An Estimation and Comparison of Human Abilities to Communicate Information through Pursuit Tracking vs. Pointing on a Single Axis

Michael Blandino[1], Edgar Berdahl[1], and R. William Soukoreff[2]

[1] Louisiana State University, Baton Rouge LA 70803, USA
[2] Toronto General Hospital Research Institute, Toronto, Ontario, Canada

Abstract. This paper describes a human subject study that compared the limits at which humans could communicate information through pursuit tracking gestures versus pointing (i.e. tapping) gestures. These limits were measured by estimating the channel capacity of the human motor-control system for pursuit tracking versus pointing along a single axis.
A human-computer interface was built for this purpose, consisting of a touch strip sensor co-located with a visual display. Bandwidth-limited Gaussian noise signals were used to create targets for subjects to follow, enabling estimation of the channel capacity at bandwidth limits ranging from 0.12 Hz to 12 Hz. Results indicate that for lower frequencies of movement (from 0.12 Hz to 1 Hz or 1.5 Hz), pointing gestures with such a sensor may tend to convey more information, whereas at higher frequencies (from 2.3 Hz or 2.9 Hz to as high as 12 Hz), pursuit tracking gestures will afford higher channel capacities.
In this work, the direct comparison between pursuit tracking and pointing was made possible through application of the Nyquist sampling theorem. This study forms a methodological basis for comparing a wide range of continuous sensors and human capacities for controlling them. In this manner, the authors are aiming to eventually create knowledge useful for theorizing about and creating new kinds of computer-based musical instruments using diverse, ergonomic arrangements of continuous sensors.

Keywords: Pursuit Tracking · Pointing Accuracy · Shannon-Hartley Theorem · Information Theory · HCI · Continuous Control · Analog Sensor.

1 Introduction

Musical practice can demand the performance of complex gestures accurately and repeatably in order to realize sound with composed attributes. Technical systems incorporated into new interfaces for musical performance often include sensors in order to afford continuous control of a parameter or a combined array of parameters that are mapped to that of musical synthesis systems. Accordingly, design of these interfaces will require a consideration of what demands of musical composition and performance can be accommodated with the sensors of the system.

Human-computer interaction (HCI) literature reflects decades of investigation into the pointing gesture for communicating information and into the relationships of target characteristics to human capability. Fitts' Law and extensions within information theory have developed knowledge of the limits of information throughput using a pointing gesture, even informing international standards for pointing devices [8, 13].

Fewer investigations of pursuit tracking with continuous control have been conducted using information theory [6, 4, 7, 1, 5]. It does not appear that any studies have directly compared these modes of communicating information using a common human computer interface with a continuous control sensor.

The ability to convey information through such a sensor is an essential part of the utility of its afforded interaction. A quantitative measure of the upper limit of what amount of information may be conveyed through a sensor is pertinent to musical performance limitations of the sensor and, further, may be important to the design of its use in this application and in others.

Beyond applications in music, it is believed that this work can be informative for design of flight control systems, video games, assistive devices, other human-computer interactions, and ergonomics.

A prior pilot study of pursuit tracking using four continuous control sensors of different modes that were not co-located with their target signals showed that channel capacities as high as 4-5 bits per second were achieved with adequate training [2]. Of those four sensors, the system including the touch strip was found to have the highest channel capacity. The human subject study of this paper furthers this work by including a higher level of training of multiple subjects and a comparison to pointing/tapping in an equivalent model and target set.

2 Model

2.1 Fundamentals

In this work, it is assumed that the subjects are aware of some target signals $X(t)$ that they want to input into a computer. Due to various effects, somewhat different gesture signals $Y(t)$ are actually registered in the computer. It is decided to model this as a communications channel as shown in Figure 1.

This model is for example suggested by prior research into human performance with airplane and related control systems [6, 7, 1]. Accordingly, the noise in the human motor control system is modeled with the signal $Z(t)$. This noise is understood to be approximately independent of the gestures being performed [7]. Such a model is suggested by research into neuromotor noise theory [9, 14]. Moreover, such signals are biomechanically filtered by the human body, which will also tend to make the noise signals look Gaussian distributed due to the Central Limit Theorem [3]. (Finally, further evidence along this vein includes the fact that errors in the endpoints of pointing tasks tend to be Gaussian distributed as well [15].)

However, the authors believe that the model requires an additional filter with impulse response $h(t)$ to model observed human behavior. Consider if it did not and imagine the case in which a subject is performing a gesture signal $Y(t)$ that approximates $X(t)$ albeit with some noise included. Due to independence of $Z(t)$ and $X(t)$, $E(Y^2(t)) = E(X^2(t)) + E(Z^2(t))$, implying that $E(Y^2(t)) > E(X^2(t))$, which will however not be the case if the user is following the target signal $X(t)$ with the same power level. Therefore, a model component $h(t)$ is needed to model how the subject's input signal component is attenuated to make room for the noise power $Z(t)$. $h(t)$ could also in some situations potentially model other dynamic effects in the subject's performance [6].

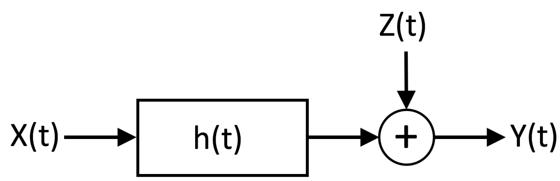

Fig. 1. A model of the user's performance in which $h(t)$ is a filter's impulse response that models the deterministic component of a user's performance, and $Z(t)$ models the random motor noise.

For the recordings made in the present study, not enough data was present to be able to robustly model $h(t)$ in detail. Therefore, using Occam's razor, and in the case of the present application, it was decided to model $h(t)$ with the constant h_0; in other words, the authors set $h(t) = h_0$.

For a given trial, if a subject is performing a gesture signal $Y(t)$ that is very close to the target signal $X(t)$, the h_0 will be close to 1.0 and the noise $Z(t)$ will have a low power. In contrast, if a subject is performing a gesture signal $Y(t)$ that is not very precisely tracking a target signal, then h_0 will be significantly closer to zero, and the noise $Z(t)$ will have a relatively larger power.

According to this model then, h_0 can be robustly estimated given even only small amounts of data. From the model, one can derive that

$$E(X(t)Y(t)) = E(X(t)(h_0 X(t) + Z(t)) = E(h_0 X^2(t)) + E(X(t)Z(t)). \tag{1}$$

Since the target signal $X(t)$ and the motor noise $Z(t)$ are uncorrelated, then $E(X(t)Z(t)) = 0$, which leads to the following:

$$E(X(t)Y(t)) = h_0 E(X^2(t)). \tag{2}$$

$$h_0 = \frac{E(X(t)Y(t))}{E(X^2(t))} \tag{3}$$

So finally, given some example data, the estimate of h_0 can be obtained by averaging as follows:

$$\hat{h}_0 = \frac{\mathrm{avg}(X(t)Y(t))}{\mathrm{avg}(X^2(t))}. \tag{4}$$

2.2 Channel Capacity for Pursuit Tracking (for Continuous Inputs)

Consider the case where the analysis is being performed on a single trial with bandwidth f_X. For pursuit tracking of continuous inputs, the channel capacity can then be estimated using the Shannon-Hartley theorem [10, 3]. For systems where the signal-to-noise ratio is constant across the bandwidth of the channel, the channel capacity at bandwidth f_X is then

$$C(f_X) = f_X \cdot \log_2 \left(1 + \frac{S}{N}\right), \tag{5}$$

where $\frac{S}{N}$ is the signal-to-noise ratio, which can be estimated as follows:

$$\frac{S}{N} = \frac{E((h_0 X(t))^2)}{E(Z^2(t))} = \frac{E((h_0 X(t))^2)}{E((Y(t) - h_0 X(t))^2)} \approx \frac{\text{avg}((h_0 X(t))^2)}{\text{avg}((Y(t) - h_0 X(t))^2)}. \tag{6}$$

2.3 Channel Capacity for Pointing (for Discrete-Time Inputs)

For pointing, the signal-to-noise ratio can be estimated in essentially the same way. A single pointing gesture operates with the channel capacity of the discrete Gaussian channel [3]:

$$C_{pointingonce} = \frac{1}{2} \log_2 \left(1 + \frac{S}{N}\right). \tag{7}$$

If sampled at the Nyquist rate (e.g. $2f_X$ pointing gestures per second for a bandwidth of f_X), then the same expression as in (5) is obtained for the net channel capacity:

$$C(f_X) = 2f_X \cdot C_{pointingonce} = f_X \cdot \log_2 \left(1 + \frac{S}{N}\right). \tag{8}$$

This correspondence, which is enabled by the sampling theorem, motivated the experimental design for the following subject test [12].

3 Subject Experiment

3.1 Apparatus

An experimental apparatus was assembled in order to compare pursuit tracking and pointing gestures using a common interface to match a co-located target signal (see Figure 2). The apparatus was comprised of a flat screen high-definition monitor of 30 cm by 47.3 cm, a Spectra Symbol 200 mm “soft potentiometer” (also known as a touch strip), an Arduino Micro microcontroller, and a 5V power adapter for reference voltage. As shown in Figure 2, the touch strip was mounted to the display surface and placed 11 cm from one short side and centered evenly between the long sides of the display.

Fig. 2. The experimental apparatus provides a display with colocated sensor for target performance.

To achieve a higher accuracy of microcontroller sampling of the sensor output, an external reference voltage was maintained through a separate 5V adapter connected to the reference pin of the Arduino Micro.

A program realized in the Cycling '74 Max application assembled and displayed the target signals onto the display and recorded the performed gesture data from the sensor as audio file data at 4410 samples per second. The application also provided instructions and control to progress through phases of the experiment.

3.2 Stimuli

Target signals were generated as bandwidth-limited Gaussian noise in two modes: pursuit tracking and pointing. For pursuit tracking gesture targets, a continuous curve with a length of 20 sec at 4410 samples per second formed the target shape (see Figure 3). These curves were formed by taking Gaussian-distributed noise sampled at 4410 Hz and filtering it by a fourth-order low-pass Butterworth filter. This filter was applied twice forwards and twice backwards, resulting in a zero-phase filter of net order sixteen [11].

For pointing gesture targets, 13 mm diamonds were presented at values sampled from the pursuit tracking curve. The signal was always sampled at twice the frequency of the bandwidth limit in an evenly spaced time interval (see Figure 4). Sampling at twice the frequency bandwidth serves to meet the requirements of the Nyquist frequency sampling rate for reproducing the original signal [12].

These Gaussian target signals were generated for 12 frequency limits that were spaced logarithmically from 0.12 Hz to 12 Hz. Two signals were prepared for

Fig. 3. For pursuit tracking, the gesture waveform moved down the screen toward the touch strip, and the subject was asked to move her or his finger along the touch strip in synchrony with the waveform as it moved downward.

each bandwidth and in the two forms of pursuit tracking and pointing. Therefore, the total number of target gestures for each participant totalled 48 gestures.

3.3 Procedure

Participants were seated in a chair of appropriate height to allow comfortable movement and free range of motion to interact with the interface. The apparatus was laid upon a work-station surface with display and attached sensor facing up, oriented with the side closest to the sensor immediately before the subject. Subject participants used an interface on the laptop device to navigate the study options and continue through its phases. There, they were directed to follow target signals of 20 second duration on the sensor apparatus.

In presentation, the two types of signals moved at the same rate from the top to the bottom of the screen to approach and travel below the sensor, crossing its axis. Targets moved at a rate of 23.6 cm per second with a total preview visibility of 2.94 seconds and post-view visibility of 0.97 seconds. The feature of implementing post-view visibility was believed to be novel, but the authors

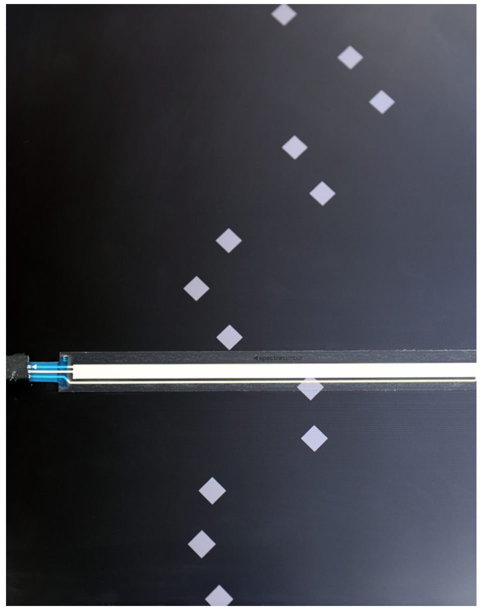

Fig. 4. For pointing gestures, diamond shapes were sampled from Gaussian targets at $2f_X$ Hz, where f_X is the bandwidth limit.

believed that it may have enabled subjects to more accurately see and follow the gesture. The range of display for the target gesture amplitude was 190 mm from a maximum value of +1.0 at the left to a minimum value at the right of -1.0.

In order to ensure a measurement of the channel capacity for participants familiar with the interface, a training phase introduced the types of gestures to the subjects in three escalating levels of difficulty. Subjects were offered the opportunity to repeat gestures in training and also to request additional gestures until they felt satisfied with their command of and familiarity with the interface.

Instructions were provided to describe the type of movements and to characterize the training difficulty levels. Three levels were provided in training for both pursuit tracking and pointing/tapping. The 0.7 Hz, 1.5 Hz, and 7 Hz bandwidth limits were presented as easy, medium, and difficult levels, respectively. For the difficult level, subjects were encouraged to make their best effort to perform the target gestures with as much accuracy as possible.

During the recorded portion of the study, the order of the 48 gestures was randomized throughout the trial in order to avoid factors that may result from learned agility or developed fatigue of participants. Participants were given the opportunity to rest, if requested.

Upon completion of each gesture trial, the guiding interface presented the option of retrying the completed gesture in case the subject felt, in their own estimation, that they could improve their performance. No performance feedback or error estimate was provided. The gesture could be repeated an unlimited number of times. When satisfied with their performance, the subject would then elect the option to accept the last performed gesture and continue to the next one.

The duration of subject trials was 35 to 40 minutes of continuous participation.

3.4 Analysis

Before conducting analysis using an information-theoretic approach, some adjustments to the data were made. First, in instances where a participant was not touching the control strip, either due to error in their use of the sensor or due to exceeding its effective sensor area, a value of -1.0 was recorded by the sensor apparatus (its rest value). Second, to compensate for errors of anticipation or delay while pointing, the beginning and ending samples for each point instance were located and extended to a midpoint between neighboring point instances, with the rationale that the modified signal was still communicating the same information inputted by the user, just transformed into a slightly different format.

Third, to account for instances where subjects were consistently late or early in the performance of the gestures, an iterative calculation of the mean-squared-error from -200 milliseconds to 200 milliseconds was conducted in relation to the target signal at 1 millisecond intervals. In the interest of finding maximum channel capacities, the most favorable delay interval within the resolution described above was tabulated and accepted as the representative value for each trial. With these adjustments, a best representation of the performed gesture is prepared for the channel capacity calculation.

Using the signal-to-noise ratio as calculated in the time domain, the channel capacity may be calculated, utilizing the bandwidth limits and the limits of human performance speeds as observed in this study. The bandwidth of the signal in the case of the human computer system is limited not only by the target design, but also by the capability of movement in time by the human participant. Where the target signal exceeded this capacity of movement, the upper limit is applied within the bandwidth component to calculate the channel capacity.

To wit, upon analysis of pursuit tracking results using the Fast Fourier Transform (FFT), the highest sustained frequency rate of movement observed was 5.6 Hz. An upper limit of 5.6 Hz was therefore applied as input to the bandwidth of the Shannon-Hartley equation for the 7 Hz and 12 Hz target results for pursuit tracking gestures. For pointing gestures, a maximum of 7.0 Hz was observed for a sustained pointing movement rate. Accordingly, a maximum of 7.0 Hz was applied to the channel capacity calculation for the 12 Hz target results for pointing.

3.5 Results

Main Subject Pool Eight subjects (1:7, female:male) from the main subject pool participated in the study. All subjects were musicians enrolled in either undergraduate or graduate music study at a research university. Subjects performed gestures with their dominant hand.

As shown in Figure 5, the mean observed channel capacity for pointing attained levels as high as 6 bits per second, representing the highest overall capacity for the subject pool. This peak channel capacity for pointing was at bandwidth limit 1.0 Hz, following a steady curve to that level and descending to the next highest capacity found near that level at 1.5 Hz.

The channel capacity of pursuit tracking similarly followed a discernible curve, clearly exceeding that of pointing capacities at 2.9 Hz and higher. Peak channel capacity for pursuit tracking was around 4 bits per second on average at bandwidth limit 2.3 Hz.

Analysis using Welch's t-test with Bonferroni correction identified any significance of differences across bandwidths between the two gesture types. It appears from these results that, with subjects having a very minimal amount of training, pointing at a lower frequency of movement allows communication of more information than pursuit tracking at such rates of movement. At 1.0 Hz, a mean of 2.6 bits/sec more information was communicated than with pursuit tracking (95% CI:1.53, 3.65; $p < 0.01$).

Under these conditions, at higher rates of movement, pursuit tracking appears to offer a higher capacity to communicate information. At 3.5 Hz, 2.4 bits/sec more information was communicated than with pointing (95% CI:1.8, 2.99, $p < 0.01$).

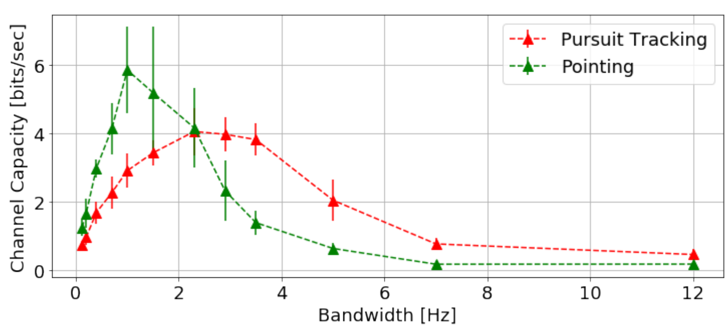

Fig. 5. Main subject pool: Estimated channel capacity across bandwidth limits f_X of target signals for pursuit tracking and pointing gestures.

A varying delay was observed for all subjects. There are several factors that could contribute to this delay. Screen refresh rates in relation to the recording of input gestures present information to the subject later than the recording.

Simple visibility of the target beneath the transparent sensor and estimation of its position under the opaque portion of the sensor could lead to some inaccuracy either before or after the recording moment. The delay of reaction to the previewed signal and delayed contact after the impulse to follow or touch the signal target point is a likely contributor to this observed delay as well.

A slackening of movement intensity was observed at the higher bandwidth limits for most participants, despite instructions of encouragement to try to follow as closely as possible or touch as many targets as possible. The seeming impossibility of following such a complex target or touching so many shapes at the rate presented was perhaps dispiriting. Fatigue could also be a factor here.

Author Data Two of the authors also participated in the study. Their data was treated separately as they had considerably more training gained during preparation of the study and apparatus design, although not as a controlled condition to prove a performance plateau. They also repeated their trials more frequently, in order to try to achieve even higher capacities. Their data is shown in Figure 6. Overall, these two authors were able to achieve higher capacities both for pointing and for pursuit tracking. The additional training appeared to provide more benefit for the pursuit tracking condition, under which the authors almost managed to catch up with their maximum channel capacities for pointing (see Figure 6).

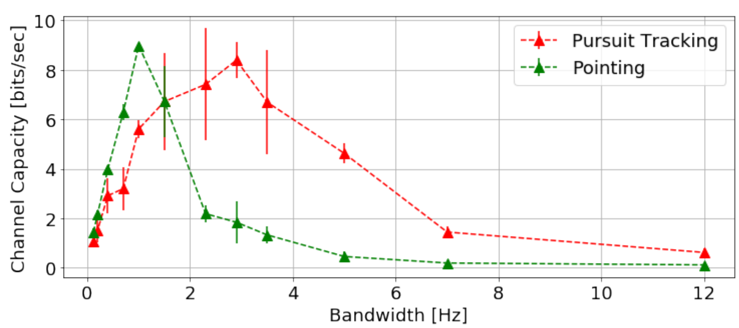

Fig. 6. Author data: Estimated channel capacity across bandwidth limits f_X of target signals for pursuit tracking and pointing gestures.

4 Discussion

In general, even with a training session component to the study design, the subjects performed as novice users compared to the authors in using the interface. Therefore, the channel capacity results should be considered maxima only for such a class of users. A more intensive training protocol, perhaps combined with a competition paradigm, could improve results and demonstrate a higher channel capacity for an advanced performer with significant practice on the interface.

Factors that could differentiate the novice from the experienced user could include a residual uncertainty due to novelty, inattentiveness during the session, and a lack of learned adaptive behavior that would assist with anticipating movement. These latter could include strategic thinking about how to best perform in light of high frequency signal components.

5 Conclusions

In summary, a comparison of pursuit tracking and pointing gestures was observed on a single analog sensor interface that was co-located with visual target stimuli. Application analysis based in information theory shows a straightforward means for evaluation of subject performance using the interface in these two ways.

In utilizing systems for applications that require higher throughput rates, composer/designers or performers can ensure that capacity is available by arranging their gestures to include pointing at a rate of 2.0 Hz to 3 Hz. Conversely, where movement of 5 Hz to 10 Hz is desired, it is clear that a higher throughput is available via a continuous control movement than via pointing.

Further investigation along these lines should include more ambitious training with interface use by subjects to seek limits beyond the novice level. Indeed, analysis of performances after memorization of the target gestures as would be the case with the performance of a composed musical work would be informative. Virtuosic levels of pointing or pursuit tracking may differ from the results found here. No feedback other than the benefits of co-location with the target stimuli were provided. Investigation of haptic, sonic, or visual feedback on the performance accuracy for subjects may demonstrate that higher capacities are possible when such information is incorporated into the human computer system.

References

1. Baty, D.L.: Effects of display gain on human operator information processing rate in a rate control tracking task. IEEE Transactions on Man-Machine Systems **10**(4), 123–131 (1969)
2. Berdahl, E., Blandino, M., Baker, D., Shanahan, D.: An approach for using information theory to investigate continuous control of analog sensors by humans. In: Proceedings of the Audio Mostly 2016. pp. 85–90 (2016)
3. Cover, T., Thomas, J.: Elements Of Information Theory. John Wiley and Sons, Hoboken, NJ, USA, 2nd edition edn. (2006)
4. Crossman, E.R.F.W.: The information-capacity of the human motor-system in pursuit tracking. Quarterly Journal of Experimental Psychology **12**(1), 01–16 (1960)
5. Doubler, J.A., Childress, D.S.: An analysis of extended physiological proprioception as a prosthesis-control technique. Journal of Rehabilitation Research and Development **21**(1), 5–18 (1984)

6. Elkind, J., Forgie, C.: Characteristics of the human operator in simple manual control systems. IRE Transactions on Automatic Control (1), 44–55 (1959)
7. Elkind, J., Sprague, L.: Transmission of information in simple manual control systems. IRE Transactions on Human Factors in Electronics (1), 58–60 (1961)
8. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of experimental psychology **47**(6), 381 (1954)
9. van Galen, G.P., de Jong, W.P.: Fitts' law as the outcome of a dynamic noise filtering model of motor control. Human movement science **14**(4), 539–571 (1995)
10. Shannon, C.: Communication in the presence of noise. Proceedings of the IRE, Proc. IRE (1), 10 (1949)
11. Smith, J.O.: Introduction to digital filters: With audio applications, vol. 2. W3K Publishing (2007)
12. Smith, J.O.: Mathematics of the discrete Fourier transform (DFT): With audio applications. W3K Publishing (2007)
13. Soukoreff, R.W., MacKenzie, I.S.: Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. Int. J. Hum.-Comput. Stud. **61**(6), 751–789 (Dec 2004)
14. Van Gemmert, A.W.A.: The effects of mental load and stress on the dynamics of fine motor tasks. Nijmegen: NICI, Nijmeegs Instituut voor Cognitie en Informatie (1997)
15. Wobbrock, J.O., Cutrell, E., Harada, S., MacKenzie, I.S.: An error model for pointing based on Fitts' Law. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1613–1622 (2008)