Virtual Spectating: Hearing Beyond the Video Arcade

Norman Makoto Su
School of Information and Library Studies
University College Dublin
Dublin 4, Ireland
normsu@ucd.ie

Patrick C. Shih
Department of Informatics
University of California, Irvine
Irvine, CA 92697-9440
patshih@ics.uci.edu

The latest in the most popular head-to-head fighting video series, Super Street Fighter IV by Capcom, now features spectating functionality. Coupled with audio chat, players can both watch and participate in matches with anyone in the world. We describe a video analysis of over 36 hours of gameplay in SSF4. Our results show that players can deftly structure audio when spectating to transform gameplay in SSF4. This new hybrid spectator sport in which boundaries between actor and audience are blurred exhibits three significant characteristics: the audio recreates the arcade culture, appropriates “invisible” but real players, and symbolically creates power relations between mute and speaking gamers. We also present exploratory evidence that suggests a starkly different experience for female gamers who wish to audibly join in virtual spectating.

Spectating, Performing, Audio, Video games, Fighting games, Super Street Fighter IV

1. INTRODUCTION
Spectating is about two key things. First, palpable excitement comes not only from watching skilled play but being immersed in a crowd of spectators who love such play. Second, play itself becomes more enjoyable when players are cognizant of their audience. Together, these two aspects create an experience that engages both players and spectators in competitive sports. While direct contact may not exist, access to varying degrees of auditory and visual acuity can establish an interplay between the player and spectator.

The viability of sports depends on supporting this relationship between player and spectator. While the sport itself draws the crowd, a variety of activities are instituted to enhance the spectating experience. In baseball, when the Los Angeles Angels of Anaheim seem to face an insurmountable deficit, the “rally monkey” on the JumboTron implores spectators to make noise and cheer for their home team. In basketball, when a member of the away team is at a foul line, fans behind the hoop hold up signs, scream, and release streamers to disrupt the opposing player’s attention. Part of the appeal of competitive sports lies in the promise of creating a connection between player and spectator.

Electronic sports (e-sports) has recently gained legitimacy as a viable spectator sport. In e-sports, video games are played competitively. First person shooter (e.g., Counter-Strike), real-time strategy (e.g., StarCraft), and head-to-head fighting (e.g., Street Fighter II) games are the customary genres of e-sports. For professional players and the events which host them, e-sports has proven to be an enormously profitable enterprise (Jonasson & Thiborg, 2010). In South Korea, the government sponsors star players (Policy Analysis & Research Team, 2007), and the production quality of tournaments is akin to a Broadway musical, complete with dramatic music, flashing lasers, and fog. StarCraft in South Korea is regularly watched on television and professional commentators explain the intricacies of the game to their viewers.

Certainly, an attractive element of e-sports is its relative accessibility. While, like any sport, mastery of a video game requires discipline and perseverance, video games allow one to immediately pick up a controller and play. When video games become too complex, they fail commercially. For example, Capcom’s Street Fighter III, though loved by professional players, proved to have such a high learning curve that it alienated many casual players from the head-to-head fighting game genre. Having learned its lesson, Capcom purposefully designed the next game in the series, Street Fighter IV, to be easy to start with but difficult to master (Su, 2010). This delicate balance in video games make them attractive as e-sports because casual players can both understand the mechanics of a professional match as well as plausibly believe that such level of skill is within reach to anyone who can play a console game.

Moreover, while gameplay had been confined to the living room amongst friends and family, the
maturation of broadband access has allowed people to expand their network of competitors. Anyone can play anyone in the world. And, anyone can watch anyone play anyone in the world. Hence, virtual gaming offers more than simple emulation of the real-life spectator/performer experience in sports or e-sports. It promises a new experience in which one can dually participate as both spectator and player in an arena. You can be in the spotlight with spectators from around the world watching you compete. And, you can be front and center, watching players dish it out on your TV screen. Virtual gaming promises to go “beyond being there” (Hollan & Stornetta, 1992) in sports.

Yet there has been sparse research focusing on the online practices of performing and spectating that arise in this new crop of competitive games that explicitly support the delineation between these two roles of performing and spectating. Past research has primarily focused on virtual gaming environments such as MUDs (O’Day et al., 1996) and MMORPGs (Nardi, 2010) that squarely place a first-person player as the continual performer. At the same time, HCI researchers have recognized the importance of designing for performer and spectator. Such work has noted the increasingly blurred line between performer and spectator for the layperson in computer-embedded environments. This research has primarily focused on physical places such as museums (Dalsgaard & Hansen, 2008) or sport pubs (Reeves et al., 2010). Purely virtual environments which feature spectator and performer roles have not been closely examined.

In this paper, we detail a video analysis of over 36 hours of recorded online gameplay with Super Street Fighter IV (SSF4), the latest incarnation of the most popular head-to-head fighting game series. SSF4 “enables” the performer/spectator experience by augmenting online matches with an undirected audio channel. In a game lobby, queued players can watch the current match and speak to both the current fighters and other spectators. Conversely, the current fighters can speak to each other and the spectators during their match. We will move beyond merely asking if the game is social; rather, we ask: how does the deft utilization of audio in SSF4 provide structures of spectating and performing? How does audio recreate and go beyond the arcade scene—replete with spectators and performers—in which the Street Fighter series has had deep roots in?

Just as past work (Whittaker, 2003) has shown the remarkable power that audio provides in tasks even without visual cues, our results similarly show that audio can completely transform the online gaming experience. While one can play SSF4 online without opting to participate in the audio channel, we demonstrate that audio enriches the experience of both mute and audible players in three significant ways: (1) Audio transforms gameplay into a new hybrid spectator sport—one in which performing and spectating are blurred beyond the game’s official boundaries—that encourages not only showmanship, but sportsmanship and mentorship; (2) audio reveals the “invisible” by uncovering collocated and mute players, allowing one virtual player to enjoy playing and interacting with these invisible cohorts; and (3) audio has surprising implications in creating unique power relationships between those who participate and those who do not participate in the audio channel.

Finally, we present some preliminary evidence which suggests that female gamers who participate on the audio channel may face challenges that prevent them from reaping the very benefits audio has in constructing a unique arena for spectator and performer.

2. METHODOLOGY

Our methodology follows past work on play that draw from a phenomenological perspective (Reeves et al., 2009, 2010). That is, we are not merely concerned with whether expected spectator-like behavior exists in Super Street Fighter IV, but rather how such behavior manifests itself through the skillful use of audio in a virtual setting. We adhere to the notion (Garfinkel, 2002) that there is work in achieving a social order that identifies one as a competent member of spectating and performing in SSF4. In order to unpack spectatorship, we must scrutinize the detailed, minute audio conversations in concert with their actual gameplay. It is in these very conversations that taken for granted structuring occurs for spectating1 to function in gameplay. As one might ask how people structure conversation openers to achieve the telephone call (Schegloff, 1986), one might ask how people structure audio chat with gameplay to achieve spectating and performing.

Like other researchers in video games (Nardi, 2010; Sudnow, 1983), we became participant observers. Only by making a concerted effort to immerse ourselves in gameplay, could we understand the unique cultural practices in audio (e.g., SSF4 lingo) when performing or spectating. Both authors created a PlayStation Network (PSN) ID and participated in the two new gameplay modes that include an audio channel:

1. In Endless Battles, lobbies of up to size eight are created and at any given time, two players compete with each other. Those who are waiting in the queue can watch the two players fight. The person at the top of the queue will challenge the person at the end of the queue. This mode emulates how gameplay is done in video arcades.

2. In Team Battles, one team of players line up in a particular order to fight a second team of players. Teams can have up to 4 players. The winner of a match continues on to fight the remaining players of the opposing team. This continues until an entire team has been eliminated. It is possible for someone to never play because their teammates (who went before him or her) eliminated the opposing team.

In both these modes, spectating is supported. Over the audio channel, players can freely talk to each other.

1 For brevity’s sake, when we refer to the demonstrable activity of spectating, it is inclusive of performing.
other. Because there is no directed chat (i.e.,
speaking to only particular players in a lobby),
anything one says will be heard by all members of
the lobby. The match is shown in real-time while players
in a lobby are awaiting their turn to play.

We approached our observations as serious game
players might. One of the authors is an experienced
(not professional) player of the Street Fighter series.
The other author is a novice to the fighting game
genre. We both tried to achieve a certain level of
proficiency with a set number of characters (referred
to as one’s “main”). Aside from dedicated training
sessions, we also read online strategy guides and
studied online videos of professional play.

Each gaming session lasted approximately 2–4
hours. They were conducted in an apartment setting
(emulating the natural environment in which gameplay
is done) and a computer lab. All sessions were
closed to outside interruptions. In each gaming
session, we recorded a mix of Endless and Team
Battles. All matches were video recorded and
transcribed. Because our focus was on spectatorship,
we discarded lobbies that did not have at least one
audio chat participant (excluding ourselves). This was
determined by first asking the prompt, “Does anyone
have a mic here?” If there was no response within
a minute, we left the lobby. Game sessions where
players had microphones enabled but refused to use
them were also discarded from our dataset.

In total we recorded approximately 36 hours of
recorded gameplay over a period of six months
since SSF4’s release on April 27th, 2010. Roughly
20 h. 30 min. of Endless Battles and 15 h. 45 min.
of Team Battles were captured. The video was
then transcribed and analyzed. For the latter,
we drew from video analysis techniques steeped
in ethnomethodological approaches (Jordan &
Henderson, 1995). However, unlike traditional video
analysis, there is no taping of “real” people. We
cannot see who controls the characters—we can only
see the characters themselves, and hear the players’
voices.

Thus, while we draw from ethnographic strategies
for data collection and analysis (coding for major
themes from our video analysis), we also apply an
ethnomethodological lens onto our data. As a result,
our work has less of a narrative, personal feel like
Sudnow’s (1983) solitary and driven account on
mastering Breakout.

3. RELATED WORK

There are two bodies of work that have relevance
to our study. The first highlights the importance of
designing for the dual role of performer and spectator
in open, physical environments. The second is the
large corpus now on video games and in particular
on online play. We observe that there is surprisingly
little overlap between these two areas of research.

Research has called attention to the importance of
spectatorship when interacting with tangible media
in open spaces. For example, performative art
incorporating “New Media” (Manovich, 2007) often
blurs the line between performer and spectator.
Museum goers cease to become mere spectators
but take on a dual role as performer and spectator.
Dalsgaard and Hansen (2008) use the term
“performing perception” to explain how the user
now simultaneously fulfills three roles—operator,
performer and spectator—in interaction. Significantly,
they note that in public spaces, “you are also an
implicit spectator of your own actions since your
own actions will be the ones that other people
are experiencing.” Reeves et al. (2010) examined
sport fan crowd behaviors. Public behavior such
as singing in pubs can induce crowds of strangers
to readily interact with each other in a contagious
fashion. At once crowds can reveal distinct groups
with distinct behaviors, but at other times the crowds
can exhibit uniform behavior. This work emphasizes
how the crowd itself serves to offer different forms
of participation rather than on how such acts may
be seen to be performative. These frameworks offer
suggestions on aspects of performing and spectating
that might be targeted in the design of technological
artifacts and their enclosures. For example, a future
challenge will be to support the fluidity of performing
and spectating in the real world.

The aforementioned work focuses on real-life
interactions in open locale. There is now a large
body of research on video games but, to the best of
our knowledge, no work focusing on the spectator
and performer experience in purely online play.
Certainly, aspects of spectatorship are touched
upon. In online games with collaborative functionality there
are always some performative activities with other
players. For example, in Counter-Strike, players can
exit the game by dramatically jumping off a building to
their death. Such humorous actions are performative
and elicit tongue-in-cheek comments from others
such as “suicide isn’t the answer” (Wright et al.,
2002). Humorous impromptu dances by characters
are also mentioned as part of the spectacle that
World of Warcraft provides (Ducheneaut et al., 2006).
Nardi (2010) argues that WoW is a new “visual-
performative medium” (p.7). Unlike traditional spaces
for performance, WoW provides a prepackaged
asmemble of visual possibilities that are easy to
start with (p.84). While her narrative interleaves
textual examples, it argues for the primacy of the
visual when participating in raids. WoW is “a kind
of theatre in which audience and performers were
one. There was plenty to look at, but at the same
time, players themselves were onstage” (p.90). These
works focus on gaming genres in which the player
remains always on stage—that is, always as the
performer. So, for example, when players die in
FPSs, they gain the ability to spectate carte blanche
on the match but are also rendered unable to
communicate with their teammates. Moreover, this
research does not touch upon the implications of
audio with performer and spectator. For virtual
worlds, text remains the more dominant medium.
Finally, while third party VOIP options are popular
for many online games, their usage has been largely
unanalyzed.
A general purpose platform, Second Life’s (SL) virtual world supports events (Boellstorff, 2010) that emulate real-life events such as wedding ceremonies, round-robin storytelling, and so on. Boellstorff (2010) focused on the debate within SL on whether voicechat should be enabled (e.g., that it would destroy a degree of anonymity valued) and noted that voice was “not used for everyday communication” (p.113). Visual aspects are again emphasized in Irani et al.’s (2008) discussion on the challenges of designing virtual artifacts in SL that visually convey presence and privacy information to audiences.

In contrast to these virtual environments and games, SSF4 is narrowly focused on competitive, one-on-one gameplay. There are no avatars, rich narratives, or crafting of virtual artifacts. Matches in SSF4 happen in rapid-fire succession. Interactions are therefore necessarily ephemeral; other means of communicating commonly used in virtual environments such as texting are not an option for supporting in situ spectator interaction. Functionally, SSF4 presents a clear delineation between who is performer and who is spectator. At anytime, one is either technically a performer or spectator but not both. Certainly, visual elements are important, but it is through the audio channel that spectatorship is “enabled” and allows the displayed center-ring match to become a virtual spectator sport. SSF4 provides the first opportunity to research virtual genres in which spectator and performer have well-defined roles (which they often do in public spaces), yet have the opportunity to intimately interact with each other (which is difficult in traditional sporting events). Lastly, our research also provides a means to understand how players might go beyond the game’s set functionality through the audio channel to instantiate a blurred, dual role for spectator and performer that other research has observed.

4. MICROPHONE USAGE IN SSF4

How common is audio usage amongst online SSF4 players? To answer this question we randomly surveyed 100 lobbies from each of the two gameplay modes (cf. Section 2). For each lobby, we recorded the number of players with their microphone turned on (without solicitation) and the total number of players present. Overall, 55 out of 225 players in 100 Team Battles mode lobbies had the voice chat capability enabled, while only 36 out of 251 players did so in the 100 Endless Battles mode lobbies we surveyed. The audio usage adoption between the Team Battles mode ($\mu = 23.3\%$, $\sigma = 0.35\%$) and the Endless Battles mode ($\mu = 13.0\%$, $\sigma = 0.26\%$) was significantly different ($p < 0.02$, $t = -2.32$).

Thus, we issue a caveat that our observations may particularly pertain to gamers who choose to participate in rooms with people speaking over the audio channel. Reasons for opting out of the audio chat may be the potential for increased lag, unwanted noise, or intrusion on privacy. We surmise that many people who have a microphone for the PlayStation 3 did not procure it for SSF4 but rather for a first-person shooter (whose sales dwarf head-to-head fighting games). Upon asking players whether they usually used a microphone in SSF4 many of them replied that they typically used it for games like Call of Duty and Battlefield 2. In our sessions, we often encountered a bandwagon effect with regards to microphone usage. Just one person participating over the audio chat would elicit others to enable their microphone. As one player remarked, “You pulled one mic out [in a lobby] now everybody has a mic [on]. That’s how it goes.” So participation in spectating may be a matter of happenstance. However, the results in this paper also has significance for lurkers—that is, those who listen in on the chatter.

5. THE AUDIBLE ARCADE

In the following sections, we use selected video transcript vignettes from our data collection in SSF4 to illustrate the emergent structures of spectating that people enact through audio. That is, what sort of spectating manifests itself in SSF4? The following prefixes are used for each vignette: S indicates spectator; P indicates someone who is playing in the current match. “Character” refers to the in-game manipulated video game fighter; “Player” refers to the actual person manipulating the character. Character names are in SMALL CAPS typography. Images are in temporal sequence from left to right. PlayStation Network (PSN) IDs are in bold italics and are anonymized.

5.1. Virtual Hype through Virtual Spectating

Past work (Su, 2010) has examined the culture of hardcore gaming in the head-to-head fighting game scene. Hype is identified as a core concept in this genre. Generally, hype is the excitement and anticipation generated by spectators and performers about gameplay. It involves elements of trash talking that define what proper play is and tutelage between players.

We observed that the audio channel to some extent recreates hype for both fellow spectators and the players. An oft-encountered behavior in our online sessions was the rollicking performance of crowds. Just as spectators layer a new acoustic in e-sports, the audio channel in SSF4 empowers users to construct their own audio on top of the in-game sound. Mimicry was common in matches. This involved shouting out imitations of the character’s noises. For example, BALROG, the boxer, yells out, “My fight money!” when he loses. DHALSIM, the yoga master, yells out “Yoga!” when stretching out his limbs to attack. GIULE yells out “Sonic Boom!” when throwing a projectile. All these utterances were routinely imitated by virtual spectators. Significantly, these imitations are exaggerated: caricatures of the character’s pitch and intonation are peppered throughout matches. This behavior is reminiscent of that displayed in Street Fighter tournaments where the crowd will, in sync, yell out the noises when a player is destroying an opponent with a well-known move (Su, 2010). Not only does this serve to rally the crowd, it also serves to demonstrate intimate knowledge of the game.
Street Fighter enthusiasts, through online strategy guides, videos, and forums, promulgate what proper play is. The language imbued in such texts admonishes and shames those who don’t follow good play. For example, Ken Masters, one of the most popularly chosen characters in SSF4, is often used by newbies. The over reliance of beginners on Ken’s signature special move, the “dragon punch” (in Japanese, shoryuken), is often criticized (Su, 2010). In online texts, these unintelligent, predictable fighters are labeled as scrubs. In parallel, we also observe that virtual spectators in SSF4 let their opinion be known through repartees. In the below session, an obviously inexperienced player is getting a lot of “lucky” hits via Ken’s dragon punch:

Endless Battle  
— Round 1 —  
S1: Whoa, tauntin’. [Reaction to taunt by P2]  
S2: Ah, man. [Disappointment at P1 losing]  
— Round 2 —  
S3: Oh! [Reaction to taunt by P2]  

Fig. 1: First taunt by P2.

Yet, at the same time, those who do have skill cannot needlessly jeer their lesser-skilled brethren. This quasi “code of honor” is part of Street Fighter discourse. In SSF4, characters can execute in-game taunts, moves that do no damage but simply mock their opponent. Taunts are risky in that they leave the character momentarily vulnerable. In Fig. 1, P2 taunts (by deftly sticking his fist out) his/Her opponent in round 1, defeats P1, and then again taunts at the start of round 2. The crowd immediately tells P1 that he cannot let his opponent get away with that. It is obvious that the crowd is on P1’s side. Once P1 defeats P2 in the second round, the crowd sarcastically asks P2 to try taunting again despite his/her loss (he does not). The crowd relishes trash talking players who display their confidence prematurely. Such interactions are emblematic of the culture of SSF4 that punishes overt braggadocio.

In the aforementioned vignette, neither player had a microphone and only the spectators spoke. However, we found that when both players had microphones, deliberate, derogatory trash talking was rare. For the most part people engaged in proper sportsmanship as one would in an arcade. When people executed clever combos or exciting setups, comments such as “nice” or “awesome” were commonly heard. Interestingly, insults and frustrations were not directed at the players but rather to the characters, (e.g., “That guy is such a bitch, Dhalsim.”). We reason that audio made the fact that there was a real player behind the character more apparent. Likewise, when facing an opponent in the arcade, players are unlikely to hurl direct insults.

In the arcade, trash talking can often involve physical gestures by the player which mimic their character. This is not possible in the virtual environment of SSF4, but players can compensate with expressive use of moves. In fact, a player can perform for the crowd moves that obviously will not hit the opponent. In the scenario at Fig. 2, the opponent is being continually hit by Dhalsim’s knee kick, his most effective “anti-air” move. After the opponent falls for it for the third time, the player teasingly executes the knee move several more times to the laughter of the spectators. Again, although the character motion was speechless, it was the performance that elicited favorable feedback from the crowd.

Fig. 2: Performing by executing three successive knees.

Nationality was also alluded to (based on your PSN geography). We encountered friendly rivalries between Canadians, Japanese, and British players. Such rivalries (Su, 2010) are commonplace in the head-to-head fighting genre (e.g., East Coast vs. West Coast, and Japan vs. USA).
The audio channel provides occasions for players to demonstrate not only their mastery with the characters but with the culture of hype behind SSF4. Thus if one were continually exposed to the audio chatter, one cannot help but become competent at spectating in the context of SSF4’s culture. Audio use and feedback provides a means for players to perform for spectators.

5.2. Improving Gameplay via Performance

According to some professional gamers, the arcade is the only way to truly level up. Professionals in the head-to-head genre have traditionally trained in arcades, reminiscent of communities of practice (Su, 2010). We would also argue that the audio channel can by itself provide a means by which performing and spectating contribute to learning.

SSF4 is an intricate game. Thus, surprises—when particular manipulations achieve unexpected effects—are commonplace. With the audio medium, players can unpack such surprises. Below, a player learns that DhalSIM has a move, Yoga Tower, that evades fireballs (it is commonly mistaken for a taunt):

**Team Battle**

P1: [DhalSIM does a Yoga Tower.] <laughs> What the hell? <laughs>
P2: Yeah that goes through fireballs. It looks like a taunt but it's actually a move.
P1: That is funny. That was hilarious. Oh my god. I actually want to tape this.

Additionally, players sometimes stop normal competitive play (even with spectators looking upon them) to demonstrate fundamentals of the game. In one session, a player asked his opponent to show him the timing of combos with Ibuki. In another session, a surprise like the one above led to a demonstration of that very move on a player’s opponent.

Spectators often provide a player with strategies to defeat their opponent. Invariably, this advice is given to the player who is presently struggling. Below, spectators give the player mixed advice, one saying that he should use his “hurricane kick” and another saying sarcastically that there is no hope (zoning can quickly refer to a particular effect and ask about its significance. Oftentimes, these spurred debates on the merits of SSF4’s design (e.g., it is too much of a defensive game, it is dumbed down) as well as what is “proper” sportsmanship (e.g., what does it mean to play “cheaply”). In the course of a single match, a performance can turn from entertainment to pedagogy.

5.3. Identifying Competence

By self-commenting on one’s actions, a player also reveals expertise. While character effects are readily viewable on the TV screen, it is difficult to comprehend whether a particular effect was intended by the player. When a spectacular sequence of events transpires, self-commenting audio can quickly clarify whether skill or luck was involved. In the conversation at Fig. 3, the player asks his opponent whether he purposefully did a move that countered a special move (the lariat). Later on (shown in the picture) a rare type of hit occurs, and the player says, “I’ll take that,” indicating that it was pure luck. Such actions reflect on the skill and experience of the player and reveal (probably not cognizantly) for the audience that he or she was not that capable to carry out such a (recognizably) dramatic move.

**Fig. 3:** P2’s character’s outstretched limbs are hit by a thrown rock from P1’s opposite side, a rare occurrence.

Significantly, intent was not always explicit. Sighs of frustration and swearing made it clear that the move the character did was not what the player desired. Such actions illustrate to the spectators that it is not their intelligence at fault but their execution skills. Just as audio may downplay events that one might ascribe to fantastic skill, it can also save one from looking incompetent—i.e., by proclaiming that one knew what to do but just lacked the proper dexterity.

6. HEARING THE INVISIBLE

In the previous section, we detailed aspects of virtual spectating that to some degree successfully emulates the behavior of head-to-head fighting that occur in living rooms and arcades. Spectating via audio provides opportunities to be part of the “in-crowd”—to demonstrate aptitude not only with the mechanics of SSF4 but the culture about it.

However, there are aspects of audio that go beyond emulating the face-to-face experience. In the following sections, we will discuss the unique aspects that audio brings to the virtual spectating experience. In particular, we will argue that audio provides a limited window into the real-world that can enrich the gaming experience.
6.1. Collocated Switching

Notably, the audio channel provides a voice for those who are physically located near the player (e.g., in a living room where the console is situated). Thus, not only are there virtual spectators via the PlayStation Network (PSN), but there are also spectators in the form of the real-life collocated spectators of the players who are playing virtually through the PSN.

Indeed, once third party ‘chatter’ becomes audible, it can instigate player switching. In team mode, one of us heard the opponent speaking to someone else. We asked who it was and he replied, “I have a twin brother, he’s actually pretty good. Yeah, this [the current player] ain’t my twin brother though. No, he’s not so bad. Hey Mike, do you want to play? Maybe you can beat him? <laughs>.” Interestingly, players often identified themselves by whether they were the owner of the PSN ID. For example, one player said, “next time we play Bob wants to go all out, your best character against his...there’s two of us, I was Abel and he was Zangief and Blanka. I’ll play and then he’ll [Bob] play you.” Here, players identify themselves and their collocated co-players for challenge. Without prompting, sometimes players would reveal “who” was really behind the character. In one match, we played a father who told us his son was playing and that he’d be playing next. As a result, I became an unsuspecting performer for spectators. Hence, background noises furnish a shared opportunity for both entertainment and learning.

6.2. Background Noise

Not only does audio provide a voice to collocated players, it also opens the door to a virtual player’s “real world.” Through chance hearings of various background noise, one can get a sense of a player’s surroundings. Often, this would provide a common point about which players would talk.

For example, one can sometimes hear the clatter of buttons and tapping of a stick—sure signs that the user has a joystick. Serious players of SSF4 use a joystick because moves are physically awkward to execute on a gamepad. Here, a player overhears his opponent’s joystick (also called fightstick):

Endless Battle
P1: [Hears the joystick.] You got the fightstick?
P2: Yeah, I have a fightstick.
P1: Yeah. I got it too but I started playing with it. It’s too hard.
P2: You just have to get used to it. Yeah, at first I couldn’t do anything but you get used to it.
P1: Hmm.
P2: It’s a lot different from the gamepad right?
P1: Yeah.

P2: Doesn’t it hurt your fingers playing the gamepad?
P1: I’m used to it. It used to hurt like my thumbs. Not anymore.
P2: Now you’ve got these thick callouses on your thumb or something?
P1: Yeah, exactly. <laughs>

Other audible events such as a baby crying, a telephone ringing, and parents’ yelling would provide fodder for jokes. At times, players would pick up and answer the phone, apparently no longer aware that everyone could hear the conversation in the lobby. In one case, a player picked up the phone and began pleading to his girlfriend over the phone. As a result, he became an unsuspecting performer for spectators. Certainly, those without a microphone could communicate via the PSN’s messaging system. However, the messaging system lacks the speed and efficacy of audio (and many people still “type” with their gamepad, not a keyboard connected to the console). Significantly, we will show that the audio channel allows one to give voice to players who lack a microphone.

7. THE POWER OF AUDIO

Perhaps the most surprising behavior we observed was how the audio channel itself created power structures amongst spectators and performers in SSF4. Those who had a microphone and would talk back had a different sort of power than those who lacked a microphone and could only listen in on the audio channel (i.e., “mute” players). For instance, we found that gameplay with audio made it more likely that people would add one to their PlayStation Network friend list and to then invite you to lobbies.

Certainly, those without a microphone could communicate via the PSN’s messaging system. However, the messaging system lacks the speed and efficacy of audio (and many people still “type” with their gamepad, not a keyboard connected to the console). Significantly, we will show that the audio channel allows one to give voice to players who lack a microphone.

7.1. Audio In-groups/Out-groups

Strikingly, spectators who actively spoke with their friends would sometimes outright exclude others (even with audio chat). One strategy of exclusion is to continually refer to the player in the third person, despite his or her attempts to proactively converse on the audio chat. In other words, one experiences a group of audio chatters ganging up on another. Below, we have a group of spectating friends mocking John’s voice and acting as if his defeat is inevitable:

Team Battle
S1: Oh, I gotta finish this newbie ass little, this John guy or Johnny or whatever.
S2: What a newbie.
S1: Even if I don’t beat him, you got him for sure. That’s guaranteed.
S3: Yeah, for sure.

Such behaviors can create a particularly antagonistic environment of play. In general, the use of third person pronouns for those without a microphone was common (in saying both compliments and insults): “No he keeps picking Cody and his Cody is not good. I want him to pick Bison.”

In addition, an interesting dynamic arose in team battles. When an opponent had a microphone, sometimes players on the opposite side with
a microphone would cheer for the opponent. Significantly, the team assignment by the PSN was of less import than what social relationship, however fleeting, developed over audio chat. For example, here we have a player who seems to have trouble deciding who to root for (P1 is on the other team and has a microphone):

**Team Battle**

S1: Come on P1! [Cheering on P1]

P1: This guy’s [P2] pretty good though. He knows his stuff.

S1: Yeah, he [P2] is pretty good. 

... 

S1: Damn, [P1 wins the round] <laughs>

S1: I kind of want to see you lose P1. <laughs>

S1: You’re [P1] gonna lose this one. [P1 loses the round]

S1: Come on P1... [Cheering on P1]

S1: I’m going for my team man, sorry.

Additionally, people with the microphone would especially join together to cheer each other against non-microphone users who would taunt (cf. Section 5.1).

### 7.2. The Mute Masses: Mixed Mediums

Most lobbies with at least two people had at least one non-microphone user. Often those users became the subject of inquiry. When the mute user was a friend of someone with a microphone, he or she would empower the friend by providing a voice. In other cases, the mute user become part of a narrative created on the spur of the moment by the audible audience.

In the following, P1 provides a voice for a friend on his team. That friend sends P1 a text message on PSN after P1 lost to P2. A mute spectator is then granted the ability to talk with others by having someone with a microphone read out-loud his text message on the audio channel:

**Team Battle**

P1: He said, he wrote down he didn’t think I’d lose to him [P2]. Cause he always sees me kick everyone’s ass...he doesn’t know that DHALSIM is my weakness.

P2: Uh-huh.

P1: Cause I don’t know how to go about getting through his attacks you know.

P2: Oh, is that what your friend was saying, your teammate?

P1: Yeah, yeah, because he’s seen me kick a lot of people’s asses really quick and shit...

Sometimes this audio-by-proxy occurred in less than flattering ways. During one team battle, a mute opponent who was continuously taunting lost to us (the researchers) in a match. He then sent a message to us with the subject header “pff...lol coward”. The empty content of the message implies that the sender knew that the message would be read immediately in the notification area (Fig. 4). Upon reading this message aloud, his team members, despite being our opponents, supported us during the match, saying, “Don’t taunt if you can’t finish him son.” We should note that we also experienced the opposite, one taunting player gave us a message saying “good shit” and that immediately changed everyone’s impression of him from a rude player to a polite one.

Having no microphone strips away any recourse to defend oneself. In some cases, mute users are accused of being troublemakers (especially if they taunt). While evidence against them may be paltry, they are likely to be targeted. Below, we observe that the team battle mode is lagging and the opponent, P2, who is taunting without a microphone is accused of lagging and faces being kicked out:

**Team Battle**

P1: Ah what the fuck.

S1: A lot of lag.

P1: Ah, come on!

P1: Come on!

P1: Damn, who’s lagging?

S2: <laughs>

P1: Damn, that’s cheating. I can’t do my move. [inaudible]

P1: Man, that was so cheap man.

S3: No kidding.

P1: God, he lags, so he gets free hits.

S4: After we go to the lobby we should kick out the lag-er. Kick whoever’s lagging. I don’t know who’s lagging though.

P1: It’s P2, he’s the one that’s lagging.

S4: Think so?

S3: Yeah, I guess.

However, when a spectator might have chosen to not use a microphone despite having one available, he/she may activate it to defend himself/herself. Here, a user turns on his mic to tell the world that he did indeed defeat his friend (not another player using his account):

**Endless Battle**

P1: Honestly, I honestly don’t think that was my friend playing. He probably had a sleeper come in and you know, “Come beat this guy for me.”

P1: [P1 turns on mic.] You jerk that was me playing!

Sometimes, lack of a microphone would actually make a person more sympathetic. One player, **Richard**, was obviously a complete beginner at SSF4. In this Team Battle, the other players were of unusually high skill. Thus, **Richard** was “holding” his team back. People noticed he had 0 battle points and called him an “amateur.” One spectator said, “What are you playing. You gotta boot him!” Soon, however, **Richard** begins to do better, despite his inexperience, and the spectators start giving him tips. In a close round, while **Richard** is running away desperately to keep alive, the spectators start yelling, “He’s [Richard] about to come back...Ultra 2 chip. Your ultra will chip him Rich!” affectionately abbreviating his name. The Ultra 2 is a special move that could’ve easily won him the match. Despite eventually losing, one spectator said, half-jestingly, “Well, you know what Richard has potential though, I think he’ll be really good one day.” Thus here the mute player becomes a shared object...
to sympathize with and the spectators—upon seeing an “underdog” do surprisingly well—immediately rally behind him. By being mute, spectators can freely attach an appropriate personality to the player to create either villains or heroes.

8. THE FEMALE SPECTATOR EXPERIENCE

In our study, we encountered no female voices (gender was sometimes difficult to discern audibly). Motivated by “breaching” activities (Crabtree, 2004) which disrupt the routine such that what is taken for granted may be reflected upon, we wanted to see the resultant behavior by introducing female voices. Because we had extraordinary difficulties finding females who were proficient at SSF4, we recruited two females confederates to speak as we played. Recognizing that this was not entirely satisfactory, we worked hard to finally recruit a professional female SSF4 player to participate in our sessions.

The very benefits that audio provide in performing and spectating for male players can turn into severe impediments for female players. While most players were congenial to the female players, those that were negative were excessively vituperative. For example, derogatory terms for females (“bitch”), sexually lewd activities (“give me a blow job”), and mockingly high pitch voices were employed on our female players.

Moreover, the topic when spectating over the audio channel was sometimes starkly different for females. All three of our female gamers encountered spectators who doubted or were confused about their gender. Below, during Team Battle dead-time, spectators awkwardly discuss our female player’s gender:

S1: Umm, P1, what...are you a dude or a chick? I’m kind of confused.
P1: Could ask the same question jerk.
S2: I can’t believe you asked that man. That was, yeah.

Other times, people assumed our players were children with high pitched voices.

Many gamers also suspected that a male was playing. Here, our female pro player is using HONDA in a match and a spectator asks about her gender in a rather disrespectful manner:

Endless Battle
S1: That ‘guy’ is good with Honda huh?
P1: What guy?
S1: The guy who’s playing!
P1: I’m not a guy.
S1: I know you’re not a guy.
S2: Your boyfriend is playing?
P1: Excuse me?
S1: Your boyfriend?
P1: I don’t need a guy to beat you.
S1: Come on.

Even relatively benign remarks revealed the assumptions by male gamers that women do not on their own play SSF4.

Endless Battle
S1: Are you 12 years old?
S2: No, I’m 19.
S1: Does your big brother know how to play?
S2: Uh, I don’t have a brother.

Thus the topic of gender through the audio channel dominates conversation and prevents “positive” activities associated with spectating on the audio channel from manifesting itself.

On the other hand, we did notice that as the pro female player’s high-level skills became apparent (far better than the researchers), she became immensely popular because she was good and female. Friend requests skyrocketed and although the chat room was largely devoid of microphones, the screen literally “lit” up with mic activation icons. Spectators remarked on the C. V I P E R “ringer” in the lobby. One spectator said, “the fact that it’s a girl [who beat me], that ain’t going to go through my mind.”

Some of these results have parallels to findings on gender in World of Warcraft. Nardi (2010) found that in some cases women were not allowed to speak on voice chat during raids because performance and attentional focus might be disrupted by the sound of feminine voices. Yee (2008) notes that gender bias has less to do with WoW’s mechanics than a game culture that codifies such stereotypes. While the results on gender here are preliminary, they nonetheless are the first to contribute to explaining why females may choose to be invisible audio-wise in a different genre where such communication isn’t strictly necessary (as opposed to WoW where communicating is key for successful play).

9. DISCUSSION & CONCLUSION

The deft spectating via audio in SSF4 by online gamers suggests several focal points for designers of virtual spectating environments (e.g., museums and competitions):

- **There are social ramifications when introducing a non-mandatory communication channel.** An optional channel can inadvertently create two distinct populations of gamers. This is not necessarily a negative—indeed, the tension created by camaraderie through the symbolic use of audio can be exciting. At the same time, it may exclude players or induce them to participate (e.g., one player messaged us saying he was going to immediately purchase a microphone).

- **Audio can augment a player with a new group of virtual/collocated participants.** Audio “uncovers” collocated players and can encourage them to participate. Indeed, for us, it was exciting to know that players on the other “side” were switching controllers and that the character we’d be playing would be exhibiting different skill sets and strategies. Moreover, audio-proxies give voice to others. One can thus “hear” extended environments unlike the arcade where one is enclosed in a single environment.

- **Simply adding audio gives large incentives for players to “perform.”** In essence, players are more likely to do showmanship and, in turn, also become part of a “performing” crowd.

- **Audio provides a powerful medium to disseminate proper play.** Via audio, the culture of Street Fighter is publicized online.
• Cultural biases may trump some benefits of audio. In SSF4, the topic of gender can completely dominate conversation, possibly discouraging participation for female gamers.
• It may be beneficial to not create direct chat functionality. Instead, simple indicators of who is talking may suffice. While the lack of directed chat precludes, for example, trying to explicitly help a particular player, it forces a greater level of participation/blurring in spectating and performing. Audio in SSF4 drives an intimate form of spectating and performing that successfully emulates particular aspects of the arcade culture—for example, the tutelage of beginners. Competent usage of audio provides structures of spectating that demonstrably show membership in SF culture—for example, through the jeering of improper taunting. At the same time, purely virtual spectating through audio provides occasions for all participants to enjoy a dual role. Despite the fact that spectator and performer are clearly delineated through the game’s mechanics, at anytime all can engage (or not) each other. As Dalsgaard and Hansen (2008) noted, a discrete delineation of spectator from performer is unsatisfying. It is not for the sake of advancement in the game that one adopts such roles, but instead to partake in the enrichment that such roles offer in gameplay.

Unsurprisingly, SSF4 is a social game. What is novel is how the social life of SSF4 renders itself visible with the skillful utilization of audio. This goes beyond simply trying to emulate e-sports, arcades, living rooms, and even organizations. What we are trying to say is that SSF4 incorporates elements of all such environments. One can play SSF4 online without any audio, but once anyone speaks, it immediately enriches both mute and audible players. Audio allows players to construct their own narrative that incorporates the culture of SSF4, the personalities (perceived or otherwise) of the players in the lobby, and the effects of the characters in battle. This power to define narratives necessarily creates tensions between those who speak and those who do not. Yet it is precisely this flexibility that makes SSF4 ever so more enjoyable when the possibility of microphone usage is introduced. Thus, spectating in SSF4 is less about watching matches but producing a vivid account from the relatively boring, animating pixels of two fighters fighting.

As Chapanis et al.’s (1972) classic study shows, audio has a striking way of being flexibly appropriated by users. With audio, SSF4 permits a marked shift in the object of focus from the characters in the match to the players—both spectators and performers—in their setting. In contrast to e-sports, MMORGPs, and FPSs, interaction between players and non-players becomes part of play. Hence, with even such a facile implementation of audio chat (no directed conversations), SSF4 becomes virtual spectating—something going far beyond the arcade milieu.

10. ACKNOWLEDGEMENTS
We thank Kara “Karafice” Leung and Sherry “Sherryjenix” Nhan for their invaluable advice in data collection and analysis. David Redmiles, Morgan Romaine, and Mike Ross also provided logistical support.

REFERENCES


