SOCIO-EC(H)O: FOCUS, LISTENING AND COLLABORATION IN THE EXPERIENCE OF AMBIENT INTELLIGENT ENVIRONMENTS

Milena Droumeva

Faculty of Education,
Simon Fraser University
mvdroume@sfu.ca

Ron Wakkary

Associate Professor,
School for Interactive Arts and Technologies
Simon Fraser University
rwakkary@sfu.ca

ABSTRACT

In this paper, we aim to conceptualize the connection between embodied interactions and the experience of understanding a dynamic auditory display response. We have termed this concept *aural fluency* and hereby continue previous work documenting in more detail the listening patterns that emerge in users' experiences with ambient intelligent environments. Aural fluency describes the acquired listening competency and focus on sonic feedback that users form over time in systems utilizing responsive ambient audio display and collaborative embodied interaction. We describe *listening positions* that characterize the concept and show the stages of aural fluency. The concept arose from the design, analysis and evaluation of an embedded interaction system named socio-ec(h)o – a project upon which we also build on from previous work in the hopes of elucidating further the complex experiences of listening attentions and thus offer insights to the field of auditory displays.

1. INTRODUCTION

The need for new concepts of how users understand and make use of their knowledge of system displays arise as tangible computing, embodied interaction and ambient intelligence systems become increasingly possible. Our understanding of the design of interaction has advanced considerably within a traditional human- computer desktop view that emphasizes visual perception and mental cognition, however there is the additional need to explore concepts related to embodied interaction, sensory perception and ambient audio- visual displays with an emphasis on social interaction or at the least multi-user settings. We see real value in adding to the emerging literature of case descriptions of collaborative and embodied interaction systems but even more critically the need to explore new concepts that emerge from in-depth empirical studies of systems.

In this paper, we aim to conceptualize and describe the connection between embodied interactions and the experience of understanding a dynamic auditory display response. We have termed this concept aural fluency, and we build on past work to not only describe better the stages and levels of aural fluency, but to also offer ways of codifying and examining them in context and over time. We believe there is a need for ideas of interaction and meaning-creation with respect to auditory perception to evolve and reflect the different reality that embodied and tangible interactive systems offer in shared, collaborative, temporally and physically persistent contexts. Aural fluency describes the acquired listening competency that users form over time in systems utilizing responsive ambient audio display and collaborative embodied interaction. The concept arose from the analysis and evaluation of an embedded interaction system named socio-ec(h)o. While we have already introduced it to the auditory display community [1], we now hope to build upon that with a discussion on listening patterns as they relate to embodied interaction and problem-solving in ambient intelligent environments.

First we'll explore the concept of aural fluency by theoretically explicating notions of listening, acoustic interpretation, and auditory training, and their relationship to embodied action and interactivity, situated cognition and perceptual dimensions of interpretation and context. We reference auditory perception and auditory display design literature, as well as the acoustic ecology and acoustic communication frameworks put forth by Schafer and Truax [2,3]. Finally we touch on a discussion of multiliteracies in educational discourse and Gilford's dimensions of fluency [4]. Using extensive qualitative analysis from speech transcripts, post-evaluation survey and video coding, we then begin to paint a picture of aural fluency as a concept that is central to the relationship between ambient system and human user through documenting listening patterns in ambient intelligent environments. We further situate these emerging concepts into both paradigms – that of auditory training and that of acoustic communication, in order to enrich our understanding of both in light of embodied interaction. We conclude with a conceptual synthesis of its role in the technological ecology of ambient, embodied interactive environments, as a vehicle for interpretation, meaning-creation and communication with the system.

2. FOCUS, LISTENING AND AURAL FLUENCY

Rather than delving into the philosophical and epistemological foundations of 'fluency' we focus on the connection between aural competency and embodied interaction in responsive environments particularly in its relationship to a distributed cognition model of problem solving. There are three paradigms that we propose will help flesh out and develop the concept of aural fluency: one comes from the auditory display design field and refers to facilitating the user's ability to derive meaningful information from sound; the other refers to patterns of [everyday] listening from the acoustic communication tradition;

and finally a perspective from education and cognitive science – four dimensions of 'fluency' as intellectual abilities.

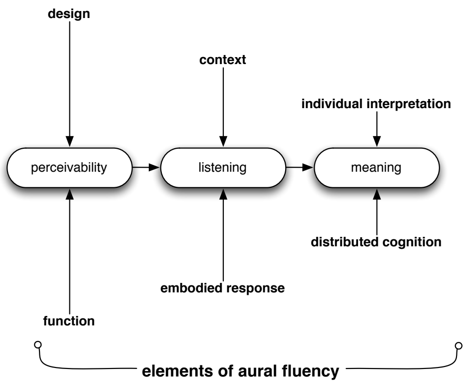

Figure 1. Elements playing a part in users establishing and cultivating an aural fluency for a system's sonic feedback.

In auditory perception research, the problem of aural training and its relationship to "perceivability" is well known. Auditory training is an essential element in areas dealing with the display of information using non-speech sound. Both in earcon/audio icon design practices [5, 6, 7] and especially in the field of sonification [8, 9], efficiency is largely defined and achieved through a careful balance between perceptual intuitiveness of the sonified information, meaningful mapping to data, and facilitation of learning how to make sense of the auditory display. Implied in this task is not only auditory perception but also auditory memory as well as associative and connotative epistemic aspects of listening. Yet there is relatively little research focusing on precisely this area of auditory display design – the nature, qualities and characteristics of how users learn to interpret complex sonic feedback. Further, there is little research into collaborative tasks that require group work and group auditory perception of feedback. This is where our project focuses and so in this paper we attempt to offer usable heuristics of the way in which collaboration, focus and listening patterns work together to constitute an "aural fluency" of a given system's feedback (see Figure 1 for our conceptual proposition).

Our design for socio-ec(h)o, albeit taking a novel direction in sonifying movement or intensity (second-level interpretations from data), is still inspired by some of the foundational literature addressing the balance of 'perceivability' and auditory training, including Flowers and Hauer's research in auditory scatter plots and graphs [10]; the works of Bonebright et al. [11] and Brewster and Brown [12] in perception of multiple data series sonified simultaneously; and Hunt and Hermann's research into interactive sonification [13]. Hunt and Hermann specifically found embodied gesture pairs well with sound feedback, and offloads some attention focus so as to facilitate comprehension from sound. In the more general field of sonification, Smith and Walker's article [14] is an example of applying contextual and peripheral auditory cues to facilitate learning of audio response in sonifications of financial information. The authors concentrate their study on the use of "additional context" elements – auditory tick-marks, axes and labels – and find it moderately helps with comprehension, however what is significant is their finding regarding the role of a reference tone in accurate value estimation by the users. In another study, referenced by Walker, Lindsay and Godfrey, the researches create a principle curve from dense scatter plots, and sonify individual data points "around the curve" as amplitude fluctuations of the main frequency, experienced as Doppler shifts [14] – thus taking advantage of the ear's natural sensitivity to the directivity of sound.

Both of these aural facilitation techniques – context-based and a continuous auditory graphing - have been shown to dramatically impact accurate value estimation, general perceivability of data, response time and accuracy of comprehension [14, 16, 17, 18, 19]. Most of these studies however do not necessarily address the *temporal* and/or *collaborative* effect of aural competency, perception and fluency when interacting with a technological system, nor do they identify discreet learning states and design for them or with them in mind. The type of auditory fluency that is needed and seems to develop in more physical, situated technological environments such as embedded interaction spaces is much more akin to everyday listening. Thus, in order to enrich our understanding of the activity of listening we look to the way it has been articulated by Truax [2] and Schafer [3] (in natural settings, from a holistic perspective), as well as Gaver [20], Bregman [21] and Ballas [22] (from an auditory perception perspective). By analysing everyday and natural acoustic environments, Truax and Schafer build an understanding of listening as an epistemic activity – a complex and dynamically shifting process, nested and interdependent on context and setting, and not simply a mechanical or a psychological process of perception (see Figure 1). Based on Schafer's typology of the natural sonic environment [3] and the World Soundscape Project's ethnographic work on local soundscapes, Truax articulates several ways of listening that he calls 'listening positions' [2]. These positions – modes or states of listening allow us to 'tune in' or 'tune out' of our sonic environment, picking out cues when needed and ignoring others. Understanding these positions then is crucial to design of auditory displays, especially ones that are nested in physical systems of embedded interaction, given that such conditions most resemble everyday contexts.

Listening-in-search is a position where the listener is actively searching either for particular cues in the environment (for example, the cocktail party effect – where we fine-tune our ears to pick out a faint familiar voice amongst a crowded noisy space) or is actively listening for any auditory cue that might be of use/meaning. Listening-in-readiness refers to a combination of non-active listening (background listening) and a pre-cognitive attention to a specific (typically familiar) sound that is expected to occur – such as a baby cry at night or the car of a familiar person. Background listening is one where we intentionally tune out the surrounding soundscape in order to focus on other modalities of feedback or tasks at hand, and finally, analytical listening is one of the listening positions that Truax attributes to the technological development of electroacoustics, recording and reproduction of sound. It is an active, deconstructive listening that occurs in situations where the user knows that sound is designed for an artistic or informational purpose, and there is a cognitive, critical engagement with the nature, characteristics and properties of sound. Lastly, an important point Truax makes about listening in everyday, physical environments is one relating psychoacoustic properties of listening to the context – namely, the fact that our ears attune, depending on the setting and our associations with it, to particular sound properties and changes [2].

Based on this brief bridging of literature on auditory training as well as characteristics of listening, we define aural fluency as a developed competency in using a system that responds through sound. In the sections to follow we hope to explicate the details of this competency and show how it may be developed by learning and acquiring key listening positions. The positions allow users to become increasingly competent and efficient in interpreting the system's response, as well as to link, in an embodied manner, their ability to affect the system through actions, and finally – to do so in a collaborative manner, as a group.

When touching on 'fluency' however, it is important to note some of the cognitive and educational connotation of this concept. Many theorists studying creativity and originality touch on 'fluency' and one of the most prominent voices is that of Gilford's 1962 [4] 'intellectual abilities' expose, where he defines four kinds of fluency (as related to literacy): word fluency, ideational fluency, associative fluency and expressive fluency. If we re-interpreted these notions to relate to aural fluency, they too, offer a usable framework for thinking about the discreet skills that players must develop in order to effectively use sonic feedback as guidance in collaborative problem-solving.

3. PROJECT BACKGROUND

Since we have already published the design details of our project elsewhere, including in the ICAD community, and the fact that we are specifically focussing here on second-level interpretive findings, we'll only provide a brief description of the project in this section. socio-ec(h)o is a prototype environment for a playful collaborative puzzle-solving activity, whose ultimate goal it is to explore the design, use and evaluation issues of embedded interaction systems and social interaction. The overall research goal is to understand how to support groups of participants as they learn to manipulate an embedded interaction space, to understand and interpret feedback reliably and effectively and achieve competency by problem-solving as a team. The research questions are numerous in a project of this nature and yet immersive and embodied interaction does not lend itself to reducible and isolatable variables that can be measured independently. Given this, our evaluation protocol focused on ecological investigation and theory-building, aiming to provide broad, yet particular set of heuristics that help describe and make sense of the different system and display components of such environments.

The socio-ec(h)o prototype is a six-level puzzle game played by a team of four people in a "black box" space with controlled immersive lighting and a surround sound auditory feedback (see Figure 2). The puzzles' solutions are collaborative whole-body physical configurations that players have to achieve and hold for a duration of time, as a team. Their movements in space are tracked by a 3-d setup of infrared cameras. Each level is represented by a unique set of light and sound feedback and the team's progression is interpreted in real time by a reasoning engine resulting in a change in the feedback's intensity intended to inform and guide players towards the right solution. The levels are progressively more challenging in terms of body states as well as in terms of transparency or 'perceivability' of feedback as represented through changes in the environment in light and audio.

Figure 2. socio-ec(h)o gameplay, Level 5. Participants work together to solve the puzzle "Gazing Over Waves" and slow down fast pulsating water streams into one coherent wave.

3.1. Auditory Display Schema

Again, since we have detailed the audio display approaches in socio-ec(h)o elsewhere [1, 23] we will just briefly explain the main innovation, which was our *intensity gradient* mapping of parametric sound change to a real-time interpreted value of the team's progress in the game. The main approach to the sonic feedback is intensity-based parametric change of a continuous ambient sound adapted from the area of sonification, however, once players reach an intensity of around 3.5 we also implemented a confirmatory signal, which they were introduced to before commencing the game (see Figure 3 below).

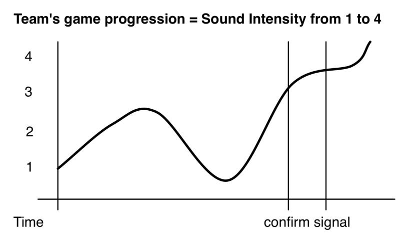

Figure 3. A schematic of a given game progress and its corresponding sound feedback mechanism.

While the base ambient sound was different for each level, attempting to *narratively* fit the puzzle and required movement at hand, the confirm signal (reward – short granulated glass clink) was consistent throughout the game levels. Figure 4 presents a screenshot of the audio engine used, where we see a 5-layer cross-fader with a selection of parametric effects attached to them, which get called up with the beginning of each level. The base ambient sounds for the levels included environmental recordings such as water, fire and abstract soundscapes such as percussion impact sounds (e.g. from marble or dice) and a polyphonic tonal drone. The way sound display fit into the game as a core mechanic was in following the natural skill mastery progression of the game – the soundscape in Level

1 was perceptually easy to interpret (an abstract musical drone where progress is mapped only to amplitude) while the soundscape for Level 4 was significantly more ambiguous (a fire of increasing intensity, as represented by five different fire crackle/bonfire sound files mixing dynamically). Thus ambiguity, perceptual and aesthetic, rather than being avoided, became a core part of the game experience itself.

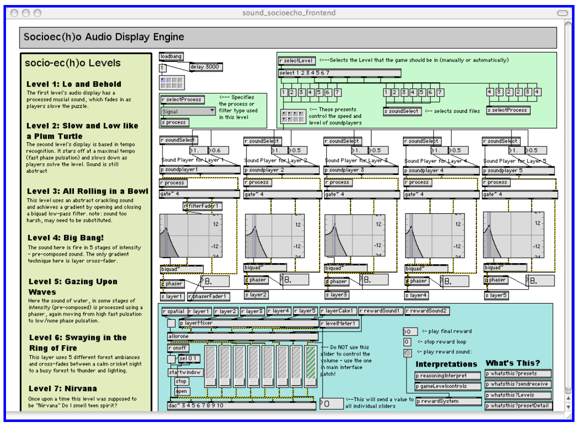

Figure 4. The audio display engine, in Cycling 74's Max/MSP software.

3.2. Study Design And Results

As mentioned earlier, this project involved many research questions and points of interest. Thus we'll hereby briefly describe the study protocol and results and then we'll focus specifically on findings related to the reception of sonic feedback. In that, we'll elaborate on our second-level analysis of stages of aural fluency, as well as exemplify the role of listening, focus and collaboration in that.

For socio-ec(h)o's evaluation we had 14 teams of four playing the game, for a total of 52 participants. Teams had an hour and fifteen minutes to complete all six levels (if they could), with a break in the middle. All study sessions were videotaped for the purpose of later video analysis. We also collected times of completion measures, verbal transcripts, and a post-activity survey. The survey instrument contained a combination of Likert-scale questions and open-ended questions relating to the effectiveness of the system's response in helping, guiding the problem-solving, and creating an enjoyable experience for users. In terms of the auditory display, we hoped to see some consistency in performance in certain levels across teams – this could point to a design success or flaw of a particular approach to sound feedback taken in that level. The survey and transcripts in turn were meant to serve as indicators of what participants thought about the design. In fact what we saw in the results is no consistency in the times of completion (ToCs) of different teams, even within the same level – for level 4 alone some teams took less than 3 minutes to solve and other teams took over 45 minutes! What is curious and cannot be explained by our initial results is participants' ratings of the effectiveness of soundscapes and audio response in different levels. According to the survey results, most participants both preferred the soundscape of level 4 and thought it worked best (this level's puzzle was titled "Big Bang" and its ambient sound base was a gradation of fire). On average, participants thought highly and positively of the system's accuracy and effectiveness with regard to sonic and other response, regardless of their individual performance. There is no correlation between performance and preference for sonic feedback. So we were left wondering how it is that teams uniformly found the sound response to be supportive and the experience positive regardless of their ability to complete the puzzles in a timely fashion or play the game "well"? The answer would have to be that something else acted as intrinsic motivation within the game, and/or there were other skills besides the ones we hypothesized about that were developing – social and team cohesion, focus on gameplay and efficacy in problem-solving together, as well as fluency in interpreting cues from the system. This question led us to explore the idea of aural fluency as a skill that developed over time and was intrinsically motivating. This concept is constituted through embodied interaction and offers its own rewards besides the game completion goal – a sense of mastery, familiarity and understanding of how to interpret and manipulate the system even if the explicit goal is not achieved.

4. LISTENING, FOCUS AND COLLABORATION

We devote the remainder of the paper to describing heuristics for aural fluency derived from our study. By way of theoretical descriptions and rich accounts we aim to demonstrate a level of transferability of our concept that goes beyond the scope of our study to apply to other embedded interaction settings, particularly collaborative AmI environments.

4.1. Listening Positions of Aural Fluency

As detailed in the theoretical section on aural fluency in the beginning, one of the paradigms that could be used to frame and explicate how and why aural fluency forms in situations of embodied interactive systems is that of listening positions [2]. Users in fact form these listening positions over time, in order to make sense of and cope with the auditory feedback presented to them. We hereby present several positions that we articulated based on our observations and video analysis; we describe them; offer vignettes from the study and then present three stages of aural fluency as design guidelines contribution. As shown in Figure 5 below, one thing we want to emphasize is the importance of what we call a "narrative transformation" as a cognitive link between listening as a perceptual activity and interpreting sonic feedback as a co-constructed, collaborative meaning-making activity.

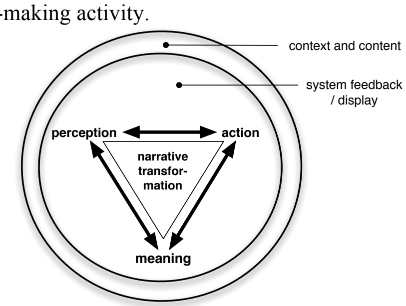

Figure 5. A schema of the interplay between perception action and meaning formation through narrative transformation.

In short, we propose that unlike individual interaction with sonic feedback, where users may employ more analytical and discriminate stances towards listening and interpreting feedback, the nature of collaborative listening has the added factor of the group listening *together*, and negotiating the shared meaning through translating sound changes *narratologically* into common references – i.e. "sounds like fire…a camp fire?..."… "we got something here.. let's keep it going" … "oh, now it's cold…" … "that's a good sign isn't it? Warmer?" "It's fast, that's good right? … no slow is better, fast is cold.", etc. It may be simply that single users interacting with an auditory display don't have to *externalize* their interpretation, however, that is precisely why it is imperative that we look into the unique way in which groups make sense of auditory displays. Following are several listening positions that we saw emerge as patterns during our observation and video coding.

4.1.1. Resetting

This is a listening position where players are often static – not moving, or not playing, listening in full attention to the sound in preparation of a strategy for an embodied configuration. It may or may not involve verbal communication about the sound feedback, but it is clearly focused on understanding and familiarizing themselves with it. In its first iteration, novice players who are taking time to familiarize themselves with the soundscape of a new level often utilize this position. As an extension of this pattern comes a process we named *auditory memory flush*, where players exit the game/space or otherwise bring feedback to its minimum in order to study it or start fresh. It is another well-known issue in sonification that several researchers have worked to address - essentially, auditory memory is fairly short and if people are asked to compare two or more sounds, or continuously compare a dynamically shifting sound in order to derive meaning, auditory memory is fragile. Eventually it gets 'full' and people start losing confidence in the just noticeable differences between different sounds or sound portions. Having a bottom-line reference tone is one way to solve this problem. In our case, the way people solved this problem for themselves was that they would knowingly go into a state or area of low intensity, in order to re-adjust their 'listening position' and start fresh with interpreting the gradient intensity change from there on.

4.1.2. Experimental Listening

In this listening position players are typically in motion and have formed embodied configurations while actively listening for a change in the soundscape that would indicate to them whether they are on the right track or not. While *resetting* is usually characterized by verbal references to the sonic response, participants in an *experimental listening* position typically indirectly reference sound. Verbal references are related to whether players are "doing something right" or not; whether the movement or position seems "hot, warm or cold" and whether (the system) "likes it" or not. The significance of this listening position to aural fluency is that it is the strongest form of communication between system and users, and thus plays the strongest role in learning and skill acquisition. Once users are more comfortable interpreting a particular soundscape – its approach to intensity and rate of change, there is a marked improvement in speed with which they are able to experiment, listen and make judgements about their progress.

Related to this 'listening to change' is the notion of *tolerance of ambiguity* - a concept that surfaced as an important feature of ambient sound feedback in ubiquitous computing spaces in general. Changing sound, especially complex changing sound, always has degrees of ambiguity as a form of feedback, however, in this situation of embodied interaction – a physical and spatial relationship to sound, participants were accommodating of less than clear auditory feedback. In respect to acoustic communication and auditory display paradigms, this concept speaks to the type of everyday listening that people are proficient in utilizing already – making sense of confusing, unclear, complex sonic situations by selectively focusing or shifting attention on different aural elements, and fine-tuning their ear to particular sound structures and qualities.

4.1.3. Narrative Listening

When examined through the lens of everyday listening practices, as developed by Schafer and Truax [2, 3], we see numerous examples of both players forming a narratological association with the soundscape – "no, no we got no fire, somebody has to keep the flame!"… as well as seamless connections to embodiment – "so the fire builds up and we're all still…what if we move towards the sound? What if…when the fire gets all crazy we move more with it?" This narratological relationship could also serve to explain why negative polarity was not a problem in Level 2, when the solutions required moving very slow (so the slower they move the slower the sound gets) but it was a problem in Level 5 where the relationship between puzzle, movement required and sound feedback was less seamlessly related, more abstract. Prior associations and familiarity with sound definitely coloured user experiences with the sonic feedback and their interpretation of its meaning, range, connotations, etc. In fact, interestingly, recognizable sounds seemed, in our preliminary tests, to present a more intuitive range and scaling (to use sonification terms) to users than abstract sounds. This principle also relates to acoustic communication theory, which discusses people's ability to naturally derive information from the structure and quality of sound at the micro level of perception, depending upon the context. In other words, it is the context that shapes listening in such a way that it can fine-tune information retrieval from subtle sound changes by recognizing patterns in the sound. As Truax points out, it is these patterns that mediate the relationship between people and environment [2].

4.1.4. Anticipatory Listening

This listening position usually emerged once players achieved the preceding listening positions and were able to anticipate sonic responses. For example, often players came close to solving a puzzle and would then hear an intermediate reward sonic cue. Through actively interacting, moving, and at the same time listening they could anticipate the upcoming reward sound. This listening position is usually accompanied by verbal references like "we almost had it!" or "we heard the sound, let's keep it up" and then verbal exclamations when the intermediate reward is activated again.

4.2. Aural Fluency Accounts

In order to illustrate the typology of listening positions in systems of embodied interaction, we offer several vignettes from our 14 sessions, of selected levels. We describe the progression and interplay of the listening positions and show screenshots of our video coded data demonstrating the relationships between listening, focus and collaboration. In other work {REF}, where we focus more on the interactional qualities of socio-ec(h)o, we have already identified two other team states as relevant heuristics that characterize the experience and success of teams' problem-solving. These are game focus (the extent to which players are focussed on playing the game) and team cohesion (the extent of collaboration within a team in any given moment). In the figures to follow, we not only present coded versions of the listening positions we have identified, but we also juxtapose them along our coding of the teams' focus and cohesion in order to paint a fuller picture of the complex process of listening in a collaborative, problem-solving, ambient intelligent environment.

Team G, Level 4 In this vignette, Team G has managed to reach level 4 however they are still developing their collaborative aural fluency, and they are thus not successful at solving the puzzle for level 4. Level 4 is in many respects a "true test" for teams since it is possible to reach this level with some luck, not just skill, and is the first level with two sequential stages of solution, which requires a heightened awareness of the system's response. Yet players spend a considerable amount of time only passively listening (in black on Figure 6 below), along with many breaks in their game focus. Not understanding why they have been unsuccessful, the players attempt to retrace their steps in this level for a while, crouching in the centre waiting "until we have something" [*resetting*] and not moving and intently listening as a means of *experimental listening*. The lack of communication about the sound [resulting form *narrative listening*] and focused experimenting may in fact be an indicator for their failure to solve the puzzle. Further, we see in the zoom-in section below that their shifting of listening attention often coincides with shifts in game focus – perhaps a greater indication of attentional issues.

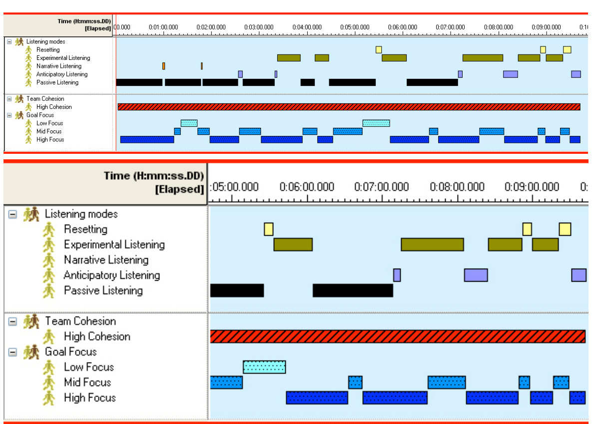

Figure 6. Video coded sections from Team G, level 4.

Team C, Level 4 This is one of the lengthy examples of level 4, taking almost 50 minutes to solve, and in actuality this team ran out of time and wasn't able to solve it. Yet the timeline more than ever demonstrates clear patterns of fluency over various aspects of the auditory feedback and as such – manipulating of system response. Since this level has two stages of completion, when teams achieve stage 1 they hear the confirmatory signal, and like many others, Team C employed a mixture of *experimental* and *narrative listening* in order to achieve stage 1; and after that they alternated between anticipatory and experimental listening stances in order to try and complete the puzzle. What is interesting to note is that once they had figured out stage 1, they employed anticipatory listening when they came together in physical formation for it as a group, without any explicit communication about listening. Only if someone moved too early they would comment: "no, we haven't heard the sound yet…" or "see, it's getting hotter here, the fire's bigger". After mastering stage 1 the team is clearly displaying aural fluency over the system's response, eliciting it with ease and precision. Every subsequent time they achieved stage 1, they continued either to anticipate the aural completion of the level poised in a passive physical position, or attempted different formations while listening experimentally. In all of these cases, listening was performed 1) while multitasking with movement and even team communication, and 2) together as a group, in an implicit (or occasionally – explicit) listening agreement. Finally, we see even more clearly here many shifts and transitions between different listening and game focus states over time, with more breaks in cohesion and durations of passive listening towards the end.

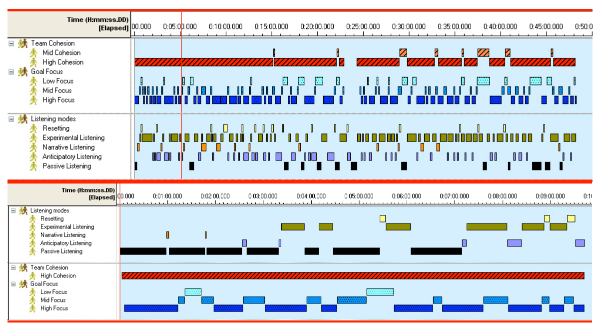

Figure 7. Video coded sections from Team C, level 4.

Team D, Level 3 In this example, Team D illustrates a mastery of aural fluency. The vignette shows a shared understanding of the listening positions: Even though there is no mention of sound a team member utters: "It got cold…" [*narrative listening*] halfway through it, suggesting that everyone has been playing dormant yet monitoring (listening to) the system's response. When they finally move ahead, a player comments while moving: "ok, this is getting higher…I think that's good" [*experimental listening*] – addressing the rising pitch of the soundscape of bouncing marbles. When the intermediate reward plays, another player immediately takes notice: "hey there's the sound – what are we supposed to do, hold it?" [*anticipatory listening*] – and he pauses for a brief moment before he continues to move like the others (see Figure 8).

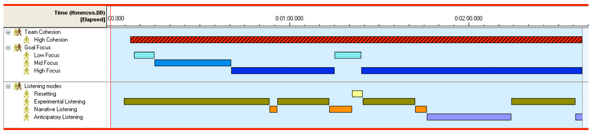

Figure 8. Video coded section from Team D, level 3.

<u>*Team A, Level 3*</u>: Key to aural fluency is that the listening positions can be seen as progressive levels of competency. Once achieved, they are utilized in a complex interplay as listeners alternate between positions. It is also clear that aural fluency undoubtedly supports collaboration and shared understandings. In this vignette, players in Team A strategize and collaborate together seamlessly supported by their shared aural fluency: Users first notice sound when they stop to strategize and pause to decide on a strategy: "let's do it to the sound of the dice…" [*narrative listening*]. When the intermediate reward comes on they are moving/playing and continue on, making a verbal notice: "oh!..there we go!" [*anticipatory listening*]. When re-evaluating their strategy they do so in motion – "it doesn't like it that way" (referring to system = sound/light) [*experimental listening*]. Again in motion, they acknowledge the sound's narrative relationship to the activity: "so how do you do dice? Oh…like rolling the dice I see…let's just move in circles"… [*experimental listening* and *anticipatory listening*]. After regrouping again to talk about strategy, they start moving and right away comment, "oh there it is...so…warmer, warmer" [*anticipatory listening*]. Another regrouping [*resetting*], and in motion again, they reflect on their embodied configuration and strategize, referring to the sonic feedback as guidance: "I don't know how small a circle we need. When it was making the sound we were much closer" [*experimental listening* and *anticipatory listening*].

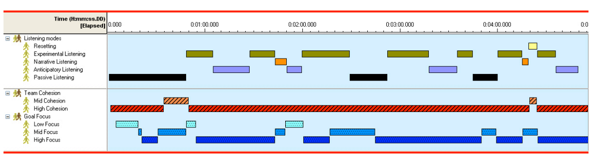

Figure 9. Video coded section from Team A, level 3.

4.3. Stages of Aural Fluency

Based on the listening positions related to listening in embodied interaction environments, and our observations, we further formulate three stages of aural fluency. There is an evident progression through the stages marked by listening positions. It is also important to note that our observations were in a collaborative setting and it is in large part integral to our description of aural fluency and its stages. We believe it was evident in the vignettes the underlying social as well as embodied characteristics of the concept. We feel that the stages provide design parameters that could in turn be used in the generative stages of a prototype to account for different and shifting competency in listening and interpreting sonic feedback of an embedded interaction system.

<u>Stage One:</u> at first users need to understand the logic and actuality of the sonic display response. For example, users need to define for themselves and their collaborators what a constant intensity is, what it means, and how is it affected by their actions, configurations and movements. *Narrative listening* and *experimental listening* are in evidence at this stage, followed by *resetting*.

<u>Stage Two:</u> In this stage of aural fluency users have created a serviceable and shared understanding of what the sound means to them and how it responds to movement on for example a gradient scale. They may however still not know what exactly affects change, or how to interpret granularities of the feedback. At this stage, there is a tolerance for some ambiguity, but there is an ongoing need to stop and reflect on the response of the system. *Experimental listening* and some *anticipatory listening* at an intermediate level with the need for *resetting* are in evidence at this stage.

<u>Stage Three:</u> In this stage of aural fluency, users have acquired enough familiarity with the particular soundscape, approach to intensity and rate of change, in addition to having accomplished a systematic approach to experimentation. They are able to very quickly shift between attempting solutions, strategizing, and experimentation since they can almost immediately tell by listening to the sound whether their new approach is favourable or not. There is less narrative listening and more focussed experimentation and exercising of players' fluency of the system's 'language'. Often in this stage, users can easily reconstruct past attempts and incrementally change them. Ease with *experimental listening*, *anticipatory listening* and the strategic use of *resetting* are in evidence at this stage.

5. CONCLUSION

Ubiquitous computing environments and spaces of embodied interaction increasingly rely on guiding and informing, ambient feedback embedded within the environment and able to serve and inform groups of participants as opposed to individual users. Understanding how people collaboratively familiarize themselves with the meaning, properties and structure of ambient feedback, and designing for various levels of expertise is what we have tried to address in this paper with the notion of aural fluency. The contribution of this paper is in articulating interaction characteristics related to listening positions and acquisitions of the overall skill of listening. Moreover, we focus on evidence of acquiring the skill of interpretation over time – a listening competency we termed *aural fluency*. As articulated here, this concept addresses not only static listening positions as characteristic but also their development overtime, with potential patterns of sequence and alternation during the embodied interaction with the ambient intelligent space. As a result of our current exploration, there are several main criteria that sound feedback for embodied systems must adhere to: it has to support embodied learning (competency building in physical and temporal conditions); and it has to respond to and manage users' attention (their listening position shifts) and still allow for effective collaboration. From the detailed accounts of playing socio-ec(h)o an emerging pattern points to the experience of learning competencies within our system as an intrinsic motivating factor and source of satisfaction with the interactive experience. The specific conditions that collaborative, problem solving ambient contexts constitute a process of acquiring competency in interpreting feedback over time, in motion, while strategizing. Often understanding sonic feedback is instant and shared, without needing explicit conversation but is supportive of implicit agreement. Users dynamically alternate between different listening positions and it's important that they don't lose or miss crucial opportunities for interpretation of the sound display and potential actions. At one time they may be listening attentively, or analytically, while in another they may be listening associatively, and in yet another – completely embodied and experimentally. These core patterns we feel constitute and elucidate the notion of aural fluency and are a critical part of ambient intelligent system design, as well as auditory display design.

6. REFERENCES

[1] Droumeva, M. & Wakkary, R. (2008) Understanding aural fluency in auditory display design for ambient intelligent environments. In Proceedings *International Conference on Auditory Displays*, 2008

[2] Truax, B. (2001). Acoustic communication (2nd ed.). Westport, CT: Ablex.

[3] Schafer, R. M. (1977). The tuning of the world. Toronto: McClelland and Stewart.

[4] Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw-Hill.

[5] Brewster, S.A., Wright, P.C. & Edwards, A.D.N. (1992). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), Auditory display, sonification, audification and auditory interfaces. Santa Fe, NM: Addison-Wesley, pp. 471-498.

[6] Walker, B., J. Lindsay and J. Godfrey. (2004). The Audio Abacus: Representing Numerical Values with Nonspeech Sound for the Visually Impaired. Presented at ASSETS '04 in Atlanta, Georgia, ACM Press.

[7] Brazil, E., Fernström, M. & L. Ottaviani (2003). A new experimental technique for gathering similarity ratings for sounds. *Proceedings of the 9th International Conference on Auditory Display (ICAD2003).* pp. 238-42.

[8] Walker, B., & Kramer, G. (1996). Mappings and metaphors in auditory displays: An experimental assessment. Presented at the third International Conference on Auditory Display (ICAD 1996), Palo Alto, California.

[9] Kramer, G. et al. (1999) "The Sonification Report: Status of the Field and Research Agenda. Report prepared for the National Science Foundation by members of the International Community for Auditory Display," International Community for Auditory Display (ICAD), Santa Fe, NM 1999.

[10] Flowers J. H., & T. A. Hauer, (1995) "Musical versus visual graphs: Cross-modal equivalence in perception of time series data," Human Factors, vol. 37, pp. 553 – 569.

[11] Bonebright, T. L., M. A. Nees, T. T. Connerley, and G. R. McGain, "Testing the effectiveness of sonified graphs for education: A programmatic research project," presented at International Conference on Auditory Display, Espoo, Finland, 2001.

[12] Brown, L., Brewster, S. et al. (2003). Design guidelines for audio representation of graphs and tables, *Proceedings of the 9th International Conference on Auditory Display* (ICAD2003), pp. 284-287.

[13] Hunt, A., & Hermann, T. (2004). The importance of interaction in sonification. Proceedings of the Meeting of the International Conference on Auditory Display (ICAD), Sydney, Australia.

[14] Smith, D. R., & Walker, B. N. (2002). Tick-marks, axes, and labels: The effects of adding context to auditory graphs. Presented at International Conference on Auditory Display (ICAD 2002), Kyoto, Japan.

[15] Hermann, T.T., P. Meinicke, and H. Ritter (2000) "Principle curve sonification," presented at International Conference on Auditory Display, Atlanta, GA, 2000.

[16] Flowers, J. H. , L. E. Whitwer, D. Grafel, C, and C. A. Kotan (2001), "Sonification of daily weather records: Issues of perception, attention, and memory in design choices," presented at International Conference on Auditory Display, Espoo, Finland, 2001.

[17] Peres C. P., and D. M. Lane, "Sonification of statistical graphs," presented at International Conference on Auditory Display, Boston, MA, 2003.

[18] Kramer, Gregory, & Walker, Bruce. (2004). "Auditory Displays." In J. Neuhoff (Ed.), Ecological Psychoacoustics. Boston: Elsevier Academic Press.

[19] Nesbitt, K. V. and S. Barrass (2002), "Evaluation of a multimodal sonification and visualization of depth of stock market data. *Proceedings of the 8th International Conference on Auditory Display* (ICAD2002), Kyoto, Japan.

[20] Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. Ecological Psychology, 5(1), 1-29.

[21] Bregman, A. (1994). Foreword. In G. Kramer (Ed.), Auditory display: Sonification, audification, and auditory interfaces. Reading, MA: Addison-Wesley.

[22] Ballas, J. (1993) Common Factors in the Identification of an Assortment of Brief Everyday Sounds. Journal of Experimental Psychology: Human Perception and Performance, 19(2), 250-267.

[23] Wakkary, R., M. Hatala, Y. Jiang, M. Droumeva, M. Hosseini. (2008) Making Sense of Group Interaction in an Ambient Intelligent Environment for Physical Play, TEI 2008: Second conference on Tangible and Embedded Interaction, ACM Digital Press