Out From Behind the Curtain: Learning From a Human Auditory Display

Peter Parente
IBM
3039 Cornwallis Road
RTP, NC 27709 USA
pparent@us.ibm.com

Gary Bishop
University of North Carolina
Sitterson Hall CB #3175
Chapel Hill, NC 27599
gb@cs.unc.edu

Abstract
In this paper we describe an approach to gathering design requirements for a software auditory display by analyzing user interactions with an ideal partner: a talking human controlling a computer. We explain the potential benefits of studying such unconstrained user interaction before detailing the design and execution of our qualitative evaluation. We report the results of our thematic coding analysis and give examples of each of the seven major user difficulties and preferences identified. To conclude the paper, we summarize the application of our results to the design of a software auditory display for common office computing tasks.

Keywords
Auditory display, requirements gathering, qualitative evaluation, formative evaluation, ideal interaction

ACM Classification Keywords
H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces----Voice I/O, auditory (non-speech) feedback, evaluation/methodology.

Copyright is held by the author/owner(s).
CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.
**Introduction**

Researchers have studied the auditory display of information at many levels. Broadbent's dichotic listening experiments [8] in the 1950s represent the first of many investigations into human perception and understanding of audio. The results from these studies spawned interest in evaluating a variety of methods for presenting information aurally. Synthesized speech [3], auditory icons [12], earcons [19,20], ambient sound [21], stereophony and spatialization [2,19,26], and concurrency [1,15,20] are some of the many techniques studied with respect to parameters such as intelligibility, recognition rate, and mental load.

The steady increase in the capabilities of computers has facilitated the integration of these techniques in software. Telephone voice response units [6,16,17], video games [11,24,29], screen readers for people with visual impairments [4,10], eyes-free displays for mobile devices [18,22,25], and sonification of data [9,13,32] are just a few of the many applications of auditory displays to human-computer interaction. Summative assessments of these and other implementations have helped refine the usability of complete, working auditory display systems.

Formative evaluations, too, have contributed valuable knowledge to the design of auditory displays. Wizard of Oz studies, in particular, have been effective at collecting user feedback early in the development of auditory displays [6]. More importantly, such studies have provided insight into how users interact with proposed designs, free from the limitations of software implementations.

---

**Studies of Constrained Interaction**

All of the aforementioned evaluation techniques constrain the interaction between the user and display in some way. Studies of human perception typically make the user a passive listener. The display encodes information in some audio form, and the user must identify, distinguish, locate, or interpret it. The user has little or no control over the auditory display.

Summative evaluations of completed systems are as flexible as the software under test. The abilities of the system are fixed by the implementation, and the user is limited to use and configuration of the provided features. Beyond the available controls, the user's only recourse for overcoming problems is to suggest improvements.

Formative evaluations give the user as much freedom as the experimental condition allows. Nevertheless, even studies involving a human “wizard”, one having much greater flexibility and intelligence than software, often limit the interaction to what a predetermined design dictates. The user and wizard must effect human-computer interaction, not human-human cooperation.

Studies of constrained interactions do provide important information about auditory display usability. However, these evaluations say little about what display technique, design, or implementation is appropriate to an application domain or desirable to a group of users in the first place. The constrained nature of these studies limits their usefulness in gathering such requirements.

---

1 In this paper, we intend the term *auditory display* to cover both speech and non-speech output.
**Learning from Unconstrained Interaction**

Studies of uninhibited interaction between a user and an auditory display can inform early design choices whereas more constrained evaluations cannot. When a user can work with an auditory display as he or she sees fit, the information requested, the manner of the request, and the preferred response offer a glimpse of user preferences, tactics, and problems free from the effects of an intended design or actual implementation. These observations can prove invaluable in "pre-design" as evidenced by the work on SpeechActs [30,31].

Conducting an unconstrained evaluation requires the removal of inhibitions from both the auditory display and the user. How this goal is accomplished can vary as long as the user has the freedom to explore methods of getting work done. This paper reports one approach to eliciting highly open-ended interaction as a means of collecting design requirements for a software auditory display.

**Our Work**

In the earliest stages of designing the Clique auditory display [23], we conducted an evaluation of unconstrained user interaction with an intelligent auditory display: a talking human expert controlling a computer, reporting on its state, and carrying out user commands. We analyzed the resulting conversations between the user and this ideal “display” to help identify requirements for our software system. The following three questions, in particular, guided our study:

- What problems do users experience when completing typical office computing tasks such as sending email, browsing the Web, managing files, and scheduling appointments when using an auditory display?
- What methods do users employ to complete these tasks?
- What must the auditory display do to alleviate these problems and support these methods?

The remainder of this paper reports the design, execution, analysis, results of our evaluation. The concluding sections discuss the impact of this study on the design of our Clique auditory display. The method of evaluating unconstrained interaction, the requirements identified for auditory display in the domain of office computing, and the solutions implemented in Clique are the contributions of this paper to the research literature.

**Study Design**

Our study was structured as a role playing scenario involving a single participant and a principal investigator (PI) in each session. The PI asked the participant to assume the role of a personal assistant to the chief executive officer (CEO) of the imaginary XYZ Corporation. In this imaginary scenario, the CEO had purchased a new, intelligent auditory display that would allow the assistant to use any standard desktop application eyes-free. The CEO was asking the assistant to evaluate the software in his or her office for thirty minutes by performing typical tasks such as answering email, scheduling appointments, and searching the Web. At the conclusion of the test, the CEO and assistant would discuss the potential of the system to keep the user working while in transit running daily company errands.
The PI next explained that he would play the part of the auditory display for this test. He would listen to commands and questions form the user, carry out requests on a standard desktop computer, and respond vocally. The PI went on to explain that the participant would speak to issue commands or ask questions, even though a mobile device might feature a keyboard or keypad instead. The PI would have access to the mouse, keyboard, and screen of the desktop computer while the participant would not.

**Scenario Rules**
The PI stated some rules to which he would adhere during the scenario. We created the rules to give the user confidence in the PI as a faithful auditory display, not one that would purposely mislead the user as part of the experiment. We kept the rules as general as possible to avoid limiting the abilities of the user and display. List 1 states these rules as reported to the participant.

**Removing Participant Preconceptions**
The PI concluded the introduction by playing three example audio clips. Each clip was a recording of a user completing office work using the PI as an intelligent auditory display. These examples served to dispel any participant notions about having to constrain their interaction with the PI as if he had limited abilities or had to act like a computer. For example, one recording demonstrated that the user could ask high-level questions such as “What other fields must I complete before sending this email?”

**User Tasks**
The tasks to be completed by the participant were initiated by email messages received during the thirty-minute session. At the start of the scenario, ten such emails were located in the participant's inbox awaiting processing. Up to four new emails arrived later at predetermined times or in response to user actions. Six of the fourteen total emails required the participant to perform non-trivial actions to fulfill their requests. The remaining emails provided information required to complete the tasks assigned by the primary six. List 2 summarizes the six tasks.

**Open-Ended Discussion**
At the completion of the scenario, the PI first asked the participant to make comments or pose questions. The PI then asked each participant to comment on the difficulty of tasks, likes and dislikes, his or her ability to master this kind of interaction, how this interaction compared with his or her everyday use of a computer, and so on.

**Study Execution**
We recruited twenty people familiar with sending and receiving email, browsing the Web, and managing files to participate in our study. Seven participants (4 males, 3 females) had a visual impairment while thirteen participants (10 males, 3 females) did not. Four participants with visual impairments (1 male, 3 females) had been blind since childhood while the remaining three developed the impairment later in life. All but one of the participants with visual impairments relied on a screen reader in their daily computer use while the exception could rely on magnification alone.

We treated the first three sessions as pilots (2 males with vision, 1 female without) in order to debug the test procedure and give the PI practice in the role of an auditory display.
auditory display. We did not include the results from these sessions in our analysis.

**Equipment and Setting**
A standard desktop computer was on-hand to assist the PI during the study sessions. Applications representative of those needed to complete the required tasks, copies of the emails assigning tasks, and mock-ups of the necessary XYZ company Web pages were pre-installed on the computer. The participant and principle investigator were seated such that participant could not read the computer screen. Participants were allowed to see the PI to reinforce the idea that they were interacting with an unconstrained auditory display. The PI consciously avoided looking at the user to prevent backchannel communication.

In each session, the PI used a standalone audio recorder to capture all interaction between the participant and the display. The recordings included both the content of the thirty-minute scenario and follow-up discussion.

**Study Analysis**
After running all of the study sessions, we transcribed the audio recordings into plain text files. The transcriptions indicated spoken dialog, pauses longer than three seconds, and any laughter or other non-speech vocalizations. Once complete, we reviewed the transcriptions for accuracy while listening to the original audio recordings again.

We next used Weft QDA [28] to code (i.e., label) segments of each transcription according to patterns of behavior seen across participants. Twenty-one thematic codes resulted from the first pass through the transcriptions. We reviewed the initial codes for the purpose of grouping them into more general categories and eliminating ones not directly related to our work on Clique. Seven major themes emerged from this grouping. Finally, we coded the text transcriptions again using only the final seven codes. Table 1 states the initial codes and their grouping into the final seven.

**Coding Validation**
To quickly validate our codings in this pre-design study, we used a cheap triangulation technique. We first recruited one former and one current computer science graduate student to assist us. The two students were familiar with our research but neither was knowledgeable about the details of this study.

We gave each validator a unique simple random sample of fifty segments drawn out of the total population of five-hundred twenty coded segments mixed with a simple random sample of uncoded segments. We also provided the validators with a list of the seven final codes, their definitions, and a few fabricated conversation segments exemplifying each. We asked the validators to label their samples with zero, one, or more of the codes per segment.

We next compared the validator codings with our own. For each segment in a sample, we counted the number matches between the validator codes and ours. We summed the number of matches and divided by the total number of codes we assigned over the whole sample to produce a simple scoring metric:

\[
\text{score} = \frac{\# \text{ overlapping codes}}{\# \text{ our codes}}
\]

**List 2.** Summaries of the six primary tasks assigned to the participant. The emails did not state these solutions directly. Rather, they read like requests from other people.

- Schedule a meeting on the CEO’s calendar on January 10th of the coming year. Reschedule all other appointments on that date.
- Forward an email stored in a local mail folder to an XYZ employee.
- Attach a local file to an email and send it to a customer. Recall the location of the file from a previous email.
- Respond to an employee saying that the CEO is busy on January 10th. Schedule the meeting after receiving confirmation.
- Download a file from a website. Attach it to an email and send it to the CEO. Recall the URL, username, and password for the site from a previous email.
- Find the direct URL to the UNC video library Web page. Email the URL to the CEO.
Using this metric, we found that 63% of the codes from the first validator and 80% of the codes for the second validator were in concordance with our labeling. These scores gave us adequate confidence to interpret our themes, though we recognized the shortcomings of this simplistic but fast approach to triangulation (e.g., differences in theme interpretation, inability of validators to suggest additional themes).

**Study Results**

Only three of the seventeen non-pilot participants completed all six tasks successfully (SM03, SM09, SM15). Out of the remaining users, six people completed five tasks (SF05, SM06, SF10, SM11, BF14, BM22), five completed four (SF07, BM08, SM12, SM13, BF23), one person completed three (BF20), and two people completed one (SM04, BF21). Five of the seventeen participants ran out of time before completing all of the tasks. The others quit early believing they had completed all of their work.

**Working Memory**

The transience of auditory information is a known source of interaction problems [14]. In this study, the transience property of audio resulted in participants repeatedly forgetting information before it could be put to use in completing a task (183 coded segments). Participants forgot information both after extended periods of time and, more surprisingly, almost immediately after hearing it. (See Transcript 1.)

Participants who recognized the limitations of their working memory coped with the problem in three ways, each one more efficient than the last. First, participants interrupted their current task, switched contexts to locate the missing information, and then resumed the stalled task with the forgotten information in mind. Second, participants guessed forgotten content with the expectation that the auditory display would correct them based on the partial information given in their guesses. Third, participants occasionally asked the auditory display to remember information and recite it later. (See Transcripts 1, 2, and 3.)

**Prompting and Discovery**

The invisibility of information is a second source of problems in auditory displays [14]. In our scenario, participants sometimes had trouble advancing the current task because of the invisibility of potential actions. When faced with this problem, users started asking very specific questions requiring yes or no answers. After asking a few such narrow questions and failing to identify a potential next step, participants began to broaden their questions.

The most common method of discovering available options, however, was to wait for display initiated prompts (40 segments). When the participant faltered for a significant amount of time, the PI asked the user a question introducing one or more novel possibilities for what to do next. (See Transcript 4.)

When asked what aspects of the interaction they found useful and useless, participants often made positive comments about the display initiated prompts and never stated that they were useless. Whether participants would hold the same opinion of less task-aware prompts likely to be given by a software system is an open question.
Visual Thinking
When struggling to find the next step toward completing a task, sighted participants sometimes introduced unnecessary visual concepts into the interaction or reported that they were trying to picture a visual interface for their current task (17 segments). Users with visual impairments mentioned nearly as many visual terms throughout the scenario (14 segments). This result seems counterintuitive, but is consistent with the way in which screen readers expose GUI concepts to users with visual impairments. (See Transcript 5.)

Some sighted participants commented on the lack of persistent visual cues as a reason for their less than ideal performance. We considered these comments as additional evidence that users were frustrated with trying to remember too much information and forgetting it repeatedly.

Content Summaries
All seventeen participants used summaries to gain an overview of large bodies of content and to reduce the amount of time spent weeding through irrelevant information (142 segments). Participants accomplished the first goal by asking questions about the quantity of a certain type of content. Users approached the second goal by asking questions intended to filter content according to a set of guidelines. A couple of participants took even more advantage of intelligent summaries by asking the auditory display to apply permanent filters to their queries. (See Transcripts 6 and 7.)

Some users commented on the usefulness of the summaries in avoiding lengthy reports and in customizing the auditory display. One user made specific comments about how such summaries would make reports from her screen reader more "succinct" and stop it from "bombarding [her] with stuff."

Navigation by Search
In addition to using summaries to sift through content, participants employed two search strategies to locate details of interest (63 segments). The first approach amounted to a keyword content or property search. The second method used references to recently reviewed content to direct the search. (See Transcripts 8 and 9.)

Unobtrusive Alerts
In the rules of the scenario, the PI stated that he would inform the participant of any important events during breaks in the conversation. These notifications often concerned the asynchronous arrival of new emails and notification of errors while users were busy with unrelated tasks (51 segments). Users sometimes reacted to these notifications by handling them immediately. Other times the same users ignored them and continued with the current task. (See transcript 10.)

During questioning, some participants expressed favorable opinions toward the notifications. One participant even mentioned that he regretted not being interrupted immediately by new mail alerts, rather than during lulls in the interaction. He said he was confused by the changes made to his inbox by the arrival of mail left unannounced by fast-paced interaction. On the other hand, one participant stated that he found all notifications pointless in this particular scenario.
**Batch Operations**

Finally, a few participants created batches of simple operations that they later applied to complete similar tasks. The use of batched actions appeared in only a few places (9 segments), but at least one participant found them noteworthy for their potential in eliminating redundant tasks. (See Transcript 11).

The fact that few participants batched actions is aligned with the results reported by Bennett and Edwards in their study of a speech interface for circuit diagrams [5]. They found that users tended to explore diagrams element by element rather than by larger structures. Users in our study mirrored this tendency toward using low-level, atomic interactions in most cases.

**Example Solutions in Clique**

We applied the findings from our study to the design of Clique, a prototype used to research improvements to the usability of auditory displays for multitasking environments. The following sections describe three features of Clique that specifically address problems and preferences identified in our evaluation.

**Supplementing User Memory**

Tasks in our study requiring use of multiple applications stressed user working memory. Both the amount of task-critical information encountered and the length of time between its discovery and use caused problems for participants.

We addressed user memory problems by providing two kinds of reminders in Clique. First, we implemented a tape-recorder feature that allows a user to store any number of spoken utterances for later review, at any time, in any context. In effect, the tape-recorder offloads some of the burden of memorization and recall from the user to the computer which has practically infinite, perfect memory. Second, we paired applications and tasks with ambient sounds of unique environments to improve user awareness of the current context and aid later recollection of information from that context [27].

**Providing Summaries**

Summaries of content were important to users in prioritizing work and avoiding unnecessary information. Indeed, because serial speech output is slow relative to viewing information on a high-bandwidth visual display, summaries were critical to the timely completion of tasks in our scenario.

In Clique, we satisfied the need for summaries by designing earcon grammars [7] and auditory icons [12] to precede, replace, or supplement more detailed spoken reports. As users browse, search, and edit content, Clique uses these relatively brief sounds to indicate the commands supported in the current context, the size of the content, the results of user actions on the content, errors, special states, and so on. Users can request spoken descriptions of the information encoded in these sounds at any time, plus further details omitted from the summaries.

**Alerting Without Forcing**

Most users in our study expressed a desire for timely notifications of asynchronous events. Whether users acted on these alerts or not was dependent on their preferences, the current task, the importance of the alert, and so on.
We enabled immediate alerts in Clique without interrupting the user by placing notifications in spatialized, concurrent auditory streams. While a user is busy listening to information about the current task, Clique can notify the user of events from other tasks by playing a sound or speaking in the secondary streams. The user is then free to take note of the background announcements, query for more details about them, or ignore them altogether as desired.

Conclusion
In this paper we demonstrated how evaluating users interacting with an unconstrained auditory display can aid in the gathering of design requirements for a software system. We presented the results of one such evaluation involving sighted and blind users working with a human auditory display for office applications. We concluded by giving some examples of how we resolved problems and satisfied preferences identified in our study in the implementation of a software auditory display.

We posit that the methods described in this paper can be useful in designing auditory displays in application domains beyond office computing (e.g., data auralization). Furthermore, we suggest that evaluating unconstrained interaction might be a valuable technique for other modalities as well (e.g., GUIs) if a human can effectively serve as the ideal interface.

Acknowledgements
This material is based upon work supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Citations