Information-Theoretic Analysis of Human Performance for Command Selection

Wanyu Liu, Olivier Rioul, Michel Beaudouin-Lafon, Yves Guiard

Information-Theoretic Analysis of Human Performance for Command Selection

Wanyu Liu[1,2](✉), Olivier Rioul[1], Michel Beaudouin-Lafon[2], and Yves Guiard[1,2]

[1] LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France
{wanyu.liu,olivier.rioul,yves.guiard}@telecom-paristech.fr
[2] LRI, Univ. Paris-Sud, CNRS, Inria, Université Paris-Saclay, 91400 Orsay, France
mbl@lri.fr

Abstract. Selecting commands is ubiquitous in current GUIs. While a number of studies have focused on improving rapid command selection through novel interaction techniques, new interface design and innovative devices, user performance in this context has received little attention. Inspired by a recent study which formulated information-theoretic hypotheses to support experimental results on command selection, we aim at explaining user performance from an information-theoretic perspective. We design an ad-hoc command selection experiment for information-theoretic analysis, and explain theoretically why the transmitted information from the user to the computer levels off as difficulty increases. Our reasoning is based on basic information-theoretic concepts such as entropy, mutual information and Fano's inequality. This implies a bell-shaped behavior of the throughput and therefore an optimal level of difficulty for a given input technique.

Keywords: Human performance · Command selection · Information theory · Mutual information · Entropy · Throughput · Fano's inequality

1 Introduction

Selecting commands remains one of the most common interactions in graphical user interfaces. Many previous studies have strived to improve rapid command selection, e.g., marking menus [10] and flower menus [1], to propose novel command selection techniques, e.g., FastTap on tablets [7], and to design command selection on innovative devices, e.g., smartwatches [11]. When an application has a large number of commands, designers often use a hierarchical navigation structure to partition the components or come up with new designs such as finger identification [5], particularly when interaction is constrained by scarcity of screen real estate.

However, apart from some menu models [2,3] that are mostly applicable to linear menus, few researchers have systematically investigated user performance

R. Bernhaupt et al. (Eds.): INTERACT 2017, Part III, LNCS 10515, pp. 1–10, 2017.
DOI: 10.1007/978-3-319-67687-6_35

in command selection. We are inspired by a recent study where Roy et al. [17] analyzed command selection data in information-theoretic terms such as transmitted information and successfully transmitted information rate, also known as throughput (TP), from the user to the computer. In the communication channel considered in [17], a user serves as the source of information with her hand as the information emitter, and transmits information to the system with the touch screen as the receiver of the coded message. The code shared by the source (the user) and the destination (the system) is the mapping of a set of touch events to a set of commands. Roy et al. hypothesized that the transmitted information levels off, as in absolute judgment tasks [13], and that TP as a function of the command's entropy is bell-shaped. As they were focused on comparing two input techniques, the authors used these measurements to illustrate the differences between two techniques. In this paper, we provide instead a theoretical analysis of these phenomena.

Information theory, first introduced by Shannon in 1948 [19], has been used in various domains including psychology [12,14,15]. In HCI, there are two major design principles that are derived from information theory: Fitts' law [6] and the Hick-Hyman law [8,9] starting from the early 1980s [16]. Fitts' [6] work was an empirical determination of the information capacity of the human motor system. Likewise, Hick's [8] and Hyman's [9] experiments assessed the cognitive information capacity in choice-reaction experiments. Fitts' law and the Hick-Hyman law are the only two surviving information-theoretic concepts used in HCI [18], despite the fact that we humans constantly send information to and receive information from computers, and vice versa.

Intrigued by the observations in [17], we aim to provide an information-theoretic analysis of user performance in command selection tasks. We thus replicated Roy et al.'s experiment [17], tailored for an information-theoretic analysis. Using basic concepts including entropy, mutual information and Fano's inequality, we provide an information-theoretic explanation for why transmitted information (mutual information between the user and the computer) should level off as the command's entropy increases, and why the rate of successfully transmitted information is a bell-shaped curve. Grounded in the fundamental principles of information theory, these formulations provide a general tool to evaluate human performance in command selection tasks.

2 Data Collection

The goal was to achieve a better undertanding of human performance in a command selection task from the information-theoretic perspective and to provide theoretical formulations for it. Similar to Roy et al.'s study [17], we consider users as information sources, who emit information with an interaction instrument, and transmit it to the system. In order to avoid the fat finger problem [20] and to collect a wider range of data, we choose to use a mouse as interaction instrument instead of the user's hand. We also assume equally probable commands, thus the input entropy (the task difficulty) is the $\log_2$ of the number of possible commands.

2.1 Participants and Apparatus

Twelve volunteers (1 female), age 23 to 31 (mean = 26.6, $\sigma = 1.9$), were recruited from our institution. All of them were right-handed and interacted with WIMP interfaces regularly.

The experiment was conducted on a Macbook Pro with a 2.7 GHz processor, 8 GB RAM and resolution at 227 pixels per inch. The software was implemented in Java and the experiment window was 600×400 pixels. The targets representing the commands were displayed at the top of the window as a row of adjacent rectangles. The total area covered by the targets was 256 pixels wide and 30 pixels high. The width of the targets depended on the experimental condition. A circle positioned 150 pixels down below the target area was used to reset the cursor position of each trial. A standard mouse was used with the same sensitivity for all participants.

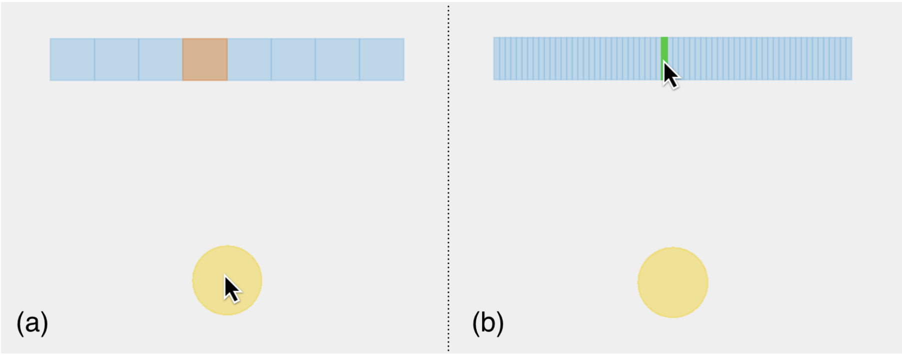

Fig. 1. (a) The cursor gets reset at the center of the circle when the trial starts in condition 8; (b) correctly selected target command turns green in condition 64. (Color figure online)

2.2 Task, Stimulus and Design

In response to a visual stimulus, participants were instructed to click on the highlighted target command (Fig. 1(a)) as fast and accurately as they could. If they correctly hit the target command, it turned green (Fig. 1(b)). Clicking on a non-target command would turn it red. In both cases the trial was complete after a single selection. The cursor was then reset automatically in the same position at the start of each trial.

Based on a pilot study, we used 4, 8, 16, 32, 64, 128 and 256 commands in the experiment, corresponding to 2 to 8 bits of information. Note that more than 7 bits of information is relatively high for normal users to process, but we wanted to push the limits of the participants. The size of the target representing each command was inversely proportional to the number of commands in the set, so that the set of target commands always occupied the same overall space.

We used a within-participant design and counter-balanced the order of the number of commands across participants with a Latin square. There were 3 replications for each block. A block consisted of presenting all targets in random order. Since we assumed a uniform distribution, each command should appear the same number of times. However, this would result in a very long and tiring selection in condition 128 and 256. In order to keep the duration of the experiment manageable, each participant had to select only 64 targets in conditions 128 and 256, but the full range was covered across all participants.

The total duration of the experiment was around 20 min per participant. In summary, the design was: 12 Participants $\times$ $(4+8+16+32+64+64+64$ Commands$)$ $\times$ 3 Replications = 9,072 trials.

3 Information-Theoretic Concepts and Notations

Before presenting the experimental results and providing the information-theoretic analysis, we define the following notations.

1. X is a random variable that takes values in $\{1, 2, \ldots, M\}$, representing the user's intended input. The number of targets is $M = 4, 8, 16, 32, 64, 128, 256$ depending on the experimental condition.
2. We assume X is uniformly distributed so that the probability $P(X = x) = \frac{1}{M}$ for all x.
3. The entropy of X is given by $H(X) = \log_2 M$ and represents the command's entropy (the task difficulty).
4. Y is another random variable that takes values in $\{1, 2, \ldots, M\}$, representing the command that is actually hit by the user, whether correct or not.
5. The error random variable is defined as:
$$E = \begin{cases} 0 & \text{if } X = Y; \\ 1 & \text{if } X \neq Y. \end{cases} \tag{1}$$
6. The probability of error $P_e = P(X \neq Y)$ representing the error rate, has binary entropy:
$$H(E) = -P_e \log_2 P_e - (1 - P_e) \log_2 (1 - P_e). \tag{2}$$
7. The transmitted information in bits conveyed from the user to the computer is defined by Shannon's mutual information:
$$I(X;Y) = \sum_x \sum_y P(X = x, Y = y) \log_2 \frac{P(X = x, Y = y)}{P(X = x) P(Y = y)}$$
Shannon's capacity is defined as the maximum possible transmitted information.
8. Throughput (TP) or information rate, in bits per second, is defined as the ratio of the transmitted information to movement time (MT), i.e., the average time to select the command for a given number of commands: $\mathrm{TP} = I(X;Y)/\mathrm{MT}$.

4 Experimental Results

In this section, we present findings on reaction time (RT), movement time (MT), error rate P_e, transmitted information $I(X;Y)$, and throughput TP as a function of the command's entropy (task difficulty).

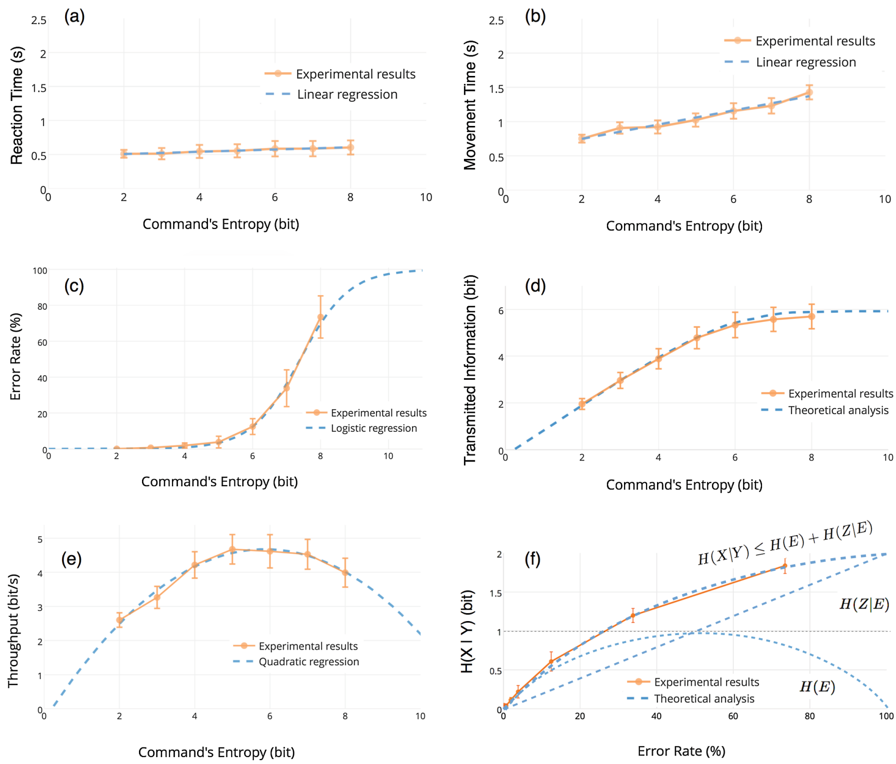

Fig. 2. Experimental results with 95% confidence intervals.

Task Completion Time. Figure 2(a) and (b) indicate that both reaction time RT and the movement time MT required to select a command are linear functions of the command's entropy. This is in line with the Hick-Hyman Law and Fitts' Law since time is proportional to task difficulty, which is logarithmic in the number of choices. We run linear regressions and find that $\text{RT} = 0.473 + 0.017 \times \log_2 M$ with $r^2 = 0.959$ and $\text{MT} = 0.540 + 0.104 \times \log_2 M$ with $r^2 = 0.968$. Task completion time is dominated by movement time MT.

Error Rate. Figure 2(c) demonstrates that when the command's entropy is small and the task is easy, users do not tend to make mistakes, hence P_e is very small. When the command's entropy increases, users make more and more errors, up to 73.5% when entropy equals 8 bits. It is obvious that when the command's entropy gets very high, the error rate would level off at 100%, as shown in Fig. 2(c). Fitting a logistic curve to the data, we obtain $P_e = 1/(1 + e^{-1.4\times(\log_2 M - 7.4)})$ with $r^2 = 0.992$.

Mutual Information. Figure 2(d) shows that $I(X;Y)$ increases gradually with the command's entropy. Similar to [17], transmitted information tends to reach an upper bound, confirming the limited capacity. The reason why it levels off given in [17] was that it is similar to absolute judgment tasks [13]. The next section offers another explanation based on information theory.

Throughput. Similar to Roy et al.'s study [17], throughput (TP) in our experiment also shows a bell-shaped behavior and reaches a maximum as shown in Fig. 2(e). Fitting a quadratic function to the data, we find TP $= -0.341 \times (\log_2 M)^2 + 1.710 \times \log_2 M - 0.146$ with $r^2 = 0.979$.

5 Information-Theoretic Analysis

In this section, we provide an information-theoretic analysis for (a) why mutual information should level off; and (b) why throughput should be a bell-shaped function of the command's entropy (task difficulty). Since the user makes errors, the output Y received by the computer is not always equal to the input X sent by the user, and therefore there is noise in the channel: Y is essentially X perturbed by a "noise" Z, which can take M possible values: one corresponding to the correct command plus $M-1$ corresponding to the possible mistakes that the user can make.

As is well known in information theory [4, Theorem 2.4.1], the mutual information is the difference between the input entropy $H(X)$ and the conditional entropy $H(X|Y)$ of the input given the output:

$$I(X;Y) = H(X) - H(X|Y) \tag{3}$$

The conditional entropy is a measure of the uncertainty about X knowing Y; but if we know Y, the uncertainty on the noise Z is the same as that on X, so we can rewrite Eq. (3) as:

$$I(X;Y) = H(X) - H(Z|Y) \tag{4}$$

Here $H(X) = \log_2 M$ represents the task difficulty. We now would like to bound the penalty term—also known as *equivocation*—$H(Z|Y)$ in the transmitted information. Since the knowledge of the output Y reduces the uncertainty on the noise Z (conditioning reduces entropy [4, Theorem 2.6.5]), we have:

$$H(Z|Y) \leq H(Z) \tag{5}$$

In words, the equivocation does not exceed the entropy of the noise. Thus it is the noise's entropy that penalizes the transmitted information.

In our experiment, users make errors as defined in Eq. (1), and we can use the chain rule [4, Theorem 2.2.1]: $H(Z) = H(Z, E) = H(E) + H(Z|E)$ where [4, Theorem 2.2.1]

$$H(Z|E) = P_e \times H(Z|E = 1) + (1 - P_e) \times H(Z|E = 0) = P_e \times H(Z|E = 1) \quad (6)$$

since there remains no uncertainty on the noise Z if there is no error ($E = 0$). Combining the above, we have that the equivocation is bounded by:

$$H(X|Y) \leq H(E) + P_e \times H(Z|E = 1) \quad (7)$$

This is known in information theory as *Fano's inequality* [4, Theorem 2.10.1].

Here $H(E)$ is given by Eq. (2) and is at most one bit (when $P_e = 0.5$). Hence making errors penalizes the amount of transmitted information by at most one bit. However, considering the second term of Eq. (7), the uncertainty on "wrong selections" $H(Z|E = 1)$ incurs an additional penalty on the amount of transmitted information: How users make errors, not just the fact that they make errors, affects the amount of transmitted information. In our case, errors are clustered near the actual target, hence the entropy of the noise is lower than if they were evenly distributed.

The relationship between error rate P_e and $H(X|Y)$ observed from empirical data matches exactly the above illustration as shown in Fig. 2(f).

We can now reason as follows:

For small M: users do not tend to make errors, $H(E) \approx 0$ and $P_e \approx 0$, therefore $H(X|Y)$ is close to zero or remains very small when the error rate is low. So $I(X;Y)$ increases with $H(X) = \log_2 M$;

For large M: we tend to have $P_e = 1$, $H(E) = 0$, users cannot make a correct selection, but the errors are clustered around the target as in pointing tasks [22]. Doubling the number of commands from M to $2M$ adds 1 bit to the command's entropy, but since the error area around the correct target is approximately the same physical size, the number of possible errors is also doubled. Hence the equivocation is also increased by 1 bit. In our data, the possible errors in condition 128 are 1–3 around the target while in condition 256 they are 1–5 around the target, which corresponds approximately to the same physical area. As a result, the amount of transmitted information $I(X;Y) = H(X) - H(Z|Y)$ is not increasing any more and levels off as illustrated in Fig. 2(d).

We can now turn to the theoretical analysis of the throughput TP. As seen above, movement time MT is a linear function of $\log_2 M$.

For small M: $\log_2 M$ is also small, and MT is dominated by the intercept, hence can be considered as approximately constant. TP increases slowly with difficulty $\log_2 M$;

For large M: MT grows linearly with $\log_2 M$, and transmitted information $I(X;Y)$ levels off. Hence TP gradually decreases as demonstrated in Fig. 2(e).

However, we should distinguish the ceiling effect of transmitted information in our case from that in absolute judgment tasks [13]. Roy et al. claimed that they have the same characteristics but in our case, the errors made by users are around the target since they can see where it is, and therefore $H(Z)$ is only a few bits. In absolute judgment tasks, although the key message is that human short-term memory has a limited capacity, we would expect that when the number of randomly ordered stimuli increases, $H(Z)$ gets close to $\log_2(M-1)$ as Y can take any value in $\{1, 2, \ldots, M\}$. If this were the case, mutual information $I(X;Y)$ should go down, instead of leveling off as $I(X;Y) \approx \log_2 M - \log_2(M-1)$ at first order when M is very large. Since the stimuli's entropy never gets very large in this type of tasks, the phenomenon is thus never observed. This would require further investigation in the context of absolute judgment tasks.

In summary, in command selection tasks, the amount of transmitted information gradually increases with the command's entropy until it reaches its capacity, and then levels off. Correspondingly, TP demonstrates a bell-shaped behavior, increasing to reach a maximum and then decreasing. This maximum (corresponding to an entropy of 6 bits, i.e. 64 commands in our experiment) provides the optimal vocabulary size for the given selection technique.

6 Conclusion and Future Work

In this paper, we provide an information-theoretic analysis of user performance in command selection tasks. The maximum in mutual information from the user to the computer indicates the channel capacity while the maximum in throughput illustrates that there is an optimal level of the command's entropy, or task difficulty, to maximize human performance for any given interaction technique.

Following Soukoreff and MacKenzie [21] who argue that people are imperfect information processors, we demonstrate that when the command's entropy increases, users tend to make more errors. Obviously, a very high command entropy is not realistic nor desirable since no interface designer nor HCI researcher would want users to make 100% errors: the design would be useless. But it is necessary and even vital to consider such cases to understand the phenomenon correctly. Armed with a theoretically justified model, one can now use it to evaluate any command selection task.

Interestingly, whether there is a time constraint also affects the amount of transmitted information. In our experiment and most other HCI experiments, participants are instructed to move as fast or as accurately as they can, sometimes both. We can imagine that if they could take their time to complete the task, the error rate would be always low, therefore the mutual information $I(X;Y)$ would always increase with $H(X)$. The theoretical formulations have shown that *how* users make errors affects the transmitted information, which is tightly related to both the experimental design and the instructions to the users. We plan to investigate this more thoroughly in future work.

Acknowledgments. This research was partially funded by Labex DigiCosme (ANR-11-LABEX-0045-DIGICOSME), operated by the French Agence Nationale de la Recherche (ANR) as part of the program "Investissement d'Avenir" Idex Paris-Saclay (ANR-11-IDEX-0003-02), and by European Research Council (ERC) grant n° 695464 ONE: Unified Principles of Interaction.

References

1. Bailly, G., Lecolinet, E., Nigay, L.: Flower menus: a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 15–22. ACM (2008)
2. Bailly, G., Oulasvirta, A., Brumby, D.P., Howes, A.: Model of visual search and selection time in linear menus. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 3865–3874. ACM (2014)
3. Cockburn, A., Gutwin, C., Greenberg, S.: A predictive model of menu performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 627–636. ACM (2007)
4. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, Hoboken (2012)
5. Ewerling, P., Kulik, A., Froehlich, B.: Finger and hand detection for multi-touch interfaces based on maximally stable extremal regions. In: Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, pp. 173–182. ACM (2012)
6. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. **47**, 381–391 (1954)
7. Gutwin, C., Cockburn, A., Scarr, J., Malacria, S., Olson, S.C.: Faster command selection on tablets with FastTap. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2617–2626. ACM (2014)
8. Hick, W.E.: On the rate of gain of information. Q. J. Exp. Psychol. **4**, 11–26 (1952)
9. Hyman, R.: Stimulus information as a determinant of reaction time. J. Exp. Psychol. **45**, 188–196 (1953)
10. Kurtenbach, G., Buxton, W.: User learning and performance with marking menus. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 258–264. ACM (1994)
11. Lafreniere, B., Gutwin, C., Cockburn, A., Grossman, T.: Faster command selection on touchscreen watches. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 4663–4674. ACM (2016)
12. McGill, W.: Multivariate information transmission. Trans. IRE Prof. Group Inf. Theory **4**(4), 93–111 (1954)
13. Miller, G.A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. **63**, 81–97 (1956)
14. Miller, G.A.: What is information measurement? Am. Psychol. **8**(1), 3 (1953)
15. Miller, G.A., Frick, F.C.: Statistical behavioristics and sequences of responses. Psychol. Rev. **56**(6), 311–324 (1949)
16. Newell, A., Card, S.K.: The prospects for psychological science in human-computer interaction. Hum.-Comput. Interact. **1**, 209–242 (1985)

17. Roy, Q., Guiard, Y., Bailly, G., Lecolinet, É., Rioul, O.: Glass+skin: an empirical evaluation of the added value of finger identification to basic single-touch interaction on touch screens. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9299, pp. 55–71. Springer, Cham (2015). doi:10.1007/978-3-319-22723-8_5
18. Seow, S.C.: Information theoretic models of HCI: a comparison of the Hick-Hyman law and Fitts' law. Hum.-Comput. Interact. **20**(3), 315–352 (2005)
19. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423, 623–656 (1948)
20. Siek, K.A., Rogers, Y., Connelly, K.H.: Fat finger worries: how older and younger users physically interact with PDAs. In: Costabile, M.F., Paternò, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 267–280. Springer, Heidelberg (2005). doi:10.1007/11555261_24
21. Soukoreff, R.W., MacKenzie, I.S.: An informatic rationale for the speed-accuracy trade-off. In: 2009 IEEE International Conference on Systems, Man and Cybernetics, SMC 2009, pp. 2890–2896. IEEE (2009)
22. 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. ACM (2008)