Hinweise bei bestimmten Interaktionssituationen sein können. Für jedes der vier Situationspaare »Positive Bestätigung« (1 und 2), »Negative Rückmeldung« (3 und 4), »Rückfrage« (5 und 6) und »Kritischer Systemzustand« (7 und 8) können menschliche Äußerungen gefunden werden, die sowohl in der Originalfassung als auch als synthetisch nachgebildete Klangereignisse sehr gut verständlich sind. Für die Interaktionssituation »Erinnerung« kann mit der angewendeten Methode kein geeignetes Klangereignis gefunden werden. Dies deutet darauf hin, dass der beschriebene Gestaltungsansatz nur auf ganz bestimmte Interaktionssituationen angewendet werden kann.

Die vorgestellte Methode der synthetischen Nachbildung bringt Gestaltungsregeln hervor, die in Musiknotation aufgeschrieben werden können und so eine einfache Klangproduktion ermög-

lichen. In zukünftigen Studien möchten wir durch eine systematische Variation der Synthesemethode noch ein tieferes Verständnis für die informationstragenden Klangmerkmale erreichen. Außerdem wäre interessant, inwieweit Klangfarbenveränderungen die Verständlichkeit der aktuellen Sinustöne beeinflussen.

## 7.0 Literaturverzeichnis

- Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1), 11-44.
- Brewster, S.A. (1998). The design of sonically-enhanced widgets. *Interacting with Computers*, 1998; 11, 2: 211-235.
- Brewster, S.A. (2002). Non-speech auditory output. In Jacko, J.A. and Sears, A. (Eds) *Human-Computer Interaction Handbook*, Chap 12, pages pp. 220-239. Mahwah, NJ.: Lawrence Erlbaum Associates.
- Buxton, W., Gaver, W., Bly, S. (1994). *Auditory Interfaces: The Use of Non-Speech Audio at the Interface*. Unfinished book, manuscript at: <http://www.billbuxton.com/Audio.TOC.html> (retrieved February 2007).
- Gaver, W. (1989). SonicFinder: An interface that uses auditory icons. *Human Computer Interaction*, 4(1), 67-94.
- Gaver, W. (1997). *Auditory Interfaces*. In M. Helander & T. Landauer & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (2nd ed., pp. 1003-1042). Amsterdam: Elsevier.
- Melcher, V. (2007). *Intuitives auditives Feedback auf der Grundlage paraverbaler menschlicher Äußerungen*. Diplomarbeit am Institut für Psychologie der Humboldt-Universität zu Berlin.
- Spath, D., Peissner, M., Hagenmeyer, L. & Ringbauer, B. (2007). New Approaches to Intuitive Auditory User Interfaces. In: Smith, M.J. & Salvendy, G. (Eds.) *Methods, Techniques and Tools in Information Design*. Heidelberg: Springer.

<sup>3</sup> öffentlich zugänglich unter <http://www.hci.iao.fraunhofer.de/de/projekte/sounddesign>