Input Latency Detection in Expert-Level Gamers

An experiment in visuomotor perception

Eryk Banatt

Stefan Uddenberg and Brian Scholl

Yale University

Department of Cognitive Science

Running Head : Input Latency Detection

Email : eryk.banatt@yale.edu

Phone : 973-945-5263

Word Count : 5205 words

Version : 4/21/17 (submitted draft)

Abstract (278 words)

As video games have evolved more and more sophisticated, a market has emerged for people to use video games as a medium for competition, called "electronic sports" or "esports." With such a market, manufacturers have started creating peripherals aimed at the highest level, much like a sports equipment manufacturer would create tennis rackets or shoes. Among these are monitors that have very short delays between an input and the screen updating with that input. The most advanced monitors have delays in the milliseconds, but some sub-communities prefer older screens over newer ones based upon the claim that the greater delay of newer monitors is noticeably distracting – there have been complaints from top-level competitors on delays of as small as 2ms, whereas most people, even in these communities, seem to not even notice delays of significantly longer than that. This project ran an experiment to assess the validity of input latency detection differences among populations, to see if there was a noticeable correlation between level of expertise and ability to detect finer and finer levels of delay. To be precise, this experiment attempted to determine if population differences in latency detection existed at all, and if they are domain-specific (i.e. gaming experts can detect it, but only in their own game) or generalizable (i.e. gaming experts are better at latency detection tasks across the board). After running an experiment, a statistically significant non-domain-specific, general perception difference was found between the expert group and the control group. These results could provide valuable insight into group differences in perception, as well as provide useful advice to competitors, tournament organizers, and monitor manufacturers with regards to what products are appropriate for events.

Keywords

Input latency; Expertise; Temporal perception; Visuomotor coordination

Background

Electronic Sports

Electronic Sports, commonly referred to as “esports” is a relatively new industry that revolves around playing video games at a professional level. As you might imagine, this is a fairly large community with roles similar to what you might find in traditional sports – Competitors, Casters, Production, Marketing, etc. As such, esports is a multimillion dollar industry with a global audience in the hundreds of millions.

Much like how traditional sports has an entire sub-industry dedicated to selling equipment (e.g. baseball bats, tennis rackets), esports too has a niche market for competitive gaming equipment that far exceeds the needs of most people. Some peripherals are heavily marketed as "gaming optimized," such as mice, keyboards, and even desks and chairs.

One particularly interesting peripheral used in this industry is the monitor. Different monitors have different amounts of delay, so players trying to maximize their advantage often pay high prices for monitors with a bit less input delay than others. Some games even have a short list of "tournament legal" monitors, since monitors with enough input delay are considered "unplayable" in a multiplayer environment. This list is often very strict, since in some games top competitors have expressed their dislike for certain monitors that deviate even a small amount from the standard (as little as 2ms).

Conventional wisdom within these communities says that stronger players have an increased ability to discern increasingly small levels of input latency compared to weaker ones (e.g. "You can't tell it's delayed by 1 frame [16.66ms] because you aren't good at the game yet"), but the upper limit of this phenomenon is highly contested even within the community. This study looks into whether this enhanced ability for input latency detection exists at all, how pronounced it is, whether it carries over to other tasks, and whether or not its correlated to player skill to any meaningful degree.

Research into temporal perception

Games, as well as film, animation, and other forms of media, provide the illusion of being continuous by showing a series of still images, or "frames" in rapid succession. You might be tempted to ask if there's an equivalent "framerate" for human vision, but the reality is a bit more complicated. While some papers set this framerate to 60hz (since this is the framerate of the common television), there have been measurements of much higher than that; one such study shut off lights for variable amounts of time, and people began to notice darkness at durations even shorter than 16 milliseconds (Watson 1986), whereas another study found people able to detect when an out-of-place image of just 13ms was inserted in between the frames of a normal film (Potter 2014). In fact, this framerate is sometimes theorized to be above 300hz in certain situations (Deering 1998).

In addition, much research has been done with regards to the perception of things happening at the same time. Since light travels much faster than sound, and because it's useful to be able to associate visual events with auditory ones, your mind takes slightly different light and sound timings and "re-synchronizes" them so that they appear to be simultaneous. This threshold appears to be at a distance of around 10 meters, or a time difference of up to 150 milliseconds, as well as up to 250 milliseconds in human speech (Spence et al. 2003). This is the so-called "Horizon of simultaneity".

In the literature thus far, though, there is a lack of research done on the horizon of simultaneity between a participant-supplied motor input and a visual stimulus, instead with all of the research being done on two differing external inputs. By extension, no research has been done to see if this changes with any sort of training. This project seeks to answer these questions.

Research suggesting expertise-based general perception changes

One study that suggests the possibility of gaming expertise affecting general perception is Green & Bavelier, 2007, in which it demonstrated that participants with experience with Action Video Games (e.g. *Starcraft*) have lower visual crowding thresholds, altering the spatial resolution of vision. As such, there is at least some precedent that gaming experience can lead to generalizable perception changes. Another such study demonstrated that action game experience reduced the magnitude of the attentional blink in both adults in children, which is otherwise considered very robust (Dye & Bavelier, 2010). Although these results may have been reached via other attentional effects, these two studies lead us to a possible hypothesis – can playing action games involving strict timings also alter the resolution of *temporal* perception? Players with action game experience, who frequently execute commands with 16-64ms windows, might have more developed temporal perception, or might simply be able to recognize delays due to an extremely precise familiarity with the lengths of time of each action in their game. Athletes too, seem to perform better on perceptual tests related to their sport

(Kioumourtzoglou et al. 1998), so it's possible that this is generalizable to groups with any sort of training in visuomotor coordination.

Research on domain-specific differences among experts

On the other end, there are a few apparent expertise-based perceptual changes that are merely domain-specific; that is to say, they're changes that don't carry over from the field that the person has expertise in. A frequently cited study demonstrated that chess masters could memorize a large number of pieces compared to a control group conditional on the fact that the positions were in positions that looked like they could have come from actual games; in positions where the pieces were placed randomly, the expert players performed roughly the same as the control group (Chase et al 1973). Although this effect has since been contested, with some authors claiming that strong players maintain some small advantage over weak players even in random positions, the effect is nowhere near as pronounced as when it is in the domain of expertise (Gobet et al. 1996). This effect was also demonstrated in snooker players, with similar results (Abernethy et al. 1994).

Super Smash Brothers Melee

The expert-level group under examination in this paper are Connecticut and New York residents that regularly attend Super Smash Brothers Melee (henceforth SSBM) tournaments. This group was chosen based on immediate availability and relative skill level, as it contains multiple professional tournament players with peak world rankings of 100 or better, a number of players with competition wins against world ranked players, and a multitude of players that travel extensively and practice many hours a week. The average age of the participants in this group was 22 (*min: 18, max: 28*), and was mostly post-college white males (*68% white, 100% male*). The control group was selected among Yale undergraduates volunteering for psychology experiments that reported to playing 0 hours of video games per week on average. The average age of this group was 19 (*min: 18, max: 21*), and was more mixed in terms of gender/ethnicity (*70% female, no single ethnicity more than 30%*).

SSBM players were selected for this experiment due to their specific circumstances that make them ideal for such a project. Since SSBM was released for the Nintendo Gamecube in 2001, it uses a standard definition 4:3 signal provided via composite video. Modern televisions and monitors rarely accept composite video, and those that do tend to have extensive delay due to needing to upscale the image to widescreen, which takes extra time. As such, at SSBM tournaments almost every television used for tournament play is a Cathode Ray Tube television (or CRT), which do not need to perform this upscaling and serve as the standard for “no added lag”. Since these televisions are no longer manufactured and are heavy and take up a lot of space, there has been a push towards some of the newer monitors with very fast response times measured as low as 2.86 milliseconds, or roughly one tenth of one frame (Laferriere, 2014). However, many players, some of them very highly ranked, still claim this small difference to be noticeable and bothersome. Even very weak tournament players seem to perceive small delays, as evidenced by the fact that virtually no players (and 0 of the top 100 players) use the otherwise popular WaveBird wireless controller, as it has a similarly small amount of delay comparable to a high-definition television. It's difficult to determine how much of this is players truly being able to detect delays of as low as 2 milliseconds, versus placebo effect, versus simply being uncomfortable with flatter, wider screens and the differences in image quality. Because of this, SSBM players are an interesting case study into the true, blinded limits of input latency detection. SSBM players are a group of potentially "trained" individuals at this task that claim to be able to notice very small delays to the point of them being distracting, as well as a group that would benefit greatly from knowledge of the upper limits of latency detection by way of selecting certain monitors over others.

Experiment 1

Avoiding Possible Confounds

The first experiment that immediately suggests itself to test input latency detection is a simple task involving software that does something upon receiving some action, like change colors when the participant clicks the mouse, and then ask the participant if there was a delay or not. However, a study like this raises a few problems and would not be well-suited for this experiment.

First, people have different monitors, all of which have wildly different input delays. In this sense, the experience of participating in the study varies wildly from person to person unless the experimenter forces them to use specific monitors, and even still the tests revolve around the specific peripherals used in testing and would be very difficult to replicate or draw reliable conclusions from.

In addition, the experience of seeing a simple stimulus, like a single color swap, might be difficult to compare to the domain-specific analogues, which use much more complicated stimuli (e.g. a character entering a certain animation after a set amount of time, as opposed to a simple blink).

To make the study more consistent among all participants, as well as avoid these potential confounds, we chose to create a device to perform this experiment, using an Arduino microcontroller and a series of LEDs that light up in sequence. This has a minimal amount of delay compared to variable screens, is inexpensive, and creates a precisely delayed, suitably complex visual stimulus upon a button input, which makes it ideal for a project of this type.

Methods

In experiment 1, participants performed a non-domain-specific test of input latency detection on a device made from simple electronics and an Arduino microcontroller (figure 1) in an attempt to answer the question of whether expert video game players have higher temporal resolution in perception compared to a control population. Participants in two different groups, separated by video game expertise (Expert Video Game Players [EVGP], and Non-Video Game Players [NVGP]), attempted to discern whether or not there was a delay inserted between when they pressed a button and when a series of lights flashed. This is done, as mentioned, with electronics and not standard computer hardware out of portability, as well as for more accurate and precise control over the magnitude of input delay being tested.

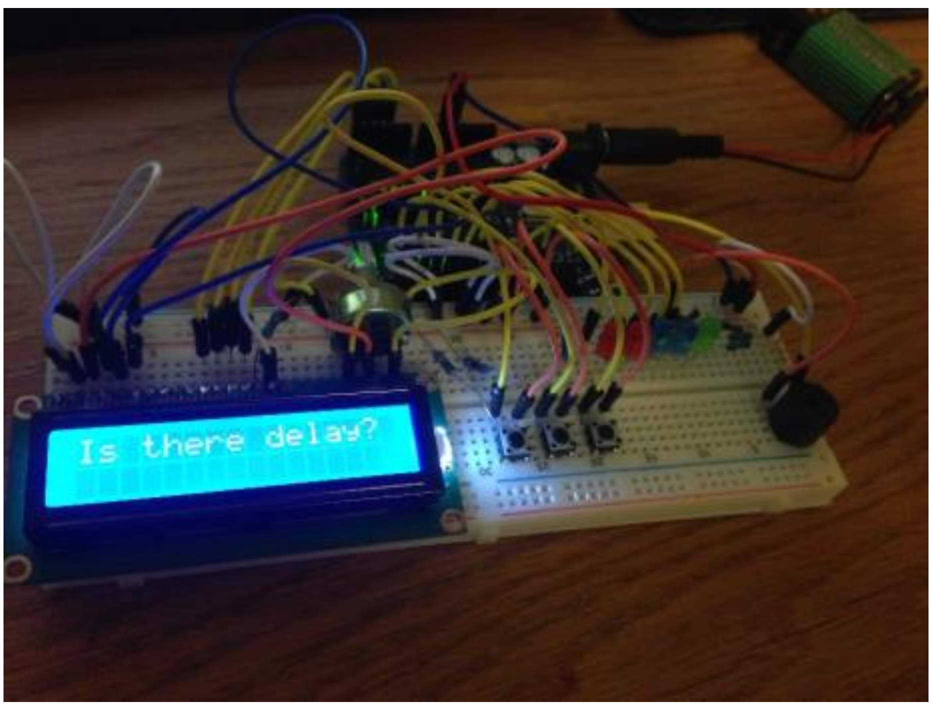

Figure 1

Participants will press the middle button to fire the stimulus, and then respond with the right ("Yes") and left ("No") buttons to answer whether or not there was a delay between when they pressed the middle button and when the stimulus appeared. After responding, they will receive no feedback about the amount of delay they actually saw.

At the beginning of the test, the participants will be given an opportunity to familiarize themselves with the device, operating under the no-delay condition as well as a long delay condition of 400ms.

Performance will be measured via two interleaved three-up one-down staircases. One staircase will start out with 1 ms of input delay, and the other will start out with a 250 ms delay. Trials will vary their input delay between the current value of the staircases, as well as 0 ms of input delay (to ensure that participants do not merely press "Yes" for every trial). Naturally, the steps taken by the staircases will become smaller as the experiment progresses. The test will end when both staircases reverse 5 times, after which the final thresholds will be shown on the display.

Following the experiment, the participant will fill out a short survey to collect information on demographics, as well as the following questions:

EVGP Questions

1. What did you think we were testing?
2. How well did you think you did on the task? (Roughly what percentage of trials do you think you answered correctly?)
3. Sometimes the delay between the button press and the lights could be quite long, and other times it could be quite short. Do you think that affected your performance on the task? In what way?
4. What is the shortest delay you think you could detect in-game? (Answering in milliseconds or frames is acceptable)
5. Did you use any particular strategies?
6. Did you find yourself switching strategies partway through the experiment?
7. How long have you been playing?
8. How often do you play online, if at all?
9. How many hours per week do you play video games?
10. How often do you travel more than 2 hours to play in a tournament?

11. On a scale from 1-10, how would you rate your own skill at this game?
12. Is there anything else you'd like to share?

NVGP Questions

1. What did you think we were testing?
2. How well did you think you did on the task? (Roughly what percentage of trials do you think you answered correctly?)
3. Sometimes the delay between the button press and the lights could be quite long, and other times it could be quite short. Do you think that affected your performance on the task? In what way?
4. Did you use any particular strategies?
5. Did you find yourself switching strategies partway through the experiment?
6. Would you consider yourself a video game player?
7. Over the last 6 months, how many hours per week do you play video games?
8. What kind of games do you play?
9. Have you participated in experiments like this before?
10. Is there anything else you'd like to share?

Results

Originally, a number of other experiments were proposed in the event that this experiment was run and no difference was found between the two groups. Such a result would have suggested that this effect, if it existed at all, was domain-specific, and further experiments would need to be designed around delaying inputs within the game.

However, running this simple general perception experiment on these two groups ended up yielding an extremely statistically significant effect ($p=0.0008$).

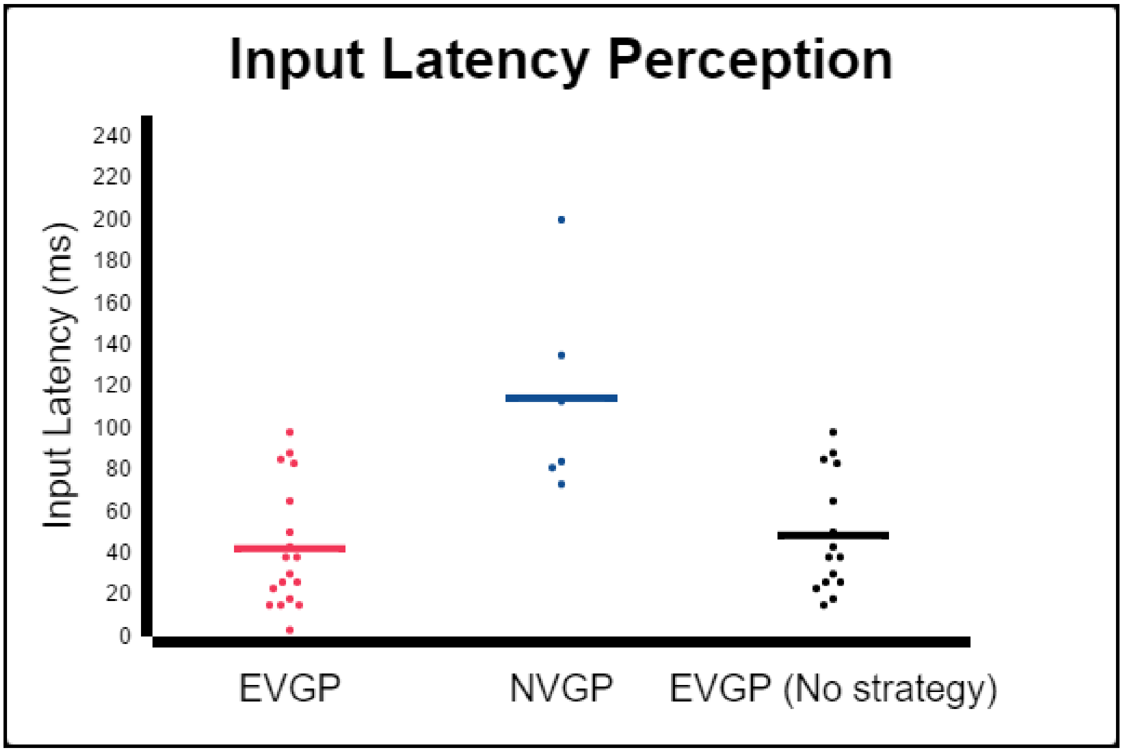

Figure 2

The experiment was run on 21 EVGP (3 excluded for not paying attention, 3 excluded for powerful strategies, N=15) and 10 NVGP (4 excluded for crashing device, N=6), with a mean of 48.4ms for the EVGP group and a mean of 114.33ms for the NVGP group, with only half of the non-excluded participants performing similarly to the worst performers in the EVGP condition.

Three EVGP needed to be excluded for devising a strategy so powerful that it allowed them to "beat" the test with extremely low values. The device would lock out button presses during the fire animation, so participants learned the fastest button timing possible in the no-delay condition and then attempted to replicate it on each critical trial. If their input would trigger two flashes, it would be a no-delay condition, and if it triggered only one flash then it must have been delayed by some amount, since their second press was locked out. As such, it turned from a perception task into a test of timing two inputs together, and the three participants got scores of 15, 15, and 3ms. It was interesting by itself that all three of these participants independently devised the same strategy, and would be an interesting study in and of itself (e.g. "Do Expert Video Game Players come up with strong strategies more often than Non-Video Game Players?").

Four NVGP needed to be excluded for crashing the software, which is a byproduct of their delays creeping so far above the upper bound that it would be unable to continue. The upper bound was set by the experimenter, who was an EVGP, with the goal that it would be "short, but long enough to be very obviously visibly delayed" – In hindsight, a serious error, considering that differences between "obviously visibly delayed" among EVGP and NVGP were precisely what was being tested in the first place, and setting the upper bound this way was not nearly conservative enough. Were these four NVGP not excluded, they would have almost certainly raised the mean even higher, but luckily a statistically significant value was able to be reached in spite of excluding them.

It is important to examine the differences in gender and race makeup between the groups, as the EVGP group was overwhelmingly white and overwhelmingly male, whereas the NVGP group was neither mostly white nor mostly male. However, examining within-group suggests that these variables have no correlation with performance on this task.

Men performed no better than women within the NVGP group, and there seemed to be no relationship between race and performance in either group.

Interestingly, there also seemed to be no relationship between relative rank of the players and their performance on this general perception task, which defies the conventional wisdom of the gaming community in question. While there may be some relationship between relative rank and the domain-specific version of this task, there seems to be no meaningful general perception difference between average tournament players and elite tournament players.

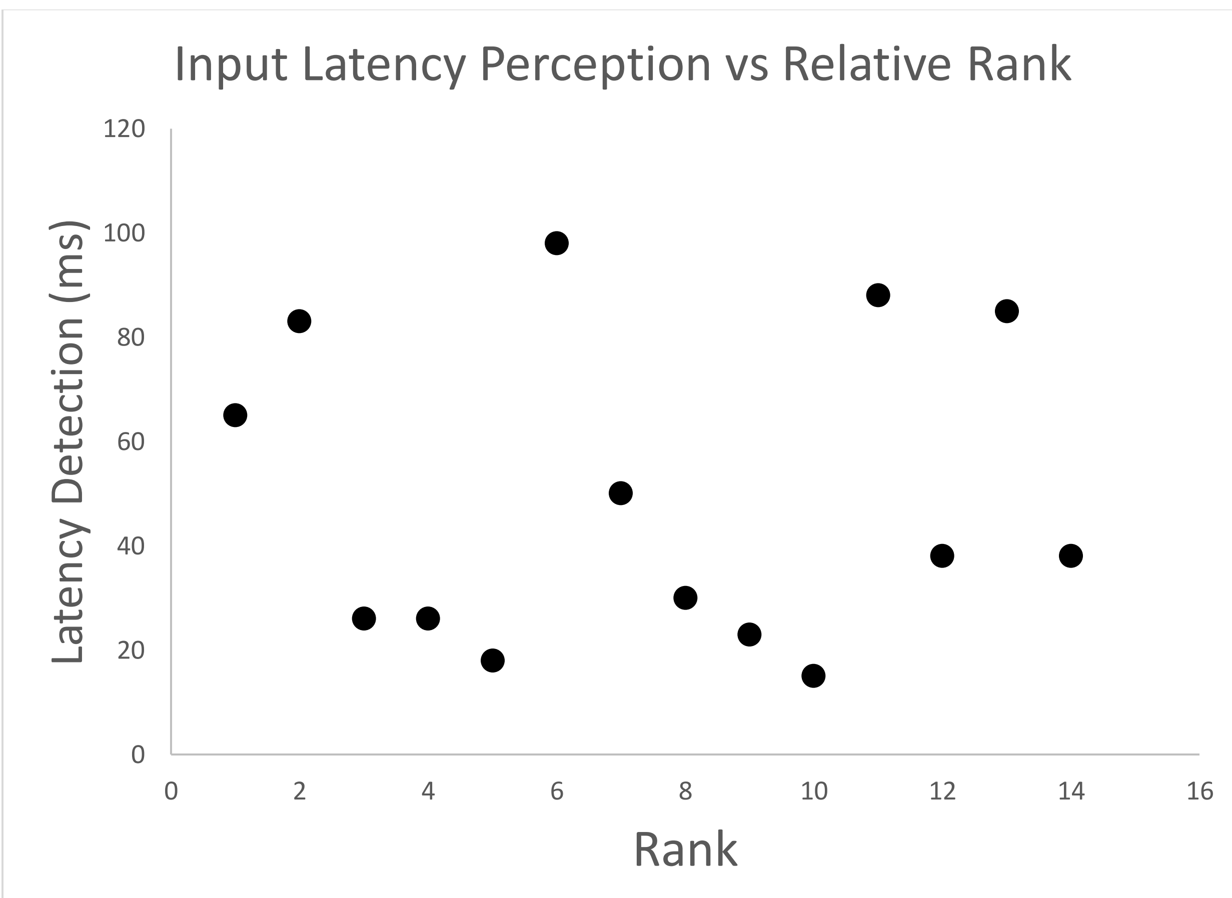

Figure 3

Discussion

There's a number of possible implications to the previous results, but overall it's pretty difficult to judge concretely due to a number of possible confounds that exist between the NVGP and EVGP groups.

The first potential reason that this result could have arisen is a self-selection effect. It is possible that people that succeed at this sort of task are more likely to find video games rewarding, and therefore seek to compete in them. As such, the correlation would be backwards – it wouldn't be that playing video games has any real effect on your perception, but rather that playing video games serves as a factor that would predict your results on a perception task like this one.

Another possible reason behind this result could be that EVGPs simply have stronger visuomotor coordination, rather than this effect arising from strictly temporal perception. Put simply, it may be that EVGPs are more comfortable pressing a button and seeing the world react, and are thereby more comfortable judging distances from a motor output to a visual input, but would be no different at judging distances between a visual input and another visual input.

A final possibility is that EVGPs have a finer temporal resolution of vision, perhaps due to familiarity with shorter timescales. For example, SSBM players are accustomed to dealing with actions that happen as quickly as 16ms apart from each other, and need to focus closely on events that happen very close together. This contrasts strongly with, say, someone whose primary hobby is reading books, where there is no pressure to separate events that happen at very short time scales. It could be that familiarity with short time scales will simply make you better at perceiving short time scales, and that this could be manifesting itself here.

These possible confounds lead to a series of potential follow-up experiments to determine the true nature of this result, which may be run after more NVGP are tested to demonstrate this effect with a larger sample size.

The first is testing to see if this effect remains if divorced from the motor component, to see if the difference in temporal perception is necessarily tied to a motor input. This can be tested with a similar experiment as the first, and with the same apparatus: instead of the delay being between the button press and the input, it would be between the flash and a second flash, with the test randomly inserting single flashes in between certain critical trials. The participants in this experiment will be asked if they saw one flash or two, the idea being if the flashes are suitably close together then they may be perceived as single flashes. If EVGP outperform NVGP on this task, it’s possible that the observed effect is due to differences in temporal resolution, rather than being visuomotor in nature. If this effect is demonstrated, then a whole host of other experiments open themselves up to possibility, in order to determine this effect’s transferability to other, more complicated visual perception tasks (like face perception).

The second proposed follow-up experiment is to run experiment 1 on a group with a similar visuomotor skillset, and compare their performance to EVGP and NVGP. A good group to use for this would be athletes, preferably ones that do not play video games. If this group performs similarly to EVGP, then it could be suggested that timescale familiarity or similar non-game-like visuomotor training could yield a similarly strong result as demonstrated in the expert gaming group, and the opposite would follow if they perform like NVGP.

A final, most ambitious potential study would be a training study. If it would be possible to take NVGP participants, and by making them play video games allow them to emerge at the end of the study with higher scores on the input delay task, compared to a group that spent that time being instructed in some arbitrary task (i.e. trivia). This would be the most costly study to run but it would be the one with the strongest implications if it were to defeat the null hypothesis. Most “Brain-Training Games” have been shown to have virtually no effect on other domains of perception or cognition, with the only soundly demonstrated improvement being at the tasks themselves, rather than any general cognition boosts (Simons et al 2016). However, we have demonstrated that Expert Video Game Players have a significant advantage over control populations at this perception task, and it certainly suggests the possibility of playing a game affecting your perception in some way.

The implications of these experiments have a number of applications beyond simply their intellectual interest – knowing the upper limits of input latency detection as it pertains to expert gamers as well as NVGPs has far reaching uses for manufacturers, marketers, tournament organizers, competitors, and so on.

For manufacturers, it would be useful to know whether or not the concerns of certain top professional video game players warrants greater R&D into manufacturing faster monitors. If players claim to be able to detect latency in a monitor that they cannot detect in a blind test, then it's safe to assume that there are other factors at play, and that putting time into developing an even better monitor might simply be time and money wasted.

For tournament organizers, it would help towards settling the debate of which monitors are acceptable to use in big events, providing a more definitive threshold with which to draw the line between legal and not legal models. A tournament organizer could be spared the effort of hunting down hundreds of outmoded televisions, or determine that the more expensive screens aren't even necessary, or the opposite – that even the best monitors are not good enough to replace the current, more conservative list of allowable televisions and monitors.

For competitors, it would settle once and for all the question of whether or not it's possible for people to reliably detect latency in certain types of monitors but not others, and potentially save them a great deal of money buying a more expensive monitor that they are physically incapable of perceiving the superiority of anyways.

For marketers, it would be a very useful to have this sort of information in order to better understand their own products as they relate to human perception. This result might allow marketers to be able to concretely recommend certain monitor types to certain audiences, based upon information about which groups are unable to perceive differing amounts of input delay. If a marketer was previously trying to market an expensive, high definition monitor to a crowd of gamers that would be distracted by extensive input delay, perhaps with this information they would be better suited marketing that product towards a community of film hobbyists, who would better appreciate its features.

The list goes on and on, and even simply the knowledge of what kind of delays humans are capable of perceiving with training could have applications in other fields, for example aviation. A fighter jet pilot being able to maneuver at seemingly imperceptible delays, perhaps even remotely, may be possible despite inherent latency if this threshold of simultaneity is high enough.

Conclusion

More experimentation needs to be done before any strong conclusions can be reached. It would be too hasty to draw any big conclusions from this, with the small sizes and the relatively different makeups of the two groups; there remain a variety of potential confounds that could have led to the result that was obtained from this experiment.

However, the pilot data is indeed very promising. So far the statistical significance of this effect is striking, and there is certainly motivation based upon the pilot data to continue along with the project to better understand the nature of this effect. We have demonstrated a rather large and robust difference in a general perception task between an expert group of video game players and people of similar age that do not play video games. This effect was demonstrated despite the complete lack of correlation with the other variables collected – age, gender, race, skill, and time spent playing all seemed to have no noticeable relationship to performance on this task within-group. As far as we could determine, "being EVGP" was the only variable that seemed to relate to performance on this task, and it affected it quite strongly.

It's difficult to say precisely what about the Expert Gamer allows them to be so much more successful than the control group at this task. It could be a familiarity with visuomotor tasks. It could be, so to speak, a "higher framerate" of visual perception. It could be a higher degree of focus, or motivation. It could be any or all of these things. What we have demonstrated is that it's *very* unlikely that it is nothing. There exists *some* difference between EVGPs and NVGPs that allow EVGPs to be much more successful on average at detecting input delays, and the future of this project is to narrow down precisely what that difference is.

In a similar vein as Bavelier's visual crowding study, it seems reasonably possible that playing video games could serve as a sort of perceptual training that allows you to improve at certain other tasks. It remains to be seen if this isn't merely a self-selection bias (for example, playing basketball does not make you taller), but it would fly in the face of what is commonly associated with gaming, like poor vision, lack of motivation, and so on. Perhaps, to be successful at games on a larger scale, you need to be more in tune with your perception than the average person, and playing games could allow you to do just that.

Citations

Abernethy, B., Neal, R. J. & Koning, P. (1994). Visual–perceptual and cognitive differences between expert, intermediate, and novice snooker players. *Appl. Cognit. Psychol*.

Andrew B. Watson (1986). Temporal sensitivity. *Handbook of Perception and Human Performance*.

Biederman, Irving; Shiffrar & Margaret M. Oct (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol 13(4).

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*.

Deering, M. (1998). The Limits of Human Vision, *2nd International Immersive Projection Technology Workshop.*

Dye, M. W., and Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Res. 50*, 452–459.

Gobet, F. & Simon, H.A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review,* 55–81.

Green C.S., Bavelier D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, 18, 88–94.

Kioumourtzoglou, E., Kourtessis, T., Michalopoulou, M., & Derri, V. (1998). Differences in several perceptual abilities between experts and novices in basketball, volleyball and water-polo. *Perceptual & Motor Skills*, 86(899-912).

Laferriere, Jas. 27 Mar. 2014. "Ugh, This TV Lags!" *Melee It On Me*. Web.

Potter, M.C., Wyble, B. & Hagmann, C.E. et al. (2014). *Atten Percept Psychophys*.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. (2016). Do “Brain-Training” Programs Work?. *Psychological Science in the Public Interest*, 17(3), 103-186.

Spence, Charles (2003). Multisensory Integration: Maintaining the Perception of Synchrony.

Stetson, Chess; Cui, Xu; Montague, P. Read; Eagleman, David (2006). Motor-Sensory Recalibration Leads to an Illusory Reversal of Action and Sensation. *Neuron.*