Clicking Circles Quickly: Determining Player Improvement in a Rhythm Game

Tommy Adebiyi

12/9/19
2.671 Measurement and Instrumentation
Tuesday AM
Prof. Peacock

Abstract

Osu! is a skill-based rhythm video game about clicking circle-shaped objects to a song's rhythm according to a "map" created by the game's active community. Players compete for the highest spots on map leaderboards and overall rankings and improve over time, so defined patterns in score improvement are valuable information for the playerbase. To quantify a player's improvement cycle on a map, the player attempted maps that were both of hard and average difficulty relative to their skill level. For each map, the player sight-read it then played it 10 to 15 more times, recording their performance each time. From fitting the data, the player's best performance is between attempts 6-9, where the player performs upwards of 26% better than the first attempts. The player also has greater improvement in consistency with average difficulty maps compared to hard maps, with a hypothesized mean difference of 1% with 95% confidence.

1. Introduction

The road to becoming an experienced rhythm game player involves many hours of practice and dedication. While some people play rhythm games such as Guitar Hero, Rock Band, and Beat Saber casually to listen and play along to their favorite songs or discover songs from artists previously unknown to them, dedicated players also play for the rush of clearing more and more difficult maps of songs and the satisfaction of being considered a high level player by the community. The skill ceiling of rhythm games with a community content creation scene is often unattainable and increases over time, so even the best players must constantly improve their playing ability in order to stay at the top of the performance ranking board.

One such rhythm game with an active community is *osu!*, which is a game players "click" circle-shaped "hit objects" to the rhythm of a song according to a preset arrangement of objects, called a map, created by the community. A screen shot of example gameplay can be seen in Figure 1, and details about the gameplay will be explained in the background section. The objects give score and add to your combo values after a successful hit. Players typically aim at the objects with a mouse or a drawing tablet, and click on the objects soon, about 500-600 milliseconds depending on map settings, after they appear with one of two keys on their keyboard.

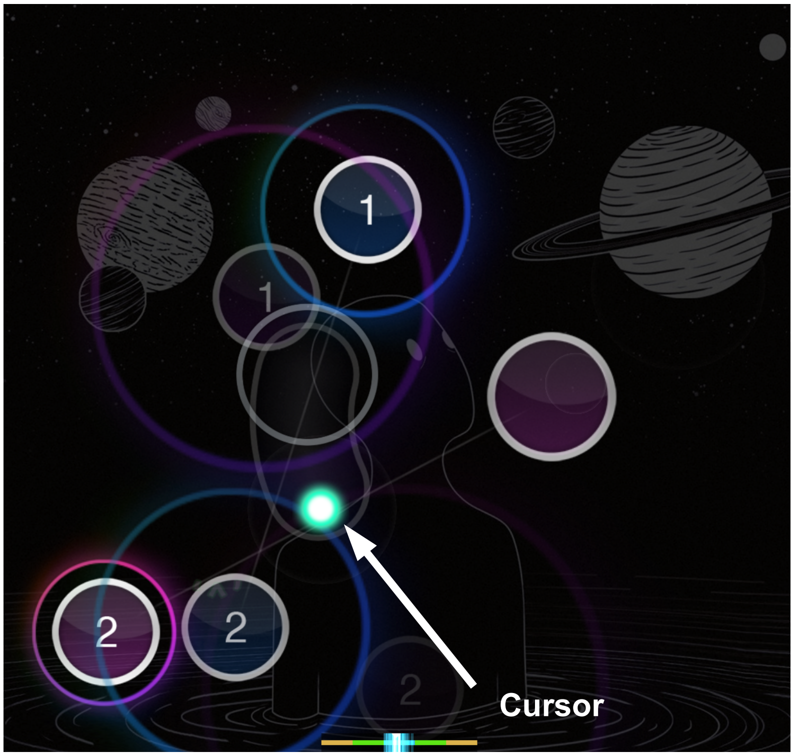

Figure 1: An example of gameplay in the game osu!. This is a picture of a very difficult map, but the basic gameplay elements are still the same: players must aim and click the circles to the rhythm of the song playing in the background. The small filled circle is the player's cursor, the circles with numbers in them are hit objects which are surrounded with approach circles that are a visual signal for when to click.

Since the maps in *osu!* do not vary with each attempt of the song, one major factor in a player's performance in the game is their memory. Despite the relative simplicity of the game, it is near impossible to predict the next pattern in a map while sight-reading, as each map is custom-made for each song and each difficulty. While having muscle memory of similar patterns observed from previous game experience could improve potential performance on a map new to the player, having prior play experience and the opportunity to develop muscle memory on a specific map has a greater influence across multiple attempts on a map.

To characterize how much of an impact map knowledge and relative difficulty has on play performance and how much of a factor fatigue is, a subject will play a set of maps that are average and challenging for their skill level based on the in game difficulty rating ("star rating*")* and their performance record on their account. For each map, they will play the song through in its entirety 10 times at the minimum. With each attempt, accuracy, highest combo, and the in game score will be tracked. This data will be used to validate that more improvement is seen in average difficulty maps compared to hard difficulty maps, and which attempts players are expected to perform their best.

2 Background

2.1 osu! Gameplay

In a typical song in osu!, players locate a hit object on their screen, aim their cursor at it with one of two common input devices, a mouse or a drawing tablet, and click on it by tapping one of two set keys on their keyboard according to a map created around a song. Hit objects include the circle, where players click on it at the location it appears on screen, the slider, where players must follow the path of after clicking on its starting position, and the spinner, where players must gyrate their cursor around the center of the screen. An example of a circle and a slider can be seen in Figure 2. An approach circle closes in on the hit object as the correct point in time for a perfect hit approaches, as seen in Figure 3. On most maps, objects appear on the screen less than 0.5 seconds before they need to be hit [1], which means the game puts strain on a player's hand-eye coordination, reactions, and hand movement speed.

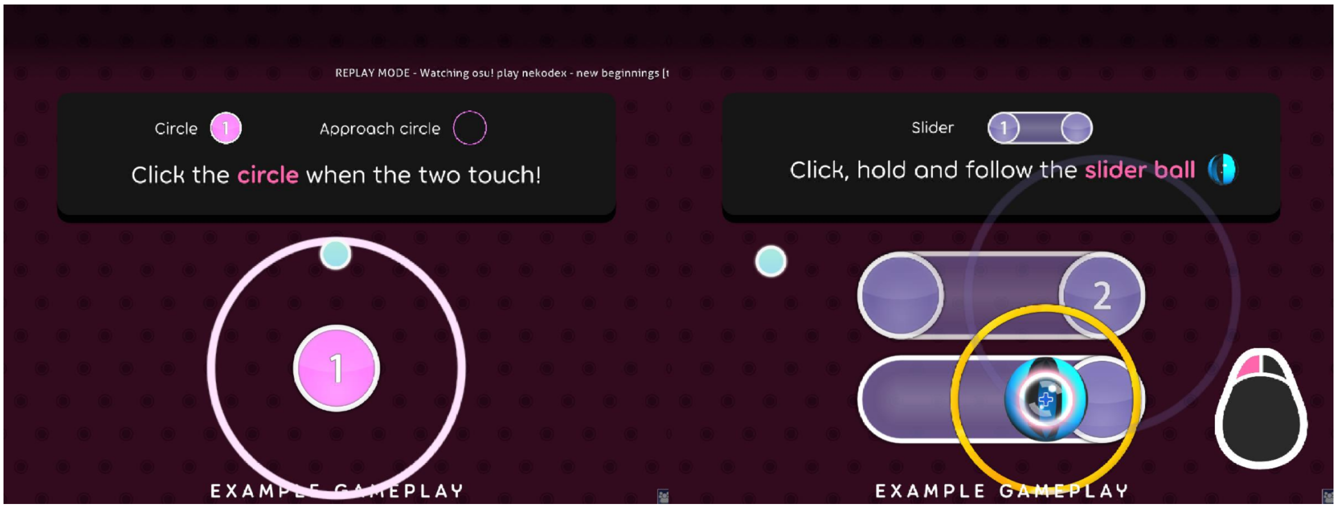

Figure 2: Examples of the circle (left) and the slider (right). The larger circle concentric with the hit object is called the approach circle, and it indicates at what time the object should be clicked.

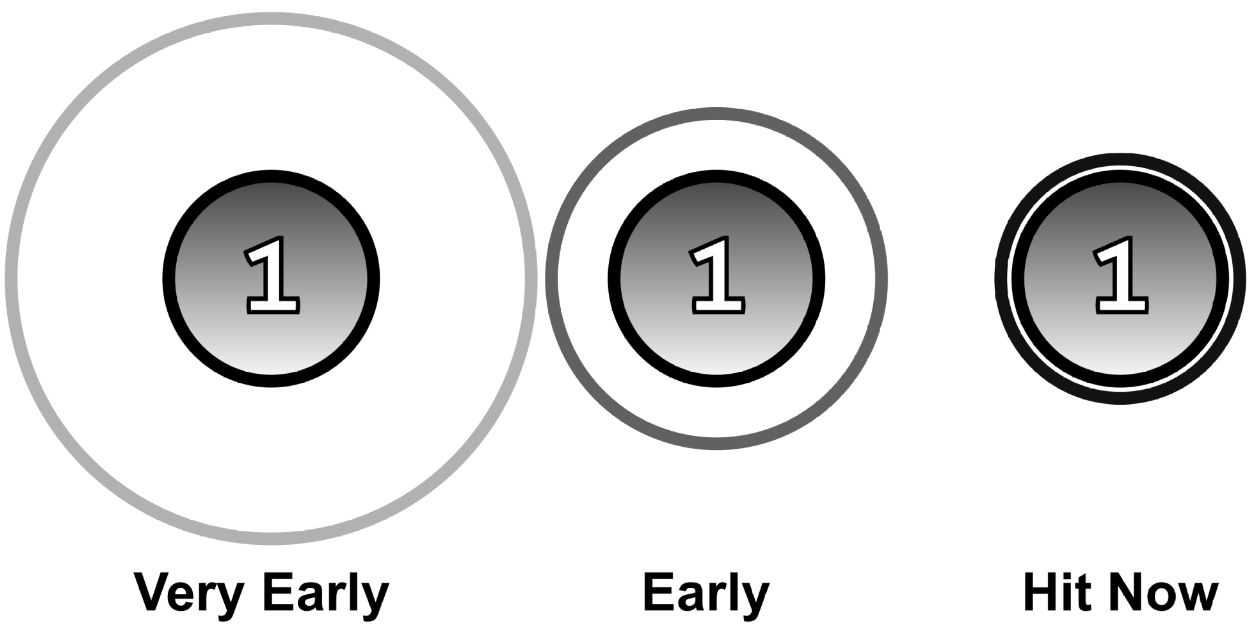

Figure 3: Example progression of an approach circle encircling a hit object. The approach circle decreases in size and increases in opacity until it matches the size of the hit object, which is when the player is supposed to tap their keyboard in order to score a perfect hit.

Maps are also composed of different patterns derived from the placement of the hit objects, which all together shape the style of a map. A "stream" is a set of closely spaced circles that draw out smooth curves on the screen and are usually matched with sixteenth notes in the music, as seen in Figure 4b. Short streams composed of 3, 5, or 7 circles are often called "bursts". This pattern requires rapid tapping on the two tapping keys on the keyboard. For example, in a song with a tempo of 200 beats per minute, which is slightly faster than the average song in the game, the player has to tap just over 13 times a second in order to accurately hit the objects, putting strain on the player's hand used for tapping the keyboard. This is no where near the limit of a player's tapping speed though. Players have been recorded hitting streams of object songs over 300 BPM with high accuracy.

Another common pattern is a "jump", as seen in Figure 4a, which is a group of hit objects mapped to eighth notes in a song that require the player to quickly span part or all of the screen to hit. Jumps strain the opposite hand, the hand used for aiming the cursor, more and the tapping hand less. Both of these patterns are acute sources of fatigue that may affect a player's performance during a long play session.

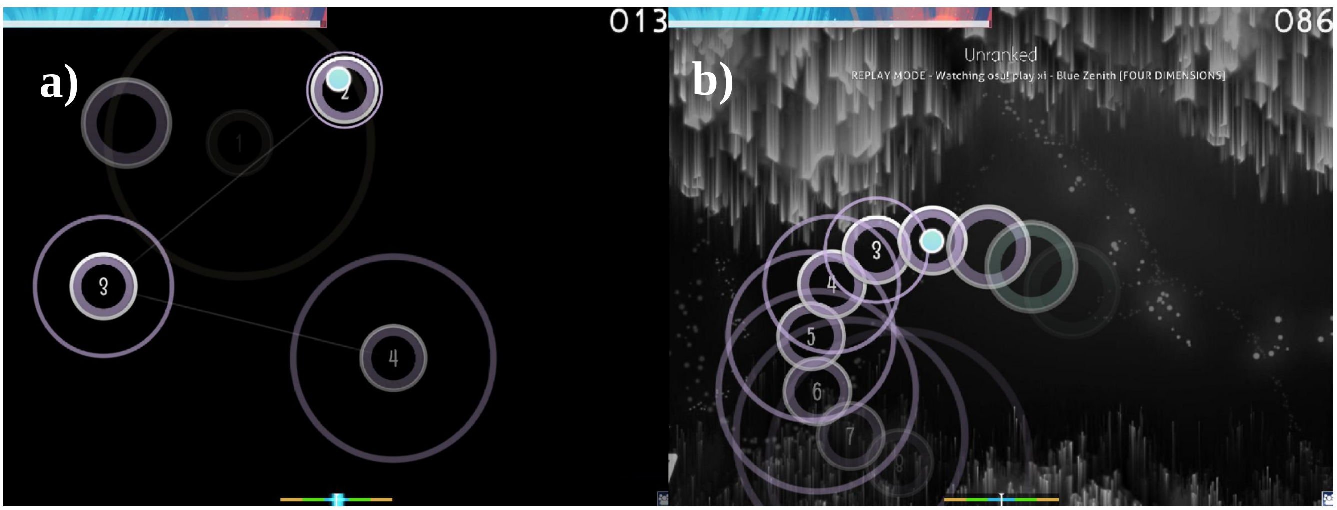

Figure 4: Examples of common patterns called jumps (a) and streams (b). The source of difficulty in both types of patterns vary. The distance between circles and nonuniform patterns in jumps are its source of difficulty, while the mechanical tapping ability and smooth aiming required for streams make it a challenge.

The ultimate goal with each map for a player is to achieve a full combo, which is when a player successfully hits every object in a map, and hit every object with "perfect" timing, which, depending on map settings, the game quantifies as typically 20-40 milliseconds off rhythm at most, in order to place highest on the leaderboards and gain more performance points on their personal profile. Missing a singular object can be extremely punishing as a result, since a miss resets combo to zero and combo has a large impact on score.

2.2 Scoring Systems

In osu!, tapping a circle on time with the music will award points and add to the combo multiplier. Missing an object will reset the combo back to 0 and award no points. In addition, the player receives a score based on how on time they hit the circle: 300 for very on time, a perfect hit, 100 for slightly off the music, and 50 for very off the music. The exact timing windows depend on the overall difficulty (OD) setting of the song. Higher OD means smaller timing windows. As seen in Figure 5, the game provides visual feedback of the score on a hit object. The game also plays a sound when the object is hit, so hits that are slightly off the rhythm will sound off rhythm. The score divided by 300 is the accuracy with which the player hit the object, so the accuracy of the whole play is the average of the accuracy of all the hit objects. Sliders also award additional combo and points based on the length of the object, but they only report one score.

Figure 5: An example of receiving scores of 100 on circles in a burst composed of stacked circles as a result of hitting the circles slightly too late.

Rather than purely using the in game scoring system, accuracy will be also be tracked. The normal scoring system multiplies the current combo by the score (300, 100, 50) achieved by hitting the object, which objects later in the combo of objects have a significantly larger impact in the final score. One miss could cause a huge difference in scores, as even 90 second maps can easily have over 500 maximum combo. Accuracy works much differently than that, as on each object a score 300 are treated as 100% accuracy, 100 is 33.33%, and 50 is 16.67%. Accuracy is the average of the individual note accuracy percentages. Therefore, accuracy provides a gauge of performance based on individual objects, score represents a player’s ability to hit a long string of objects. Figure 6 shows an example results screen that reports these statistics.

One more important statistic that the game tracks is called *unstable rate* (UR). It is the standard deviation of the hit timing error on each hit object in milliseconds times 10. This can be useful for tracking tapping consistency over the length of a song [2].

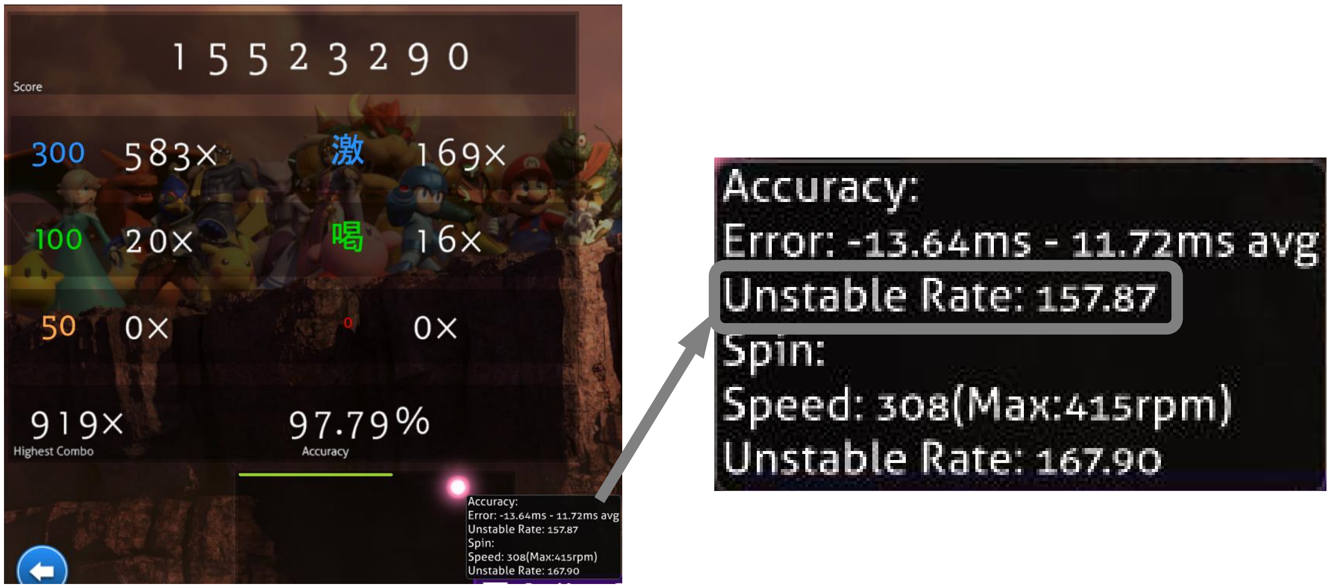

Figure 6: A results screen showing score at the top, scores on each individual object in the middle, combo and accuracy near the bottom, and unstable rate in the small box next to the cursor. The box is enlarged for detail below the image, and in the box, the unstable rate number under Accuracy (not spin) is the important statistic. Accuracy, combo, score, and unstable rate are all statistics that this experiment will track across plays.

2.3 Map Difficulty and Star Rating

Star rating is a measure of a map's difficulty. It is determined by algorithms that recognize patterns in maps that are particularly straining to the player and takes into account their relative difficulty. It also considers variables set by the map creator that are not hit objects such as approach rate, the length of time between a hit object appearing and the perfect time to hit it, circle size, and overall difficulty, which affects the timing window (higher overall difficulty means the "on time" hit window spans less time, for example) [3].

2.4 Minimizing Reaction Times

The fast-paced nature of osu! means that tracking each object by purely central vision can be tiring and confusing. Therefore, more performance can be extracted by using muscle memory, also known as motor memory, and reactions based on objects in the periphery of vision. Research has shown that after some training, muscle memory can be used to point at targets outside of a subject's vision. In said experiment, further improvement in location accuracy was seen after the three sightless runs after which improvement was negligible [4]. In addition, human eyes are very strong at recognizing movement in their periphery. While the eye's peripheral vision is much worse at perceiving visual detail, it is still able to judge the location of objects relative to itself and recognize basic patterns such as colors and simple shapes [5], similar to osu!'s hit objects! In the pursuit of minimal reaction time and optimal osu! performance, one must combine muscle memory and their peripheral vision, which are expected to require some time to adapt to a new song.

3 Experimental Design

Osu! is primarily played on PC/Mac/Linux and the inputs are transferred via a keyboard for tapping and a drawing tablet or mouse for aiming. The setup can be seen in Figure 7. It is composed of a Windows PC, an XP Pen G640 drawing tablet for translating hand movements to x-y aim on the screen and a mechanical keyboard with Outemu Brown switches for tapping inputs. The osu! client provides a replay file that tracks all the movements of the cursor and all the tap timings in milliseconds. This data can be made human readable after being parsed by external community-made software.

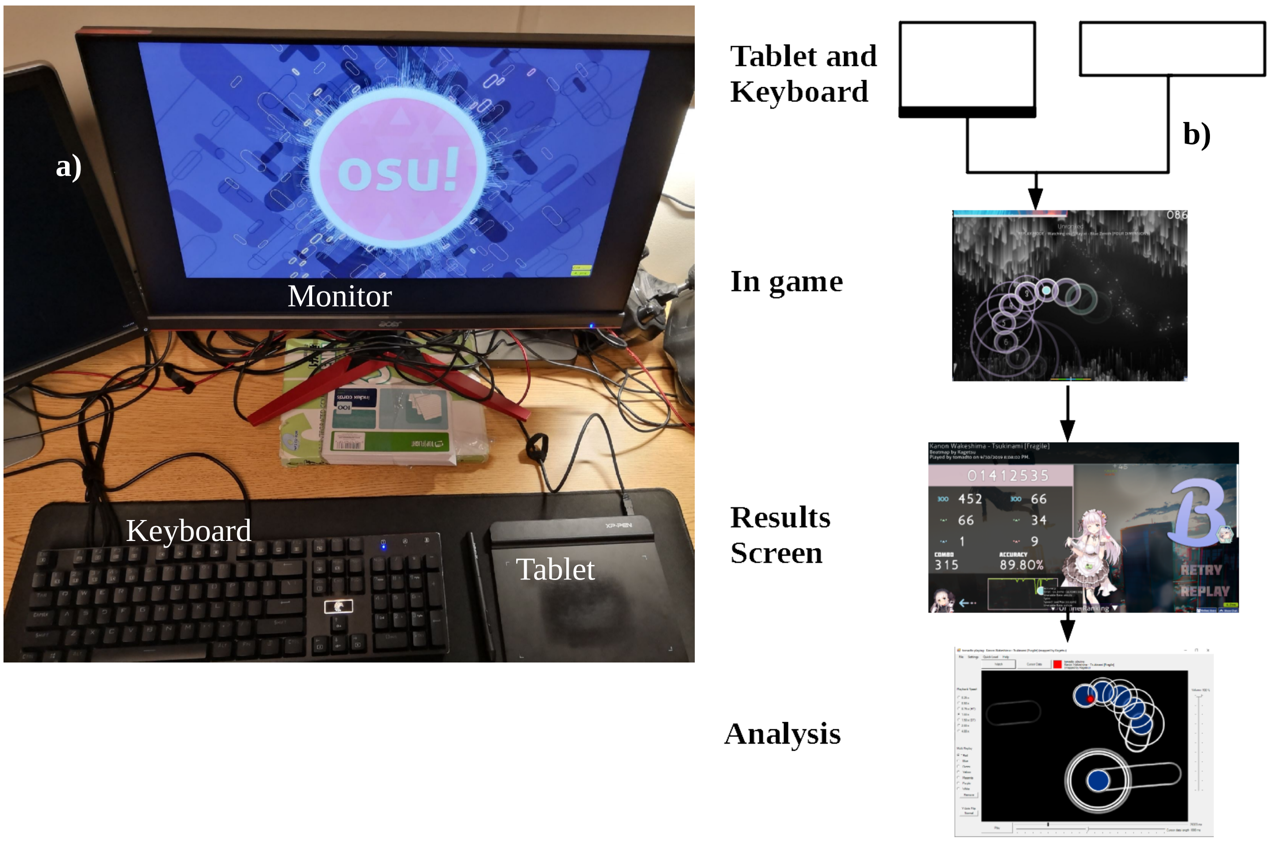

Figure 7: a) A picture of the setup used for experiments. The tablet is used for aiming at objects, and the keyboard is used for tapping on them. Objects appear on the monitor in time with music. b) Flow chart of data collection and analysis process. Data from the play as a whole is available on the results screen after completing the map. This can then be used to run analyses in Excel and MATLAB.

For each trial, a map was arbitrarily picked within the rough star rating defined as either hard or average, which for this data is 5.5* to 5.9* for hard and 4.5 to 4.9* for average difficulty. The player must not have played the map before in recent memory, but prior to playing the map, the player is allowed to listen to the song in order to minimize the effects of varied previous song familiarity. Maps were to be roughly 90-120 seconds in length. (not to be confused with the length of the song). The sight-reading attempt provides a gauge of the effect of map knowledge but was also used to quickly qualitatively evaluate whether the map was a good fit for the player in terms of interest in song and structure of the map for the experiment. Maps that did not fit well (e.g. maps with extremely unusual patterns, songs that the player found not enjoyable to listen to or maps that were overly frustrating to play) were not recorded.

In the data below, six maps were played all by the same player. Three of the maps were categorized as a hard difficulty for the player while three of them were of average difficulty for the player. The maps are also labeled in the figures ahead based on the style of map which can be either stream, burst, or a hybrid of the two styles. As stated in the background, the three statistics for each attempt in each map tracked will be score, accuracy, and unstable rate. These can be simply recorded from the in game results screen seen after each attempt.

4 Results and Discussion

With these trials, the goal is to quantify the improvement patterns of performance of players on maps of average and hard difficulty for the player, and determine if there is a significant difference in these patterns of improvement between hard and average maps. Prior research done by others related to using muscle memory to aim at a single target at a time found that there was noticeable improvement until the third attempt. However, a map in osu! involves aiming at many targets over a couple minutes of play time, not just one target. Therefore, it is predicted it that it will take longer than three attempts for a player to learn a map enough before it becomes pure muscle memory.

4.1 Unstable Rate (Consistency)

One metric that measures a form of performance and improvement is unstable rate (UR), which, as mentioned in the background, is a measure of tapping consistency in a map. It will also be referred to as simply "consistency" in the future. Lower UR is better UR. Figure 8 plots the unstable rate of all six maps and all their attempts. All the average difficulty maps had a significant fit equation which can be seen in Table 1, and all of the fits indicated improvement in consistency for about the first 7 attempts, with the player becoming upwards of 19% more consistent compared to the first attempt. Afterwards, since two of the fits are quadratic, the improvement turned into a worsening of performance. Since these attempts were done in rapid succession, this could be due to the player fatiguing physically. The data support that streams and bursts cause more physical strain since it involves tapping fingers rapidly and with control.

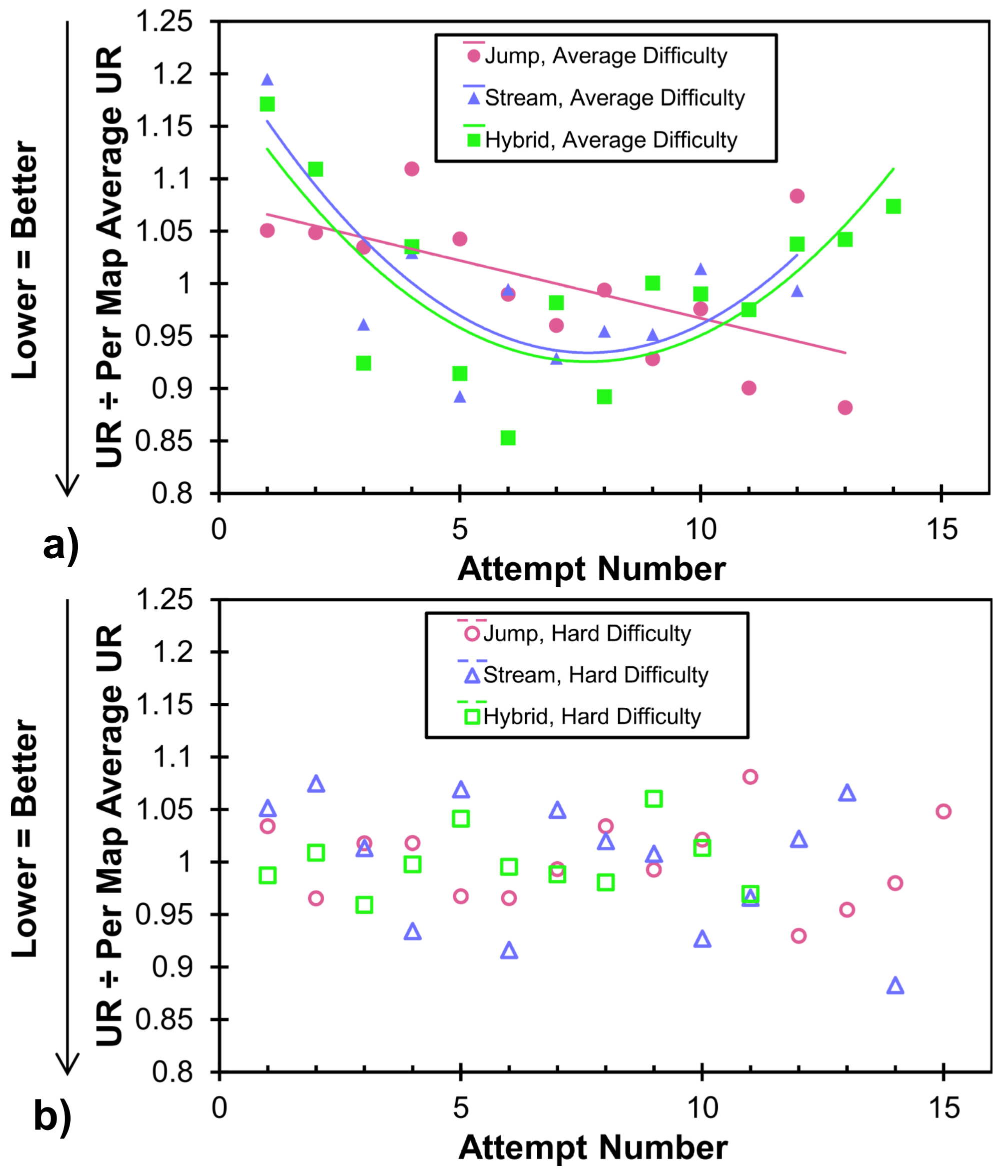

Figure 8: Plot of unstable rate divided by the average unstable rate across all attempts on one map versus the attempt number on each map. 8a shows all of the average difficulty maps and 8b shows all of the hard difficulty maps. The division was done to put all the data on the same scale across maps while preserving the data shape from the raw data. The maps with statistically significant fits had their fit lines plotted. The plot shows a rough trend in change in UR over time. Some of the fits show that some attempts reach a minimum within attempts 5-10 which is upwards 26% lower than the first attempt before leveling out or trending to an increase in UR, while the average difficulty jump map is has a linear fit, indicating that streams and burst are more fatiguing physically.

Table 1: Table of the equations for the fit lines in figure 8.

	Quadratic	Linear	Constant
Jump, Average Difficulty	--	-0.0110±0.0094	1.077±0.074
Stream, Average Difficulty	0.0050±0.0031	-0.076±0.041	1.23±0.12
Hybrid, Average Difficulty	0.0046±0.0024	-0.070±0.037	1.19±0.12

When performing student t-tests based on the data plotted above, two populations had significant differences in mean. Figure 9 shows that across all 6 maps, attempts 6-10 were 1.6% more consistent than attempts 1-5 with 95% confidence. Attempts greater than 11 were not significantly different in consistency from all the other groups of attempts. Figure 10 shows that within attempts 6-10 across all maps, the player was more consistent on the maps of average difficulty versus the maps of hard difficulty.

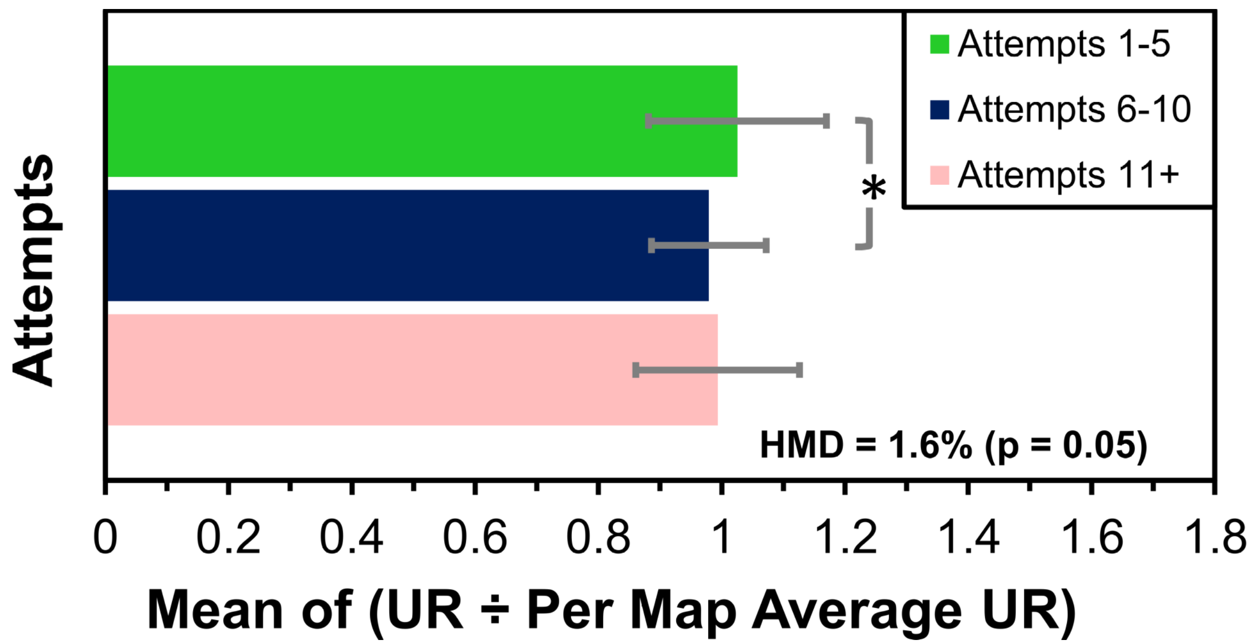

Figure 9: Bar graph of Average (Mean) UR in 3 bins of attempts across all maps with uncertainty error bars. T-tests showed that there was a significant difference in mean performance between attempts 1-5 and 6-10. The hypothesized mean difference with 95% confidence was 1.6%. This means that the player improved in consistency after the first 5 attempt. There was no significant difference in consistency between attempts 11+ and the other bins of attempts, indicating that the player lost consistency, possibly due to fatigue.

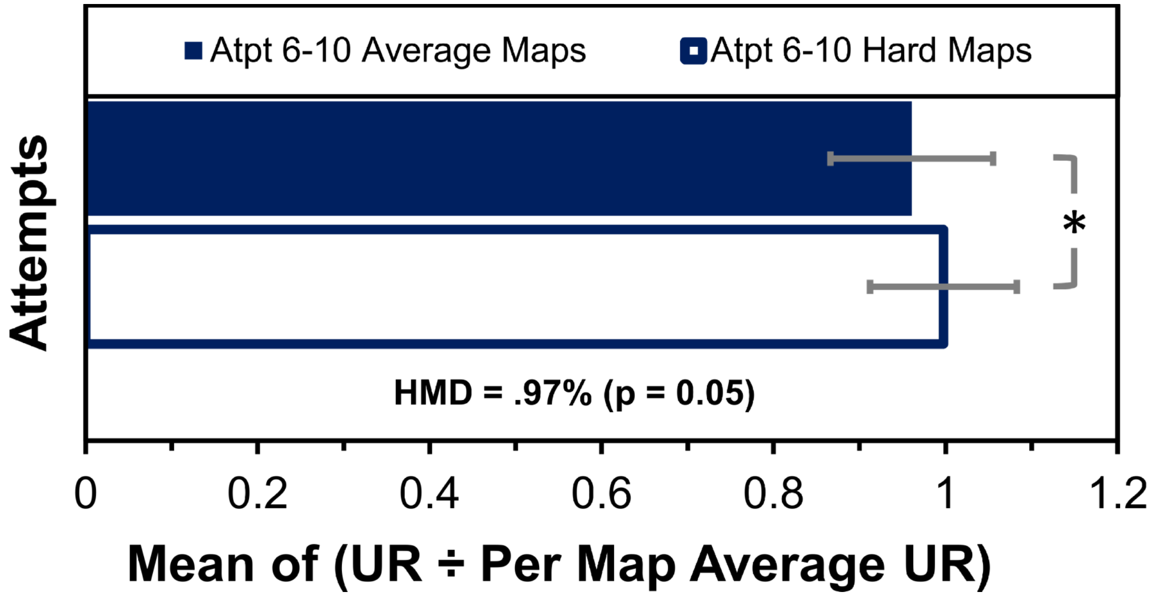

Figure 10: Bar graphs of the Average (Mean) UR of attempts 6-10 in all the maps, grouped into hard and average difficulty maps. A t-test showed that there is a significant difference in performance between hard and average maps at this range of attempts. The HMD was 0.97% with 95% confidence. This shows that a player has a greater ceiling in level of consistency on maps that are at their level of play.

4.2 Score

Figure 11 shows a scatter plot of all the scores from all the attempts across all the maps, using the same division used in UR to align the data with each other. Two hard maps and one average difficulty maps had significant fits as seen in Table 2, but all with different variations of fit, which means it hard to draw a solid general conclusion based on this plot. The drastic concave up quadratic fit for the average difficulty jump map indicated that the player may have ran into a mental barrier, becoming frustrated at the map to the point where it affected performance, before improving in performance with later attempts again. scatter in the data could be attributed to the influence of combo on score, and the maps with significant fits had a tight spread of combo (which could mean that the player consistently scored a very high or very low combo across all the attempts, usually the latter with hard maps). One area that could be investigated is the increase in spread of data in the later attempts in comparison to the earlier attempts.

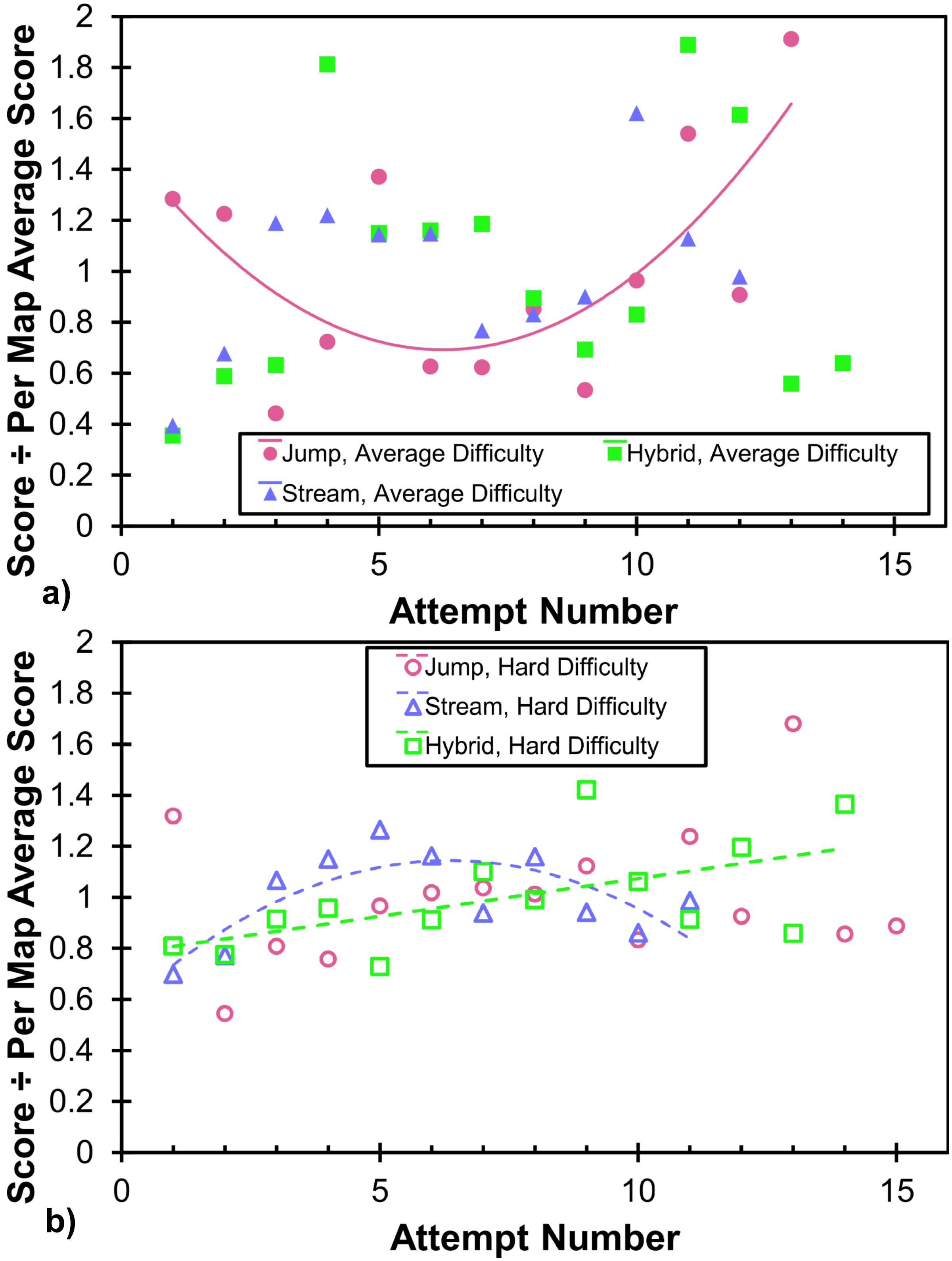

Figure 11: Plot of score ÷ per map average score versus attempt number. 11a shows the three scores of the three average difficulty maps, and 11b shows the hard difficulty maps. Maps with statistically significant fits had their fit lines plotted. Score generally increased over time,

however in the case of the average difficulty jump map, the player's score became as high as 46% worse than the first attempt before improving again. This could be due to a more complex mental cause, essentially the player being frustrated at the game. The two statistically significant fits for the hard difficulty maps show a general pattern of player improvement over time, however, the hard stream map score drops over time, again indicating player physical fatigue.

Table 2: Table of the equations for the fit lines in figure 11.

	Quadratic	Linear	Constant
Jump, Average Difficulty	0.021±0.017	-0.26±0.25	1.52±0.76
Stream, Hard Difficulty	-0.014±0.010	0.18±0.12	0.57±0.33
Hybrid, Hard Difficulty	--	0.030±0.025	0.78±0.22

As shown in figure 12, when performing student t-tests, the only conclusion for a significant difference in means occurred between the first 5 attempts and attempts past 11 for all the maps. With 95% confidence, the player achieved 3% higher scores for attempts past 11 in comparison to attempts 1-5. This does prove that the player indeed improves over time primarily in ability to hold a long combo of objects.

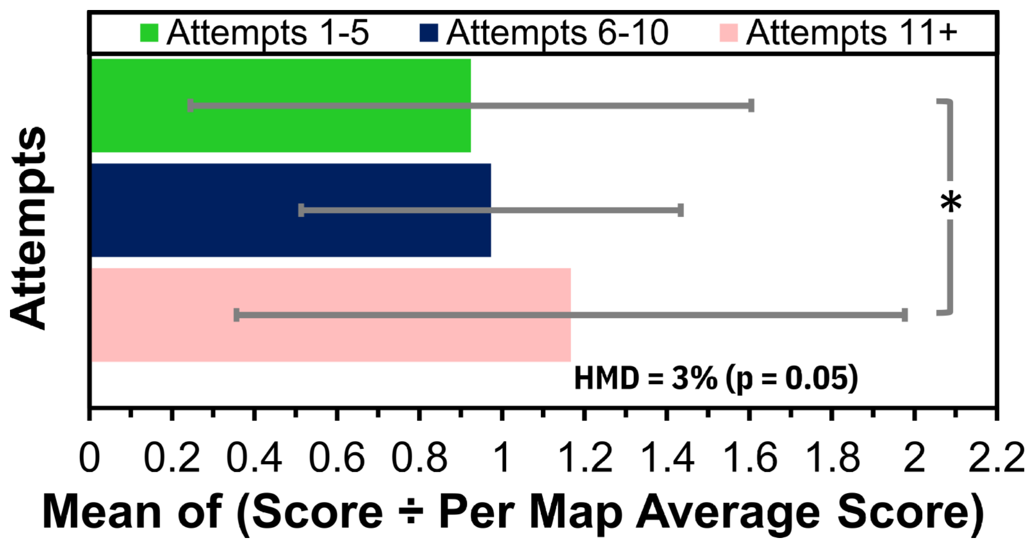

Figure 12: Bar graphs of the mean of binned attempts of score from the plot in figure 11. Colors of the bar have the same meanings as in figure 9. A T-test showed that there is a statistically significant difference in mean score between attempts 1-5 and attempts past 11 for across all maps, showing that the player improves over time in their ability to hit a long string of objects. The HMD was 3% with 95% confidence.

4.3 Accuracy

Figure 13 shows the scatter plot of (tapping) accuracy versus attempt number (using the same division to align all the maps together). From the plots and fits with equations in Table 3, accuracy shows a general trend upwards from the first few attempts peaking at at most a 5.9% increase of accuracy from the first attempt. Much like consistency, the performance drops off after 6-8 attempts for 2 of the fits, and the distribution of accuracy seems to spread at the end like score. T-tests produced no significant hypothesized mean differences with 95% confidence between different sets of attempts or different difficulties within each set of attempts from accuracy, so there are no bar graphs of binned attempts. Purely from the fits, the player therefore improves in accuracy before fatigue takes over in maps with some streams and lowers accuracy in the longer term for average difficulty maps.

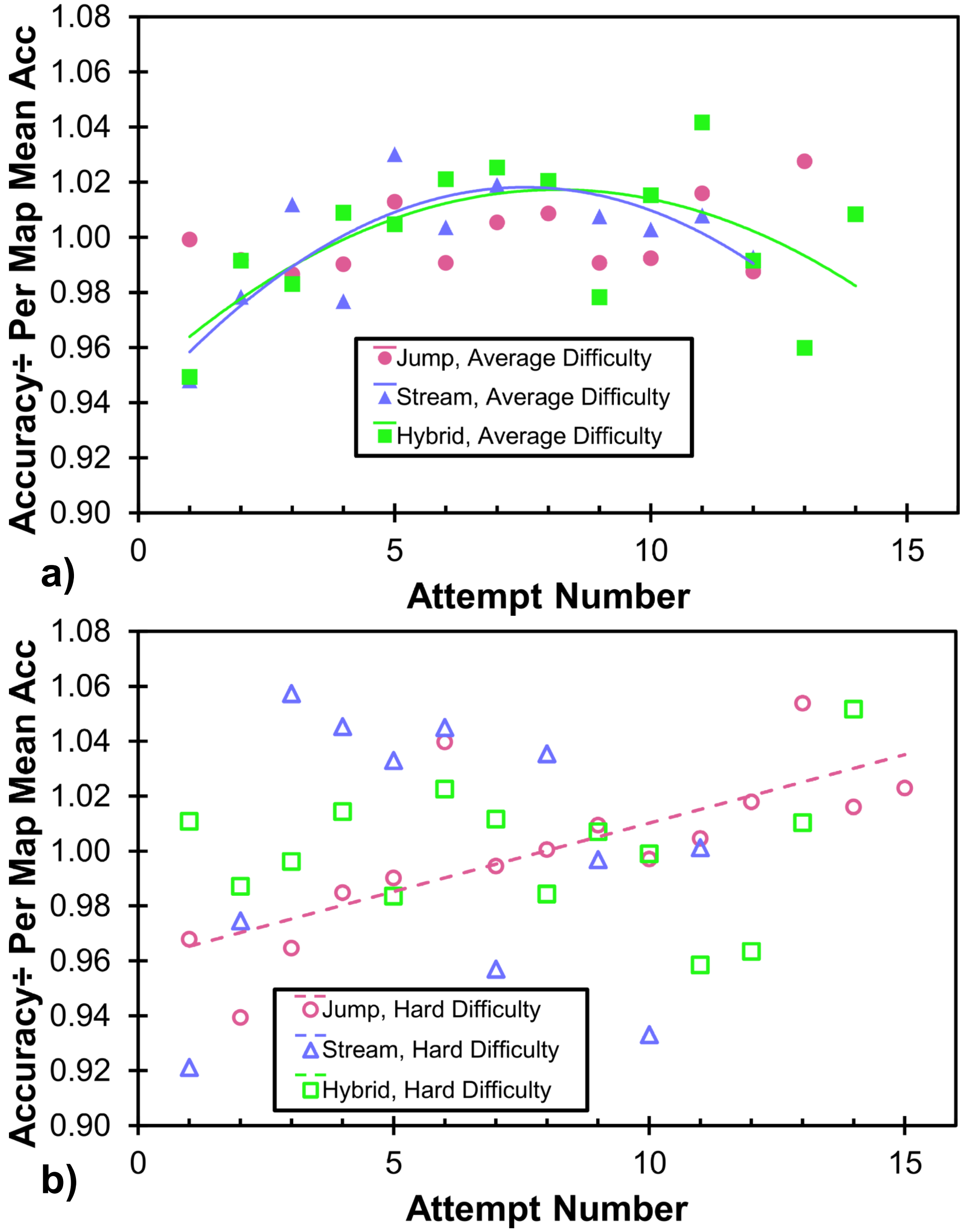

Figure 13: Plot of Accuracy divided by Mean Accuracy per map as in the other plots. 13a is all of the average difficulty maps and 13b is all of the hard maps. Statistically significant fits are shown. As with the other statistics, there is a general pattern of positive performance over time with player improving accuracy as much as 5.9% compared to the first attempts, with the maps with streams/bursts showing a drop in performance in the long term from fatigue.

Table 3: Table of the equations for the fit lines in figure 13.

	Quadratic	Linear	Constant
Jump, Hard Difficulty	--	0.0050±0.0025	0.960±0.023
Stream, Average Difficulty	0.0050±0.0031	-0.076±0.041	1.23±0.12
Hybrid, Average Difficulty	-0.00103±0.00088	0.017±0.014	0.948±0.044

4.4 Discussion

Across almost all the data, there was a general pattern of improvement for about the first half of attempts before performance dropped. Whether the map had stream sections generally indicated how much performance dropped, with jump maps often having a positively sloping linear fit. Fatigue in the tapping hand therefore negatively affects the player’s performance after many attempts in a map.

Score would be the best indicator of fatigue in the aiming hand, as a mis-aimed object will reset combo and greatly affect score, however, barring the maps with a lot of streams, score improved over time, even after worsening in the beginning, indicating that fatigue in the aiming hand does not affect performance in the same manner that fatigue in the tapping hand does.

There were minimal differences in the performance patterns between hard and average difficulty from the analysis above. The only statistically significant difference was seen in UR, where the player had greater consistency in attempts 6-10 across all maps (HMD of 0.97% with 95% confidence), showing that the ability to comfortably read maps shows signs of having an effect on performance over time between the two difficulties of maps.

Having multiple types of maps at each difficulty probably contributed to the unexpectedly high level of scatter in data. While useful at seeing if there was an impact of multiple styles of maps on performance, it would have been more beneficial to focus on one type of map in order to develop less scattered data and present stronger conclusions. Additionally, standardizing human variables such as the time of day, level of fatigue, amount of sleep, and many other factors could have strengthened the data too, but since osu! is a game focused on very precise levels of play, the data scatter was inevitable.

5 Conclusions

In terms of determining a pattern of performance across multiple attempts of a map, a player’s best performance is between attempts 6-9 as derived from quadratic fits and t-tests. Depending on the statistic referenced, the player performs upwards of 5.9% to 26% better than the first attempt. T-tests on unstable rate support this conclusion, where the player had a significant increase in consistency with a HMD of 1.6% with 95% confidence.

There are signs that there is more of a chance of improvement with maps that are at the player’s skill level, with a t-test showing that the player had a HMD of 0.97% in consistency between hard and average attempts in the same attempts bin, with the player having more consistency with average difficulty maps. However, further testing should be done to evaluate the validity of this conclusion.

Players are then expected to reach peak performance at around the 6th-9th attempt in rapid succession of a two-minute long map. They are also expected to improve more on a map that is at their skill level, emphasizing the importance of being able to read and understand the patterns in a map quickly and confidently, which is objectively easier to do with a map at the player's skill level, towards performance.

The main mode of improvement to this experiment would be more data. First take more data from maps of one style, then take more data from one player from all kinds of maps, then more data from many players. Additionally, more attempts on a map could be useful to verify long-term performance. It would be interesting to determine whether the player drops off or continues to improve in performance after 20, 30, or even 50 attempts, albeit the amount of time required to play a map that many times could introduce some unwanted variables.

Another method of evaluating performance would be to examine key missed circles in a map, finding how far off the player was, and analyzing how that changes with attempts.

6 Acknowledgments

The author gratefully acknowledges Dr. Barbara Hughey, Prof. Thomas Peacock, and Thalia Rubio for advice and support and *The Leftovers* for map suggestions.

7 References

[1] "Beatmap Editor / Song Setup," osu! [Online]. Available: https://osu.ppy.sh/help/wiki/Beatmap_Editor/Song_Setup#approach-rate. [Accessed: 26-Sep-2019].

[2] "Accuracy" [Online]. Available: https://osu.ppy.sh/help/wiki/Accuracy. [Accessed: 27-Sep-2019].

[3] "Changes to Osu! Star Rating / Performance Points" [Online]. Available: https://osu.ppy.sh/home/news/2019-02-05-new-changes-to-star-rating-performance-points. [Accessed: 27-Sep-2019].

[4] HEIDE, B. J., and MOLBECH, S., 1973, "Influence of After-Movement on Muscle Memory Following Isometric Muscle Contraction," Ergonomics, **16**(6), pp. 787–796.

[5] Lou, C. I., Migotina, D., Rodrigues, J. P., Semedo, J., Wan, F., Mak, P. U., Mak, P. I., Vai, M. I., Melicio, F., Pereira, J. G., and Rosa, A., 2012, "Object Recognition Test in Peripheral Vision: A Study on the Influence of Object Color, Pattern and Shape," *Brain Informatics*, F.M. Zanzotto, S. Tsumoto, N. Taatgen, and Y. Yao, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 18–26.