Human-Computer Interaction Design based on Interactive Sonification – Hearing Actions or Instruments/Agents.

Mikael Fernström, Eoin Brazil

***Abstract*— this paper outlines a number of steps for human-computer interaction design using sound as representation - auditory icons. The design process is based on listening tests, gathering free-text identification responses from participants. These responses and their classifications can then suggest how accurately the sounds can be identified and possible metaphors and mappings of sound to human action and system status.**

Finally, we conclude with a practical design example, providing a pesudo-haptic user experience through sound to convey information about software-defined user interface components.

***Index Terms*— Interactive Sonification, Interaction Design. Auditory Icons, Actions, Instruments, Agents, Ecological Acoustics, Everyday Sounds, Identification, Classification.**

I. INTRODUCTION

It is somewhat a mystery that sound is not used more extensively in human-computer interaction design as we in the real world can hear a lot but, perhaps, the auditory reality is too "invisible" and therefore auditory interface design appears to be difficult. There is really no reason, apart from accessibility for visually impaired users, to add sound to human-computer interfaces that are already optimised for the visual modality, as it does not really contribute to system usability.

In novel human-computer interaction paradigms [1-3], such as ubiquitous, pervasive, wearable and disappearing computing, interactive sonification might offer possible alternatives to the otherwise dominant visual displays, freeing up our eyes to see the surrounding world or do what small visual displays do not do well. In brief, using sound for representation in interaction design is useful for attracting attention to events or locations, for non-visual communication in general, including speech, alarms, notification and feedback. Sound is less useful for continuous representation of objects (as we have no ear-lids), for absolute readings (as most people perceive auditory dimensions such as pitch, loudness and timbre as being relative), for fine-detailed spatial representation, and also is problematic in noisy or noise sensitive environments.

Designing interactive sonifications for human-computer interaction requires that a number of issues are addressed. We have to consider where and how sonification is appropriate. As designers we must take into account the users' capabilities, while carrying out tasks in real environments, and also consider that surrounding noise levels might mask the system's sounds. If sound is considered to enhance interaction, ways of creating and testing auditory metaphors need to be explored and to what extent the use of sound contributes to the users' performance and subjective quality of use. To be able to design with sound, we need a high-level understanding of what and how we hear (e.g. Gaver [4, 5]). While there are an extensive amount of studies on the perception of musical sounds and speech, relatively little is known about other kinds of non-speech sounds and in particular so called everyday sounds.

In the new interaction paradigms previously mentioned, we can question how we think about interaction and human activity. In previous work on auditory interfaces, ranging from Gaver's Sonic Finder [6] to Brewster's hierarchical earcons (e.g. [7]), human action has to a large extent been thought of in a discrete way, e.g. like kicking a football, where a user action starts a process that then completes without any further user control. This view might be appropriate when pushing buttons or flicking switches. An alternative view is action as a continuous flow, e.g. a pen stroke, where we continuously move a pencil on a surface, relying on our learnt gesture through proprioception, as well as haptic, visual and auditory feedback. This latter view is becoming important as many of the new input devices such as pens, digitisers, cameras, etc., are capable of detecting quite complex human actions.

Still, at the core of our design space, a fundamental problem is how to classify and select sounds to be suitable for a particular interaction design. Depending on our intended users, tasks and context there is initially a broad continuum in this design space, ranging from concrete to abstract representations, i.e. from auditory icons to earcons [8, 9]. If we are designing for casual everyday use we probably need to consider concrete forms while if we are designing for highly specialised domains (such as cockpit or process control applications) where our users will be selected and trained, having high performance requirements, we might need to focus on psychoacoustic issues such as

This work was supported in part by Enterprise Ireland under the Basic Research Programme Multimodal Browsing project and by the European Commission under contract IST-2000-25287 ("SOb - the Sounding Object").

Mikael Fernström is with the Interaction Design Centre, Department of Computer Science & Information Systems, University of Limerick, Limerick, Ireland (e-mail: mikael.fernstrom@ul.ie).

Eoin Brazil is with the Interaction Design Centre, Department of Computer Science & Information Systems, University of Limerick, Limerick, Ireland (e-mail: eoin.brazil@ul.ie).

detection accuracy and time, perceived urgency. In the latter case the design space can be more abstract [10]. In our work on human-computer interaction design with auditory representations in the Sounding Object[1] project, we have concentrated on the former domain of auditory interfaces, for ubiquitous computing and casual use.

II. Exploring What people hear

While it is assumed that everyday sounds have inherent meaning, learnt from our everyday activities, hearing such sounds in isolation without context can be quite confusing. The sound of a single isolated footstep can for example be heard as a book being dropped on a table. Interestingly, this problem is somewhat similar to how linguistic homonyms work, i.e. words of the same spelling or sound as others, but with different meanings [11]. To further develop our understanding of what people think they hear we decided to conduct listening tests, an approach also used by many other researchers e.g. [12, 13]. We made high-quality (44.1 KHz, 16-bit) recordings of 104 everyday sounds (durations between 0.4 and 18.2 seconds) and had 14 postgraduate students listen to the recorded sounds in random order using headphones, responding in free-text format to what each sound was. In most cases the descriptions given were quite rich. Based on Ballas' method for calculating *causal uncertainty* [14-16], the responses were sorted and categorised as well as evaluated if correct or not. From the responses we extracted actions and instruments/agents segments of the texts, i.e. what the participants thought were the objects/materials making the sound and how the objects interacted. In our preliminary analysis we found that 77% of the actions and 71% of the instruments/agents were correctly identified.

We also noted that the resulting data set, with all responses, categorisations and measurements of causal uncertainty and accuracy could be used for suggesting possible use of sounds in interaction design, somewhat similar to Barrass' method of collecting stories about when sound is useful in everyday life [17]. From a designer's point of view it is interesting to note that the data set used in this way contains information about how people describe everyday sounds as well as measurements of how accurate the identification of those sounds are.

Looking at the sequencing between different sounds, we can see that sometimes the order of sounds affects the identification, particularly for sounds with high causal uncertainty. For example, the sound of falling rain is often randomly identified either as rain or frying bacon. If the sound before the rain sound is a birdcall, the response is more often rain. If the sound before the rain sound is clashing of ceramic plates and cutlery, listeners tend to respond that it is the sound of frying and the location is a kitchen. Similar observations are discussed by Ballas and Howard [14] suggesting that there are several parallels between how we process language and everyday non-speech sounds.

III. Sounding objects

With the findings of the listening tests, we can start to suggest possible auditory representations and metaphors for interaction design. Based on Barrass' [17] *TaDa* approach we can do a task and data analysis that allows us to select sounds that can communicate the dimensions and directions that give users adequate feedback about their actions in relation to the system and the status and event of the system. We then need to create ways so that the system can produce the selected sounds and finally evaluate the resulting design with users [18]. If we were to just play sound files for representing user actions, they would always sound the same and never (or seldom) be, for example, expressive, e.g. to be mapped to the user's detected effort or the size of the objects involved. This was one of the issues addressed by the Sounding Object project.

The Sounding Object project explored new methods for physically inspired modelling of sounds for sound synthesis. Our work was initially largely informed by ecological acoustics and for example Gaver's work on auditory icons [4, 5, 19]. We also worked towards *cartoonification* of sound models, i.e. simplifying the models while retaining perceptual invariants. The models were implemented in *pd*[2] and tested in a number of ways ranging from perceptual experiments to artistic performance. Compared to ordinary sound files, sound objects can provide 'live' sound models that can be parametrically controlled in real-time. The difference can also be described as *fully formed* objects versus *evolutionary* objects [20]. For a fully formed object all its parameters are known when activated while for an evolutionary object parameters may vary throughout a continuous use or activation of the object. This implies that sound objects, which can be used both as fully formed and evolutionary objects, are well suited for auditory representations for new interaction paradigms with continuous control by user actions.

With *cartoonification* (or as some prefer, *caricaturisation*) the sound models can be made computationally more efficient than fully realistic models; hence they are more likely to be applicable for 'thin' platforms such as wearable or handheld computers. For example, in two of our prototypes using various click sounds to provide feedback for soft-button actions and friction-like sounds for pen stroke actions, a sound object approach only required half the storage/memory space compared to similar implementations with pre-recorded sound files [21].

With auditory icons there is also the possibility that users sometime could have difficulties determining if sounds are produced by real events or by artificial means, and *cartoonification* might help users to differentiate the artificial sounds from the real, while still communicating relevant system features and feedback through parametric control of perceptual invariants.

Being able to parametrically control sound models in real-time can also, potentially, help to make sonifications less annoying. With prerecorded sound files, sounds used in an auditory interface always sound exactly the same but

[1] www.soundingobject.org

[2] www.pure-data.org

with sound objects and parametric control we can vary properties of the sounds, for example mapping the size of objects or the effort of actions, so that small objects or actions make small sounds and large objects or actions make large sounds.

Revisting the overall results from the Sounding Object project [22, 23] it is interesting to note that all the sound models developed throughout the project point towards an epistemology different to Gaver's trichotomy of primitives of *solids, liquids and gases* [5]. An alternative emerging view indicates that the primitive classes might be better understood if we think of the world of sound producing events as *impacts, frictions and deformations*. The simulated material properties of the objects involved in such interactions are controllable through parameters passed on to our sound models.

The analysis of listening tests as previously described also indicate that actions are better identified than instruments/agents (77% versus 71% in our listening tests). This might suggest that interaction design using auditory representations should focus on mapping human activity to actions rather than objects. However, it is sometimes possible to communicate objecthood through actions.

IV. EXAMPLE: AUDITORY SOFT-BUTTONS

An implementation with auditory soft-buttons was explored in a prototype we developed on a Xybernaut[3] wearble computer. We only used the touch detection on the touch device, not the visual display. See Figure 1, below.

Figure 1: Xybernaut touch screen device

A number of soft-button areas, with different layouts, were defined. When users moved their fingers on a button area a simple friction sound model was used. To emphasise the boundaries of each button, an impact model was used for producing entry and exit click sounds. In an informal test with three users we found that the users were able to feel their way around the touch device and make drawings of the layout of soft-buttons. See Figure 2, below.

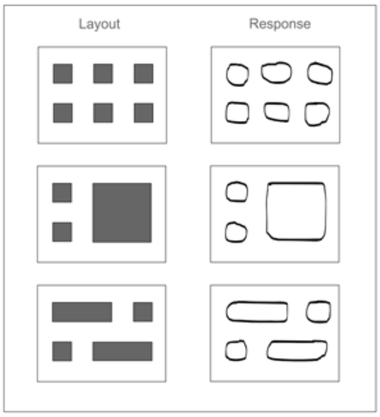

Figure 2: Examples of user responses to various soft-button layouts

This indicates that this kind of auditory representation of soft-buttons allow users to have a pseudo-haptic experience supporting the development a mental model of the layout of the device [21].

A similar pseudo-haptic approach was investigated by Müller-Tomefelde [24] who in one of his demonstrations communicated differences in surface texture through friction-like sounds in a pen-based digitizer application. In the commercial world, Apple Computer's *Ink* application for handwriting input through a digitizer tablet also attempts to enhance the user experience through the use of friction-like sounds as feedback to the user's pen strokes with a stylus.

V. FUTURE RESEARCH

More studies are needed on what people hear and everyday sounds to increase our understanding of the perceptual and cognitive processes involved. In particular, studies of the effects of combinations of different auditory icons in sequence or in parallel seems to be lacking.

We have found the *pd* environment and Sound Objects a highly productive approach for prototyping sound designs for interactive sonification but for fully integrated applications we need to seriously consider if a set of sonification primitives can be integrated with an operating system. This can in turn result in the development and verification of reliable toolkits for systems developers, similar to what is available for graphical user interfaces today. There is also a need to educate and support interaction designers so that they can open up their creative thinking towards interactive sonification, e.g. that it is possible to provide continuous feedback in real-time for gesture-based devices.

All components in human-computer interfaces also have aesthetic properties. It is probably possible to design sonifications that are psychoacoustically correct and quite efficient but very unpleasant to listen to. As suggested by Eric Somers [25] we need to draw upon the knowledge and ideas of *Foley* artists (sound design for film, radio and television) as well as lessons learnt from various theories of acousmatic music.

[3] www.xybernaut.de

VI. SUMMARY

In this paper we have discussed some novel approaches to interactive sonification in human-computer interaction design, in particular for casual use and ubiquitous and wearable applications. We suggest that such designs should be based on the results from listening tests with possible metaphors being extracted from users' descriptions of everyday sounds. The listening tests can also provide guidance in our understanding of how combinations of auditory icons can be interpreted by users. Finally, we describe an example of using Sound Objects for interaction design and suggest areas of future research.

ACKNOWLEDGEMENTS

We would like to express our thanks to all the postgraduate students at the University of Limerick who participated in our experiments. Thanks to Xybernaut GmbH for giving us access to their wearable computers.

REFERENCES

[1] E. Bergman, "Information Appliances and Beyond." San Francisco: Morgan Kaufmann, 2000, pp. 385.
[2] B. Shneiderman, *Leonardo's laptop: human needs and the new computing technologies*. Cambridge, MA: MIT press, 2002.
[3] H. Rheingold, *Smart mobs: the next social revolution*. Cambridge, MA: Perseus books, 2003.
[4] W. W. Gaver, "How do we hear in the world? Explorations of ecological acoustics," *Ecological Psychology*, vol. 5, pp. 285-313, 1993.
[5] W. W. Gaver, "What in the world do we hear? An ecological approach to auditory source perception," *Ecological Psychology*, vol. 5, pp. 1-29, 1993.
[6] W. W. Gaver, "The Sonic Finder: An Interface that Uses Auditory Icons. The Use of Non-Speech Audio at the Interface," presented at ACM CHI '89, Austin, Texas, USA, 1989.
[7] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "A Detailed Investigation into the Effectiveness of Earcons," in *Auditory Display: Sonification, Audification and Auditory interfaces*, G. Kramer, Ed. Reading, MA, USA: Addison-Wesley Publishing Company, 1994, pp. 471-498.
[8] S. A. Brewster, "Non-speech auditory output.," in *The Human Computer Interaction Handbook*, J. Jacko and A. Sears, Eds. USA: Lawrence Erlbaum Associates, 2002, pp. 220-239.
[9] W. W. Gaver, "Auditory interfaces," in *Handbook of Human-Computer Interaction*, M. G. Helander, T. K. Landauer, and P. Prabhu, Eds., 2 ed. Amsterdam, The Netherlands: Elsevier Science, 1997.
[10] R. D. Patterson and A. J. Datta, "Extending the domain of auditory warning sounds: creative use of high frequencies and temporal assymetry," in *Human Factors in Auditory Warnings*, N. A. Stanton and J. Edworthy, Eds. Aldershot, Hants, UK: Ashgate, 1999, pp. 73-90.
[11] J. A. Ballas and R. T. Mullins, "Effects of context on the identification of everyday sounds," *Human performance*, vol. 4, pp. 199-219, 1991.
[12] N. J. Vanderveer, "Ecological acoustics: Human perception of environmental sounds," in *Dissertation Abstracts International. 40/09B, 4543. (University Microfilms No. 8004002).* 1979.
[13] W. W. Gaver, "Everyday listening and auditory icons." San Diego: University of California, 1988.
[14] J. A. Ballas and J. H. Howard, "Interpreting the language of environmental sounds," *Environment and Behaviour*, vol. 1, pp. 91-114, 1987.
[15] J. A. Ballas, "Common Factors in the Identification of an Assortment of Brief Everyday Sounds," *Experimental Psychology: Human Perception and Performance*, vol. 19, pp. 250-267, 1993.
[16] J. A. Ballas, "What is that sound? Some implications for sound design," presented at Design Sonore, Paris, 2002.
[17] S. Barrass, "Auditory Information Design," in *CSIRO*: Australian National University, 1997, pp. 306.
[18] B. Walker and G. Kramer, "Mappings and metaphors in auditory displays:An experimental assessment," presented at Third International Conference on Auditory Display (ICAD-96), Palo Alto,CA, 1996.
[19] W. W. Gaver, "Using and Creating Auditory Icons," in *Auditory Display: Sonification, Audification and Auditory interfaces*, G. Kramer, Ed. Reading, MA, USA: Addison-Wesley Publishing Company, 1994, pp. 417-446.
[20] W. Buxton, W. Gaver, and S. Bly, "Auditory interfaces: the Use of non-speech audio at the interface," in *Unfinished*, 1994.
[21] M. Fernström, "Sound objects and human-computer interaction," in *The Sounding Object*, D. Rocchesso and F. Fontana, Eds. Firenze: Mondo Estremo, 2003, pp. 29-40.
[22] D. Rocchesso, R. Bresin, and M. Fernström, "Sounding Objects," *IEEE Multimedia*, vol. 10, pp. 42-52, 2003.
[23] D. Rocchesso and F. Fontana, "The Sounding Object." Firenze, Italy: Mondo Estremo, 2003.
[24] C. Müller-Tomefelde, "Sounds@Work - Akustishe Repräsentationen für die Mensch-Computer Interaktion in kooperativen und hybriden Arbeitsumgebungen," in *Elektrotechnik und Informationstechnik*. Darmstadt: Technischen Universität Darmstadt, 2003.
[25] E. Somers, "Abstract Sound Objects to Expand the Vocabulary of Sound Design for Visual and Theatrical Media," presented at ICAD2000, Georgia Institute of Technology, Atlanta, Georgia, USA, 2000.

Mikael Fernström is Director of the Master's program in Interactive Media at the University of Limerick, Ireland. His main research interests are interaction design, sound, music, and electronics. Fernström studied electronic engineering and telecommunications at the Kattegatt technical college in Halmstad, Sweden, and has an MSc in human-computer interaction from the University of Limerick.

Eoin Brazil is a research assistant at the Interaction Design Centre at the University of Limerick, Ireland. His main research interests are interaction design, sound, software visualisation and music retrieval. He has a BSc and an MSc in Computer Science from the University of Limerick.

Readers may contact Mikael Fernström at the Interaction Design Centre, Computer Science Dept., Univ. of Limerick, Limerick, Ireland; mikael.fernstrom@ul.ie.