Scaffolding Player Location Awareness through Audio Cues in First-Person Shooters

Colby Johanson  
University of Saskatchewan  
Saskatoon, Canada  
colby.johanson@usask.ca

Regan L. Mandryk  
University of Saskatchewan  
Saskatoon, Canada  
regan.mandryk@usask.ca

ABSTRACT
Digital games require players to learn various skills, which is often accomplished through play itself. In multiplayer games, novices can feel overwhelmed if competing against better players, and can fail to improve, which may lead to unsatisfying play and missed social play opportunities. To help novices learn the requisite skills, we first determined how experts accomplish an important task in multiplayer FPS games – locating their opponent. After determining that an understanding of audio cues and how to leverage them was critical, we designed and evaluated two systems for introducing this skill of locating opponents through audio cues – a training system, and a modified game interface. We found that both systems improved accuracy and confidence, but that the training system led to more audio cues being recognized. Our work may help people of disparate skill play together, by scaffolding novices to learn and use a strategy commonly employed by experts.

Author Keywords
First-person shooter; scaffolding; training; audio cues; games; expertise development; novices

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI)

INTRODUCTION
Digital games are complex, requiring mastery of a diverse set of skills [33] in order for a player to become proficient; and this process of learning to master the skills necessary to overcome in-game challenges is also part of what makes games fun to play [19, 29, 37]. Although games sometimes directly instruct players on how to acquire and hone their skills, it is more common that players are left to experientially acquire and refine these skills through play itself [27, 41]. This experiential learning works best when the game designer has full control over the challenges a player encounters – for example, by revealing new skills in an optimal order, controlling the pacing of difficulty and rewards, or adapting the game to meet the skill of the player through dynamic difficulty adjustments [25, 34].

In multiplayer games, however, designers don’t have the luxury of control over the game’s challenge; the difficulty is primarily determined by the skill of other players [34]. If the skills of the player are poorly matched, experiential learning is ineffective [27]. This is in addition to an unsatisfying play experience: i.e., one that lacks suspense [1], is frustrating [22], and results in the weaker player feeling overwhelmed or anxious [9]. Although choosing to play with people of the same skill level would solve these problems, there are situations in which people with diverse skill choose to play together – for example, when parents play with their children or friends play socially, one person may be considerably better at the game than the other.

Many systems have attempted to address in-game skill discrepancy, including asymmetric roles in team-based games – which allows dissimilarly skilled players to play on the same team [47], match making – which allows similarly skilled strangers to play together [34], and dynamic skill assistance – which allows dissimilarly skilled players to play together, with weaker players receiving assistance from the system [7, 8, 20, 46, 47]. While these systems can help to minimize the problems of skill discrepancies in certain situations, none actually address the underlying issue – the weaker player hasn’t learned the skills required to succeed in a multiplayer game. We therefore need to provide scaffolding to facilitate and accelerate skill development in the skills needed to become a better player.

In this paper, we focus on helping players acquire the skill of being aware of an opponent’s location for improving performance in multiplayer first-person shooter (FPS) games. Opponent location awareness is an important skill because it allows a player to know where to look to find their opponent, which permits them to begin the aiming process earlier or to make a decision to avoid the opponent until they are adequately prepared – a player with poor awareness risks being targeted by opponents unexpectedly.

Hypothesizing that experts are skilled at maintaining awareness of opponents outside of their immediate field of view, our first goal was to determine the techniques or heuristics experts use to achieve opponent location awareness. Our second goal was to apply this knowledge to design and evaluate systems for instructing novices on this skill.
To determine how experts maintain awareness of opponent locations under varying conditions, we first launched an online experiment in which expert FPS players were presented with videos of in-game scenarios. Participants were asked to provide a location estimate of their opponent’s in-game location at the end of the scenario, listing the audiovisual information they used to inform their estimate. We confirmed that experts were generally able to locate the opponent in the scenarios, whereas the novices were not. The greatest differences occurred in the scenarios in which auditory information was required to locate the opponent. In these scenarios, the novices struggled to make sense of what they were hearing or recognize which sounds were important. The experts, on the other hand, had no trouble with this. They were able to directly connect the sounds they were hearing to locations in the game’s environment, and commented on the importance of audio cues to succeed at the task. With this in mind, we focused our efforts on scaffolding an understanding of audio cues as the method for maintaining awareness of unseen opponents.

We designed two approaches for scaffolding audio cue interpretation – a training-based approach in which novices learn to recognize cues through a standalone game, and an interface-based approach in which iconic representations of the sounds were shown on the game’s interface, known as the heads-up display (HUD). In a second online experiment, we evaluated both approaches and found that both improved the novice’s confidence and accuracy, but that those who underwent the training-based approach – instead of the HUD icons approach – were also able to identify more audio cues, and were slightly more accurate.

Scaffolding expertise development in the skills for success can help novices avoid the frustration of experiential learning in a setting in which it is ineffective, i.e., when the skill of the opponent is vastly superior. Our work takes a step towards helping people of disparate skill play together in the long-term, by scaffolding novices to learn and use a strategy commonly employed by experts. Scaffolding skill development can open up social play opportunities for novices who would not otherwise be able to play with their more expert friends. Enabling opportunities for social play may allow people to use games as a way to enact their friendship (building closeness through play-based activity) [20], help people feel more connected to one another [40, 49], and satisfy players’ need for social connection [12].

**RELATED LITERATURE**

We present concepts of learning and skill development as applied to games and interfaces, followed by an overview of skills general to games and skills specific to FPS games. Finally, we look at current options for learning FPS skills.

**Learning and Skill Development**

A particularly influential model of skill development which has been applied in both user interface design [10, 23] and games [36] is Fitts and Posner’s [16] three stages of development – consisting of the cognitive, associative, and autonomous stages. During the cognitive phase, the user begins forming conceptions about the task and discovers what exactly needs to be done. It is slow, attention-demanding, and other tasks can interrupt its execution [38]. Learners may consider many strategies before settling on one [10]. The associative phase is where the user pays more attention to how to perform the task and begins to improve in executing the task. After extensive practice, the user has reached the autonomous phase where they are able to perform the task automatically. For complex tasks requiring unobvious strategies, the practice is deliberate and just beyond a person’s competence level [2, 14].

Another important model of skill development is experiential learning [28]. This model is based on the idea that learning is a continual process “whereby knowledge is created through the transformation of experience” [28], or more simply, that one’s experiences impact learning. Kolb argues that learning occurs primarily when an outcome does not match one’s expectations – the learner must re-form their ideas to not violate expectations.

**Learning in Games**

Games are considered to be exceptional learning tools; they take complete novices and train them to perform long, complex, and difficult tasks, while being fun [18, 19, 29]. Many have attempted to explain how this is accomplished.

Kiili [27] proposed a model for game-based learning environments based on experiential learning theory, flow theory, and game design. Kiili argues that learning in games primarily occurs in the flow channel, as otherwise anxiety or boredom [9, 27] are present. Therefore, to encourage experiential learning, a game’s challenges should be matched to the player’s current skill level, increasing as they improve. Kiili’s experiential gaming model contains two main phases: the ideation phase and the experience phase. The ideation phase is when a player generates potential solutions and ideas. These ideas improve as the player begins to consider the constraints of the system. In the experience phase, the player will test and observe the outcomes of the solutions generated during ideation. After testing, the player will begin “reflective observation” [27, 40] and return to the ideation phase to devise a new idea.

Gee [18, 19] has explored how good games promote learning. For example, he argues that they are good at promoting problem solving and understanding. Gee provides a number of examples and many of these examples support the ideas presented by Kiili’s model or Fitts and Posner’s three stages of development. For example, by presenting players with problems early on that encourage a way of thinking to allow players to form better guesses when they face harder problems (assisting with the ideation phase), by presenting problems that are just beyond a player’s comfort level (reaching mastery and staying in a state of flow), and by extended practice of skills until they are nearly automatic, followed by tests of mastery (autonomous stage reached through deliberate practice).
Additionally, both Gee and Kiili stress the importance of clear goals and feedback in order to promote learning.

An alternative approach to experiential learning is direct instruction, which can take the form of tutorials [34, 41]. Andersen et al. [6] tested the impact of tutorials on games of differing complexity. They found that providing tutorials for complex games increased play time significantly. Simple games, however, were not impacted by the presence of a tutorial; the mechanics could be discovered easily through experimentation and so a separate tutorial was not helpful. Therefore, considering Kiili’s model of experiential learning [27], learning is more effective in games which have been designed to allow a player to discover game mechanics on their own, presumably while maintaining the maximum amount of time in the flow channel [9].

**In-Game Skills**

In developing a scale to evaluate games, Norman [33] identified six important in-game abilities. Three of these are relevant to FPS games. Perceptual-motor abilities include perceptual speed, pattern recognition, object identification, simple and choice reaction time, tracking, targeting, timing, rhythm, and response mapping. Cognitive processing abilities involve processing or interpreting information communicated through text and speech and hidden object detection. Information utilization abilities involve managing in-game resources, using information from memory, and filtering out irrelevant information. We list the relevant factors as we examine FPS skills.

**In First-Person Shooters**

The defining skills of the FPS genre are those of moving and shooting (both perceptual-motor abilities). Specifically, of navigating a 3D environment while simultaneously pointing at objects in that environment [3, 48].

Another way of identifying the skills required to play FPS games is to look at the ways in which players are cheating. The most common techniques involve the use of “aimbots” or “wall hacks” [31, 50]. Aimbots help the player point at opponents (perceptual-motor), often by locking the targeting reticle onto the nearest opponent [31], whereas a wall hack helps the player locate opponents (cognitive processing). The simplest form of a wall hack allows a player to see their opponents through walls [50] and advanced versions reveal further information on the status of the opponents. The role of aiming and assistance to aiming in multi-player FPS matches has been explored extensively by Vicenio-Moreira et al. [46, 47], who discovered that aim assists alone are insufficient to balance FPS games in a competitive multi-player setting. They combined aim assists (i.e., a minimal aimbot) with differential damage dealing and location awareness assistance (i.e., a minimal wall hack) to balance game outcome, even between novice-expert pairings [46].

Experts have managed to learn the skill of locating opponents, either experientially or from an online guide. This is done through recognizing a sound (information utilization), then estimating where the sound is coming from (cognitive processing) [17]. The use of informative sound design in games has been studied previously. For example, Jørgensen [26] argues that sound can be used as an information system that allows players to pick up more data than in the visual system, and Stockburger [42] describes how players can make a decision to seek out the source of “acousmatic” (unseen) sounds.

Using sound effectively is just one instance of how a player can make use of available information to gain an advantage over their opponents. Two further examples, specifically from the *Quake* series, are item timing and threat assessment. Item timing is a technique used by experts to estimate when an important item will be available again (‘respawning’ in FPS terminology) by paying attention to the time an item was taken [17, 32] (cognitive processing and information utilization). Threat assessment is used by experts to determine whether to flee from or fight an opponent. Conroy et al. [11] found that when players were given all of the available information about an opponent’s status compared to their own (including health, current weapon, and distance away), they agreed on the reaction to the situation and its threat (information utilization). However, the actual in-game responses matched the agreed-upon reactions only when experts were playing.

**Learning First-Person Shooters**

Some FPS games – particularly single-player ones – include tutorials that can be played separately. Two representative examples include *Half-Life*’s (1998) classic hazard course and *Counter-Strike: Global Offensive*’s (2012) weapon course. Both of these examples are designed to teach the player the basics of the game’s systems, including moving, shooting, and interacting with objects in the environment. Neither of these tutorials go further than this, and instead expect the player to play the game to experientially gain the rest of the requisite skills. If a novice attempted to play a competitive multi-player game after undergoing only this minimal training, they may be overwhelmed [9].

The choices for training beyond what current tutorials provide are currently limited to online guides and tools designed to improve one’s reaction time and aim [4, 5]. Another – now unavailable – option was “FPS Trainer” [43, 44], which attempted to teach players about “timing, positioning, and appropriate weapon selection” [43] through real-time in-game feedback systems. Unfortunately, the efficacy of this tool is unknown.

We next describe our studies that discover how experts locate opponents and scaffold novices to do the same.

**STUDY 1: DETERMINING EXPERT BEHAVIOR**

Our first goal was to understand how experts locate their opponents in FPS games and whether their approaches vary depending the circumstance. From this, we can develop tools for scaffolding development of these skills.
To accomplish this, we conducted an online experiment presenting different in-game scenarios to FPS players of varying skill levels. The scenarios were presented as short video clips in which you watch from the perspective of one of the two in-game players (in a resolution of 1280x720 and a constant bit rate of 4000 kbps). For each scenario, we asked participants to indicate where they believed their opponent’s location was. All scenarios were from the FPS game *Quake Live*, which was chosen due to the variety of possible in-game scenarios as well as its exaggerated audiovisual effects.

**Study Procedure**

After providing informed consent, participants filled out a background questionnaire asking about their gaming history, frequency, and expertise with FPS games – both in general and with *Quake Live or Quake 3 Arena*. Expertise levels were self-rated, and participants chose from novice, experienced, or expert. Participants then were given instructions describing the elements of the game’s heads-up display (HUD) and how to read the 2D top-down map.

The video scenarios were then presented. For each scenario, participants watched the short (between 8 and 14 seconds) video clip and answered questions about the other player’s location at the end of the video. They were asked to describe in words where they thought the other player was, and click that location on a 2D top-down representation of the game’s map. For both location estimates, they were asked to give a confidence rating between 1 (meaning “I don’t know where the other player is”) and 5 (“I know where the other player is”). They also had the option of selecting “I don’t know,” which allowed them to not answer and automatically assigned the lowest confidence rating. Participants were asked to list the cues used to determine the opponent’s location – they completed statements prefixed with “I saw”, “I heard”, or “I noticed” from a drop down menu and an accompanying text field. Listed items were automatically made available for later scenarios.

At the end of the experiment, after completing every scenario, we gave participants the opportunity to provide comments and to list the cues they believed were important. We also allowed participants to review their performance, by visualizing their selections and the correct opponent locations on maps and giving tips on which cues they could have used to come to the solution. We included this post-task review process to encourage participation, promoting the experiment as a way in which participants could test – and learn to improve – their opponent locating skills.

As we were interested in determining differences between experts and novices, we recruited participants through a variety of online sources, including announcements on gaming forums, the bulletin board at the University of Saskatchewan, and social media. In total, 92 people completed the study; 8 were excluded from the analysis for not watching all the videos to completion, leaving 84 participants (77 male, 7 female; all over the age of 18; the majority aged between 18 and 25). For *Quake Live* expertise, 38 (45.2%) self-identified as novice, 30 (35.7%) as experienced, and 16 (19%) as expert.

**Scenarios**

We presented a variety of carefully chosen scenarios to participants. These scenarios were chosen to be sufficiently different from each other so that we could see what strategies expert players used under various circumstances.

The introductory scenarios were straightforward to act as a baseline and to allow participants to familiarize themselves with the process. For these, the opponent’s location was revealed before the end of the video. The scenarios that followed were chosen to emphasize different approaches to locating the opponent. In three scenarios, the participant was expected to listen and understand audio cues, and in two they were expected to see and interpret visual cues. Two scenarios are described below in order to illustrate the differences between auditory and visual scenarios.

**Auditory-Dependent Scenario:** The point of view (POV) player walked towards a jump pad and looked up to check for a trap. The opponent is then heard jumping and picking up five armour shard items, indicating their location.

**Visually-Dependent Scenario:** The POV player encountered the opponent in a hallway. Damage was exchanged, but the POV player dealt more damage than the opponent. The opponent then retreated through a doorway and fired a few defensive rockets towards the POV player. The nearby yellow armour item will respawn and be available shortly. These cues indicate that the other player is likely heading to the yellow armour.

**Data Analyses**

In our analyses, we grouped the scenarios by the type of cue that was primarily responsible for understanding the opponent’s location, resulting in auditory scenarios, visual scenarios, and introductory scenarios.

To determine the differences in accuracy between expertise groups in the visually- and auditory-dependent scenarios, we inspect heatmaps of the location responses. The heatmaps are based on kernel-density estimation, as described by [30]. We also examined the written location responses, looking for patterns or similar responses which might indicate a common strategy for locating opponents. We also present the following metrics:

**Guess Counts:** Count of whether or not participants chose to make a location guess for each scenario in the group, represented as a percentage of the total number of possible guesses, to accommodate different expertise group sizes.

**Confidence Rating:** Participant confidence in their guesses collected using a 5-pt scale (5 is more confident).

**Cue Counts:** Count of identified cues prefixed with “I saw”, or “I heard”, represented as an average of counts over the scenarios in the group.
Table 1. Descriptive statistic results by scenario group and expertise. Numbers in brackets indicate standard deviation.

<table>
<thead>
<tr>
<th>Scenario Group</th>
<th>Guesses</th>
<th>Confidence</th>
<th>I Saw</th>
<th>I Heard</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intro.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>1.00 (.00)</td>
<td>4.97 (.13)</td>
<td>1.0 (.22)</td>
<td>1.0 (.73)</td>
</tr>
<tr>
<td>Experienced</td>
<td>0.95 (.20)</td>
<td>4.95 (.20)</td>
<td>0.9 (.29)</td>
<td>1.1 (.65)</td>
</tr>
<tr>
<td>Novice</td>
<td>0.95 (.23)</td>
<td>4.41 (.71)</td>
<td>1.0 (.44)</td>
<td>0.6 (.49)</td>
</tr>
<tr>
<td><strong>Vis.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>1.00 (.00)</td>
<td>4.16 (.63)</td>
<td>0.9 (.56)</td>
<td>0.6 (.47)</td>
</tr>
<tr>
<td>Experienced</td>
<td>0.92 (.23)</td>
<td>4.05 (.63)</td>
<td>0.7 (.49)</td>
<td>0.8 (.66)</td>
</tr>
<tr>
<td>Novice</td>
<td>0.89 (.29)</td>
<td>4.13 (.73)</td>
<td>1.0 (.51)</td>
<td>0.5 (.58)</td>
</tr>
<tr>
<td><strong>Aud.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert</td>
<td>1.00 (.00)</td>
<td>4.92 (.19)</td>
<td>0.2 (.21)</td>
<td>2.4 (.75)</td>
</tr>
<tr>
<td>Experienced</td>
<td>0.94 (.20)</td>
<td>4.61 (.66)</td>
<td>0.2 (.17)</td>
<td>2.3 (.75)</td>
</tr>
<tr>
<td>Novice</td>
<td>0.68 (.39)</td>
<td>2.89 (.92)</td>
<td>0.1 (.17)</td>
<td>1.4 (.77)</td>
</tr>
</tbody>
</table>

To compare the performance of the expertise groups, we conducted a one-way analysis of variance (ANOVA) on confidence ratings for each scenario type separately, using pairwise comparisons (least significant different) to examine differences between the expertise groups. The guess and cue counts, which do not meet the assumptions of ANOVA, we used the nonparametric Kruscal-Wallis test, with the Mann-Whitney U for pairwise comparison [15].

**Scenario Results**

Table 1 shows the descriptive statistics. We group the results by scenario type.

**Introductory Scenarios**

Within the introductory scenarios, expertise did not yield a significant difference in guess counts ($\chi^2=0.71, p=.700$).

There was a difference in the confidence of guesses based on expertise ($F_{2,81}=3.7, p=.029$). Post-hoc tests reveal that novices were less confident than experienced ($p=.035$) or expert ($p=.022$) players; there was no observed difference between experienced and expert players ($p=.580$).

There was a difference in the number of audio cues used, depending on expertise ($\chi^2=9.52, p=.009$); novices used fewer audio cues than experienced ($p=.003$) and marginally fewer than expert ($p=.055$) players. There was no observed difference between experienced and expert players ($p=.630$). There was no observed difference in the number of visual cues used by expertise ($\chi^2=1.18, p=.553$).

**Visually-Dependent Scenarios**

In the visual scenarios, expertise did not yield a significant difference in the number of guesses ($\chi^2=3.78, p=.151$), or in the confidence of the guesses ($F_{2,81}=0.55, p=.579$).

There was no significant difference in the number of audio cues used, based on expertise ($\chi^2=3.7, p=.156$). There was, however, a difference in the number of visual cues used based on expertise ($\chi^2=7.275, p=.026$), in which novices listed more visual cues than experienced ($p=.022$), but not expert ($p=.99$) players; there was no observed difference between expert and experienced players ($p=.457$).

The heatmaps (Figure 1) show a difference in the pattern of the location guesses, in which novices tended to underestimate the opponent’s distance, though they understood the correct direction.

For the written responses of the previously described example scenario with the respawning yellow armour, many expert (75.0%) and experienced (60.0%) players mentioned how the opponent is heading towards the yellow armour, while few (5.3%) novices did.

**Auditory-Dependent Scenarios**

Within the three auditory scenarios, there was a significant difference in the number of guesses based on expertise ($\chi^2=1.18, p=.553$); novices made fewer guesses than experts ($p=.003$), but not experienced players ($p=.139$). There was no observed difference between experienced and expert players ($p=.083$).

There was a difference in the confidence of guesses based on expertise ($F_{2,81}=41.8, p<.001$); novices were less confident than both experienced ($p<.001$) and expert ($p<.001$) players. There was no observed difference between expert and experienced players ($p=.255$).

There was a difference in the number of audio cues used, depending on expertise ($\chi^2=23.3, p<.001$); novices used fewer audio cues than experienced ($p<.001$) or expert ($p<.001$) players. There was no observed difference between experienced and expert players ($p=.623$). There was no significant difference in the number of visual cues used ($\chi^2=1.55, p=.460$).

There were also differences in the pattern of location guesses, as seen in the heatmap (Figure 1). In the scenario shown, novices failed to connect the sound of the teleporter to the opponent’s location, and instead guessed that the opponent was in the water. Expert and experienced players had no issue identifying the teleporter exit location and specifying that the opponent was there.

Written responses from experts tended to describe the other player’s location based on the origin of the last sound they heard. For example, in the auditory scenario described earlier in which the opponent is near the armour shards, the majority of expert (93.8%) and experienced (80.0%) players described the other player’s location as being near the armour shards, whereas novices (10.5%) tended not to. The same pattern exists for the other two scenarios in this group and overall, expert (91.7%) and experienced (75.6%) players mostly made note of the revealing cue to determine the opponent’s location – describing the opponent as being...
near the rail gun, teleporter exit, or amour shards due to the associated audio cue they heard – whereas novices (5.3%) did not. The heatmaps (Figure 1) confirm how this approach allowed participants to locate the opponent.

**Importance of the Cues**

In a final survey, participants were asked, “of the cues (sounds, visual information, etc.) you noted, which ones do you think were the most important?” Many participant responses described how unique audio cues can reveal a player’s location or ranked the importance of the various cues. Those who ranked the cues tended to indicate that visual cues were the most important. Visual cues indicate that an opponent is nearby and could damage the player, making visual cues both salient and dangerous. While audio cues may be perceived as less important, when they reveal the opponent’s location, the player can make the choice to engage their opponent in combat or avoid them. Stockburger describes this choice as being able to “actively visualize or acousmatise” [42] the source of the sound.

**Summary and Discussion of Study One**

From the results, it is apparent that the greatest differences in the responses were during the auditory scenarios, with novices noting fewer cues and making fewer guesses while feeling less confident and giving less accurate guesses.

The notable difference between the auditory scenarios and both the introductory and visually-dependent scenarios, aside from accuracy, is the relatively higher confidence. Novices were just as confident as the expert or experienced players in the visually-dependent scenarios, suggesting that novices are more confident in utilizing visual information than auditory information.

That expert and experienced players often described the other player’s location as being near a landmark (such as an item pickup) is noteworthy, because it indicates that they were using a common strategy when locating opponents. When a player touches or interacts with these landmarks, an audio cue can be heard. This allows one’s position to be pinpointed. Additionally, these landmarks also have strategic significance [17, 32], meaning players are more likely to be moving to or from them. These factors make these landmarks critical, and as a result, both expert and experienced players have learned to tell when their opponents are near them.

With the knowledge that novices lacked confidence and performed worse in the auditory scenarios than in the introductory or visual scenarios, as well as the importance of the audio cues associated with landmarks, we set out to formulate ways of assisting novices to leverage audio cues.

**STUDY 2: HELPING NOVICES LEVERAGE AUDIO CUES**

We designed and evaluated two scaffolding systems to help novices use audio cues for location awareness of opponents.

**System Designs**

Our two systems used different approaches: a training-based approach, and an interface-based approach.

**Training-Based Approach**

The task of learning sounds in a game is comparable to the task of learning a language’s vocabulary – in both cases you are learning to associate words and sounds with objects. This is a well-recognized real-world problem for which there are a variety of solutions. One of these solutions is Duolingo, a successful [45] online app developed with gameful design principles [13] to “… give everyone access to a personalized education in a scalable way” [35]. The design of Duolingo heavily inspired the design of our system, in which we wanted users to learn to listen to the game, identify the sounds they were hearing, and apply this knowledge as they play.

We separated the audio cues into sets. There was one set for movement sounds, one for weapon-related sounds, one for health-related sounds, one for armour-related sounds, and another with the leftover miscellaneous sounds. Participants had to learn a set of sounds before moving on to the next set. To reinforce learned cues, we included cues from previous sets in the later sets. Every audio cue had an associated icon and description.

Before we trained a user on a particular set of audio cues, we provided an explanation for why the cues were important. An example of one of our explanations was: “As a player moves through the environment, they generate sounds depending on what they are doing. This information can be used to help you locate your opponent.” During this time, they also had the opportunity to familiarize themselves with the audio cues by listening to them along with seeing their associated icon and description.

The first stage of the training consisted of straightforward questions where a description of an audio cue was presented and the user had to choose the correct icon and sound combination. To reinforce learning, choosing an option always played the sound associated with that object. Before moving onto the next stage, the user had to answer every question correctly at least once, an approach which was used throughout the training.

The second stage consisted of two different types of questions. One type involved choosing a sound based on an icon and description while the other was the reverse, i.e., choosing an icon and description based on a sound.

**Interface Approach: Visualizing Audio Cues**

As an alternative to the training approach, we also explored the idea of modifying the game interface to help novices understand audio cues in the context of gameplay. This was inspired by evidence in Study One that the confidence of novices increased in the presence of visual cues, that visual cues were recognized more often, as well as previous work by Holloway et al. in visualizing FPS audio for hearing impaired gamers [24].

In our approach, we modified the open-source Quake Live demo playback tool, WolfcamQL, and added a region to the player’s HUD for audio cue icons, to visualize the
opponent’s sounds on the HUD with an icon corresponding to that sound. The HUD icons don’t reveal any additional information to a player who already understands the audio cues, but should make highlight the presence of an audio cue and trigger the participant to consider it. Although there are many salient cues for prompting recall, we use iconographic representations of the cues as our goal is to highlight the presence of audio information and help players to learn to associate that information with the behaviour of their opponent.

Study Procedure
To evaluate the two scaffolding approaches, we reused the same online study system that was used in Study One. We crossed the 2-factors (Training and Interface) in a between-subjects design, creating four groups. The procedure of the study itself was identical to Study One, aside from the addition of training and the alternative interface with auditory icons in the HUD. Training took place before any scenarios were presented, and the alternative interface was handled by replacing the videos in Study One with videos containing the visualization of the audio cues.

As our goal was to assist novices, we did not want to recruit many experts. Therefore, participants were recruited through Amazon’s Mechanical Turk website, rather than through a variety of online sources, including the gaming forums that experts visit as in Study One.

A total of 232 participants completed the study; 22 were removed either for not watching the videos to completion or failing to make a guess on both introductory scenarios – leaving 210 participants (131 male, 79 female; all over the age of 18 with the majority being between the ages of 26 and 30). Over the four conditions, 51 completed the version without any scaffolding, 53 completed the version with the audio cue HUD icons, 54 completed the version with training before the scenarios, and 52 completed the version with both the training and audio cue HUD icons.

To test the retention of the training after two weeks, we invited those who completed the training version of the study to complete a re-test of the final round of the training, which consisted of the second stage of training for all the learned cues. We opened this retest for only 25 participants (7 female, 18 male; majority aged 18-25). Participants were collected based on how quickly they responded to the call.

Data Analyses
We use the same measures as in Study One – guess count, confidence rating, and cue count. To show accuracy in the auditory scenarios, we present heatmaps for all three scenarios and examine the written responses for indicators of accuracy. Like the expert responses from Study 1, the indicators were mentions of landmarks or their audio cue.

We analyzed the measures for differences between the four training/interface groups. For the confidence ratings, we performed a univariate ANOVA with two between-subjects factors for each scenario type (introductory, audio, visual) separately, using pairwise comparisons (least significant different) to examine between-group differences. For the counts (guesses, I heard, and correct notes) we were limited to non-parametric tests, and thus use the Mann-Whitney test separately for both the training and HUD icons factors; we did not test for interaction effects between the two factors because of the limitations of non-parametric tests [15].

Results
We present the data in three parts. We first present the efficacy of the training system and how well participants were able to learn the sound cues, followed by the results for the scenarios, and finally the re-test of the sound cues two weeks after training.

Efficacy of the Training System
We present the descriptive statistics of the final performance (Table 2), for the final round of training, which included all 13 sound cues, for both the final two question types (icon from sound and sound from icon) and an average of the two. We present performance as a ratio of the number of correct responses over all responses to handle the different numbers of trials as a result of adding wrong responses back into the queue (see system design).

We found our training system to be effective for teaching users the audio cues, with the final overall accuracy being 96%; the worst performing cue resulting in 91% accuracy and the best performing cue resulting in 100% accuracy.

Scenario Results
Table 2 shows the descriptive statistics, and Table 3 shows the results of the statistical tests.

Introductory and Visually-Dependent Scenarios: We trained players to recognize audio cues, thus we expected that our systems would not improve accuracy or confidence in the introductory or visual scenarios. There were no main effects of training or HUD icons on either the number of location guesses participants made or the confidence in their guess (see Table 3). There were no interaction effects between training and HUD icons for the confidence for the introductory scenarios \((F_{1,206}; p=.490)\) or visually-dependent scenarios \((F_{1,206}; p=.843)\). For the introductory scenarios, there was a significant main effect of training, (but not HUD icons), on the number of audio cues listed, in which training increased the number of

![Figure 2. The HUD with audio cue icons. The icons indicate that the opponent has been heard walking and jumping.](image-url)
The other player went through the teleporter.

Was the guess?

When they made a guess, were they more accurate?

Were participants more willing to make a guess?

Were they more confident in their guesses?

For scenario 5, neither scaffolding approach appeared to make a difference in accuracy. This may be due to the ambiguity of the teleporter’s exit location, as teleporter exists were not labelled on the map as the item pickups were. Several responses by participants indicate that this is the case, e.g., “the other player went through the teleporter to it’s (sic) destination, but I am unsure of the destination on the map,” or “I know they exited a teleporter, but those are not really labelled on the map.”

There were significant main effects of training and HUD icons (see Table 3) on whether or not participants correctly identified the opponent’s location in terms of a landmark or its audio cue in their written responses, in which both approaches increased the number of such responses.

Were they more confident in their guesses?

For scenario 3, those who received no scaffolding performed poorly while those who did were more likely to identify the opponent’s correct location based on the item pickup audio cues they heard.

The other player went through the teleporter.

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?

Did you see

When they made a guess, were they more accurate?

Did you take a guess?

Did you have a guess?
for descriptive statistics). We found that participants retained their understanding of the previously trained cues, although there was a slight drop in the correct identification of the pickup cues that were easily confused with each other. The average accuracy fell from 95% to 92% for pickup cues and 97% to 96% for the other cues.

**Summary and Discussion of Study Two**

We found that both training and the introduction of HUD icons helped. They both improved the accuracy of the location guesses made (Figure 3), as well as the quality of the written responses. In addition to this, they also increased the participant’s confidence. However, training resulted in additional improvements that were not present with the introduction of HUD icons alone. Participants listed hearing more audio cues and made marginally more guesses. In addition, participants in the training condition who re-tested retained their knowledge of audio cues.

**GENERAL DISCUSSION**

The scaffolding systems we provided assist in the learning process and lead to rapid improvements. The systems primarily help the participant in the early stages of the learning process; practice is still required to master the skill. We look at how each system assists with learning in the context of Kiili’s and Fitts and Poster’s models.

The training system primarily helps players during the ideation phase in Kiili’s model; an understanding of the sounds they are hearing allows them to incorporate that knowledge into potential solutions to the problem of locating opponents. In the same way, it would assist with Fitts and Posner’s cognitive phase and help players settle on a strategy for locating enemies. During the associate phase, the player – having already learned how to differentiate sounds and understand how they can be applied in the game – can focus on improving.

The modified HUD interface system also assists with the ideation and cognitive phases. However, it does so by visually highlighting the information; the user may not know what it means, just that it may be relevant to the problem. Therefore, it is doubtful whether it would be beneficial during the associative and autonomous phases, because the modified HUD could result in a dependence of the interface rather than the correct approach of listening to the sounds.

For both systems, whether or not the player is able to reach the associate phase will depend on whether or not they are able to successfully practice the skill of identifying and using audio cues. A training system similar to our scenario-based test system could potentially provide an environment for practicing cue recognition.

**Types of Audio Cues**

It became apparent that some audio cues were more revealing of information than others. For example, an item pickup sound reveals a player’s exact location, while footsteps reveal a player’s approximate location. We have identified three categories of audio cues that expose a player’s position, shown in Figure 4.

The first category, fixed-point, refers to those audio cues that originate from a static point within the game environment. These include such cues as item pickup sounds, teleporter sounds, or more organic sounds such as doors opening or elevators moving. The next category, fixed-area, includes those sounds which can be connected back to a fixed area within the environment. This category is made up of sound cues created by the environment itself, such as splashing water, rustling bushes, or the transition from one ground type to another (e.g. stepping from dirt to concrete as the player enters a building). The final category, unfixed, is for sound cues that cannot be definitively linked to a specific location, and includes sound cues such as gunfire, footsteps, and jumping.

![Figure 4. The different categories of audio cues and the relationships between requisite knowledge and the degree to which the other player’s location is revealed.](image)

As the amount of experience required increases, the information that is revealed decreases (see Figure 4). The fixed-point cues only require a player to know the specific point on the map the sound originated from along with what that sound is, and this reveals the exact player location. Fixed-area cues require further map knowledge (e.g. locations of doors or pools of water), and in some cases require a player to listen for changes in sound cues (e.g. moving between ground types). When understood, fixed-area cues are able to reveal the area an enemy is in, rather than the exact point on the map. Unfixed cues reveal even less about a player. They require a player to have familiarity with the level’s passageways and to pay attention to the changes in sound (e.g. increase or decrease in volume) and allow a player to narrow down the number of routes the opponent may be taking or whether they are heading towards or away from them.

**Further Benefits of Audio Cues**

Considering the online strategy guides [17, 32] and Conroy et al.’s [11] work in understanding threat assessment in FPS games, it is clear that many skills build upon an understanding of audio cues. For example, a central part of the strategy in arena FPS games like Quake is paying attention to the status of important items. In order to secure an advantage, an expert player needs to be aware of when the items were picked up last so they can try to collect them when the next respawn. Additionally, paying attention to whether or not an item was collected by an opponent can reveal more about that opponent’s status, including approximately how much health and armour they have and what weapons they have. A player who is aware of these
variables can then make strategically sound in-game decisions, improving their performance.

LIMITATIONS AND FUTURE WORK
The approach of ascertaining what makes someone an expert in a game and training for those skills is, to our knowledge, original. However, our specific approach led to a number of limitations, and there are still many unanswered questions.

Testing online allowed us to gather data from participants with a range of skill levels; however, it also led to some limitations. The lack of a 3D environment could have led to confusion in interpreting the top-down 2D map and the lack of time pressure during the task may have resulted in reduced anxiety [9]. Other limitations of our approach include the narrow focus on the audio cues present in Quake Live, and the self-rating of player expertise, which, though potentially unreliable, was used due to the lack of an accurate system to objectively measure it.

The next step is to look into using a typical interactive 3D FPS environment, both to confirm that the scaffolding approaches continue to work, and to evaluate whether it eventually leads to in-game performance improvements.

Other future research directions include:

Scaffolding other skills present in multiplayer games. We looked at a scaffolding single FPS skill, yet there many skills in multiplayer games for which experiential learning is ineffective. Scaffolding may be used to introduce players to skills that can be learned independent of each other, such as aiming and navigation in FPS games [47, 48], as well as skills that build upon existing ones, such as strategic decision making and identifying threats [11], which build upon the use of audio cues.

Providing skill assistance to facilitate practice. Our system was designed to assist players during the initial phases of skill acquisition; they still require practice in order to master the skill. Based on Kiili’s [27] game-based experiential learning theory, we may be able to provide skill assistance for the skill not being practiced to optimize the challenge level and maximize the time spent in the flow channel, where learning primarily occurs.

Determining the relationship between motivation to succeed or interest in the genre and the efficacy of scaffolding approaches. Motivated players wishing to learn under their own volition might behave differently than those who complete the training simply because it is a paid experiment. In this study, we limited our characterization of players to their expertise; future work should incorporate additional characteristics of players and should validate our findings with players who are intrinsically motivated to learn the audio cues and improve their FPS skills.

The impact of scaffolding on a player’s enjoyment. As stated previously, mastery over a game’s challenges is an important part of what makes games enjoyable [18, 19, 29].

Scaffolding should allow for quicker mastery over in-game challenges; however, future work should address whether or not scaffolding players – rather than having them learn experientially – changes play experience.

CONCLUSION
The current reliance on experiential learning within the FPS genre has resulted in novices who may feel anxious and overwhelmed when they compete against tougher opponents [9], which could hinder their ability to learn the correct strategies to succeed [27]. Therefore, we investigated ways of scaffolding novice’s skill development for a critical FPS skill – locating opponents.

We found that the differences in how experts located opponents compared to novices occur in situations where only auditory information is present. Experts have learned how to easily make use of audio information to locate opponents, whereas novices lacked confidence, did not understand the sound cues they were hearing, and were ultimately unable to locate the opponent.

To help novices learn to use audio cues, we introduced two prospective scaffolding approaches (HUD icons and external training) and tested how they changed the responses over the same set of scenarios. Both approaches increased confidence, shifted the way that novices reported on an opponent’s location to closer resemble the responses of experts, and improved accuracy. The training system emerged as being the superior scaffolding approach; in addition to the improvements that the HUD icons gave, training also helped novices identify more audio cues and further improved the accuracy of their responses.

Our work takes a step towards helping people of disparate skill level play together in the long-term, by scaffolding novices to learn and use a strategy commonly employed by experts. Scaffolding skill development can ultimately open up social play opportunities for novices who would not otherwise be able to play with their more expert friends, helping people use games to stay connected with others.

ACKNOWLEDGEMENTS
We thank NSERC for funding, the Interaction lab at the University of Saskatchewan, and our participants.

REFERENCES


