# A WHISPER IN THE WOODS — AN AMBIENT SOUNDSCAPE FOR PERIPHERAL AWARENESS OF REMOTE PROCESSES

*Fredrik Kilander, Peter Lönnqvist*

Department of Computer and Systems Sciences
Stockholm University/KTH
Forum 100, SE-164 40 KISTA, Sweden
{fk,peterl}@dsv.su.se

## ABSTRACT

The concept of a weakly intrusive ambient soundscape (WISP) is presented as a means to provide a peripheral awareness of processes beyond a user's immediate attention. The WISP is a component in a larger environment for ubiquitous computing, centered around a conference room scenario. The experiences from a demonstration prototype indicate that the choice of sounds and the intensity of their presentation can greatly influence the way the WISP is perceived. The work relates in various ways to the sonification of data and audio-based techniques for maintaining peripheral awareness.

## 1. INTRODUCTION

Audio cues in computer interfaces has become quite common in recent years, for the most part due to a general proliferation of sound support in modern personal computers. However, with the exception of entertainment like games, music and video, sound in computer software is almost always used to emphasize events that are fairly obvious anyway: login, logoff, errors, warnings, email notifications and calendar reminders. The combined effect of sight and sound is by design intended to attract the user's attention to the event, and thereby disrupt the current activity. The human user is regarded as a mechanical, task-switching entity that must be interrupted in order to respond to some event.

Human beings are usually very good at building mental models of their environments. In familiar surroundings, like at work or at home, we often have a fairly accurate picture of the activities in our immediate vicinity. We infer the doings of our collegues or family members, even though we can not observe them directly. Sound plays an important role in the creation of such models, because it carries a number of subtle and familiar cues. For example, footsteps, the sound of a copier, the tinkling of glasses, rushing water or a creaking door all aid to distinguish between the anticipated and the unexpected. Less obtrusive and remote sounds also play an important role. We are usually not aware of the hum of the ventilation system or the distant din of conversation, until they suddenly change in intensity or our name is mentioned [1].

In our work on ubiqutious service environments for meetings and collaborative work [2], we have introduced the concept of a *Weakly Intrusive Ambient Soundscape*, or WISP [3]. A WISP caters for a physical space, such as a conference or common room, or a personal work area. In the WISP, states and events in the computational and physical environment are presented as subtle and non-intrusive sound cues. The listening experience is intended to convey intuition rather than interruption; each signal should be sufficiently non-intrusive to be accepted without disturbing the focus of the task at hand, while distinctive enough to be separable from other cues.

From practical experience with our prototypes and the experiences of everyday life, we believe that the success of an artificial audio ambience is coupled to two important properties of human cognition: learning and anticipation. Upon its introduction, the sounds in the WISP are unfamiliar encounters and therefore by their very novelty, craving for attention. The meaning of a particular effect must also be learned, or it will be regarded as meaningless noise. However, just as the steady ticking of a clock quickly slips from our attention, anticipated sounds fade into the background of our consciusness. The effect that we hope to capitalise on, is the subtle difference between the anticipated and the perceived. For example, varying the rate of the ticking of the grandfather clock, or the almost inaudible conversation that turns into an emotional argument.

## 2. THE FUSE ENVIRONMENT

The ambient soundscape created by the WISP is intended to be part of a larger setting, in which computers, software and facilities for graphic and audio presentation interact with the users in a coordinated fashion. We have used the phrase *Future Ubiquitous Service Environment* (FUSE) for this particular kind of environment, and our first prototype was therefore promptly nicknamed fuseONE [2].

### 2.1. fuseONE

In fuseONE, we dressed a small conference room with a few selected appointments, nowadays more often than not found in similar locations. The physical space contained a full bandwidth videoconference system, a SmartBoard [4] combined projection screen and pointing device, a cordless keyboard and mouse, a set of iButton identification devices [5], a VGA projector and a server computer. Hidden loudspeakers were driven by the server computer's audio output ports.

The computing space of fuseONE consisted of the department systems on the local area network, the Internet, standard application software and our own applications and middleware.

A gathering in fuseONE usually followed a similar pattern. The participants gathered with their laptops around the conference table and identified themselves through the iButtons. Running one of the fuseONE software components, the laptops displayed representations of the other participant's, the projection area and other services. Users were then able to send files to each other, to the projector or interact with other components.

Although a WISP was never deployed in the fuseONE environment, it illustrates the kind of environment we experiment in; a relatively small, shared room where users focus on a common task but are still aware of peripheral processes and can receive unobtrusive notifications of external events. This lead us to our next prototype, fuseTWO, which dealt mainly with the issue of notifications.

### 2.2. fuseTWO

The fuseTWO prototype was created to study how notifications can be presented to users while managing the amount of intrusion created by the delivery. The notifications were small text messages, addressed to specific users.

The physical setting was, just as in fuseONE, that of a conference room, but now extended to contain a number of notification terminals. The notification terminals were of two kinds, public and private. Four Compaq iPaq palmtop computers implemented the private terminals with a side-scrolling banner. The public terminals varied in size and appearance and consisted of a simulation of a wall-mounted analog clock that superimposed text on the face of the clock, a side-scrolling banner on a wall projection screen and a WISP installation.

All terminals (with the exception of the projection screen banner) employed sound. The private terminals could emit a sharp, annoying ring similar to that of a telephone and the wall-mounted clock could chime. The WISP could not reveal the text of a notification, but had much more dynamic range and choice of sound character than the other terminals.

Notifications were sent to the terminals by a notification router. The router software examined each notification and based its routing decision on the recipient, the privacy of the notification and the availability of private or public notification terminals.

Another important component in the fuseTWO prototype was a rotary knob mounted in the middle of the conference table. The knob would allow the users to continuously set the level of intrusiveness accepted by the group. The level of the knob was monitored by the router, and in the minimum end position, no notifications were allowed to reach any terminal. As the knob gradually was turned towards the maximum position, more terminals were employed with an increasing amount of intensity. The wall-mounted clock began to chime and the palmtop computers would ring. Technically, the router informed the terminals of the current knob setting and the terminals adapted their behaviour according to built-in thresholds.

### 2.3. Wisp in use

The WISP terminal service that was used in the fuseTWO prototype, reacted to a notification by playing a sound associated with the recipient of the notification. The WISP also responded to changes in intensity by altering the playback volume and by switching between three levels of reverb. The highest intensity level featured a dry sound, without any artificial reflections. The medium level had a moderate amount of reverb added to it, and the low intensity level was almost completely saturated by reverb, giving the effect of a far sound being heard in an empty garage.

The fuseTWO prototype was on display in a booth for two days at an international science project review, during which time a number of interesting experiences were made. We will here remark only on those associated with the WISP terminal.

During the exhibition the WISP used four distinct sounds. This was because the setup involved exactly four pre-arranged users. Three of the sounds were of a fairly short duration while the fourth was longer. The short sounds consisted of a female giggle ($1.3$ seconds), a golfboll dropping in a tin cup ($1.2$ seconds) and a car engine starting ($3.3$ seconds). The fourth sound was a fairly long thunderstorm sequence ($15.3$ seconds) and did not receive any reverb treatment.

The sounds were choosen through a combination of factors. We wanted to offer a variety in their nature, to make each sound clearly distinguishable from the others. We also wanted to move away from singular events towards a true ambience, which is why we added reverb to simulate a surrounding space and included the long thunderstorm.

The WISP had, through the versatility of the support for sound in Java (SDK 1.3), the ability to play several sounds simultaneously. In order to prevent sounds from being triggered too close to each other, we introduced a form of traffic shaping in the WISP, to the effect that there was at least a $1$ second delay between successive playbacks. This allowed the short sounds a space in which they could be recognized, while still play back overlapping each other.

In use, one of the most interesting observations we made, was that the sounds very quickly became familiar and blended into the general atmosphere of the exhibition hall (which was quite noisy without our contributions). The high rate of random messages generated solely for demonstration purposes may well have contributed to this effect, but in spite of this the sounds easily became annoying and impossible to ignore when the volume was turned up too high.

At medium intensity levels, the reverb added to the sounds helped to create an artifical ambience, although one observer felt that the reverb made the sounds feel colder and less pleasant. At low intensity levels, the amount of reverb and low playback volume made the sounds almost inaudible, indistinguishable and different to separate from the general noise in the room.

The strongest reactions, received from casual visitors to the prototype, was to the giggle and the thunderstorm. Several people said they liked the giggle. In a least one case, the giggle effect caused quite a substantial emotional response in that the listener thought to recognize the recording as herself (which was not the case). As for the thunderstorm, several people indicated that it gave a cozy atmosphere.

The other two sounds, the golfball and car engine were met mostly with indifference. The sound of the ball falling into the cup was difficult to recognise (except possibly for avid golfers). The car engine was easily identified, of good quality, but seemed to lack character.

One person with a hearing disability, found that the WISP sounds interfered with his hearing even at low volume, because it raised the general noise level around him which was in turn emphasized by his hearing aid equipment.

## 3. A WISP IN A SHARED SPACE

The deployment of a WISP in a shared space, like a meeting room or kitchen, is escorted with issues of how to join the preferences of several people in what essentially is a broadcast medium. For example, there must be some subscription and routing mechanism which allows the WISP to receive events for the users present in the space it caters for. In our prototypes we used iButtons as a localisation device.

Having detected the user's presence, the WISP must then retrieve the user's preferences, such as event mappings and sound parameters. These mappings may conflict with those of other users present, for instance if two people has associated different events to one and the same sound cue. There appears to be two basic solutions for this. In a centralized model, each user must select sound cues from a single resource. The WISP does not allow customized sounds to be introduced by users, and has therefore the ability to recognise a clash between user's choices. In a distributed model, each user is encouraged to personalise the aural cues, to such an extent that it becomes unlikely that two users share identical cues.

The centralized model is easy to implement, but will most certainly leave some people dissatisfied. The distributed model, while offering the highest degree of freedom, suffers from the possibility that people avoid the customisation step because it is perceived as tedious. It also has no real protection from reintroducing conflicts should people exchange popular sounds, and it cannot prevent obnoxious auralisations or willful abuse.

## 4. RELATED WORK

The arena of research in which the WISP is situated is intersected by at least two central topics. These can be characterised as the sonification of data for presentation purposes on the one hand, and support for peripheral awareness on the other.

Sonification, or audification as it is also sometimes called, deals with the issue of converting non-audio data into sound. The purpose is to create a sonic environment that reflects the data in such a way, that the listener can identify or target specific events or locations in the data where there are interesting features. The output of the sonification can be intended for the immediate attention of the listener or as an ambient background, depending on the application.

Below follows some examples of work where the audio was intended for the listener's immediate attention. We begin with Childs [6] who synthesised music to aid the analysis of computational fluid dynamics. The output was used to locate convergence behaviour and low activity areas in the massive amounts of data generated by the simulations.

Chafe and Leistikow [7] transformed packet round trip times on the Internet into audible sounds, to experience network performance in real time and to analyse trends in vast collections of log data.

Internet statistics were also gathered and sonified by Hansen and Rubin [8], who studied web and chat sites, and converted the activity there into a number of drones, string sounds and tones.

Bederson and Druin [9] constructed an automated tour guide which supported ad-hoc paths and pauses in a museum tour, with personal, auditory addresses at selected exhibits. Like many other techniques, the tour guide depended on wearable equipment for the reproduction of sound, whereas WISP relies on loudspeakers concealed in the physical environment.

Hudson and Smith [10] demonstrated an interesting encoding technique in a system for electronic mail preview using non-speech audio. In their system, incoming mail was analysed for size, recipients and content, and then encoded into a sentence of audio symbols, carefully designed to be non-intrusive yet distinctive. The effort of loading semantics into a language of noises is very interesting, because human are often adept at learning such languages, just as we learn Morse code or the semantics of movie music. With learning and adaptation the need to attend a sound is reduced and this seems to argue for the possibility of loading WISP with a favourable set of semantics.

Mynatt et al [11, 12] developed Audio Aura, a system with wearable playback and personal positioning facilities. The system was designed for task and calendar reminders, email status and information of the activities of colleagues. In particular, a user departing one location could leave a lingering, audible aura which the system would pick up and play back to other users if they approached the spot in time. Depending heavily on localisation, Audio Aura could provide audio which was relevant not only in the general context of the recipient, but also semantically connected to the physical environment.

Sawhney and Schmandt [13, 14] have worked on Nomadic Radio, another system in which the user wears the sound generation equipment. Just as Walker and Brewster [15] they employ techniques for three-dimensional sounds, by which another dimension of semantics can be introduced by positioning the perceived sound in relation to the listener's head. Nomadic Radio is very sophisticated and hosts a range of features; seven levels from silence, through ambient background information to insistent signalling. In addition, tactile and spoken commands may be given to the system via the wearable. Nomadic Radio also monitors the user's activities and speech in order to minimise intrusiveness. In relation to a WISP, Nomadic Radio is designed to follow and interact directly with its user at almost all times. It is much more like a personal assistant shaped like a collar, while WISP attempts to be a more ghost-like entity. A particularly interesting result from Nomadic Radio is that user's preferred to have ambient background information continuously audible, as it reassured them that the system was operational.

The topic of peripheral awareness has seen substantial interest, particularly in the field of computer-supported cooperative work, where the participants in a working group may be physically distant yet closely connected by communication technology. In these settings, it becomes important to maintain a sense of awareness of the activities of the other people in the group are doing, without having to monitor their activities closely. This can be done, for example, by the sonification of events generated by people or the transformation and re-rendering of data from video and audio pickups. For WISP, we are of a mind that if such mechanisms are useful, they should be able to convey an intuitive sense of any graspable process.

Pedersen et al has studied peripheral awareness in the AROMA system [16, 17]. They looked at the use of abstract representations as presence indicators, in order to find a middle ground in which the recipient was not unduly disturbed and the monitored person's integrity was intact. They measured the activity and people present in a room, and displayed this information using audio, visual and haptic (touch) presentations. One particularly interesting result was that the abstractions caused problems of interpretation for the recipients. With too many semantic mappings the abstract representation was unnoticed or rejected as too hard to learn. Another important find was that users tended to overinterpret the symbolism in the display when it involved familiar elements.

An interesting audio treatment was performed by Smith and Hudson using the concept of a low-disturbance, non-speech sound [18]. Essentially, the sender's voice is sampled, encoded and resynthesised at the receiver. The major point, however, is that the decoded sound at the receiving end is unintelligeble, thus preserving the privacy of the sender, while still being recognisable as a speech activity.

Barra et al [19] applied sonification to monitor a HTTP server. To counter the repetetive and potentially tedious task of listening to machine-generated sounds, they allowed the sounds generated by events in the HTTP server to be mixed into an audio stream of the operator's choice.

Gaver, Smith and O'Shea [20] found in the ARKola bottling plant simulation that auditory icons helped collaboration and diagnosis among workers in a shared and complex task. By virtue of its many different sounds, the simulation automatically forced the user to place certain sounds in focus. One result of this was that the simple cessation of a sound did not always convey the information that the corresponding process had stopped.

Ishii et al reports on ambientROOM [21, 22, 23], a specially designed room in which lighting, air, video and sounds together form an ambience laden with content. The use of natural, ambient sounds carrying information makes ambientROOM the earlier work most closely related to WISP. However, whereas ambientROOM explores many dimensions as conduits for awareness, WISP only concerns itself with ambient sound.

## 5. CONCLUSIONS AND FURTHER WORK

We are in no doubt that the creation of an artifical audio ambience is meaningful in some circumstances. Our own experiences from our prototypes and the results or others in the areas of sonification and peripheral awareness inspire us in this regard. Success, however, appears to be dependent upon the careful deliberation on several factors.

For example, which processes and events should be allowed to generate sound? Information that is related to the activities of people has a life-like quality and meaning, but are obviously sensitive with respect to privacy. The monitoring of mechanical activities such as network or server performance easily runs the risk of being monotonous. In a shared environment, one recipient may listen with interest while others find themselves exposed to an incomprehensible noise.

What should be the nature or the generated sounds, in order to achieve peripheral awareness and minimise intrusion? From our fuseTWO prototype, we are inclined to believe that easily recognisable and natural sounds stands the greatest chance of being accepted as a part of the environment. In particular, a continuous background murmur is probably more easily ignored than a singular sound, and it also continuosly reassures the listener that it is operative.

Voices, and voice-like sounds earns a special recognition in our mind because we are very accustomed to hearing and interpreting voices. For the same reason, voice-based sounds can be experienced on many levels; emotional, cultural and social. A whisper in the woods can be just as provocative as a symphony orchestra crescendo.

Future work will see the development of our third prototype. Just as in the earlier prototypes, the setting involves a conference room with a host of services and interaction resources. A WISP will be one of them and we hope to further explore the relationship between noise and a meaningful ambience in this setting.

## 6. REFERENCES

[1] John R. Anderson, *Cognitive Psychology and its Implications*, W. H. Freeman and Company, fourth edition, 1995.

[2] Patrik Werle, Fredrik Kilander, Martin Jonsson, Peter Lönnqvist, and Carl Gustaf Jansson, "A ubiquitous service environment with active documents for teamwork support," in *Ubicomp 2001: Ubiquitous computing: international conference; proceedings*, D. Abowd, Gregory, Ed. 2001, vol. 2201 of *Lecture notes in computer science*, pp. 139–155, Springer-Verlag.

[3] Fredrik Kilander and Peter Lönnqvist, "A weakly intrusive ambient soundscape for intuitive state perception," in *Continuity in Future Computing Systems*, J. Doherty, G., M. Massink, and D. Wilson, M., Eds., Library and Information Services, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, 2001, pp. 70–74, The Central Laboratory of the Research Councils.

[4] "Smartboard," SmartBoard is a product of SMART Technologies Inc.
(http://www.smarttech.com).

[5] "ibutton," iButton is a product of Dallas Semiconductor Corp.
(http://www.ibutton.com).

[6] Edward Childs, "The sonification of numerical fluid flow simulations," in *Proceedings of the 7th International Conference on Auditory Display*, Helsinki University of Technology, Espoo, Finland, 2001, Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, http://www.acoustics.hut.fi/icad2001/proceedings/index.htm.

[7] Chris Chafe and Randal Leistikow, "Levels of temporal resolution in sonification of network performance," in *Proceedings of the 7th International Conference on Auditory Display*, Helsinki University of Technology, Espoo, Finland, 2001, Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, http://www.acoustics.hut.fi/icad2001/proceedings/index.htm.

[8] H. Hansen, M. and B. Rubin, "Babble online: applying statistics and design to sonify the internet," in *Proceedings of the 7th International Conference on Auditory Display*, Helsinki University of Technology, Espoo, Finland, 2001, Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, http://www.acoustics.hut.fi/icad2001/proceedings/index.htm.

[9] B. Bederson, B. and A. Druin, "Audio augmented reality: A prototype automated tour guide," in *The ACM Human Computer in Computing Systems conference (CHI'95) conference companion*, 1995, pp. 210–211.

[10] E. Hudson, Scott and Ian Smith, "Electronic mail previews using non-speech audio," in *CHI'96 conference companion*. 1996, pp. 237–238, ACM.

[11] Elizabeth Mynatt, D., Maribeth Back, Roy Want, and Ron Frederick, "Audio aura: Light-weight audio augmented reality," in *Proceedings of the 10th annual ACM symposium on User interface software and technology*, New York, 1997, pp. 211–212, ACM Press.

[12] Elizabeth Mynatt, D., Maribeth Back, Roy Want, Michael Baer, and B. Ellis, Jason, "Designing audio aura," in *Conference proceedings on Human factors in computing systems*. 1998, pp. 556–573, ACM Press/Addison-Wesley Publishing Co.

[13] Nitin Sawhney and Chris Schmandt, "Nomadic radio: Scaleable and contextual notification for wearable audio messaging," in *Proceedings of the CHI 99 Conference on Human factors in computing systems: the CHI is the limit*, New York, NY, USA, 1999, pp. 96–103, ACM Press.

[14] Nitin Sawhney and Chris Schmandt, "Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments," in *ACM Transactions on Computer-Human Interaction (TOCHI)*, New York, NY, USA, 2000, pp. 353–383, ACM Press.

[15] Ahsley Walker and Stephen Brewster, "Spatial audio in small screen device displays," Tech. Rep., Department of Computing Science, University of Glasgow, Scotland G12 8QQ, 2000.

[16] Rønby Pedersen, Elin and Tomas Sokoler, "Aroma: abstract representation of presence supporting mutual awareness," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '97)*. March 1997, pp. 51–58, ACM Press.

[17] Rønby Pedersen, Elin, "People presence or room activity supporting peripheral awareness over distance," in *Conference proceedings on Human factors in computing systems*. 1998, pp. 283–284, ACM Press/Addison-Wesley Publishing Co.

[18] Ian Smith and E. Hudson, Scott, "Low disturbance audio for awareness and privacy in media space applications," in *Proceedings of the third ACM international conference on multimedia*. 1995, ACM.

[19] Maria Barra, Tania Cillo, Antonio De Santis, Teenie Matlock, Ferraro Petrillo, Umberto, Alberto Negro, Vittorio Scarano, and Paul Maglio, "Personal webmelody: customized sonification of web services," in *Proceedings of the 7th International Conference on Auditory Display*, Helsinki University of Technology, Espoo, Finland, 2001, Laboratory of Acoustics and Audio Signal Processing and the Telecommunications Software and Multimedia Laboratory, http://www.acoustics.hut.fi/icad2001/proceedings/index.htm.

[20] W. Gaver, W., B. Smith, R., and T. O'Shea, "Effective sounds in complex systems: The ARKola simulation," in *Proceedings of CHI'91*, April 28–May 2 1991.

[21] H. Ishii and B. Ullmer, "Tangible bits: Towards seamless interfaces between people, bits and atoms," in *Proceedings of Conference on Human Factors in Computing Systems (CHI '97)*. March 1997, pp. 234–241, ACM Press.

[22] H. Ishii, C. Wisnecki, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin, "ambientroom: Integrating ambient media with architectural space," in *Proceedings of CHI '98*, New York, 1998, pp. 173–174, ACM.

[23] Craig Wisneski, Hiroshi Ishii, Andrew Dahley, Matt Gorbet, Scott Brave, and Paul Yarin, "Ambient displays: Turning architectural space into an interface between people and digital information," in *Proceedings of the First International Workshop on Cooperative Buildings (CoBuild'98)*. Febrary 1998, Springer.