NAIL087 - Fitt's law

Petr Houška

2. 2. 2018

Fitt's law

Fitt's law is an information theory based predictive model for the speed of human perceptual-motor system. It is based upon a hypothesis that the time to complete an interaction is directly and linearly proportional to the amount of information required[1] by the task. And that the information required is a logarithm of the ratio between the task's amplitude and tolerance. In other words that interactions of the perceptual-motor systems each have a certain fixed information capacity.

For example, given a task of clicking on a target with a finger, where D is the distance from the starting point to the center of the target and S is the width of the target measured along the axis of motion, the time required according to Fitt's law is:

$$Time = a + b * ID$$

$$ID = log2(\frac{D}{S})$$

The original paper also introduces the index of performance, nowadays usually called throughput TP, that essentially denotes the information capacity of the subject's specific perceptual-motor system.

$$TP = \frac{ID}{Avg(Time)}$$

It is important to note that the law only considers very simple, already learned interactions that are limited by neither physical nor higher cognitive abilities. While it has originally been shown for direct hand manipulations it holds true under a variety of conditions: with different limbs - including eye gaze, intermediate devices such as with a computer mouse/stylus, environments, and user populations. The rule has many applications ranging from human-computer interaction research to the design of assembly lines.

Experiment

Our experiment stays true to the idea of Fitt's law and aims to measure the time required to do a very simple and easily learnable interactions of a sensory-motor system. We present the user with with a black screen and a randomly positioned circle (target) of a random size. The goal for the user is to click the target circle as fast as possible, afterwhich the target disappears, and a new circle (home) appears in a default position. After the user clicks the home circle, a new target circle appears again. This is repeated many times to mitigate all influences but the limitations of the sensory-motor system which we're trying to measure.

The result of our experiment is a dataset that - among other - contains following variables for each clicked random circle: the distance from home to the target, the size of the target, the angle between home and the target, and the time between home has been clicked last time and the target circle has been clicked - which corresponds to the time required to move from default position to the target.

To mitigate the fact that a subject might get better at the task as he/she would learn it each participant was first asked to practise for ~5 minutes. A computer mouse was chosen as the input method.

---

[1]Also known as the index of difficulty (ID).

Internal parameters

The experiment consists of 500 target circles presented in a random order. A config file specifies 4 different sizes (100, 200, 300, 400), 5 different distances (100, 200, 300, 400, 500), and 5 potential angles (15, 30, 45, 60, 75). Each of the options is uniformly represented, creating 100 unique combinations. For future work a different config in which the ID instead of its underlying components would be uniformly distributed could be considered.

While repeating all combinations 5 times could lead to participant's fatigue and thus influence the result it - at the same time - enables us to control for random influences such as the mouse not registering correctly, or being interrupted. 500 iterations was, thus, chosen as reasonable middle ground.

External parameters

The experiment was conducted on a 14" laptop with a screen resolution of 1600x900. The mouse used as the input method was Microsoft Arc mouse. It is important to note here that all but one participant were unsatisfied with the ergonomics of the mouse. While it is possible that the the input device being uncomfortable could influence the results (e.g. increase standard deviation of the result timings) the Fitt's law should hold true even under these conditions.

All of the four subjects were in their twenties and consider themselves very proficient with computers. All but one (participant 2) had no prior experience with using that particular mouse that was used as the input method. All subjects complained about boredom during the experiment. One participant (4) suffers from OCD and mild anxiety disorder. The experiments were conducted in reasonably calm, well lit, and generally non-distracting environments during early afternoon. An average session took ~15 minutes, without the training.
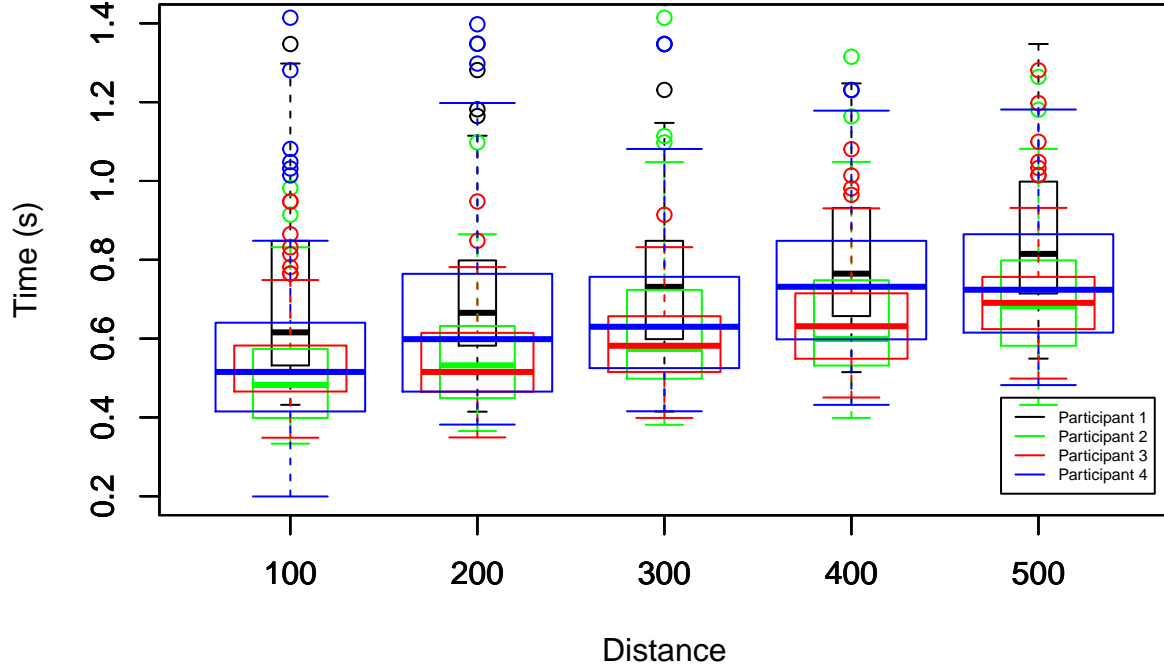
Results

With the exception of participant's 4 data all timings fell within the window of $0.3 - 1.8$ seconds. Participant's 4 case contained 11 outliers outside the window of $0 - 2$ seconds with the maximum being 6.9 seconds. Generally participants 2, 3 were fastest, 1 and 4 a bit slower and with more variance, and 4 with the highest number of outliers. None of these differences were high enough, however, to show any serious problems with the experiment implementation.

Visualising the relationship between size, distance, direction and the target variable - time, some clear patterns emerge. The relationship between size and time seems to be decreasingly hyperbolic. It is also interesting to note that the spread of the timings gets lower with higher size. With distance the relationship looks increasingly hyperbolic. The direction seems to be independent and doesn't show any notable differences in result time based on the target's angle to the default position. That hypothesis is further supported[2] by low correlation between time and direction ($\sim + - 0.05$)

Following graph shows the relationship between distance and Time. While there's a clear trend, the noise within the data is rather large across participants and different distances. The situation is very similar for size as well. Therefore we shouldn't assume any model will fit our data perfectly.

[2]There could still be not-directly correlated relationship, however.

Fitt's model

To evaluate how well Fitt's law describes our data we compute ID and then do a linear regression with time as the variable we want to explain. Following table shows both a and b variables obtained by linear regression and the resulting adjusted R^2 statistic for each participant's model.

	a	b	AdjR2
p1_id	0.64	0.10	0.28
p2_id	0.47	0.11	0.46
p3_id	0.50	0.09	0.49
p4_id	0.55	0.13	0.26

As we can see, the model parameters, especially b which denotes the ratio between time and "information needed to complete the task", are pretty close to each other. a is higher for participants 1 and 4 which is consistent with them being slower on average. Their models also show significantly lower $adjR^2$ metric and thus lower fit. That is also consistent with our previous observation that these participants have higher variance in their click times.

Following graph shows the data (points) and our Fitt's law based models (lines) for participants 1 and 2. We can see that both models explain the overall trend and the lower $adjR^2$ for participant 1 is caused by larger variance in time data and not by a different trend not-explained by the Fitt's law.
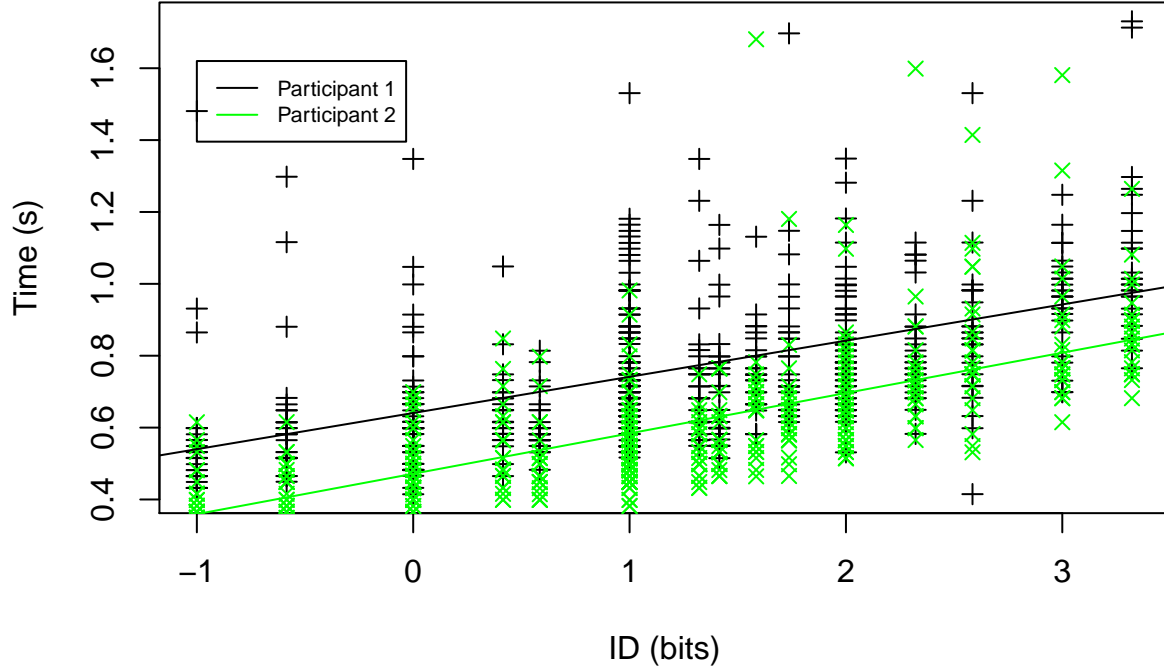
Other models

To compare other models them with the Fitt's based one we've chosen participants 1 and 2 as they represent the two more distinct group within our dataset. A linear ($Time = a + b * Size + c * Distance$), hyperbolic ($T = a + b * 1/S + c * 1/D$), two polynomial ($T = a + b * S + c * D + d * S^2 + c * D^2$ and similarly for 3rd degree polynomials), and polynomial variation of Fitt's ($T = a + b * ID + c * ID^2$) models were computed. Following table shows the adjusted R^2.

Participant	Fitts	Fitts2	Lin	Hyp	Poly2	Poly
Participant 1	0.276	0.286	0.304	0.277	0.313	0.316
Participant 2	0.458	0.477	0.480	0.459	0.490	0.489

The basic Fitt's model shows the worst adjusted R^2 metric, followed closely by the hyperbolic models. While the polynomial version of Fitt's model is better it is still worse than all of other models. The polynomial ones have an obvious edge over the linear one and of those two the more complex fairs better. The difference between poly2 and poly3 isn't large, however. It is also worth noting that all of the differences between models are consistent over both participants suggesting they're not just a product of random noise.

Conclusion

While it could seem surprising that the Fitt's based models were the worst it is important to remember that unlike the other models, they essentially use just one explanatory variable - the ID. And even though the

ID is computed using the two variables the other models use (Size, Distance) some information - naturally - gets lost. That combined with the fact that the differences were not large actually proves that ID and Fitt's law in general explains the data reasonably well.

Therefore the win of a polynomial model isn't surprising. Nor is it, given the observed relationships between Size and Time and Distance and time, surprising that the difference between poly2 and poly3 was almost negligible. The rather low explanatory power of the hyperbolic model was more unexpected.

Limitations

There're two concrete limitations of this experiment. The first one was the usage of an uncomfortable mouse that most probably contributed to larger than usual variance among gathered data. This issue could easily be mitigated in the future.

The second one is that to get enough data the experiment must be repeated enough times which, unfortunately, leads to participants' fatigue and larger variance among data. In essence it's hard to get both enough data and also ensure the subjects keep trying their absolute best. This issue is harder to solve but could be improved via better incentive for the participants to do well, for example through a gamification.

Future work

There's a number of potential extensions to this experiment. From different data acquisition as discussed in the experiment design section to more elaborate data analysis. Having more participants we could for example cluster their data according to their timings and look at the predictive power of models computed from a larger number of participants. Different types of models could also be examined.