# An Approach for Using Information Theory to Investigate Continuous Control of Analog Sensors by Humans

**Conference Paper** · October 2016

DOI: 10.1145/2986416.2986450



David Baker
Louisiana State University

# An Approach for Using Information Theory to Investigate Continuous Control of Analog Sensors by Humans


Edgar Berdahl
Center Computation & Tech.
and School of Music
Louisiana State University
edgarberdahl@lsu.edu

Michael Blandino
Honors College and CCT
and School of Music
Louisiana State University
mblandi@lsu.edu

David Baker
School of Music
Louisiana State University
dbake29@lsu.edu

Daniel Shanahan
School of Music
Louisiana State University
dshanahan@lsu.edu



## ABSTRACT

For applications in HCI and sonic interaction design, the accuracy with which humans can continuously control analog sensors is investigated. The field of information theory suggests that a human together with a user interface can be modeled as a communication channel. Specifically, the Shannon-Hartley theorem implies that the channel capacity/throughput can be estimated by asking human subjects to "perform" gestures that match idealized, bandlimited Gaussian "target gestures." Then, the signal-to-noise ratio of the recorded gestures determines the channel capacity/throughput. This approach is tested on human users alternately operating one of four simple analog sensors.

In contrast with prior work in HCI, joint probability density functions do not need to be estimated nor must geometrically "impossible" gestures be eliminated from consideration. Suggestions are made for creating knowledge about user interfaces that could potentially transmit an enhanced amount of information to a computer.


## CCS Concepts

**•Human-centered computing → User models; Pointing; Empirical studies in HCI;** *Graphical user interfaces;* Virtual reality; **•Applied computing → Sound and music computing;**


## Keywords

User interface design, throughput, continuous control, channel capacity, Shannon-Hartley theorem, mutual information, information theory, sound and music computing





*AM '16, October 04 - 06, 2016, NorrkoÌĹping, Sweden*





ISBN 978-1-4503-4822-5/16/10. . . $15.00

DOI: http://dx.doi.org/10.1145/2986416.2986450


## 1. INTRODUCTION

In the field of sonic interaction design, custom user interfaces are often created for controlling sound in real time. Applications include digital musical instruments, sound art installations, games, sonification systems, context-aware devices, other industrial designs or even apparatus for experiments in psychology or cognition.

Interaction designers can choose from a wide variety of sensors for controlling sound. These sensors include discrete or "digital" sensors such as buttons, rotary encoders, or keyboards as well as continuous "analog" sensors such as potentiometer knobs, touch strips, sliders, distance sensors, force sensors, linear faders, among many others. Interaction designers typically learn about the various practical advantages and disadvantages of these sensors, but it is may also be useful to quantitatively consider the human ability to control them. Using perspectives from information and telecommunications theory, this paper examines the capacity of a human to continuously transmit analog information flowing into a computer.

### 1.1 Overview

Historically, in the broader study of Human-Computer Interaction (HCI), humans have been assumed to mostly engage in discrete-time interactions. The efficiency of these interactions has been estimated in terms of the **throughput** ($\lessapprox$ 5 bits per second for pointing [17], $\lessapprox$ 8 bits per second for typing in English [14], etc.). Information theory and telecommunications theory have successfully been applied for studying these discrete, target-oriented interactions and relating them to the observed throughputs [1]. For example, Fitts' Law provides an index of difficulty measured in bits that can model the time it takes humans to point at target areas using a particular user interface [17]. Due to its wide applicability and adaptability, Fitts' Law has been employed to analyze various systems in HCI [3, 2], inspiring a continuing wave of quantitative studies in HCI throughput [17].

However, to the authors' knowledge, such information-oriented quantitative approaches have not previously been extended to study human control of continuous sensors, without requiring detailed modeling of probability density functions. The straightforward approach proposed in this work

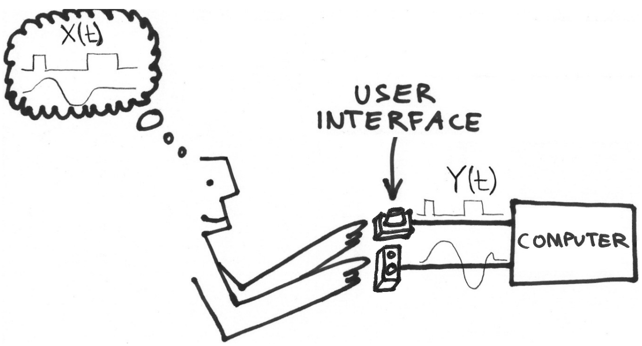


**Figure 1: Model of a user controlling sound using continuous analog sensors (© Edgar Berdahl).**

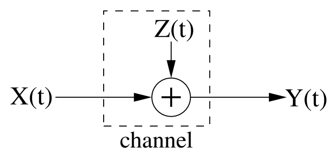


**Figure 2: Traditional model of a bandlimited communications channel from information theory.**

could be useful for studying musical interfaces, since continuous control is particularly helpful or perhaps even crucial for realizing expressive music interaction [18], and other interfaces of interest could benefit as well such as for gaming, aircraft, boats, teleoperated robots, etc.

## 2. MODEL OF USER'S INPUT

The authors propose to model the user's input as shown in Figure 1. It is presumed that the user is aware of *target signals* [19] that are to be transmitted to a computer via the sensors. As shown in a thought bubble connected associated with the user's mind in Figure 1 (left), these signals are represented by $X(t)$.

Secondly, it is considered that the user's goal is for the *sensor signals* $Y(t)$ to match the target signals $X(t)$ as closely as possible. In other words, ideally the user would be able to have the target signals registered perfectly by the computer ($Y(t) = X(t)$), but in practice, it is not feasible for a human user to completely accurately operate a continuous, analog sensor in real time.

## 3. CONNECTIONS WITH INFORMATION THEORY

### 3.1 Considering the User and Sensors as Forming a Communications Channel

Taking the model from Figure 1 into account, the user and sensors acting together can be considered to serve as a *communications channel*, essentially an entity that communicates the target signals $X(t)$ to a computer, which receives the sensor signals $Y(t)$. Imperfections in $Y(t)$ can be modeled as an additive noise signal $Z(t)$ such that

$$Y(t) = X(t) + Z(t) \tag{1}$$

as illustrated in Figure 2. Causes of the signal $Z(t)$ include noise in human cognition, noise in the human motor control system, physical and electronic sensor noise, and analog-to-digital quantization noise, etc.

### 3.2 The Channel Capacity

The channel capacity $C$ of a channel is an important quantity, as it describes the maximum number of bits that can be transmitted over a communications channel [4, 8]. In general, determining the channel capacity $C$ for a specific channel may be quite challenging—this value is the maximum throughput that can be achieved given any scheme for storing meaning in the $X(t)$ signals and for any possible decoder that extracts meaning from $Y(t)$. Formally, the channel capacity is defined as the following:

$$C = max_{p(X)} \quad I(X;Y), \tag{2}$$

where $I(X;Y)$ denotes the mutual information between $X(t)$ and $Y(t)$ and $p(X)$ is the probability density function for $X(t)$ [4, 8]. (2) shows that generally speaking, all possible signal sets for $X(t)$ described by $p(X)$ may need to be considered for a given channel.

### 3.3 Prior Work and Challenges

It's possible that the complex nature of (2) explains why relatively few attempts have previously been made at estimating the channel capacity for continuous control of analog sensors. Oulasvirta et al. investigate this question in the context of 3D motion capture interfaces and touchscreens, targeting the related application of gesture recognition [12, 16]; however, the authors appear to be overestimating the channel capacity, according to its formal definition [4, 8]. The reason for this is that Oulasvirta et al. do not have a way of discounting certain input signals that may be impossible for users to perform–for example, it is not possible to put the human fingertips in any randomly selected arbitrary position in Cartesian space. However, since Oulasvirta et al. are counting such orientations, the estimated channel capacity is greatly inflated and no longer directly comparable with traditional throughput measurements [12].

Furthermore, modeling $I(X;Y)$ generally is quite challenging because accurately estimating the joint distribution $p(X,Y)$ may require a very large amount of data, particularly when both $X(t)$ and $Y(t)$ are continuous. Or, if a smaller amount of data is employed, the numerical estimation of $I(X;Y)$ and therefore also $C$ will be biased. For example, in 1960 that approach was applied in the context of studying human control of a steering wheel. The bias in $I(X;Y)$ and $C$ was believed to be on the order of 10%, but no reference was available for experimental comparison [6].

In contrast, the present work proposes an alternative approach for estimating the channel capacity by avoiding the need to estimate $p(X,Y)$ and $I(X;Y)$. This is facilitated by verifying the hypothesis that the channel shown in Figure 1 can be modeled as a bandlimited Gaussian channel, and in that case, choosing $X(t)$ to be Gaussian-distributed maximizes the mutual information $I(X;Y)$ [4]. For this reason then, the following approach is proposed:

- conduct a pilot study in order to evaluate what range and density of channel bandwidths $f_X$ should be tested (see Section 5),
- generate independent and identically distributed (i.i.d.) Gaussian noise signals,

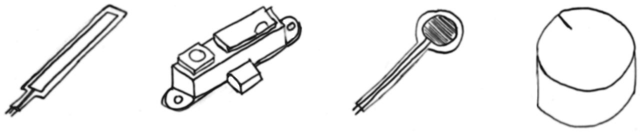

**Figure 3: From right to left, the sensors used in the study: 10cm-long "Softpot" touch strip, Sharp distance sensor (4cm - 30cm), Interlink force-sensing resistor (0.5"/1.27cm in diameter), and knob connected to a one-turn linear-taper potentiometer.**

- low-pass filter them at a series of cutoff frequencies $f_X$ to obtain the target signals $X(t)$,
- using a human subject test, record the corresponding sensor signals $Y(t)$ in response,
- estimate the signal-to-noise (SNR) ratios for the trials,
- apply the Shannon-Hartley theorem to estimate the channel capacities $C$, taking the maximum over the tested channel bandwidths $f_X$ for each sensor.

# 4. HUMAN SUBJECT TEST

Four simple continuous sensors with analog output were selected from the sensors available at Sparkfun Electronics. These sensors have been commonly employed for sonic interaction design. The sensors were an 10cm-long "Softpot" touch strip, a Sharp GP2Y0A41SK0F distance sensor (4cm - 30cm), an Interlink force-sensing resistor (0.5"/1.27cm in diameter), and a one-turn linear-taper Panasonic potentiometer (EVU-F2AF30B14) (see Figure 3). A human subject test was designed in order to estimate the channel capacity in bits of novice users continuously controlling the sensors.

It was hypothesized that these communication channels would be demonstrated to be bandlimited Gaussian channels if bandlimited Gaussian target signals $X(t)$ were employed. Therefore, the $X(t)$ target signals were generated using i.i.d. Gaussian-distributed random white noise, which was lowpass-filtered at the cutoff frequencies 0.25Hz, 0.5Hz, 1Hz, and 2.5Hz.

14 undergraduate students in music were recruited to participate in the human subject test. The focus was on designing a simple pilot experiment that would not be overly challenging for the test subjects and that could be expanded upon for future studies. For this reason, the sensors were presented sequentially to the participants. For each sensor, the subject was first trained on the sensor in the matter of a few minutes. In the training, the subject was asked to control the sensor in such a way that the sensor signal $Y(t)$ followed a visually presented target signal $X(t)$ for twenty seconds. Then, following the training session, the subject was randomly presented with four other target signals $X(t)$ filtered at the bandwidths 0.25Hz, 0.5Hz, 1Hz, and 2.5Hz. For each of these target signals, the subject "performed" the target signal until she or he was satisfied with how well her or his sensor signal $Y(t)$ matched $X(t)$. To help compensate for learning and fatigue effects, different test subjects were presented with the sensors in different orders. For more information on the procedure, please see the following video: `https://vimeo.com/167634709`

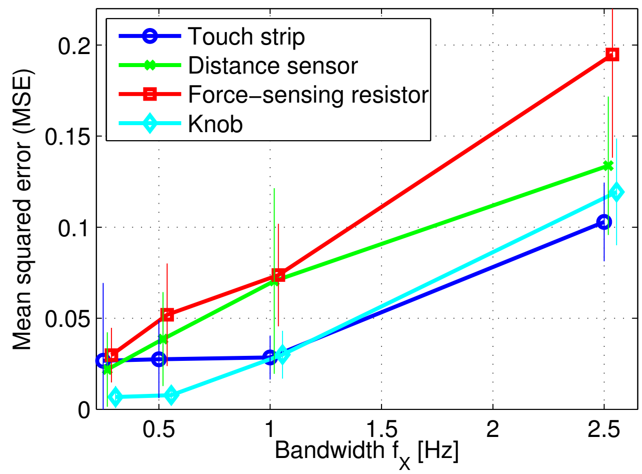


**Figure 4: Mean squared error of the test subjects operating each of the sensors at various bandwidths. The error bars indicate one standard deviation above and below the estimated means. (Note: Trials were only completed at the bandwidths 0.25Hz, 0.5Hz, 1Hz, and 2.5Hz, but slight horizontal offsets were introduced above in order to prevent the error bars from overlapping.)**

One subject elected to quit the test before completing all of the trials. Because this subject's data also appeared to be very noisy overall, perhaps due to a potential lack of focus during the test, it was decided to eliminate that subject from the pool. The other 13 subjects completed all trials, yet errors were observed with the touch strip sensor. Users tended to "fall off" of the side or bottom of the touch strip, which would cause the sensor signal $Y(t)$ to rapidly approach full-scale. Therefore, as user error was not under test in this experiment, and as it was desired to be able to use as much of the recorded data as possible, all $Y(t)$ samples within 4% of either end of the sensors' ranges were discarded.

## 4.1 Mean Squared Error

The mean squared error (MSE) $E(Z^2) = E((Y-X)^2)$ estimated for the trials was plotted in Figure 4. Two-way analysis of variance indicated that both the bandwidth in Hz and the sensor type had a statistically significant effect overall (p-value of 0.05).

Paired t-tests were additionally conducted for each bandwidth $f_X$ to compare if the different sensors resulted in statistically different MSE. Because the goal was to evaluate if some comparisons could be made statistically (and not if *all* comparisons could be made), the Bonferroni correction was not applied). Of the 24 comparisons, 13 were statistically significant (p-value of 0.05). Similarly, paired t-tests were conducted for each sensor to compare if changing the bandwidth $f_X$ had a statistically significant effect. 17 of the 24 comparisons were deemed statistically significant (again p-value of 0.05 and no Bonferroni correction).

The force-sensing resistor and distance sensor resulted in the largest MSE on average. Fundamentally different than the other sensors, the force-sensing resistor might have been more challenging for test subjects to control because it aimed to measure the force applied rather than the position or angle. The distance sensor might have been problematic because the test subject's hand was not restricted to be on axis with the sensor, complicating the relationship between the sensor output and the position of the hand. Compared with the other sensors, both the force-sensing resistor and the distance sensor were quite nonlinear, which may also have contributed to the difficulty in controlling them accurately given only a few minutes of training time.

As expected [17, 15, 10], the faster moving (e.g. higher bandwidth) target signals tended to be harder to follow, as evidenced by the generally larger MSE shown at higher bandwidths (see Figure 4, right). However, the low MSEs in Figure 4 (left) do not indicate a maximal amount of information transmission since the signals are so slowly changing, motivating the need for estimating the channel capacity.

## 4.2 Channel Capacities

The empirical distribution of $Z(t) = Y(t) - X(t)$ was studied for each (bandwidth, sensor) pair and appeared to be approaching Gaussian distributions. This was not surprising due to the Gaussianity of the target signals $X(t)$ and the central limit theorem [11, 4], but it needed to be verified. Then, since the target signals $X(t)$ were also originally chosen to be bandlimited Gaussian noise, the channel in Figure 1 was determined to formally be a bandlimited Gaussian channel [4]. Moreover, the mutual information $I(X;Y)$ was already maximized by virtue of $X(t)$ and $Z(t)$ being Gaussian-distributed over the bandwidth $f_X$. Therefore, the channel capacity $C(f_X)$ could be estimated using the Shannon-Hartley theorem [4]:

$$\hat{C}(f_X) = f_X \cdot \log_2 \Big(1 + S\hat{N}R(f_X)\Big), \tag{3}$$

where $S\hat{N}R(f_X)$ was the observed signal-to-noise ratio of the channel when operated at bandwidth $f_X$. Accordingly, $SNR(f_X)$ could accordingly be estimated using the following ratio:[1]

$$SNR(f_X) = \frac{E(X_{(f_X)} Y_{(f_X)})}{E(Z^2_{(f_X)})}. \tag{4}$$

Accordingly, the test subject data from Figure 4 was analyzed and presented in Figure 5. Two-way analysis of variance indicated that both the bandwidth in Hz and the sensor type had a statistically significant effect overall (p-value of 0.05).

Paired t-tests were additionally conducted for each bandwidth $f_X$ to compare if the different sensors resulted in statistically different capacities. Slightly fewer comparisons were statistically significant compared to Section 4.1, but this is expected now that the means appear to no longer increase monotonically with the bandwidth $f_X$ (see Figure 5). Of the 24 comparisons, 10 were statistically significant (p-value of 0.05, no Bonferroni correction). Similarly, paired t-tests were conducted for each sensor to compare if changing the bandwidth $f_X$ had a statistically significant effect. 11 of the 24 comparisons were deemed statistically significant (p-value of 0.05, no Bonferroni correction). More test subjects would be needed in order to obtain more statistically significant comparisons; however, as discussed below, the authors would like to increase the amount of training in future subject tests, which seems likely to increase the channel capacities and reduce the standard deviations (see Figure 5).

[1]The careful reader will note that the signal power needed to be estimated using $E(X_{(f_X)} Y_{(f_X)})$ instead of $E(X^2_{(f_X)})$ in order to compensate if test subjects applied a gain other than one to the target signal $X(t)$ in producing the sensor signal $Y(t)$.

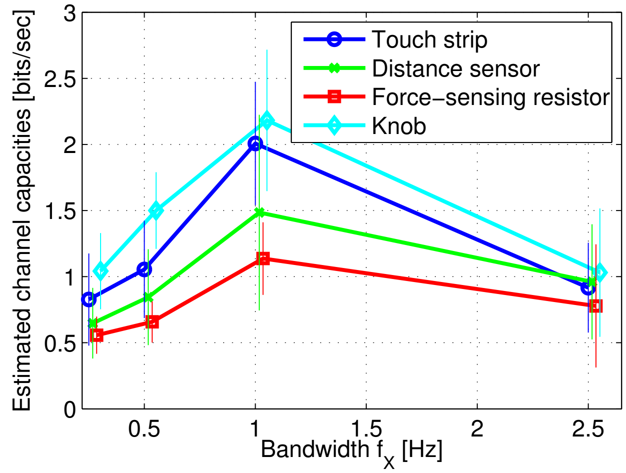


**Figure 5: Channel capacity estimates for a novice user operating each of the sensors across a range of bandwidths. The error bars indicate one standard deviation above and below the estimated means. (Note: Trials were only completed at the bandwidths 0.25Hz, 0.5Hz, 1Hz, and 2.5Hz, but slight horizontal offsets were introduced above in order to allow for visual inspection of all the error bars.)**

As expected, although the increase in bandwidth $f_X$ had the potential to increase the channel capacity estimates, it also decreased the signal-to-noise ratio, which could also cause the channel capacity estimates to decrease. With this data set, the highest channel capacity estimates were found at a bandwidth of $f_X = 1$Hz. Viewing the corresponding data points in Figure 5, it follows that the test subjects would have been able to communicate more information on average using the touch strip and knob, at a bandwidth of 1Hz, than in the other conditions.

The channel capacity across all bandwidths $f_X$ could be estimated using the following:

$$C = \max_{f_X} \hat{C}(f_X) = \max_{f_X} \left( f_X \cdot \log_2 \Big(1 + S\hat{N}R(f_X)\Big)\right). \tag{5}$$

Although some variance was present, this data suggested that, overall on average, the test subjects could achieve a channel capacity estimate of as high as about **2 bits per second** for the touch strip and knob (see Figure 5), and somewhat less for the distance sensor and force-sensing resistor.

However, it was possible that the channel capacities could have been higher if measured at some bandwidths in between 0.5Hz and 2.5Hz. Moreover, it was hypothesized that the channel capacities were likely further limited by the small amount of training (e.g. ≈5 minutes) that the test subjects received.

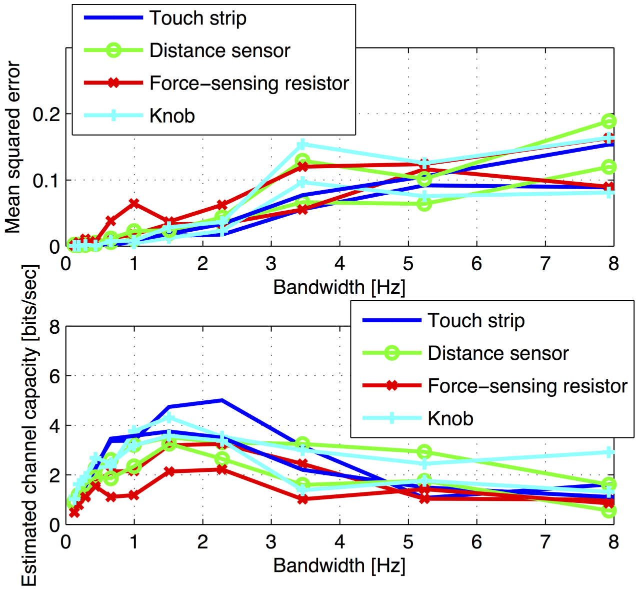


**Figure 6: Mean square error (above) and estimated channel capacity measured for the first two authors.**

## 5. FOLLOW-UP TEST

In order to try to shed some light on the shape of the bandlimited channel capacity estimate curves in more detail, the first two authors decided to measure themselves at twelve bandwidths, logarithmically spaced in between 0.125Hz and 12Hz. Besides checking more frequency points, the authors had had much more training than the recruited human subjects, so it was expected that the channel capacities would increase. Although the data points corresponded only to single trials in Figure 6, the structure appeared to be a plausible extension of Figures 4 and 5.

Subjectively, the bandwidth setting at 8Hz seemed to be too fast for the human motor control system to keep up with [13].[2] This was evidenced by the particularly elevated mean squared error at high bandwidths (see Figure 6, top-right).

The peaks of the estimated channel capacities occur in the neighborhood of 1.5Hz to 2Hz, with the touch strip and knob sensors. Interestingly, this data suggests that **the channel capacity estimates for these sensors for sufficiently trained users may extend as high as 4 to 5 bits per second or even higher**. Although these data points correspond to single trials, they fall within the general pattern of the neighboring trials (see Figure 6, bottom). Since the channel capacities are higher than in Figure 5, it seems quite clear that it is very important for users to be trained in accurately operating the sensors and to carefully choose the right range and density of bandwidths $f_X$ to test.

## 6. CONCLUSIONS

Evidence was presented of channel capacities for control of a single, continuous sensor as high as 4 or 5 bits per second. These results are in agreement with prior experimental work using one specific interface [6]. These channel capacity estimates are very competitive with common discrete-time interfaces, which provide throughputs of between 5 and 8 bits per second. Therefore, it is quite plausible that, through continuous control of multiple sensors simultaneously, trained human users might be able to transmit more information to a computer than with discrete-time sensors.

This result is in accordance with the historical development of acoustic musical instruments, which has resulted in designs such as the violin. The violin is revered because it affords users continuous control over multiple expressive parameters simultaneously–pitch, bow pressure, bow tilt, bow position along the string, bow velocity, etc., all via a single, unified interface. Moreover, via these kinds of traditional musical interfaces, abstract high-level information must be encoded. For example, research in music psychology has examined the musician's ability to communicate emotion effectively. Gabrielsson and Juslin found that the intended emotion of a performer has significant effects on aspects of tempo, dynamics, timing, and spectrum, and that these effects were similar across performers [5, 9]. Interestingly, listeners were able to decode these effects quite accurately. Consequently, barring visual communication and prior knowledge of listeners, the information for communicating emotion must be encoded within the sensor signals $Y(t)$ depicted in Figure 1 for a given musical instrument user interface under test.

In summary, a relatively simple procedure for estimating the channel capacity of a human user operating a continuous analog sensor has been proposed and tested. This method requires much less data than previously investigated methods, and the results are more general: the *channel capacity* is estimated for human-control of a given analog sensor rather than the *throughput* for a specifically designed user interface using the analog sensor. And the throughput (aka. the mutual information) is always lower-bounded by the channel capacity [4].

The approach presented in this paper is made possible by employing bandlimited **Gaussian target gestures** as the signals $X(t)$. It is very important for test subjects to be sufficiently trained in accurately operating the sensors. Also the the range and density of bandwidths $f_X$ to test must be selected with care. In future work, the authors hope to provide higher estimates of the channel capacities for these sensors by employing a more rigorous training procedure. It is also planned to estimate the channel capacity of a wider series of computerized sensor systems as well as the violin.

## 7. REFERENCES

[1] J. Accot and S. Zhai. Performance evaluation of input devices in trajectory-based tasks: An application of the steering law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 466–472, New York, NY, USA, 1999. ACM.

[2] X. Bi, Y. Li, and S. Zhai. Ffitts law: Modeling finger touch with fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1363–1372, New York, NY, USA, 2013. ACM.

[3] X. Bi, B. Smith, and S. Zhai. Multilingual touchscreen keyboard design and optimization. *Human-Computer Interaction*, 27:352–382, 2012.

[2]Nonetheless, faster speeds up to 30Hz have been demonstrated to be achievable in very specific nonlinear environments relating to performance of drum rolls [7].

[4] T. Cover and J. Thomas. *Elements Of Information Theory*. John Wiley and Sons, Hoboken, NJ, USA, 2nd edition edition, 2006.

[5] A. Gabrielsson and P. N. Juslin. Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of music*, 24(1):68–91, 1996.

[6] E. R. F. W. Grossman. The information-capacity of the human motor-system in pursuit tracking. *Quarterly Journal of Experimental Psychology*, 12(1):01–16, 1960.

[7] A. Hajian, D. Sanchez, and R. Howe. Drum roll: Increasing bandwidth through passive impedance modulation. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2294–9, Albuquerque, New Mexico, April 20-25 1997.

[8] N. Jayant and P. Noll. *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall, Upper Saddle River, NJ, USA, 1984.

[9] P. N. Juslin. Emotional communication in music performance: A functionalist perspective and some data. *Music Perception: An Interdisciplinary Journal*, 14(4):383–418, 1997.

[10] C. Mayheld and S. Moss. Effect of music tempo on task performance. *Psychological reports*, 65(3f):1283–1290, 1989.

[11] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. E. K. Smith. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological Review*, 95:340–370, 1988.

[12] A. Oulasvirta, T. Roos, A. Modig, and L. Leppänen. Information capacity of full-body movements. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 1289–1298, New York, NY, USA, 2013.

[13] G. Rasmussen. Human body vibration exposure and its measurement. Technical Report 1, Brüel and Kjaer, Naerum, Denmark, 1982.

[14] C. Shannon. Prediction and entropy of printed english. *Bell System Technical Journal*, 3:50–64, 1950.

[15] T. Sheridan and B. Verplank. Human and computer control of undersea teleoperators. Technical report, Department of Mechanical Engineering, Massachusetts Institute of Technology, Boston, MA, USA, 1978.

[16] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos. User-generated free-form gestures for authentication: Security and memorability. In *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '14, pages 176–189, New York, NY, USA, 2014. ACM.

[17] R. W. Soukoreff and I. S. MacKenzie. 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.

[18] D. Wessel and M. Wright. Problems and prospects for intimate musical control of computers. *Computer Music Journal*, 26(3):11–22, Sept. 2002.

[19] D. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7-8):1317–29, Oct.-Nov. 1998.