Auditory Perspective of Interfaces: An Overview

Simon Liu
University of Auckland
20 Symonds Street, Auckland
wliu080@aucklanduni.ac.nz

ABSTRACT

Current paradigms of interacting with computers are predominantly visuals-centric. However, the data end-users are expected to deal with are ever-increasing, and making full use of other senses can alleviate the burden of representing the entirety of data via visuals alone. Research into the auditory perspective of interfaces is a fairly new field, with strong focus on psychoacoustics, parameters of sound in design and sonification in exploring data sets, with a lot of strong basis in designing for the visually impaired.

Focus has been put into novel applications of a more sound-centric system of interfacing with computers, but more research needs to be done in looking at unified ways of sonification and also into core aspects of audio displays as a means of complementing existing displays. Despite advancements in technology, as this field is quite new, current experimental research typically are designed to confirm various aspects of sound in their feasibility for use as a natural progression from just sound used in displays as auditory icons or earcons.

Other areas of research look into the use of sound in dealing and coping with large amounts of data, and models and frameworks with which to design a system for enabling a smooth transition to such a model.

This report will aim to provide a brief and cursory look into several auditory research papers, covering three core aspects of the field of research in auditory human-computer interaction that are geared towards typical end-users, and not just for the visually impaired.

Author Keywords

Sonification; model; listening; auditory display; interface

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces: Auditory (non-speech) feedback

General Terms

Human Factors

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
CHI'12, May 5–10, 2012, Austin, Texas, USA.
Copyright 2012 ACM 978-1-4503-1015-4/12/05...$10.00.

INTRODUCTION

The focus of this paper will be to give a more detailed overview into three key areas of research in usage of auditory concepts in interfaces – the studies and use of auditory icons/earcons/spearcons; use of audio and sonification in exploratory data analysis; and studies into proposed models and frameworks with regards to auditory interfaces and sonification design. Focus will not be in solutions geared just towards the visually impaired, as more and more recent research consider its applications for all general end-users.

The terms auditory icon, earcon and spearcon are well defined and accepted in use - auditory icons, audio sounds that have natural, real world mappings; earcons, audio sounds for events without a real world mapping; and spearcons, which is speech speed up to the point of sounding as an earcon. With the possible exception of spearcons, these concepts are not fairly new, but a lot of research is being undertaken to gauge the effectiveness of its usage, determine models in designing such sounds in a way that contextually works for the majority of users and comparative looks into how well auditory icons, earcons and spearcons fare in various experimental conditions.

Sonification in data analysis is being pursued for its' potential in being more effective in conveying and allowing for easier pattern recognition in large sets of data. Models are being developed and proposed to aid sonification – determining what aspects of sound map to the data and how it can best be used to explore such data. Research is being made into how an audio environment itself can help someone perceive and deal with data. An aspect of this paper will be to look into a proposed model for sonification based data display and a study into the impact of sonified graphs and the resulting levels of comprehension.

A common current across many papers relating to the auditory perspective of interfaces is an acknowledgement of the need for more research into more concrete frameworks and models to unify various aspects of auditory & sonification design. Covered in this paper will be a look into various methods and frameworks with regards to auditory display design.

AUDIO CUES

The impact of the various types of audio cues have long been studied, but as each type is more geared towards different situations, comparative studies tend to have difficulty in creating an experiment that has strong real-world links yet at the same time is unbiased towards any of the three main audio cue categories. Typical approaches involve abstracting the problem to the point where all three categories appear to apply – a comparative study on the three audio cue types with respect to navigational performance in auditory menus by Walker et al. [7] discovered that there was no real-world equivalent to match menu items such as "auditory icons cannot be reliably created for items such as 'Select Table'", thus relegating the experiment to menu items such as Bird, Dog, Horse, etc., which have less application than the ideal testing on common menu items such as File, Open, Save, etc. Applications have already been designed to aid the design and creation of audio cues – one such example is the AWESOME web-based tool, by Fagerlönn [3]. However, it is fairly primitive in scope, and instead of actually facilitating user manipulation of the sound files, only the illusion of this control is given, with the sounds pre-recorded for the settings. This however, limits parameters – "three sound parameters in three levels will require 27 sounds … four sound parameters in four levels will require 256 sounds." [3].

Auditory Icons

Typically, research done on auditory icons is comparative, being compared against the other types of audio cues, as they are more established, being the natural assumption for audio cue types, whereas types such as earcons are less intuitive in their efficiency and have far more parameters and area for growth.

Earcons

Usage of earcons tends more towards the area of sonification and its potential in data analysis. As earcons have no 'natural' ties, arbitrary mapping systems [7] can be used, depending on the scenario and situation. Whilst this theoretically provides earcons with greater flexibility than auditory icons, it also means more research needs to be done into ways of sonifying data, what parameters of sound lend itself best to being mapped [6] and to also take into consideration that a different scenario, such as its use in navigating menus may mean a completely different mapping schema is required. Common parameters varied with earcons are in rhythm, timbre, pitch and intensity, but the effects that each parameter has on users' perception of the earcon are still being documented and studied. A study by Brewster et al [2] was conducted on purely earcons compared against "unstructured bursts of sound" where users would be presented with a display of icons with an accompanying earcon. Later on, the participants were played an earcon and had to provide information on it. What was discovered was that in one phase (sounds for icons) "rhythms were ineffective but in phase II they produced significantly better results" [2], with the suggestion for more research into the effects of rhythm. In general however, earcons with some form of underlying metaphor tend to fare better, although as tested by Brewster et al [2], the difference between standard and musical earcons was not statistically significant.

Spearcons

A growing amount of research is being conducted into spearcons, though not for data display. Spearcons are similar to the text-to-speech that is provided for visually impaired users. Despite spearcons being touted as "created by speeding up a spoken phrase until it is not recognized as speech", there is an overlap in spoken speed between those pronounced by applications such as JAWS for proficient users and slower spearcons, some spearcons being only 'compressed' to 50% original length, or otherwise a two times speed up without loss in pitch. Studies have found that spearcons are very well suited for aiding the usability and navigation of hierarchical word-based structures [7]. In the study by Walker et al [7] on spearcons in a menu-based environment, participants were made to select a menu item, but the menu was only provided audibly – either speech only, auditory icon + speech (after a pause by the user on the same menu item for too long), earcon + speech or spearcon + speech. Despite Walker et al. not being able to test on familiar menu items, it was noticed that spearcons created from natural 'families' of menu items, such as 'Save' and 'Save As', would also be "acoustically similar at the beginning of their sounds", whilst still being unique (in the case of 'Save' and 'Save As', the spearcons would differ in length). It was also noted that there is a noticeable improvement with spearcons over that of speech, though not statistically significant [7], but Walker et al confer that usage of audio cues in menu traversal "is not typically intended to speed up overall performance" but for usability and provision of context during menu traversal. Whilst future use of spearcons may not replace visual icons and text, due to the fact that they are based on speech, spearcons could possibly be readily dynamically generated, requiring little developer effort but at the same time aiding in usability of interfaces.

DATA ANALYSIS

An important branch of research into novel aspects of the auditory perspective of interfaces is with using sound to portray, traverse and analyze large amounts of data. To facilitate this end, research has been done on the feasibility of extracting pattern/data from sonification [4], and research continues to work towards a viable model for designing sonification displays [6]. In Saue's proposed model for interaction in exploratory sonification displays [6], Saue presents the concept of the inherent time aspect of sonification and the mapping of non-temporal to temporal domain, known as temporalization. Four different classifications of time scales are presented to facilitate a model mapping – spectral time, local variations in timbre on a short (<50ms) scale; rhythmic time, relative change to events locally (<2s); event time, "irregularly spaced singular events" (>2s) and ambient time, always present no change or slow change. The metaphor of 'walking' as the means of traversal through large data sets is proposed, with an extended example of walking through a forest and dealing/analyzing the sounds heard as justification to the decision. A notable challenge prevalent in this type of sonification display for data sets is the fact that "sound is inherently a temporal medium" [6], developers must consider how to deal with a user constantly traversing the data and when the user is perfectly stationary. Just as in a visual representation of large quantities of data, constant traversal of data set may mean that it becomes hard to later go back to points of interest. Saue suggests two options for this: "Mark the spot with a visual marker for instant callback" or "Make a path through the spot, so that we can retrace it later" [6]. The design consideration for when the user is perfectly stationary is whether to constantly maintain the sonification at that point in time, or whether to loop back the past few moments in the time traversal, but in an obvious way. Saue accedes that the model is far from being implemented but has the future goal of modeling a data set with a soundscape mimicking the conditions of the scenario itself – "a virtual seismic sonification display could be modeled as a subterranean cave in which the listener moves around" [6].

Another approach to researching the feasibility of sonification is studying perceptibility of sonifications. A recent study (2011) on the ability of users, with minimal training, to visually represent changes in pitch on a graph was conducted by Patel et al [4]. This study was conducted with a strong basis on existing research into "comprehension rates and reaction times of listeners" presented with sonifications and also matching of sonifications to its corresponding visual representation. Their goal to "identify the factors that drive or influence a listener's ability to accurately comprehend and process a sonification" [4] was met through participants being instructed to draw a line graph, visual representation of a sonification played to them, on a tablet device. Unlike audio cues, which are typically short in length, the ability to comprehend changes in the sonification are more related to data analysis on an audible field, and tested sonifications were all "15 seconds in length" [4]. From the results an "overall recognition accuracy rate of 76%" was obtained. Whilst this was statistically significant compared to the null hypothesis of 50% accuracy (with p-value of 0.0000), a 76% rate does not give high hopes to the reliability of being able to accurately convey information in a sonification display. However, more research will need to be done to determine whether this value changes in different scenarios, such as different sonification method use, or asking participants to only look for a certain pattern in the data. Another result obtained from the experiment by Patel et al, is that overall, a participant's ability to recognize the correct magnitude of the sonification given that they recognized that a tone was being played was 89% [4]. More research into ways and the corresponding efficiency with which a user perceives various sonifications will play a big part in designing a framework for designing models to deal with auditory data display of all scenarios.

MODELS AND FRAMEWORKS

The number of proposed models and frameworks in this field of study are far too numerous to devote just a section of a paper to, and indeed, there are papers which are devoted in reviewing and providing suggestions towards suitability of models and frameworks [1]. For the purposes of giving a broad overview, only a few models and framework styles will be outlined, from Brazil [1].

Model – Repertory Grid

This model is based on the idea that "the world can be seen through the contrasts and similarities of events" [1]. Direct elicitation is used to obtain a description of an event on a per participant basis. Based on the judgments and descriptions formed a more personal and relational view of the event can be ascertained. Despite relying on direct user specification, "studies have shown that only within a small number of sounds does any large degree of dissimilarity exist". With the data obtained in a repertory grid, various results can then be generated.

This method is not just used in auditory interaction design, as the model itself is generic enough to be applied to uses in food technology or even medicinal fields.

Model – Similarity Rating / Similarity Scaling

The basis of this model relies on participants being asked to scale events based on acoustic dimensions, with the difference between similarity rating and similarity scaling being whether this rating occurs on a multi-dimensional (scaling) level or simply a mono-dimensional (rating) level. This method allows for gauging of each participants way of forming links and patterns, and provides insight into why the particular user may "confuse particular sounds or types of sound" [1].

Model – Ear-witness Accounts

A play on the term, 'eye-witness accounts', the use of ear-witness accounts behaves similar to its namesake – an attempt to provide a more authentic experience of the audio as it was heard, as opposed to a "simple recording of an auditory event…not a true reflection of the experiences of a listener present" [1]. The documentation of such an event is specific and detailed, not an idealized account of the sound.

Model– Saue's Temporal Exploratory Sonification Model

The temporal exploratory sonification model focuses primarily on the aspect of temporalization of data, the mapping from "a non-temporal to a temporal domain" [6], and its subsequent sonification for use in exploratory data analysis. The model links 4 different types of sounds to 4 classifications of time scale and the subsequent types of tasks that relate to it. This attempts to map the structure and facets of a data set to a soundscape, where the types of sounds reflect the nature of the data set. Saue intuits that a user would then be able to naturally analyse the data as a sound environment as though one would naturally detect patterns in a natural, real-world environment.

Framework Style – TaDa

TaDA, task and data analysis, refers to the construction of sound in a manner that only conveys useful information, and this gathered by the analysis of the problem space. There are three main analysis parts that are performed with TaDA: "the first being the task analysis of the question, the second is the information analysis of the answers, and the third part is the data characterization of the subject" [1], with the objective in mind to determine the core properties of the problem and also the linking between activities in the problem space.

Framework Style – Narrative Sound Artefacts

The narrative sound artifact framework relies on a narrative-centric approach for the design and creation of sound artefacts. Brazil lists Back's micro narratives and Hug's design oriented approach [1] as tried approaches to the Narrative Sound Artefact framework. The micro narrative approach treats the concept of stories being a small event, and can be combined in a modular fashion. A design process is used on the stories, "metatopics", and can be combined by the user to create artefacts. A sonic elaboration process is then used to come to a prototype, and this is then reviewed. This focus allows for a more dynamic approach to the creation of artefacts and allows for faster changes to scenarios.

Framework Style – Pattern Based

Pattern based frameworks are not unique to the auditory field, and are common in most disciplines, including software engineering. The principles behind pattern-based frameworks rely on the idea of reuse – that common sub-problems tend to arise, and by careful documentation of solutions to such problems, patterns facilitate good design practice whenever a similar problem is recognized. The *paco* framework outlined by Brazil [1] focuses on the aspect of context space as a means of categorization of design patterns. This allowed for "multi-dimensional classification according to context of use" and aided other designers to match problem to pattern.

SUMMARY

The area of auditory human-computer interaction research is largely focused on perception of audio and sound, sound as a means of data exploration, models and frameworks for the design and elicitation of sonifications and audio cues.

Research into audio cues – auditory icons, earcons and more recently spearcons are still ongoing, as there are too many aspects of sound and scenarios where audio cues are useful still to be considered.

Research into data analysis are in part hampered by research into how to best sonify data and slightly lacking development into frameworks that are best suited to such a generic scope as data analysis. Granted, more research is occurring in this branch given that most studies have shown reasonable success in proving the feasibility of sound as useful aid, if not stand alone tool, in handling with large data sets.

Future Work

There is much to be done in the field of auditory interfacing as it is a growing but yet largely untapped aspect of computing, and a general consensus is that as more and more information handling is expected of end-users, more ways of utilizing sound as a human-computer interaction dimension will be required to enable such a level of growth [4]. There appears to be a growing demand for research into developer more robust models for the design, evaluation and deployment of sonifications and various implementations of audio cues. Doing so will enable more rapid and robust development of applications geared towards implementing novel sound interaction paradigms.

Another aspect of research potential should be in the area of user perception and handling of sounds, with a focus on discovering which aspects of sound lend more to a user's pattern detection, for data analysis, as well as how to handle or train user accuracy in use of, or pertaining to acknowledgement of different sound cues.

REFERENCES

1. Brazil, E., A Review of Methods and Frameworks for Sonic Interaction Design: Exploring Existing Approaches. *Proc. Intl. Conf. Auditory Display,* (2010), 41-67 http://www.springerlink.com/content/y8q105lq4w6832g6/fulltext.pdf
2. Brewster, S.A., Wright, P.C., Edwards, A.D.N., An Evaluation of Earcons for Use in Auditory Human-Computer Interfaces. *Conf. Human Factors in Computing Systems*, (1993), 222-227 http://dx.doi.org/10.1145/169059.169179
3. Fagerlönn, J., Liljedahl, M., Designing a Web-Based Tool that Informs the Audio Design Process. *Proc. Intl. Conf. Auditory Display,* (2010), 68-79 http://dx.doi.org/10.1007/978-3-642-12439-6_4
4. Patel, N.S., Hughes, D.E., Studying Listener Comprehension of Sonifications through Visual Replication. *I/ITSEC,* (2011) http://ntsa.metapress.com/app/home/contribution.asp?referrer=parent&backto=issue,123,150;journal,1,22;linkingpublicationresults,1:113340,1
5. Ramloll, R., Brewster, S., An Environment for Studying the Impact of Spatialising Sonified Graphs on Data Comprehension. *IProc. Sixth Intl. Conf. Information*

Visualization, (2012), 167-174 http://dx.doi.org/10.1109/IV.2002.1028773

6. Saue, Sigurd, A Model for Interaction in Exploratory Sonification Displays. *Proc. ICAD*, (2000), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.1629&rep=rep1&type=pdf
7. Walker, B.N., Nance, A., Lindsay, J., Spearcons: Speech-Based Earcons Improve Navigation Performance in Auditory Menus. *Proc. ICAD*, (2006), 63-68 http://ntsa.metapress.com/app/home/contribution.asp?referrer=parent&backto=issue,123,150;journal,1,22;linkingpublicationresults,1:113340,1