Sound Intensity Gradients in an Ambient Intelligence Audio Display

Milena Droumeva
School of Interactive Arts and
Technologies,
Simon Fraser University,
Surrey, B.C. Canada
mvdroume@sfu.ca

Ron Wakkary
Associate Professor
School of Interactive Arts and
Technologies,
Simon Fraser University,
Surrey, B.C. Canada
rwakkary@sfu.ca

Abstract
This paper describes the prototype of a real-time responsive audio display for an ambient intelligent game named socio-ec(h)o. The audio display relies on a gradient response to represent and anticipate player action. We describe the audio display schema, and discuss results of our current experimentation in guiding player actions through types of audio feedback, for creating sound recognition, perceptions of change and sound intensity.

Keywords
Ambient intelligence, auditory display, game, play, responsive environments, sound design

ACM Classification Keywords
J.5 [Arts and Humanities] H.5.2 [information Systems]: interfaces and presentations H.5.5 [Sound and Music Computing]: Systems

Introduction
Our current research involves the design of an ambient intelligence (AmI) environment for a multi-user physical game known as socio-ec(h)o. Our overall goal is to explore approaches to group communication and skill acquisition in AmI environments. AmI requires meaningful and clear yet ambient feedback to users.

Figure 1. Participants playing socio-ec(h)o

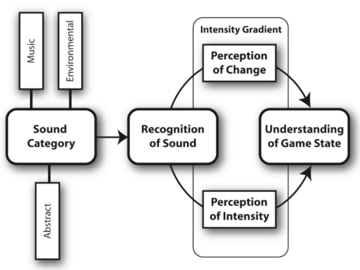

Figure 2. A model of our sound and perception approach

We believe that gradient responses can provide this type of feedback. We have developed an audio display schema, and experimented with gradient response and intensities in the audio display in order to confirm *and* anticipate users' actions.

Background

Sound is an important channel through which humans perceive their natural or designed environments. There is growing literature on auditory perception and design for task-oriented, highly computerized environments [1, 2, 10], virtual audio and spatialization [3] and sonification of information clusters [1, 6, 13], yet relatively few of these approaches to sound design have found their way in the design of sound for games [5, 9]. As a game platform, socio-ec(h)o draws on a variety of existing sound design patterns [1], as well as classical psychoacoustics, investigations into perception of environmental sound, and spatialization [8]. Our sonic design benefited also from the acoustic communication approach developed by Schafer and Truax [11], where sound is seen as a ubiquitous communication channel between listener, soundscape and a physical (and cultural) environment.

Design approach

Ambient intelligence is the embedding of computer technologies and sensors in physical environments that combined with artificial intelligence, respond to and reason about human actions and behaviours within the environment. The socio-ec(h)o game is played by a team of four players and features six levels of increasingly difficult word puzzles solved by coordinated body positions and movements. The environment is responsive to players' actions through abstract light and sound (figure 1). Players' movements are tracked using a Vicon motion capture system (www.vicon.com). A custom reasoning engine was developed to track the game state and infer players' actions. For more technical details of our prototype see [12].

Our goal was to create an ambient feel and function that is in keeping with AmI and supports group interaction. We felt that the interface required a gradient rather than direct response to players in order to best represent game states and anticipate players' actions. This approach poses a challenge for the audio displays. Together with the visual display, the audio display must create the aesthetic feel, represent current game state, and guide future player actions through abstract and ambient real-time sonification. After a number of design workshops, we arrived at an approach for the audio display that provides a continuous, ambient response along a gradient.

In its simplest form, the approach can be described as dynamic soundscapes that are not only recognized by players, but in turn interpreted to determine the state of the game (figure 2). The soundscapes are made of sounds that can be categorized as *musical*, *abstract* and *environmental*. One question is what combination of sound categories would be recognized and perceived as meaningful in the context of the game?

The meaningful sounds aid an understanding of the game state and the affect of players' actions. We felt we could modulate this understanding by increasing or decreasing the rate of intensity of or changes in the sound along a gradient. An example of this is the game of "hot and cold", where players use words along the continuum of hot to cold to signify the proximity of another player to a solution. The question is what sound techniques would work best to signal increased intensity and change along a gradient?

Audio display schema

The audio display consists of three components: a *real-time ambient sonification* that has a different soundscape for each level of the game; an *anticipatory feedback* sound to signal when all participants are working together towards the goal; and a *confirmatory feedback* sound, which signals the completion of the goal and the progress to the next level. The latter two feedback signals are consistent through all levels, so as to create a coherent expectation of success. These responses work in conjunction with each other. The gradient intensity is applied to the real-time ambient sonification component and builds up to the two types of feedback.

For example, players in socio-ec(h)o must achieve a specific goal at each level, which is a coordinated configuration of movements and body positions. For example, an answer to the word puzzle "lo and behold"

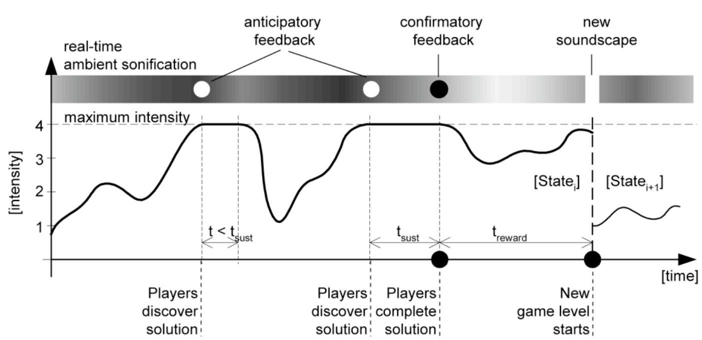

Figure 3. Mapping of the intensity gradient and audio display to sample game play

was for all players to crouch low and be still. Further, they must hold this configuration for a short period of time, typically 4-6 seconds. This ensures that the actions were purposeful and knowingly completed rather than an accidental formation. In figure 3 we illustrate players' actions and the mapping of intensity levels and the audio display components. Reading from left to right, the players explore different combinations of movements and actions. As some players discover a combination, the intensity of the real-time sonification increases, in this case to a value of 2 out of a maximum of 4. As more players discover the solution, the intensity level briefly drops but then increases to the maximum level of 4. At this point, an *anticipatory feedback* sound occurs signaling to the players that they've done something right and if they keep at it they will complete the solution. However, we see that the players are unable to hold the combination and the intensity level rapidly decays to 1 (in our observations players often stop after an increase in intensity or anticipatory feedback in order consider what they just did). The players quickly resume the configuration and the *anticipatory feedback* sound occurs again. This time the players complete the solution by maintaining the configuration for the required duration, and the *confirmatory feedback* sound occurs. The real-time sonification fades away then back as it transitions to the next level. Here, a new soundscape is composed and the intensity level is reset to a minimum.

Current progress

To date we have developed the socio-ec(h)o prototype and have implemented preliminary user testing. This includes two three-hour sessions with eight participants, and an additional two-hour session with four other participants. The group included three females and nine males ranging in ages from twenty-one to fifty-nine. Each session began with a warm-up to introduce the concept of puzzles solved through physical action and support through implicit responses. Each team of four played two levels followed by discussions. After all levels were completed or a total of two hours of interaction (60 minutes in the short session), the game was stopped and an open-ended interview was conducted.

Results

We've organized our results by first discussing what combination of sound categories and audio processing techniques created recognition, followed by a discussion of techniques that best achieved an intensity of sound, gradient of change, and feedback. While visual display was part of the interface we gave greater attention to the role of audio.

Sound Category

The first two game levels featured an abstract musical soundscape. Level 3 had an abstract non-musical sound (crackling of rocks), and the last three levels had environmental soundscapes. Level 4 and 5 were represented by fire and water respectively, while level 6 featured a more complex soundscape of a forest that gradually turned from a calm night to a full-fledged storm. The socio-ec(h)o's soundscapes evolved from one-dimensional, internally coherent sounds, to complex ecologies. In the course of our testing, it became clear that the more abstract sounds (such as a crackling of rocks, or highly processed musical abstractions) were hard for participants to recognize. As a result, these sounds posed difficulties in determining the level of change in the gradient response. In the discussion after levels 2 and 3, players commented that they could tell the sound display was intensifying, but found it hard to tell "by how much." In contrast, after levels 4 and 5 (featuring water and fire sounds) players reported a noticeable positive difference in the "sharpness" of the ambient response. The environmental sounds were more recognizable, and players had an easier time interpreting changes in them. Players also found environmental sound more contextualized and richer in narrative. As one participant commented: "the sound of fire didn't just get louder, it was different, more intense, and I automatically associate it with a positive thing, with success – building a big fire!" It seems, if players have prior experience with a sound, they have a sense of its inherent scale of intensity, i.e. they know what full-blown fire sounds like, as opposed to a faint crackle. This accounts for the difficulty with unfamiliar abstract sound. The players lacked a sense of scale and therefore found it hard to interpret the gradient of change.

Perception of intensity

socio-ec(h)o uses sound that is coherent in its basic characteristics of pitch, rhythm and timbre within the same game level, and represents success rate in the game by *intensifying* these basic sound characteristics. In Table 1 we detail the range of audio design approaches we experimented with to achieve gradient sound display in each game level.

The *Layer fader* is a simple cross-fader between 5 layers of sound which could be specifically chosen or pre-composed to represent increased intensity of the same group of sounds (i.e. dripping water gradually changing into a fast river stream, over 5 steps). The *Low-pass filter* affects the perception of intensity through adding and subtracting frequencies from the sound's spectrum. The result is a sense of brightness or muffled sound that is quite recognizable. The *Phaser* approach relies on running a select sound source through a simple phaser, which varies the sound's *phase* – i.e. its perceived tempo heard as a pulsation. This approach was made even more complex with culturally associative environmental sounds such as water. Intensity here was related inversely to tempo – shifting from a pulsating beat to full, continuous sound at the completion of the goal state.

As research in classical and contemporary psychoacoustics suggest [8], amplitude change, followed by pitch change and tempo change are the most readily and easily perceived sound variance characteristics. Thus game levels are based on these approaches (and combinations thereof) in order of their "ease of perception". For example, the bi-quad filter's effect was effective as an approach to gradual change

Table 1. Audio display methods used in levels

Level	Method	Description of Effect	Modality
1	Amplitude	Sound levels slowly come up	Volume
2	Phaser Layer fader	Tempo goes up, crossfade of 5 sounds	Tempo timbre
3	Layer fader Pitch shift	Crossfade of 5 sounds	Timbre Pitch
4	Layer fader Low-pass filter	Sound crossfades and muffled-to-bright	Timbre Association
5	Phaser Pitch shift	Sound goes up in tempo and pitch	Tempo Pitch
6	Layer fader	Crossfade of 5 sounds	Timbre Association

our design workshops, yet didn't work as well in preliminary testing. We believe the reason for this is the abstract nature of the sound category used with this effect. We still believe low-pass filtering could be effective in representing a gradient, only with less abstract sounds. The phaser effect (slowing down or speeding up of a pulsating beat) produced a perception of change, yet participants reported it hard to the gauge the gradient scale. When prompted during discussion, participants suggested pitch shifting from low to high as a clear way to represent rising intensity. This reinforced one of the approach we had already incorporated in the system.

Perception of change

8-channel spatialization was used to create a perception of change in the soundscape. This was done by gradually moving the sound in space, the audio display cross-faded sounds of different intensity in and out, giving the players the impression that sound is "going away" or "getting closer" to their relative position in space. Participants commented positively on the sound's spatial movements, saying it made the feedback appear "more crisp" and it helped them interpret change easier. Participants responded the most positively to a combination of environmental sound (whether intensifying fire or going deeper into the forest) and a multi-channel diffusion. All participants commented positively on the immersion quality of the play space.

Anticipatory and Confirmatory Feedback

An issue that became clear from participants' comments was the importance of the anticipatory and confirmatory feedback sounds. As players put it, they wanted to know when they were "on the right track," but they also wanted an indication of when they "got it," in order to "celebrate after all the hard work." Recognition of the sound, as well as overcoming masking, i.e. sound has to be heard clearly over the ambience were key issues. The sounds were random pitch variation of two soft abstract sounds (granulated tapping of glass). We played the anticipatory and confirmatory feedback sounds to all participants before the game began. We found that prior listening was sufficient in creating recognition of sounds when heard again later in the game.

Conclusion and future work

In this paper we have outlined a model of sound design that uses an intensity gradient approach to signifying system and game information to players. We have presented a design approach to interactive sound that we feel is a good fit with AmI systems.

AmI environments rely heavily on a meaningful and clear, yet ambient response. In socio-ec(h)o we introduce a new concept of providing intensity-based audio display with a gradient approach. In this respect we were encouraged by the results of our preliminary tests. Even when the display was less clear, participants were still able to use the audio as a reliable feedback system throughout the game. In addition, specific approaches, such as using environmental sound and certain audio processing techniques proved very effective in contextualizing the game and providing game state information to players. Future work will include further study of the relationship between the audio processing of different sound categories, and the auditory perception of recognition, intensity and change.

References

[1] Adcock, M., Barrass, S. Cultivating Design Patterns for Auditory Displays. *In Proc. ICAD 04* (2004)

[2] Back, M. Sound Design for Brain Opera's Mind Forest: Audio for a Complex Interactive System. *In Proc. DIS* 97. ACM Press (1997). P.23-25.

[3] Brungart, D. S. Near-Field Virtual Audio Displays. *Presence, 11*(1), (2002). 93-106.

[4] Butz, A., Jung, R. Seamless User Notification in Ambient Soundscapes. *In Proc. IUI 05*, ACM Press (2005) p.320–322.

[5] Drewes, T. M., Maynatt, E.D., Gandy, M. Sleuth: An Audio Experience. *In Proc. ICAD 2000* (2000).

[6] Hunt, Andy, & Hermann, Thomas. (2004). The Importance of Interaction in Sonification*. In Proc. ICAD 04*, Sydney, Australia (2004).

[7] Lorstad, H., d'Inverno, M., Eacott, J. The Intelligent Street: responsive sound environments for social interaction. *In Proc. ACM 2004*, ACM Press (2004) p.155-162.

[8] Neuhoff, J. Ed. *Ecological Psychoacoustics*. Boston: Elsevier Academic Press. (2004).

[9] Rober, N., Masuch, M. Leaving the Screen: New Perspectives in Audio-only Gaming. *In Proc. ICAD 2005* (2005). P.92-98.

[10] Shinn-Cunningham, B., Ihlehem, A. Selective And Divided Attention: Extracting Information From Simultaneous Sound Sources. *In Proc. ICAD 04* (2004).

[11] Truax, B. *Acoustic Communication*. Ablex Publishing. 2nd Ed. (2001).

[12] Wakkary, R., Hatala, M., Lovell, R., Droumeva, M. Ambient Intelligence Platform for Physical Play. In Proc. ACM Multimedia, ACM Press (2005).

[13] Wenzel, E. M. Spatial sound and sonification. In Kramer G. (ed), Auditory Display, SFI Studies in the Sciences of Complexity, Proc. Vol. XVIII, Addison-Wesley, (1994).