Hearing is Believing: Evaluating Ambient Audio for Location-Based Games

Jason Kurczak, T.C. Nicholas Graham, Claire Joly
School of Computing
Queen's University
Kingston, Ontario, Canada
{kurczak, graham}@cs.queensu.ca

Regan L. Mandryk
Department of Computer Science
University of Saskatchewan
Saskatoon, Saskatchewan, Canada
regan@cs.usask.ca

ABSTRACT
In location-based games, players act as their own avatar within in the physical world. Such games are increasing in popularity due to the wide adoption of smartphones that contain the location sensors necessary for their play. When played on a small, hand-held display, location-based games have two problems: users may be distracted, possibly leading to accidents; and players must map the display contents to the physical world, possibly reducing their sense of immersion. In this paper, we present the results of a study showing that *ambient audio* - a continuous stream of audio representing an entity in space - can replace visual displays for navigation tasks in location-based games. We find that ambient audio reduces player performance, but increases their sense of immersion in the virtual world and increases player safety.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: User Interfaces—*Auditory (non-speech) feedback, evaluation/ methodology*; H.5.1 [**Information interfaces and presentation**]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*

General Terms
Experimentation, Human Factors

Keywords
Ubiquitous game, location-based game, ambient audio

1. INTRODUCTION
Location-based games overlay a virtual world over the physical world, allowing players to move and interact within a physical environment such as a university campus or the corridors of a building [7]. Examples of location-based games include action games such as Human Pacman [2], treasure-hunting games such as PiNiZoRo [14], Capture the Flag games [19], and an increasing number of commercial games such as Parallel Kingdom. Location-based games are experiencing a surge of popularity as they can run on the modern smartphones that millions of people carry every day [6].

A significant issue with location-based games is that they require players to attend to a display showing the virtual world, diverting their eyes and attention from possible dangers in the physical world – players may risk tripping on steps or walking in front of a speeding bicycle. This problem can be mitigated by requiring players to wear virtual reality glasses so they can view the virtual and physical worlds at the same time. Such glasses are expensive, and require accurate low-latency tracking of the player's position and head-orientation, functions not provided by smartphones.

In this paper, we propose *ambient audio* as an alternative approach for overlaying a virtual world onto the physical world. Using ambient audio, players hear the activities of virtual entities rather than seeing them. For example, a player might hear a roaring-water sound representing a virtual waterfall. The sound is presented using spatial audio, so that a waterfall to the player's right *sounds* like it is to the right, and the waterfall sounds louder as the player approaches it.

Ambient audio differs from speech or alert-based audio applications such as GPS navigation systems. Rather than listening for commands ("turn left now"), the player hears a continuous stream of sound from the virtual entities within the environment (roaring waterfall). Since the sound streams are continuous, players can easily shift focus to events in the real world (e.g., avoiding bumping into a passing pedestrian) without losing context. The roaring waterfall sound will still be there when the player returns attention to it.

In its simplest form, ambient audio can be implemented using standard smartphones and a stereo headset. As a modest enhancement, a low-cost head tracking system such as TREC can be used to allow positional sound to track head movement [6].

Ambient audio is most directly applicable to games involving navigation between entities. Sounds can be used to represent waypoints, friends and enemies, and phenomena such as explosions or geographical features. It can be combined with other modalities such as visual displays and spoken voice – for example, players might navigate to a destination using ambient audio, then (while stationary) engage in a

Full presentation, ACE'2011 - Lisbon, Portugal

mini-game using the Smartphone's display.

In this paper, we compare ambient audio to visual displays for ubiquitous games. To illustrate the concept of ambient audio, we introduce the the novel *Growl Patrol* game. We present a study showing that in this game, ambient audio leads players to have a greater sense of presence in the virtual world than when using a hand-held visual display, and that when using ambient audio, players are less likely to collide with obstacles in the physical world. Players' game scores were higher with a visual display, but overall players expressed preference for the ambient audio version.

The paper is organized as follows. We first introduce the Growl Patrol game as an illustration of ambient audio. We then present our study evaluating the comparative effectiveness of ambient audio and visual displays, and finally use the study to reflect on the strengths and weaknesses of the ambient audio approach.

2. RELATED WORK

Ambient audio has been used in location-based interfaces as early as the second world war, where stereo audio beacons were used to guide planes to night landings on unlit runways [9]. Since then, both ambient and spoken sound have been used to enhance location-based games and mobile interfaces in general. In this section, we review existing literature in the use of audio for location-based games, other types of mobile interfaces, and investigations into the effectiveness of audio and visual interfaces.

2.1 Ambient audio in location-based games

Ambient audio has been used in a small number games. These include *SoundPark* [10], *Songs of the North* [3] and *Viking Ghost Hunt* [8].

SoundPark is a team game that is played outdoors, requiring players to "capture" sounds and use them to construct a song [10]. "Hunter" players use earphones to hear sounds spatially in the environment, and must gather them up and bring them back to home base in order to assemble them into a musical arrangement.

In Songs of the North, players perform quests to find items and rescue characters. They kill monsters and cast spells using a virtual drum on the screen [3]. Players ambiently hear other characters, enemies' attacks, and the drum spells of other players. The authors found that players were not likely to use the audio interface unless they were explicitly told of its importance. Songs of the North uses external speakers with no volume control, making the game socially awkward to play.

In Viking Ghost Hunt, players find "paranormal" zones using audio cues triggered by GPS location [8]. Once a zone is found, players interact with the screen to decode secret messages. Ambient sound helped to create an immersive and emotionally engaging experience, where 70% of players either agreed or strongly agreed with statements characterizing the game's audio as immersive.

These three examples provide early indications that ambient audio can be used to help players perform navigational tasks in location-based games, that ambient audio can mesh well with screen-based mini-games played when the player is stationary, and that the use of ambient audio can contribute to immersion. In the one game where ambient audio was not a success, the use of external speakers was identified as a pitfall.

2.2 Ambient audio in other mobile applications

Ambient audio has been used in other applications intended for mobile use, with goals of reducing the demand on users' visual attention or of increasing the application's immersive experience.

Several applications apply ambient audio to navigation tasks, with the goal of allowing users to focus visual attention on a demanding foreground task (e.g., driving) while navigation instructions are conveyed via audio. Holland *et al.* created the *AudioGPS* system as a minimal-attention navigation system, using spatial audio cues to highlight waypoints of the path to be followed [4]. Initial informal evaluation of this system resulted in positive feedback.

Stahl later created and tested the ambient *the Roaring Navigator*, for identifying and navigating to landmarks in a zoo environment [13]. Representational sounds are played at the location of each landmark or animal in the zoo. All of the participants using the system were able to complete a wayfinding task around the zoo, and most rated the interface easy to use.

The *Noisy Planet* is an ambient audio application usable by tourists in an unfamiliar city [6]. Users hear the locations of points of interest. As they approach an interesting location, they can look at their smartphone display and see information about the location. Noisy Planet was intended to allow users to experience the sights of a city without the intrusiveness of verbal navigation instructions.

2.3 Effectiveness of audio interfaces

Ambient audio has the promise of helping users navigate more safely while increasing their immersion in a virtual world overlaid over the physical world. The above examples indicate that systems based on these principles can be practically built. So far, however, there has been little empirical evaluation of how effectively ambient audio conveys information, and whether, when compared to visual interfaces, they improve users' safety, immersion and enjoyment. We now review those experiments that have been performed to-date.

Pirhonen *et al.* tested a user's ability to control a touch-screen portable music player while walking around a set of obstacles [11]. Participants using gestured-based input with audio feedback experienced a smaller workload compared to those using the touch-screen GUI, and also tended to walk at a pace much closer to their preferred walking speed.

Walker and Brewster compared the use of visual and spatial audio interfaces to monitor a background task while performing a demanding visual foreground transcribing task [16]. The background task required that users monitor file transfers that periodically required user intervention. Participants used either a standard visual progress bar or a novel spatial audio progress bar that could display transfer progress and rate. Users' performance in both the foreground and background tasks was better in the audio condition; users were significantly more likely to correctly handle the file transfer task, and were also able to type at a faster rate.

Takeuchi and Sugimoto compared audio and visual interfaces for navigating city streets [15]. *CityVoyager* is a system that analyzes behaviour while performing everyday shopping, and then guides users to recommended shops based on their shopping history. A simple beeping "metal-detector"-like interface was used to indicate recommended shops, and was played over a speaker. When users had access to both the visual and audio interfaces rather than the visual alone, they scrolled the map eight times less, and visited the recommended shops more often. When the audio cue beeps became very frequent near shops, however, users reported being more distracted by the interface than when using the screen alone.

In addition to these studies on the effectiveness of ambient audio, there have been notable studies on the comparitive effectiveness of spoken-word audio in navigation tasks. Walker *et al.* found that car drivers using a spoken-audio guidance system had fewer navigational errors than those using a visual mapping device, and did not slow down as much as in the visual case [17].

Sodnik *et al.* also tested drivers in a simulator, using steering-wheel mounted controls to navigate a menu in order to perform predetermined tasks while driving [12]. Drivers who had spoken-word representations of the menu system drove more safely than those using visual displays, with 60% fewer driving penalties assigned on average during the tests.

Collectively, these studies show that ambient audio interfaces can improve users' performance when they need to split their attention between two tasks. They also show that audio in general can lead to lower error rates. These hint that ambient audio may be useful in location-based games. No study, however, comprehensively compares ambient audio versus visual displays in terms of their effects on performance, safety and immersiveness. In chapter 4 of this thesis, a first step is taken towards filling this gap.

3. EXAMPLE: GROWL PATROL

We now present *Growl Patrol* as an example of how ambient audio can be used within a location-based game. Growl Patrol has been implemented using the TREC toolkit [6], and has been play-tested by tens of people.

Growl Patrol is played outside, using GPS to track the player's real-world location. The premise of the game is that a number of small animals (cats, dogs, and birds) have cunningly escaped from the local pet shop, and the player is charged with bringing them back. These animals run around the play area until they are picked up, at which point the player can drop them off at the pet shop. There is also a hungry tiger nearby, who attempts to steal any animals that the rescuer is carrying. The player's goal is to catch and drop off as many of the animals as possible in the time allowed, without losing any to the hungry tiger.

Figure 1: Growl patrol gameplay: players hear animals overlaid on the physical world.

Growl Patrol uses an interface based on ambient audio. Players hear the animals through stereo headphones. Each animal is represented by a constantly repeating sound – birds twitter, cats meow, and dogs bark. Sound is spatialized on a 2D plane; if the animal is to the player's left, the sound will appear to be coming from the left, and the closer the animal is to the player, the louder it will sound.

Figure 1 shows a player outside playing Growl Patrol. He is wearing stereo headphones to "hear" the game. His position is tracked via a GPS, and his head orientation is tracked with a gyroscope and compass attached to his hat. The figure shows that the player hears a bird tweating to his right. The map in the top of figure 1 (not part of the game) illustrates what the player hears: the bird is close by to the right, and the tiger is in the distance to the left.

The player typically runs towards one of the available animals, while attempting to dodge the tiger. The tiger takes on a different sound depending on whether it is patrolling (low growling sound) versus actively chasing the player (fierce roaring). Once an animal is caught, an audio homing beacon starts up to indicate the location of the pet store.

Growl Patrol uses ambient audio to provide a virtual overlay on the real world, using sound to locate the animals and pet shop. This approach can improve immersion versus a hand-held visual display. Rather than having to look at a map and work out how it corresponds to the physical world, the animals' sounds are overlaid directly onto the real world. As one player said, "it's like you're in a park trying to catch animals, as opposed to staring at the little map trying to catch animals."

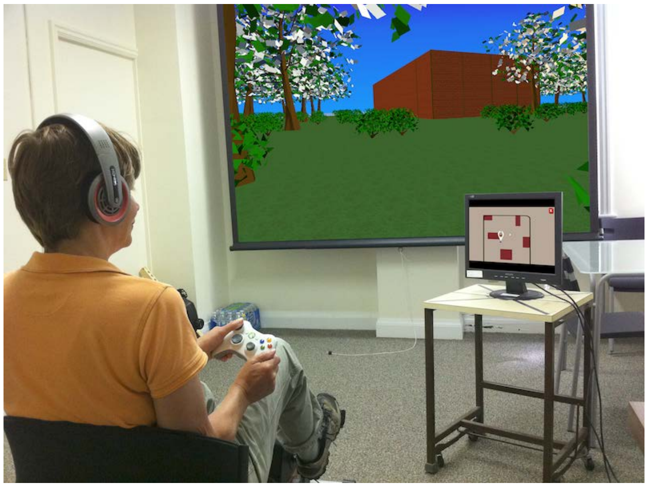

Figure 2: Indoor experimental equipment.

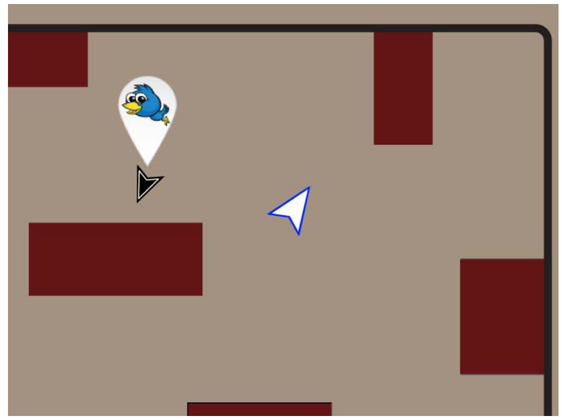

Figure 3: An example of the radar map used in the visual condition. This map is shown on a display on the player's right side.

The use of an audio-based interface can improve player safety versus a visual interface, as players' vision is free to view the world around them. The use of ambient audio (versus alert-based or spoken audio) allows players to drop their attention from the game when the real world intrudes. When the player returns their attention to the game, no important information has been missed, allowing the player to immediately continue chasing the next animal.

Ambient audio is, however, limited both by humans' ability to accurately localize a sound source, and by the inaccuracy of algorithmically spatialized sound. Growl Patrol requires players to chase progressively faster animals and to get within a 10 m radius in order to catch them. Specifically, sound is spatialized using a generalized approximation of a human head-related transfer function (HRTF), providing lower accuracy than real-life sound [1]. Additionally, players find it difficult to determine a sound's distance based on volume alone. Therefore, while our anecdotal experience indicates that Growl Patrol's ambient audio increases immersion and safety versus a hand-held display, it comes at the cost of increased difficulty in gameplay.

4. COMPARING AMBIENT AUDIO AND VISUAL DISPLAYS

We tested the hypotheses that when compared to hand-held visual displays, ambient audio can increase immersion in location-based games while increasing their safety, possibly at the cost of in-game performance. The experiment compares an audio and visual version of a location-based game inspired by Growl Patrol.

Testing Growl Patrol in an outdoor area on the Queen's University campus would risk unpredictable interruptions to the game by pedestrians or vehicles, and would introduce exogenous factors such as inaccuracy of current localization sensors. Therefore, we used indoor equipment to create an experience analogous to playing outdoors. The experimental setup is shown in figure 2. A large projector screen is used to display a 3D outdoor scene from a first-person perspective. This represents the "physical" world. The player uses a stationary bicycle to control forward and backward movement in the "physical" world. Like the real physical world, this displayed environment does *not* show the animals. Instead, the position of the animals is indicated either using spatial audio or a top-down map screen. An Xbox 360 gamepad is used to turn the player left and right in the 3D park environment. The sound is provided through open-type supra-aural headphones.

4.1 Recruited participant groups

24 participants were recruited from the student population at Queen's University. Participants' ages ranged from 19 to 44, with a median age of 25. 12 "experts" and 12 "novices" were recruited. Experts had at least 25 hours of experience playing video games that use radar maps, or currently play such games at least 3 hours per week. Novices had fewer than 10 hours of lifetime experience with video games that use radar maps.

4.2 Conditions

We tested two conditions: an ambient audio interface and a visual interface. In the ambient audio condition, players hear animals through their headphones. In the visual condition, they see animals on an overhead map provided on a side display. In both conditions, players use the equipment shown in figure 2. Players pedal to navigate a representation of the physical world shown on a large screen. Players must chase and catch a series of animals. Three animals are used: cat, bird, dog, and then back to cat. The animals are not shown on the large screen, so players must locate them with the ambient audio or the side display map, depending on the condition.

Once a player has caught an animal, a "level up" tone sounds to indicate the catch, and the next animal appears in the game. Each successive animal moves faster than the previous one, progressively increasing the difficulty of the game.

The virtual world contains shrubs, trees and buildings, shown on the large display. If a player bumps into one of these obstacles, an "Ouch!" sound is played, and the screen blanks for 0.5 s.

Both versions of the game were instrumented to log the player's position, score, collisions and speed.

4.2.1 Ambient Audio Condition

The player hears the animal to catch through stereo headphones. The animal's position is spatialized in the 2D plane. The side display of figure 2 is not used in this condition.

The animal's position is spatialized assuming that players look forward at the large display. This removes the possibly confounding factor of sensor error in head tracking.

4.2.2 Visual Condition

This version of the game shows the animal's location on a small screen located to the player's right. The virtual world is presented as a top-down map, similar to the radar maps in games such as World of Warcraft or Halo. An example of this map is shown in figure 3.

Each type of animal is represented by a different icon, and the map shows significant features like buildings and the perimeter of the play area. The player is represented by an arrow in the centre of the screen that rotates to point in the player's direction of travel, while the map scrolls beneath this arrow, centered at the current location.

To mimic a hand-held device's screen for the visual condition, a 17" flat-panel monitor with a 4:3 aspect ratio was used. It was placed 138 cm away from the player to approximate the size of a smartphone's screen at a typical viewing distance. (Kato *et al.* report that users hold mobile phones on average 35 cm away from their eyes [5]. A 17" display at 138 cm from the user occupies the same visual angle as a 3.5" display at 35 cm.) The screen was located at eye-level, to the side of the main projector screen at an angle of 28 degrees to avoid obstruction of the projector screen.

4.3 Procedure

At the beginning of the trial, the participant completed the Witmer and Singer Immersive Tendencies questionnaire [18] and a custom questionnaire given to determine participant expertise with video games and associated equipment.

Next, participants tested each of the two conditions. The order of conditions was randomized so that each condition was performed first by half of the participants.

For the first condition, players were given a short explanation of the game rules, and a summary of either the sounds (in the audio condition) or graphics (in the visual condition) used to provide feedback in the game. Participants played for a five minute training period, during which they were free to ask questions to clarify the game rules and interface.

Next, the participant played the game again for five minutes. The participant was not permitted to ask questions of the experimenters. On completion, they completed the Witmer and Singer Presence questionnaire [18], and a custom questionnaire regarding their experience with the game.

This process was repeated for the second condition.

After completing both conditions, the participant completed a final custom questionnaire comparing the two interfaces. Finally, they completed a semi-structured interview discussing these preferences between the interfaces.

5. RESULTS

In this experiment there were three primary types of measures to examine and compare between the two conditions. We considered player performance, player presence, and player preference.

5.1 Data Analyses

Our experiment was a single factor (interface: visual or audio) design with one between-subjects factor (expertise: novice or expert). After determining that the order of presentation of interface had no effects on any of the dependent measures, we conducted a Repeated Measures Multivariate Analysis of Variance (MANOVA) on the data from the Presence questionnaire, the participants' scores, speeds, and number of collisions, with interface as a within-subjects factor and expertise as a between-subjects factor. α was set at .05.

To examine differences between the two conditions on the non-parametric questionnaire data, Wilcoxon signed-ranks tests for 2-related samples were used.

5.2 Player Performance

Player performance in the game was measured in terms of the number of animals caught (score), whereas player performance in terms of navigation was measured by the average speed of player (speed) and the number of collisions with environment obstacles (collisions). There were main effects of interface on the score ($F_{1,22}$=18.8, p≈.000, η^2=.46), the speed ($F_{1,22}$=14.1, p=.001, η^2=.39), and the number of collisions ($F_{1,22}$=28.9, p≈.000, η^2=.57). In all cases the values were higher when using the visual interface than when using the audio interface (see table 1 for the means and variances).

Although we observed a significantly greater frequency of collisions when using the visual interface than when using the audio interface, this could be an artifact of the fact that players generally progressed further in the game when using the visual interface, therefore encountering more challenging gameplay. On average players progressed through 9.2 levels (range:3-14) when using the audio interface and 11.9 levels (range:4-19) when using the visual interface. If we balance the number of levels for each player, (i.e., remove data for levels completed in one condition, but not the other by player), a pairwise t-test reveals that there are still significantly more collisions when using the visual (mean=.070, SD=.055) interface than when using the audio interface (mean=.030, SD=.037) (t23=3.9, p=.001).

There were main effects of expertise on score ($F_{1,22}$=9.10, p=.006, η^2=.29) and speed ($F_{1,22}$=12.5, p=.002, η^2=.36), but not on collisions ($F_{1,22}$=.01, p=.922, η^2≈.000). Experts achieved higher speeds and higher scores than novices (see table 2).

There were no significant interaction effects of expertise and interface on any of the performance-related dependent measures.

Table 1: Performance measures between two conditions

Measure	Audio		Visual	
	mean	SD	mean	SD
Score	8.33	3.09	11.25	3.94
Collisions	10.13	11.35	27.42	18.55
Speed (m/s)	5.34	1.55	6.32	2.07

Table 2: Performance measures between two groups

Measure	Novices		Experts	
	mean	SD	mean	SD
Score	8.13	3.41	11.46	1.74
Collisions	19.04	12.23	18.5	14.62
Speed (m/s)	4.81	1.38	6.83	1.42

5.3 Player Presence

Presence is represented as four measures: involvement, sensory fidelity, adaptation/immersion, and interface quality. There was a significant difference in the measures of sensory fidelity ($F_{1,22}$=12.3, p=.002, η^2=.36), adaptation/immersion ($F_{1,22}$=8.5, p=.008, η^2=.28), and interface quality ($F_{1,22}$=6.8, p=.016, η^2=.24) depending on interface. Players felt that the audio interface was more immersive and provided higher sensory fidelity, whereas the visual interface had a higher interface quality (see table 3 for means and variances). There was no effect of interface on involvement ($F_{1,22}$=3.5, p=.073, η^2=.14).

There were no significant effects of expertise on any of the presence-related measures, and no significant interactions of expertise and interface on the presence measures.

5.4 Player Preference

In questionnaires administered after using each interface, we asked players a number of questions related to their play experience (see table 4 for means and variances). Players reported that their experience using the audio interface was more fun than using the visual interface (Z=2.8, p=.005) and that they found it easier to avoid obstacles when using the audio interface (Z=3.5, p≈.000). In addition, players noted that the visual interface made it more difficult to play (Z=2.8, p=.010), and that they stopped pedalling more often when using the visual interface (Z=2.3, p=.023). All other questions did not result in significant differences.

By considering the significant results for experts separate from those for novices, we found that both experts and novices reported that it was easier to avoid obstacles using the audio interface (experts: Z=2.4, p=.016; novices: Z=2.6, p=.010). And both experts and novices reported that the audio interface was more fun, although only marginally so for experts (experts: Z=1.9, p=.058; novices: Z=2.0, p=.041). Only novices reported that it was harder to play using the visual interface (Z=2.3, p=.022); experts did not report this difference (Z=1.3, p=.204). The same trend was true for results of stopping pedalling more often using the visual interface, only marginally so (experts: Z=1.4, p=.156; novices: Z=1.8, p=.064).

Table 3: Presence measures between two conditions

Measure	Audio		Visual	
	mean	SD	mean	SD
Involvement (/84)	58.25	12.13	55.42	10.74
Sensory Fidelity (/42)	27.38	4.79	22.71	7.78
Adaptation/immersion (/56)	44.92	7.34	40.54	7.55
Interface quality (/21)	6.42	2.67	7.33	2.76

6. DISCUSSION

The results of the experiment are consistent with our hypotheses: players received better scores in the visual condition, but felt a stronger sense of presence and fewer collisions in the audio condition. We now discuss these results.

6.1 Performance of ambient audio

Players' performance was worse using the ambient audio interface, as indicated by lower scores and lower pedaling speed. Participants' comments in interviews indicate that a primary cause of this is that spatialized audio is less accurate than a visual map for locating animals. One participant states "In the map it was easier to track the animals and to anticipate where they're going to. The current location was easy to find in the auditory game, but where they were going wasn't, and so it was more a game of following, so it became harder to catch them in the auditory game." Another says "[The map has] easier information to rely on than the audio, particularly when you're close to the animal and it switches from right to left and it keeps moving around, you're spinning around in circles looking for something you can't see."

An interesting secondary effect is that players reported feeling more stress to perform in the visual condition. One participant reports: "I think that with the visual map with the animals, there was kind of pressure on me, that I have to catch these animals." Other participants stated " If I knew on the visual task that the animal is across the map, that I have to go all the way over there and run really fast, it made that more stressful." and "When I could see how fast the animal was going... it was like, 'well now I have to pedal faster' ". This increase in stress and sense of urgency may also have contributed to the higher pedaling speed observed, and hence higher score.

6.2 Safety

We consider number of collisions with obstacles as an indicator of the safety of the interface. The results indicate that the number of collisions was significantly higher in the visual condition. This is consistent with our hypothesis that when players shift their attention from the (vitrual) world to a secondary display, they have more difficulty seeing obstacles, and therefore are more likely to hit them.

Feedback from participants is consistent with this hypothesis. One participant says of the audio condition, "I didn't have to keep jumping my eyes off the map to see where I was, and with how tightly [spaced] the some of the obstacles were, it was easier for me to just focus on avoiding obstacles than jumping back and forth." Another said, "I could focus all my visual attention on the main screen, so I knew where I was going all the time. The map on the side was a lot more distracting."

Table 4: Player preferences between two conditions (7-point Likert scale)

Question	Audio		Visual	
	mean	SD	mean	SD
When an animal was far away, I could easily tell which direction it was in	5.00	1.69	5.63	1.56
I often had to stop pedalling when (listening to the ambient audio) / (looking at the map)	4.13	2.09	5.21	1.44
I found it easy to avoid the obstacles in the world while locating the animals	4.58	1.77	2.79	1.28
I felt rushed during the game	3.13	1.75	3.54	1.84
It was hard to use the (ambient audio) / (map) to locate the animals	2.83	1.55	2.75	1.62
I thought that the game was fun to play	5.67	1.40	5.00	1.62
When I was close to an animal, I had no trouble finding it	4.17	1.52	4.75	1.45
There wasn't enough time to finish the game	2.25	1.22	2.46	1.56
I thought that the game was hard to play	2.79	1.47	3.79	1.50

As a secondary effect, it is possible that the additional stress that players reported in the visual condition (as discussed in section 6.1) may have resulted in higher pedalling speed, which could have contributed to increased collision rate.

6.3 Immersion

Players reported higher sense of presence in the audio condition. We hypothesized that this was due to the audio being overlaid directly onto the world, whereas in the visual condition, the world view and map view are not directly connected. Players comments lend weight to this hypothesis. One participant said: "In the map version, I was tearing myself away from the real world screen to look at the map... it was more engaging to look at the main screen and think, 'Hey I could actually be there', as opposed to thinking 'I need to find out where I am so I'll look at this other screen'." Another said that the audio version "involved me more personally in the environment".

Interestingly, several players reported that immersion was reduced in the visual condition by the absence of any sound at all. For example, "hunting around looking for silent animals doesn't seem as realistic", and "If there was some random rustling of leaves in the background that was not related to the animals but just happened to be there, I probably would have found it more immersive." This implies that even when the ambient audio is not required as the main navigational mechanism, it could be included regardless to contribute to players' sense of presence in the virtual world.

6.4 Expertise

We hypothesized that prior expertise with games that use radar views could counteract the reduced immersion and safety in the visual condition. In fact, our results showed both novices and experts experienced higher sense of presence with the audio version, and both found it easier to avoid obstacles in the audio version. This is a positive result, indicating that the immersion and safety benefits of ambient audio carry over to gamers who are expert with current interface styles.

Some differences were seen. Both novices and experts rated the audio condition as more fun, but novices more strongly. Perhaps most interestingly, novices found the visual condition harder to play, but experts did not. These findings indicate that experts may be more comfortable with the limitations of using a secondary display.

6.5 Limitations of Study

Our study considered only one game, the simplified version of Growl Patrol. We argue that since the game consisted solely of the task of locating and chasing a virtual animal, the results are likely generalizable to navigational tasks in any location-based game. Further studies with different games would be beneficial in confirming this claim of generality.

By carrying out the study indoors, we gained several benefits: we were able to monitor players and collect detailed data more easily, we removed the problem of accurately localizing players' positions and head orientations, and we removed the possibility of external interruptions to the game. The use of this controlled environment provided us with considerably better confidence in our results than would have been possible with an unconstrained exterior environment.

Our experience is not limited to the indoor version of the game, however. We have implemented Growl Patrol for play outside (as was shown in figure 1). Anecdotally, we have found the results to be similar to that observed in our formal study; however, problems with locating animals in the virtual condition were exacerbated by inaccuracies in the positional sensors. The GPS used to monitor position is updated only once per second, leading to appreciable error if the player is moving quickly. A compass and gyroscope were used to monitor head-orientation. The compass is inaccurate when the player is moving, and the gyroscope is inaccurate when the player rotates their head quickly. Sensor hardware is continually improving, but hardware limitations must be considered when implementing ambient audio today.

An important limitation of ambient audio not considered in this paper is that there is a limit to how many audio streams players can track. Further research is required to establish how many audio streams can be effectively distinguished. Anecdotally, we have found the number to be three or four with current sound spatialization technology. This implies that visual interfaces can convey more information than ambient audio interfaces. For example, in our study, the secondary visual display showed the locations of buildings and the extent of the play area, neither of which were available to players using the audio interface. Our results show that ambient audio is useful for navigation-style tasks and for maintaining situational awareness while mobile. However, some forms of information (such as the full details of the map) must necessarily be visual in form.

Ambient audio is most useful when the player is moving. For parts of the game where players are stationary, the full richness of a visual display can be used. For example, the PiNiZoRo game [14] combines a navigational task (which could be supported by ambient audio), and a set of visual mini-games that are played while stationary.

Our core lessons for game designers are therefore that ambient audio games should not require overly precise navigation, should not require players to track large numbers of information sources, and should not require mobile players to access information that is not easily representable in ambient audio form. Despite these restrictions, ambient audio is directly applicable to any location-based game that relies on navigational tasks, and can help increase players sense of presence in the virtual world, while improving their safety.

7. CONCLUSIONS

Location-based games played in an exterior environment have become increasingly popular, particularly with the advent of smartphones that contain the necessary sensors allowing their play. Location-based games on such devices have two problems, however: requiring players to look at a screen while moving in the physical world may result in accidents, and presenting the use of a hand-held screen provides a disconnect between the virtual and physical worlds.

In this paper, we have evaluated the use of ambient audio to address these two problems. To illustrate the concept of ambient audio, we have presented *Growl Patrol*, a location-based game allowing players to catch virtual animals in the physical world. To show the effectiveness of ambient audio, we have reported on a study comparing an ambient game interface to a visual interface. The study showed that ambient audio improved players' sense of presence in the virtual world, and increased players' safety (as measured by number of obstacles they bumped into.) Not suprisingly, players' game performance (as measured by score) was lower in the audio condition. These findings held for both novice and expert players. From these results, we conclude that ambient audio is a promising interface style for games involving navigation in the physical world.

8. ACKNOWLEDGEMENTS

We would like to thank Jonathan Segel of Alcatel-Lucent for his support and many ideas related to this project. This project was supported by the GRAND Network of Centres of Excellence, and by NSERC Strategic Project Grant 365040.

9. REFERENCES

[1] D. R. Begault. Challenges to the successful implementation of 3-D sound. In *Audio Engineering Society Convention 89*, pages 864–870, 1990.

[2] A. D. Cheok, K. H. Goh, W. Liu, F. Farbiz, S. W. Fong, S. L. Teo, Y. Li, and X. Yang. Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. *Personal Ubiquitous Comput.*, 8:71–81, May 2004.

[3] I. Ekman, L. Ermi, J. Lahti, J. Nummela, P. Lankoski, and F. Mäyrä. Designing sound for a pervasive mobile game. In *ACE*, pages 110–116. ACM, 2005.

[4] S. Holland, D. R. Morse, and H. Gedenryd. Audiogps: Spatial audio navigation with a minimal attention interface. *Personal and Ubiquitous Computing*, 6:253–259, 2002.

[5] S. Kato, C. S. Boon, A. Fujibayashi, S. Hangai, and T. Hamamoto. Perceptual quality of motion of video sequences on mobile terminals. In *SIP*, pages 442–447, 2005.

[6] J. Kurczak and T. N. Graham. TREC: platform-neutral input for mobile augmented reality applications. In *EICS*, pages 283–288. ACM, 2011.

[7] C. Magerkurth, A. D. Cheok, R. L. Mandryk, and T. Nilsen. Pervasive games: bringing computer entertainment back to the real world. *Comput. Entertain.*, 3:4–4, July 2005.

[8] N. Paterson, K. Naliuka, S. K. Jensen, T. Carrigy, M. Haahr, and F. Conway. Design, implementation and evaluation of audio for a location aware augmented reality game. In *Fun and Games*, pages 149–156. ACM, 2010.

[9] M. Peden. *A Thousand Shall Fall: The true story of a Canadian bomber pilot in World War Two*, pages 157–159. Stoddart, 1988.

[10] R. Pellerin, N. Bouillot, T. Pietkiewicz, M. Wozniewski, Z. Settel, G.-S. E., and J. R. Cooperstock. Soundpark: Exploring ubiquitous computing through a mixed reality multi-player game experiment. *Studia Informatica Universalis*, 8(3):21 pages, 2010.

[11] A. Pirhonen, S. Brewster, and C. Holguin. Gestural and audio metaphors as a means of control for mobile devices. In *CHI*, pages 291–298. ACM, 2002.

[12] J. Sodnik, C. Dicke, S. Tomazic, and M. Billinghurst. A user study of auditory versus visual interfaces for use while driving. *International Journal of Human-Computer Studies*, 66(5):318 – 332, 2008.

[13] C. Stahl. The roaring navigator: a group guide for the zoo with shared auditory landmark display. In *MobileHCI*, pages 383–386. ACM, 2007.

[14] K. G. Stanley, I. Livingston, A. Bandurka, R. Kapiszka, and R. L. Mandryk. PiNiZoRo: a GPS-based exercise game for families. In *FuturePlay*, pages 243–246. ACM, 2010.

[15] Y. Takeuchi and M. Sugimoto. A user-adaptive city guide system with an unobtrusive navigation interface. *Personal and Ubiquitous Computing*, 13:119–132, 2009.

[16] A. Walker and S. Brewster. Spatial audio in small screen device displays. *Personal and Ubiquitous Computing*, 4:144–154, 2000.

[17] J. Walker, E. Alicandri, C. Sedney, and K. Roberts. In-vehicle navigation devices: Effects on the safety of driver performance. In *Vehicle Navigation and Information Systems*, pages 499 – 525, 1991.

[18] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoper. Virtual Environ.*, 7:225–240, June 1998.

[19] K. Xu, S. J. D. Prince, A. D. Cheok, Y. Qiu, and K. G. Kumar. Visual registration for unprepared augmented reality environments. *Personal Ubiquitous Comput.*, 7:287–298, October 2003.