Abstract
This paper describes a study designed to tease out two dimensions of arcade gameplay: scale of input interface and performance environment. In order to probe how each dimension affects enjoyability and engagement, we employ multiple measures, including galvanic skin response (GSR), distracter tests, Likert scale surveys, and in-game telemetry. To examine the effect of scale, we compare experiences playing on a touchscreen wall and playing on a tablet PC. To examine the effect of public versus private gameplay, we compare experiences playing alone and playing in front of a small audience. Our results suggest that larger-scale input devices are both more enjoyable and more engaging. The data was less conclusive for public versus private gameplay.

Keywords
Arcade gameplay, input devices, touchscreens, galvanic skin response, physicality, gameplay performance
Introduction
In recent years, games like Dance Dance Revolution (DDR) and Guitar Hero have generated a resurgent interest in arcade game interfaces. On home-based, personal systems, games are played via small, system-specific controllers or a standard mouse and keyboard. Multi-player games on personal systems often involve some degree of public performance. However, avatars and pseudonyms result in a somewhat anonymous gameplay experience. Moreover, player and audience are not always collocated. Arcade games, by contrast, boast novel input devices such as toy guns, car simulators, snowboards, dance pads, and musical interfaces. As a result, arcade gameplay is frequently more physical. Groups of friends often play together or compete against each other, engendering a more social, performative gameplay environment.

We argue that these two dimensions – scale of interface and performance environment – lead to very different gaming experiences. However, it is unclear whether one factor is more dominant than the other, or whether the two factors affect gameplay experience in different ways. As home and arcade gameplay experiences converge in public arenas like Internet cafes and in emerging consumer devices like the Nintendo Wii, navigating these two dimensions will become increasingly important for game designers.

Related Work
At least in terms of output devices, research suggests that larger-scale interfaces are more engaging. Patrick et al. conclude that larger-scale interfaces increase immersion by “tricking a person’s perceptual systems into thinking they are really there” [6]. In order to examine the emotional effects of videogames, several researchers have utilized physiological indicators [5, 8]. It is worth noting that many of these physiological examinations of gameplay experience have focused on game content. Thus, we think there remains a need to use physiological indicators to examine not only the experience of the game itself, but also the experience of the interface and gameplay environment.

One particularly useful physiological indicator for quantifying gameplay experience is galvanic skin response (GSR). GSR measures skin resistance and conductivity via the eccrine sweat glands. Ravaja et al. argue that GSR is a better index of arousal than heart rate since it is only influenced by the Sympathetic nervous system. Mandryk et al. likewise conclude that normalized GSR is correlated with fun, and “reflects both emotional responses [and] cognitive activity” [5].

A more qualitative measure of engagement is Mihaly Csikszentmihalyi’s concept of “flow” – the state in which “people become so involved in what they are doing that the activity becomes spontaneous, almost automatic; they stop being aware of themselves as separate from the actions they are performing” [1]. Steven Poole, among others, has argued that gameplay, as “a kind of high-speed meditation,” is perfect example of flow” [7]. Thus, we infer that some type of distracter test – testing the player about information external to the game world – will be a good measure of player engagement.

Hypothesis
Based on related work and our own personal experience, we expected that larger-scale input devices
would be both more enjoyable and more engaging. We also expected that private play would be more engaging than public play since an audience provides distractions.

**Experiment Design**

Our study consisted of sixteen undergraduate subjects (5 female) from Stanford University, randomly split into four distinct condition groups of four students each: 1) tablet PC, alone 2) tablet PC, in front of an audience 3) touchscreen wall, alone 4) touchscreen wall, in front of an audience. We also ran twelve pilots to solidify our study design and debug the game. In each session, a subject was asked to play Squares Attack, a simple game developed for this study. Subjects were given two minutes to become familiar with the game using standard mouse input. Subjects were then asked to play the game using stylus input for seven minutes while wearing GSR sensors on their non-dominant hand. For the large-scale interface conditions, participants played on a 70” Smart Board™ touchscreen wall; for the small-scale interface conditions, participants played on a 10.4” Compaq™ TC1000 tablet PC. In the public play conditions, both the experimenter and an observer watched for the entire duration; in the private play conditions, the subject was left in the room alone while playing the game. To minimize differences between the wall and tablet conditions, users played while standing.

**The Game: Squares Attack**

The objective of the game is to drag a white circle around the screen, avoiding collisions with moving red squares. For every red square that the white circle hits, the player loses points; for every two seconds that pass collision-free, the player gains points. In addition, players must diffuse yellow bombs that randomly appear on screen before they explode, or points are lost. After seven minutes of gameplay, the game ends and instructs the user to remove GSR sensors.

We decided to develop our own game so that we could log gameplay data such as score, percent of screen used, and total distance. We designed a simple game to level the playing field between expert and novice gamers. To encourage stylus use and active play, the game was designed to be fast-paced. By creating our own game, we could replicate the same sequence of game events by seeding the random number generator with the same value. Because a player cannot “lose,” all participants played for the full duration.

**Evaluation Metrics**

Four evaluation metrics were used in the study. Galvanic skin response was used to evaluate enjoyability, while a distracter test was used to evaluate engagement. In addition, self-reported Likert scale scores and in-game telemetry were used to reinforce collected data.

**Galvanic skin response (GSR)**

We developed a simple and inexpensive sensor device for measuring GSR. The GSR sensor circuit was built using a d.Tools board and provided schematics [4, 9]. Three 20k ohm pots allow adjustments to the current that flows through a subject’s body, serving as a calibration mechanism among individuals. Attached to the circuit are two finger probes. This design was inspired by Gasperi’s Velcro-backed probes [2]. GSR data was sampled at 50 hertz, which generated a data point every 20 ms. Sensor data was logged by
Exemplar, a sensor prototyping tool developed at Stanford University [3].

Distracter test
Over the course of the game, a series of characters (q, c, a, w, r) were presented as watermarks in the center of the screen. Center placement was used to create similar ocular angles between wall and tablet conditions. We assert that this is a good measure of engagement because players in a state of flow will be less likely to notice background information dissociated from the actual gameplay.

Self-reported Likert survey scores
After completing the session, subjects completed a short survey consisting of eight questions, each based on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). The questions asked for subjective reporting on engagement, enjoyability, past gaming experience, perception of skill and performance, and preference for public versus private play.

In-game telemetry
Score, percentage of the screen used, and total distance dragged were recorded in order to gauge if performance affected the gameplay experience.

Results

GSR data
GSR readings were first normalized to account for individual physiological differences, as well as the more physical nature of wall gameplay as compared to tablet gameplay. We use the following normalization:

\[
\text{Normalized GSR} = \frac{\text{Mean} - \text{Min}}{\text{Max} - \text{Min}}
\]

The mean was calculated over all data points for a given individual. In essence, the normalized GSR value represents a percentile score that describes the level of the participant’s sustained arousal, as compared to their personal maximum.

On average, touchscreen wall users exhibited higher GSR percentile score (57.5%) than tablet PC users (43.6%). Using a one-sided t-test, the result was statistically significant (\( p = 0.05 \)). Because of this increase in emotional arousal, we infer that the touchscreen wall is more enjoyable.

Participants in the public condition exhibited higher GSR scores (54.2%) than private condition participants (46.9%), but this result was not statistically significant (\( p = 0.161 \)).

Distracter test data
On average, touchscreen wall users remembered fewer letters (0.5 letters) than tablet PC users (1.75 letters). Using a one-sided t-test, the result was statistically significant (\( p = 0.037 \)). Thus, we infer that wall users were better able to achieve a state of flow, and were therefore more engaged.

Participants in the public condition remembered fewer letters (0.75 letters) than private condition participants (1.5 letters), but the result was not statistically significant (\( p = 0.146 \)).

Self-report data
On average, touchscreen wall users reported enjoying the experience more (3.9 self-rating) than tablet PC users (3.4 self-rating). Using a one-sided t-test, the result was statistically significant with a \( p \)-value of
This result corroborates the GSR data, which also suggests that the touchscreen wall was more enjoyable.

Similarly, touchscreen wall users reported that they were more engaged (4.5 self-rating) than tablet PC users (4.13 self-rating). Using a one-sided t-test, the result was statistically significant with a p-value of 0.062. This result corroborates the distracter test data, which also suggests that the touchscreen wall was more engaging.

Participants in the private condition reported that they were more engaged (4.5 self-rating) than public condition participants (4.13 self-rating). Using a one-sided t-test, the result was statistically significant with a p-value of 0.062. This result stands at odds with the distracter test data (private condition participants were able to remember more letters). However, that result was not statistically significant.

**Telemetry data**

Wall users scored worse than tablet users by a significant margin (p-value = 0.019). They also rated themselves as less skilled, although the margin was not significant. Thus, greater enjoyment in the wall gameplay experience cannot be explained by frustration with performance level or lower score.

Telemetry also shows that participants in the private condition scored better than public condition participants by a significant margin (p-value = 0.060). Perhaps more interestingly, however, participants in

<table>
<thead>
<tr>
<th>Gameplay Condition</th>
<th>GSR Percentile</th>
<th>Enjoyed Game*</th>
<th># Letters Recalled</th>
<th>Engaged in Game*</th>
<th>Easy to Learn*</th>
<th>Good at Game*</th>
<th>Score</th>
<th>Screen Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Private Wall</td>
<td>0.487</td>
<td>4</td>
<td>0.75</td>
<td>4.75</td>
<td>4.75</td>
<td>3.25</td>
<td>18138</td>
<td>67.67</td>
</tr>
<tr>
<td>Private Tablet</td>
<td>0.451</td>
<td>3.25</td>
<td>2.25</td>
<td>4.25</td>
<td>5</td>
<td>3.75</td>
<td>20913</td>
<td>74.64</td>
</tr>
<tr>
<td>Public Wall</td>
<td>0.663</td>
<td>3.75</td>
<td>0.25</td>
<td>4.25</td>
<td>4.25</td>
<td>3.25</td>
<td>16613</td>
<td>47.36</td>
</tr>
<tr>
<td>Public Tablet</td>
<td>0.361</td>
<td>3.5</td>
<td>1.25</td>
<td>4</td>
<td>4.75</td>
<td>3.5</td>
<td>18688</td>
<td>69.01</td>
</tr>
</tbody>
</table>

* = self-reported on Likert scale (1=Strongly disagree; 2=disagree; 3=neither disagree or agree; 4=agree; 5= strongly agree)

Table 1. Aggregated data for various testing groups. Data sources include normalized GSR score, self-reported responses via Likert survey, distracter test based on letter recall, and in-game telemetry.

![Figure 5: Enjoyment and engagement in tablet vs. wall play. Self-reported scores corroborate the GSR data and distracter tests results.](image-url)
the private condition utilized a larger portion of the screen space when dragging the circle (71.1%) compared to public condition participants (58.2%). Using a one-sided t-test, the result was statistically significant with a p-value of 0.029.

Limitations and Future Work

The study suffered from several issues that might limit how generalizable the results are. First, the game ran about 40 seconds longer on the tablet PCs than it did on the touchscreen wall. Thus, it is possible that tablet users simply had a greater window of time in which to become bored. Second, even though the participants were randomly assigned, users in the public tablet condition classified themselves as more novice gamers when compared to the other three conditions. Indeed, the difference was significant. As we expand the scope of this study, we hope that larger sample sizes will minimize such abnormalities.

In future iterations of the study, we hope to examine a range of “public” gameplay environments. For example, two experimenters comprise a very different audience than a large crowd of rowdy peers. In fact, it is quite possible that public versus private settings would yield more significant differences if the public condition better replicated a real arcade environment.

Finally, we hope to run the experiment on a variety of different game genres in order to confirm that these results are not merely particular to Squares Attack. From a game design perspective, it would be useful to know whether different genres require different handling of interface scale and performance environment.

Acknowledgements
We would like to thank Björn Hartmann for all his help with building the Galvanic Skin Response sensor, and Scott Klemmer, David Akers, and Lilly Irani for their support and guidance throughout the process.

References