High Frame Rate Psychophysics: Experimentation to Determine a JND for Frame Rate

Elizabeth DoVale

Motion Picture Science, Rochester Institute of Technology (RIT)
1 Lomb Memorial Dr, Rochester, NY 14623 USA
erd1847@rit.edu

Abstract—**In past research at Rochester Institute of Technology (RIT), a temporal texture space was defined. It was used to classify what attributes contribute to the perceived motion quality of stroboscopic reproductions. One of these attributes was determined to be the content's captured and displayed frame rate. Using knowledge from this and other past frame rate research and psychophysics, a comprehensive experiment was set up to see if a just noticeable difference (JND) could be found for changes in frame rate. Thresholds were tested starting with base conditions of 24 FPS, 48 FPS, and 72 FPS due to their significance to the motion picture industry. After a verification pilot study, 77 observers in total participated in the experiment. The collected data points to a resulting JND threshold between 26 FPS and 28 FPS for the 24FPS test point, but a significantly noisier signal was determined for both the 48FPS and 72 FPS test points. Explanation for this nosier signal could be a lack of consistent detectability in human perception at these higher frame rates. This experiment is presented as one in a series of studies into stroboscopic motion perception at RIT.**

I. Background

A. Inspiration for Work

High frame rate (HFR) research is not new to RIT, and there are a number of research efforts around the motion picture industry regarding HFR technology. Sean Cooper posed the concept of 'temporal texture' space [1]. Temporal texture space is a way to plot, model, define, and quantify the human visual perception of motion content. Cooper believed that adjustment of a candidate content set's field-of-view, object velocity, and frame rate would help determine the dimensions of the temporal texture space. Cooper ran an experiment addressing these theories, testing a number of live action video sequences that were chosen to cover the motion gamut. Frame rate emerged as a principal component across two of three identified perceptual dimensions (based on Cooper's use of a multidimensional scaling algorithm), sparking interest and desire for further experimentation. Concerns of noise in the data were raised, making it unclear if the other two content parameters corresponded to other orthogonal perception dimensions or if these dimensions related to other parameters not explored. The noise in the data could have been introduced by small inconsistencies in the recorded content (since each video clip was recorded separately, the on-screen talent's performance suffered from notable variation) or due to potentially small inconsistency in the playback of content on the display. Cooper's theories are the inspiration behind all of the HFR research that has occurred at RIT [3].

Elizabeth Pieri decided to pursue some of these thoughts [2]. The bulk of her research revolved around a vocabulary study. Previously, the motion picture industry referred to the hyper-real motion perception often seen in HFR content as the 'soap opera effect' due to the fact that most soap operas are filmed at 30 frames per second (FPS) and have extensive depth-of-field when compared to standard cinema (24 FPS) employing shallower depth-of-field. Pieri ran an initial experiment to ask participants to describe the look of clips at varying frame rates. Collecting words that appeared most often in initial user responses, she then ran an experiment where a separate set of participants quantitatively ranked the words after watching clips at different frame rates. Pieri noticed trends for certain words as frame rate varied. This common vocabulary helps to ensure more accurate communication with others regarding HFR content and perceived motion quality.

Outside the work of key researchers such as Trumbull [7] [8], Watson [5] and Daly [6], much of the current motion picture industry's experimentation with HFR involves subjective single stimulus observation. Most of the time, research occurs when a studio or filmmaker is interested in shooting a film in a non-standard frame rate. A group gets together in a theatre to watch some test clips that were shot at varying frame rates, and they assess which clips they found aesthetically pleasing. As more films consider the possibility of shooting in HFR, the importance of conducting scientific study into human perception of motion quality becomes more crucial. What if humans perceive frame rate nonlinearly (similar to the function of perceived lightness versus luminance, CIE L*)? What if humans don't perceive a change in frame rate above a certain point? These are all questions that inspired the research discussed in this paper and hopefully this work will motivate more questions like these so that future research will occur across the industry.

B. Psychophysics and JND

A visual Just Noticeable Difference (JND) occurs at the point where a change in stimulus intensity becomes just perceivable to the human visual system [9]. In order to find a JND threshold, there are a number of different psychophysical experiment types available. The Method of Constant Stimuli involves presenting a fixed set of content at varying stimuli intensity levels in a random order, and forcing an experiment participant to respond whether or not they detect the stimulus. A two-interval forced choice experiment consists of showing the experiment participant two samples with different stimuli intensity (or conversely, one sample with modified stimulus and one at a baseline control condition), and forcing them to choose which content clip contained the manipulated stimulus. The data collected from a two interval forced choice experiment is plotted such that the percentage of correct responses, or probability of detection, is a function of stimulus intensity. This plot yields a psychometric function. For this experimental set up, 75% is the ideal JND threshold. This can be calculated using equation 1, where p is the corrected probability of actual detection (in this case 50% as we want half of the population to legitimately detect a change in stimulus), C is the underlying random chance of success (50%; two-interval forced choice means the participant has a 50% chance of 'guessing' correctly), and p' is the raw measured percentage of detection

.

$$p = (p' - C) / (1 - C) \quad (1)$$

II. Experiment Set-up

An experimental test bed has been designed and improved throughout the HFR research studies conducted at RIT. A PC computer was built in order to provide hardware suitable for HFR playback up to 144 FPS via a dedicated uncompressed graphics stream in OpenGL. The computer is hooked up to an ASUS PG278Q LCD G-Sync monitor [4]. G-Sync monitors allow for applications to control the refresh rate of the monitor. The various frame rates desired for testing can be matched by the refresh rate of the display.

A two-interval forced choice experiment was designed. A playlist of video content pairs at various stimuli intensity was created and played back in a random order. In a dark surround, the experiment participant, sitting two picture heights away from the screen, watched the two video clips and then was asked to pick the clip that had more of the stimulus present. Figure 1 contains an image of the experiment set-up.

Fig. 1. Experimental setup, experiment taken in a dark surround with the participant at 2 picture heights away from the monitor

Smooth, fluid, and *sharp* were the words used in instructing the experiment's participants during stimulus identification due to their positive correlation with increase in frame rate, as found in Pieri's research. Notably, the participants were never instructed to identify which clip had a higher frame rate. This same verbiage was also used in a post experiment survey. The survey was used to obtain qualitative input from the experiment participants.

Since it is important to know how accurately the system is playing back frames, a validation test was completed prior to the start of the experiment trials. A Sony FS700 was used to capture the monitor's playback at a rate of 960 FPS. Each of the stimuli frame rates was validated, meaning their playback was found to be accurate within a few percentage points. Also, The LCD transition characteristics from buffer refresh were measured and shown to exhibit a small cross-fade artifact blamed on the liquid crystal state switching. The system performed well but the executed experiment results obviously reflect the exact behaviour of the test bed with the played content and can't be generically extended to other display dynamics.

For content, a 3D animated clip was chosen due to concerns raised from Cooper's research regarding inconsistent character performance in live action clips. A frame from the content clip can be seen in figure 2. Due to how the content clip was exported from Maya, there is no motion blur within the frames. Also, each frame was played only once; there were no multiple refresh flashes as is used in some motion picture display technology.

Fig. 2. 3D animated content clip frame, content used in the experimental study

III. Results and Discussion

A. *Pilot Study*

A pilot study was needed in order to find an estimate of the range of frame rates to test for each JND threshold, thus verifying the stimuli set for the user experiment. The two testing points chosen were 24 FPS and 48 FPS. These points were selected due to their relevance to the motion picture industry; 24 FPS being standard cinema capture and 48 FPS being used in some recent films (ex: *The Hobbit*). 48 FPS is also twice as fast as the standard cinema frame rate and the refresh rate of a twin bladed film projector, providing an interesting point of comparison.

Two HFR researchers participated in the pilot study for both the 24 FPS and 48 FPS test points. The comparison stimuli for 24 FPS were 26 FPS, 28 FPS, 30 FPS, 32 FPS and the stimuli for 48 FPS were 52 FPS, 56 FPS, 60 FPS, and 64 FPS. The experiment comprised a 2-interval forced-choice test where each pair was shown 12 times, and the order of presentation was shown in both permutations (ex. 24 FPS followed by 26 FPS and 26 FPS followed by 24 FPS). This included the controls (ex. 24 FPS vs 24 FPS), which are mainly used to identify the participants' tendency to pick either the first or second shown video clip disproportionately.

The first observer is an industry expert. They have been doing visual research with HFR for three years. In figure 3, a plot of observer one's psychometric function for 24 FPS can be seen, and in figure 4, a plot of their psychometric function for 48 FPS can be seen. The blue dots are the percentage of times the higher frame rate stimuli was correctly selected relative to each corresponding test frame rate. The red line is a sigmoidal curve fit done in Matlab using the 'sigm_fit' function. It is clear that observer one is proficient in spotting visual changes in motion quality. In both the 24 FPS and 48 FPS psychometric functions, they cross the 75% threshold, the green dashed line, exhibiting a JND of +1.5 FPS for 24 FPS and +4 FPS for 48 FPS.

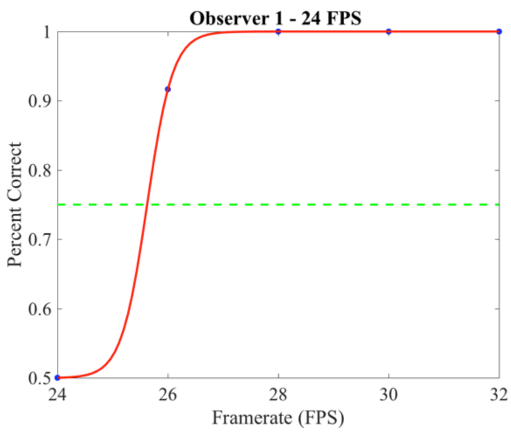

Fig. 3. Pilot test data: observer one's psychometric function comparing content against 24 FPS (24 FPS vs 26 FPS, 28 FPS, 30 FPS, 32 FPS)

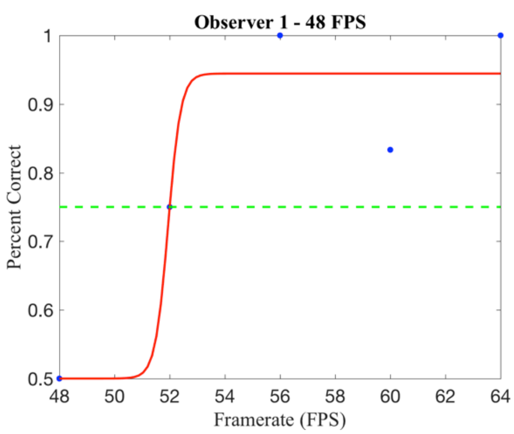

Fig. 4. Pilot test data: observer one's psychometric function comparing content against 48 FPS (48 FPS vs 52 FPS, 56 FPS, 60 FPS, 64 FPS)

Observer two is less experienced, but would also fall into the expert category. In figure 5, a plot of observer two's psychometric function for 24 FPS can be seen. In figure 6, a plot of their psychometric function for 48 FPS can be seen. It is important to note that the psychometric function for observer two is hand sketched due to a fitting error in the sigmoid fit function. Observer two required a higher stimulus in order to cross the 75% JND threshold line, +3 FPS at 24 FPS and +10 FPS at 48 FPS. This is most likely attributed to lesser expertise or a slightly lower motion sensitivity as compared to that of observer one.

The combined results from both pilot participants were also assessed. These graphs can be seen in figure 7 and figure 8. From the combined collected data from this pilot study, a conclusion that the full population JND thresholds would fall within the listed test ranges was confirmed. The determined comparison stimuli for 24 FPS to be used in the full study were 26 FPS, 28 FPS, 30 FPS, 32 FPS and the stimuli for 48 FPS were 50 FPS, 52 FPS, 54 FPS, 56 FPS, 58 FPS, 60 FPS, 62 FPS, 64 FPS, and 66 FPS. Both of the pair ranges cover approximately the same frame rate range as the pilot study, but an interval of 2 frames was chosen for the 48 FPS test as compared to the interval of 4 frames that was used in the pilot study in order to increase the precision of the collected data for the larger group.

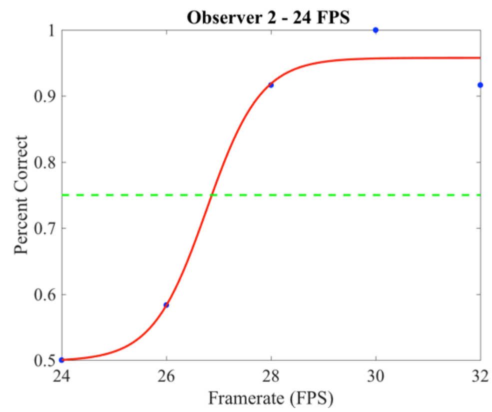

Fig. 5. Pilot test data: observer two's psychometric function comparing content against 24 FPS (24 FPS vs 26 FPS, 28 FPS, 30 FPS, 32 FPS)

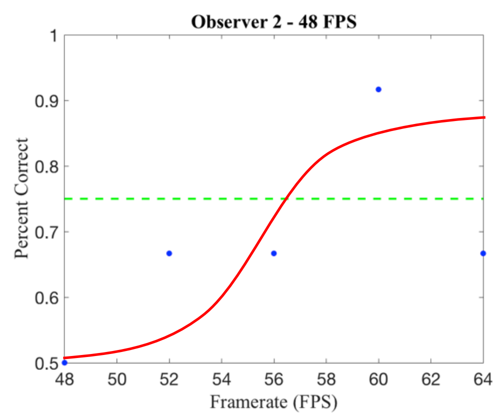

Fig. 6. Pilot test data: observer two's psychometric function comparing content against 48 FPS (48 FPS vs 52 FPS, 56 FPS, 60 FPS, 64 FPS)

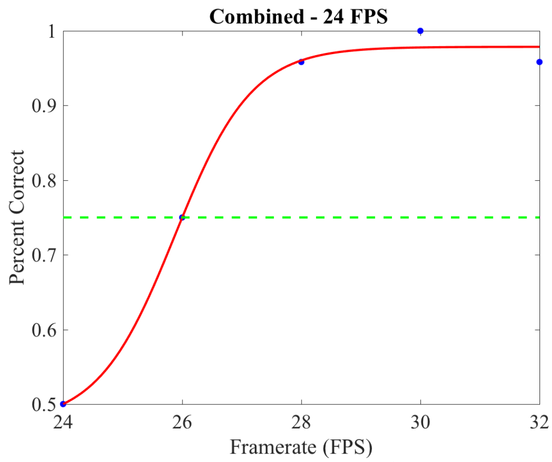

Fig. 7. Pilot test data: combined psychometric function comparing content against 24 FPS (24 FPS vs 26 FPS, 28 FPS, 30 FPS, 32 FPS)

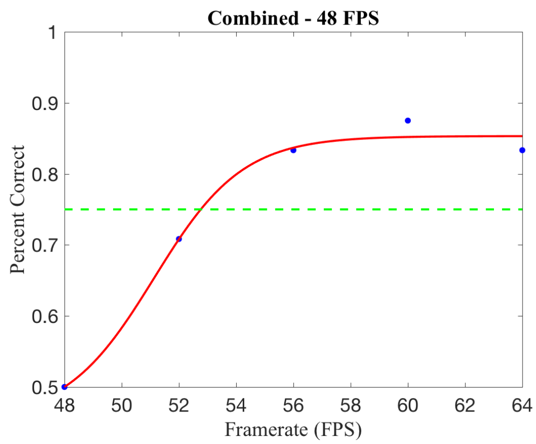

Fig. 8. Pilot test data: combined psychometric function comparing content against 48 FPS (48 FPS vs 52 FPS, 56 FPS, 60 FPS, 64 FPS)

C. *Experimental Study*

The experimental study mimicked the execution of the pilot study previously described except that each pair was shown only two times to each observer. The order of presentation included both permutations (ex. 24 FPS followed by 26 FPS and 26 FPS followed by 24 FPS). 72 FPS was chosen as another baseline testing point since it is three times standard cinema frame rate, thus allowing for even spacing between three different testing points (24 FPS, 48 FPS, and 72 FPS). The comparison stimuli for 72 FPS were 80 FPS, 88 FPS, 96 FPS, 104 FPS, 112 FPS, and 120 FPS.

77 individuals participated in the experiment. Population statistics can be seen in figures 9 through 12. In summary, population was majority female with the age distribution heavily centered around college-aged participants (19 to 23). The expertise distribution, in figure 9, was determined using the participant's response to two ranking questions regarding experience manipulating video content and experience analysing image quality.

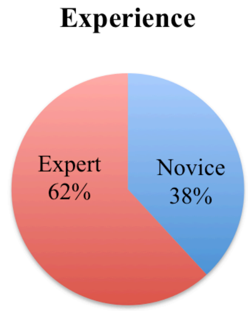

Fig. 9. Experiment population data: experience level of population (77 total) based off participant's response to two ranking questions regarding experience manipulating video content and experience analysing image quality

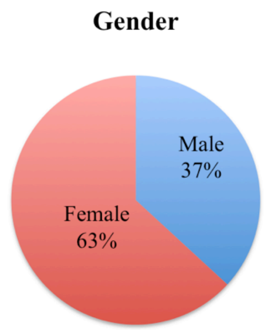

Fig. 10. Population data: gender distribution of population

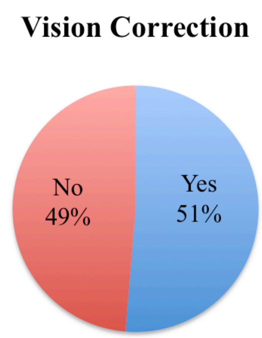

Fig. 11. Population data: vision correction (do they need glasses or contacts?)

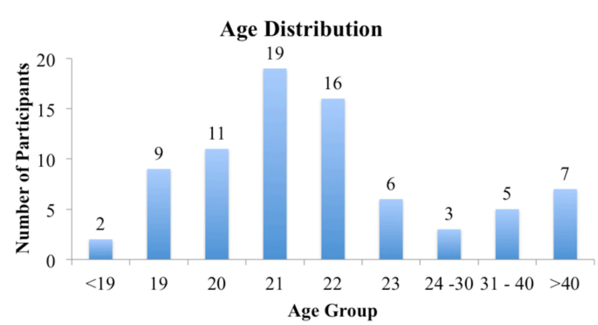

Fig. 12. Population data: age distribution of population, majority of the population falls into the college age range (19 to 23)

In figure 13, a plot of the population's psychometric function for 24 FPS can be seen. The blue dots are the percentage of times the higher frame rate stimuli was correctly selected relative to each corresponding test frame rate. The red line is a sigmoidal curve fit. In the 24 FPS psychometric function, the full population crosses the 75% threshold, the green dashed line, exhibiting a JND between +2 FPS and +4 FPS.

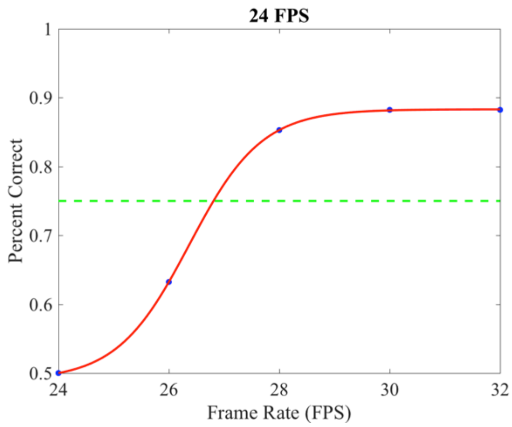

Fig. 13. Experimental data: full population psychometric function comparing content against 24 FPS (24 FPS vs 26 FPS, 28 FPS, 30 FPS, 32 FPS)

In figure 14, a plot of the population's psychometric function for 48 FPS can be seen. The data points in the 48 FPS figure are significantly nosier, and only one test point appears just above the 75% JND threshold line. A potential reason for the noise is due to the experience level of the participants. Based on verbal and written feedback from the observers, many found it hard to see a difference between the presented clips. This frustration may have caused a disconnect, leading a number of participants to lose interest or to just start guessing. A 50/50 'guess' then gets added to the results of all those who may have actually seen the intended stimuli (like the expert group), causing the whole data set to shift downward on the plot or to render noisier. To help examine this potenitial source of error, the novice group and expert group were separated. This can be seen in figure 15 and figure 16, respectivly. A visual difference in the noise can be observed between the novice and expert plot. Even so, there is not a huge difference between the two data groupings in JND determination. Perhaps the frustration affected both groups of participants, or perhaps the experiment was too challenging. The data does show evidence of the start of an expected psychometric sigmoid and so a universal lack of detectability at any frame rate above 48 FPS doesn't seem fully plausible (as would be confirmed in the work of Cooper and Pieri). Due to the frustration evident from post interviews and analysis of participant qualitative response, lowering the corrected probability threshold could help adjust for the participant frustration. If a 70 % JND threshold line is used (signifying that 40 % of the population legitimately saw the difference), a JND between +10 FPS and +16 FPS exists for the expert participants at 48 FPS.

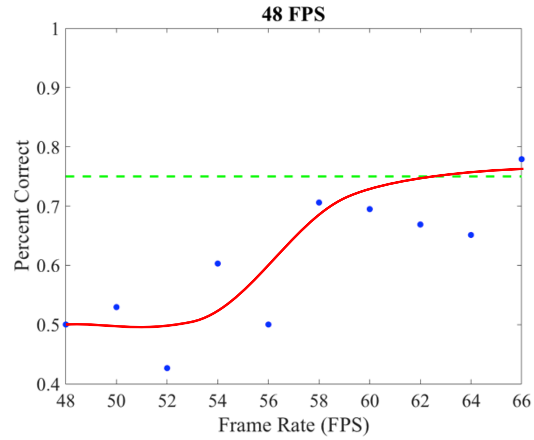

Fig. 14. Experimental data: combined psychometric function comparing content against 48 FPS (48 FPS vs 50 FPS, 52 FPS, 54 FPS, 56 FPS, 58 FPS, 60 FPS, 62 FPS, 64 FPS, 66 FPS)

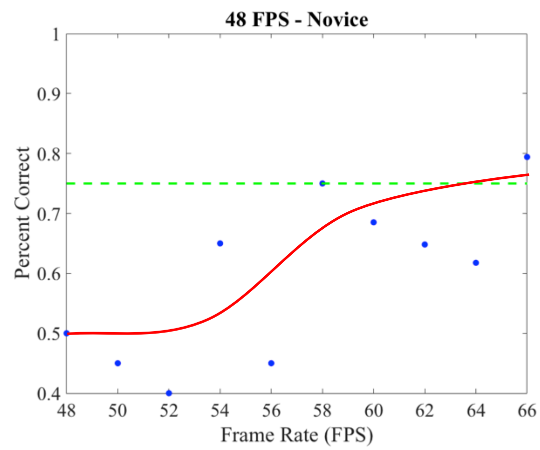

Fig. 15. Experimental data: novice psychometric function comparing content against 48 FPS (48 FPS vs 50 FPS, 52 FPS, 54 FPS, 56 FPS, 58 FPS, 60 FPS, 62 FPS, 64 FPS, 66 FPS)

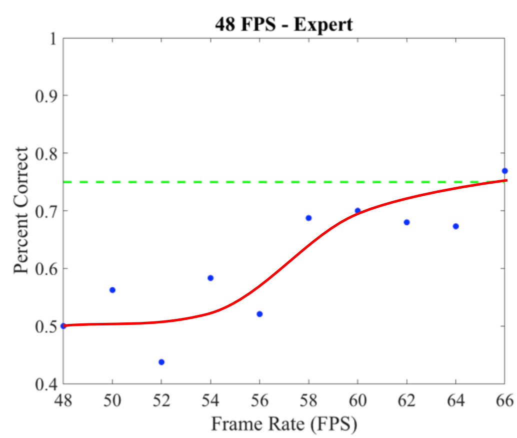

Fig. 16. Experimental data: expert psychometric function comparing content against 48 FPS (48 FPS vs 50 FPS, 52 FPS, 54 FPS, 56 FPS, 58 FPS, 60 FPS, 62 FPS, 64 FPS, 66 FPS)

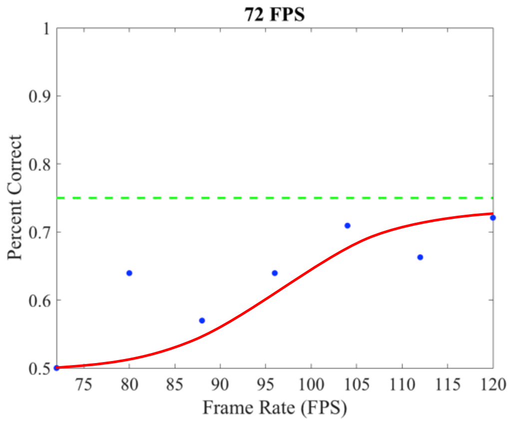

Fig. 17. Experimental data: combined psychometric function comparing content against 72 FPS (72 FPS vs 80 FPS, 88 FPS, 96 FPS, 104 FPS, 112 FPS, 120 FPS)

In figure 17, a plot of the population's psychometric function for 72 FPS can be seen. A lot of the data trends seen in the 48 FPS data appear once again in the 72 FPS data though even more exaggerated. An interesting thing to note is that the 72 FPS data has relatively poor detection rates even though it has a large change in framerate (+48 FPS at the extreme). A JND threshold percentage shift to 70% yields results of between +24 FPS and +36 FPS for detectability. Explanation for this nosier, flatter signal could be a lack of consistent detectability in human perception at these higher frame rates, now in a potentially different regime from the 48 FPS results. There is a possibility that above a certain point (i.e. above 72 FPS), the human visual system might not be able to adequately perceive a change in framerate.

IV. Conclusions

Based on the experimental data collected, a JND threshold for 24 FPS was determined to fall between +2 FPS and +4 FPS. A JND threshold was hard to find for 48 FPS due to noise in the data, which was investigated under the assumption that more novice viewers may have become frustrated with the lack of noticeable change. Upon splitting the 48 FPS data into a novice and expert group, this initial theory isn't a strong explanation. Other posible explanations for this noise could be difficulty in the experimental task, boredom, habituation and ultimately lack of attention. A lowering of the identification threshold could help adjust for the data skew given these possibilities. The example of a 70 % JND threshold line was used (signifying that 40 % of the population saw the difference), a JND between +10 FPS and +16 FPS was found for the expert participants at 48 FPS. 72 FPS suffered from similar data noise but overall the data has relatively low detectability even though it has a large change in framerate (+48 frames). It was theorized that the explanation for this nosier signal could be a lack of consistent detectability in human perception at these higher frame rates.

For future work, it would be interesting to experiment with a double flash, similar to that employed in a twin bladed film projector, to see its effects on the perception of the JND threshold for frame rate. It would also be beneficial to develop and expand the test bed to allow for additional parameters, such as evaluating comparisons between different amounts of motion blur or shutter angle in content, and testing both animated and live action clips. Also, turning the experiment into a game may be beneficial in keeping participants engaged and preventing frustration.

References

[1] Cooper, Sean. "Towards A Psychophysical Approach to Modeling Temporal Texture in Cinema". RIT thesis (unpublished) 2015.

[2] Pieri, Elizabeth. "The Effect of Frame Rate on the Human Visual System". RIT thesis (unpublished) 2016.

[3] Cooper, Sean, Elizabeth Pieri, and David Long. "Is There an Uncanny Valley in Frame Rate Perception?." *Annual Technical Conference and Exhibition, SMPTE 2016*. SMPTE, 2016.

[4] Ghodrati, Masoud, Adam P. Morris, and Nicholas Seow Chiang Price. "The (un) suitability of modern liquid crystal displays (LCDs) for vision research." *Frontiers in psychology* 6 (2015).

[5] Watson, A. "High frame rate and human vision: A view through the window of visibility," *SMPTE Motion Imaging Journal*, 2013.

[6] Daly, Scott, et al. "A psychophysical study exploring judder using fundamental signals and complex imagery." *SMPTE Motion Imaging Journal* 124.7 (2015): 62-70.

[7] Trumbull, D. "Motion picture system," *United States Patent No. 4,477,160*.

[8] Trumbull, D. "Conversion of high to low frame rate motion picture films." U.S. Patent No. 4,889,423.

[9] Wolfe, Jeremy M., et al. *Sensation & perception*. Sunderland, MA: Sinauer, 2006.

Acknowledgments

I would like to thank the following individuals for all of their help and support throughout the completion of this research: Elizabeth Pieri and Sean Cooper for starting HFR research at RIT; Brendan Kirchbaum and Jared Athias for their help with the animated content; David Long, Ricky Figueroa, and Gabe Diaz for being a helpful advising committee; Tanvi Raut for her coding assistance; And finally, the RIT SOFA Film Cage for their the equipment rentals and help.