What you touch is what you get: self-assessing a minimalist tactile sensory substitution device

Luca Brayda∗ Claudio Campus† Monica Gori‡

Robotics Brain and Cognitive Sciences department
Istituto Italiano di Tecnologia
Via Morego 30, 16163 Genova, Italy

ABSTRACT
In this work we study the links between subjective and objective measures of both performance and task difficulty, with a minimalist tactile sensory substitution device. We considered both psychophysical variables and subjective measures in a haptic discrimination task performed by blindfolded sighted subjects. We show that task difficulty significantly affects haptic sensitivity, perceived performance and perceived difficulty. Moreover subjects are able to predict their own performance. Results seem not to depend on fatigue. Therefore our device seems to show that ”what you touch is what you get”. Our device is thought to deliver tactile maps with touch: visually impaired subjects can potentially take advantage of independently evaluating their own performance.

Index Terms: H.5.1 [Multimedia Information Systems]: Artificial, augmented and virtual realities—; H5.2 [User Interfaces]: Haptic I/O—; K4.2: Social Issues—Assistive technologies for persons with disabilities

1 INTRODUCTION
Most studies assessing the efficacy of haptic devices attempt to measure performance, meaning with that the ability of the actuation system to deliver a set of haptic stimuli which are perceivable, distinguishable from each other, and easily interpretable [11].

Specifically, several interfaces were developed to substitute visual representations with a pseudo-visual equivalent acquired by other senses, by exploiting the so-called sensory substitution process [17]. For example, it was shown that it is possible to encode spatial information using actuation forces or vibrational cues statistically perceivable for the average subject [10, 14].

However, researches have rarely evaluated how much users were actually aware of their better or worse sensitivity [12, 8]. This is a main issue for sensory-deprived people using haptic interfaces: if sensory substitution devices are conceived to be usable as rehabilitation tools, the ability of self-evaluate performance is at least as important as the performance itself. A correct self-evaluation reduces the need of expert human evaluators and facilitates mobile or home-based applications.

User feedback is generally given through subjective tests [6]. Instead, jointly analyzing objective performance and task difficulty, and how they are perceived, could help in providing a more complete and reliable evaluation of an instrument. If users are correctly integrating information from a novel device, subjective evaluations of difficulty and performance should converge to their objective counterparts. In this case the instrument ensures that ”What You Feel Is What You Get” [13].

∗e-mail:luca.brayda@iit.it
†e-mail:claudio.campus@iit.it
‡e-mail:monica.gori@iit.it

Figure 1: Top-view, bottom-view and side-view of the tactile mouse (TAMO) device.

In this work we study the links between subjective and objective measures of both performance and task difficulty, within the training phase of a minimalist visuo-tactile sensory substitution device, called the Tactile Mouse (TAMO).

The mouse-shaped device is aimed at exploring 3D virtual maps with minimal tactile information; the third dimension (height of virtual objects) is given haptically by the system, while the other two (length and width) are determined by the active hand-arm movement of the subject. This study is focussed on the self-judgement of the single haptic degree of freedom. We try to answers to the following research questions:

1. What is the effect of a haptic task difficulty on performance? If such effect exists, increasing objective difficulty should entail worse objective performance. The same should hold in the subjective domain.

2. Can subjects reliably evaluate themselves while using the TAMO device in a haptic discrimination task? If so, the subjective measures should predict the objective counterpart.

2 THE DIGEYE SYSTEM
TAMO is a haptic device aimed at delivering virtual maps with touch to both sighted and visually impaired subjects. Unlike previous similar prototypes [16], it vertically displays information by means of one single tactile equivalent of the pixel, called taxel [15]. Other single-taxel devices have been proposed, which however exploit only tangential skin deformation [5] or are wearable [18]. TAMO is part of the DIGital EYE DIGEYE system, in which the TAMO is coupled with a tablet and used as an absolute positioning system to deliver height functions and thus 3D maps (see [2] for details).
corresponding to a rotation of the lever of height $h_i$ (as in Fig. 3, blue lever), where $h_i$ is one of the possible virtual heights, $i$ is the number of steps of the motor for each height variation, $R$ is the maximum range of degrees attainable by the motor and $B$ is the precision of the stepper motor. The device allows to rotate the lever by paces of a fraction of a degree, namely increasing $i$ by 1 translates in increases of $180/255 = 0.706^\circ$. The constant $R = 90$ reflects the fact that the lever can rise from $0^\circ$ to $90^\circ$, while the constant $B = 128$ comes from the 7 bits available to command the step motor. $i = 25$ was set by the experimenter to cover most of the angle range (i.e. testing the lever when it is almost horizontal up to when it is almost vertical under the finger). For $h_i = 1$ the angle spanned by the lever is $\alpha_1 = 18^\circ$. It is then possible to obtain three equispaced reference stimuli angles $\alpha_0 = \{\alpha_1 = 18^\circ, \alpha_2 = 36^\circ, \alpha_3 = 54^\circ\}$, one for each $h_i$.

3.2 Experimental protocol

Fig.3 shows the experimental protocol. Subjects were requested to compare two tactile stimuli presented in random order by the lever of the TAMO, choosing which of the two heights was higher in a two-interval forced choice task (2IFC). The stimuli were a constant reference stimulus at a given $h_i$ and a comparison stimulus at $h_{i+\Delta}$. The reference stimulus was noticeable because it was kept higher than the absolute thresholds of elevation known for tactile map symbols [7].

Eleven values of $\Delta$ were chosen (one value for $\Delta = 0$ and ten in a $i$ range between 1 and 25) and logarithmically spaced to ensure an estimation of the psychometric function as reliable as possible. Appropriate ranges were computed in the pilot study. We tested three value of $h$ according to a latin square scheme to avoid possible statistical biases due to the order of the three constant stimuli.

Each period of stimulation was started and stopped by the synthetic voice with ad-hoc orders (the synthetic voice said "Down!" 1 second before each stimulation and "Up!" 1 second after each stimulation) given to the subject. As well, pauses of 1 second were inserted in-between stimulations. After two stimuli were given (internal loop of Fig.3), the system waited until the subjects made the forced choice, by answering to the question "Which of the two stimuli was higher?". Subjects had to press either of two buttons, one assigned to the answer "First", one to the answer "Second".

Subjects rested for 25 seconds every 25 trials, to minimize the effects of tactile numbness. The reference was repeated for 110 stimuli/session (external loop) for each $h$. The total experimental duration was about 45 min. Therefore, a total of 14 subjects · 3 sessions/subject · 110 stimuli/session = 4620 stimuli were recorded.

3.3 Analysis

We considered the following variables:

$\Delta$, the amplitude of the initial haptic stimulus (reference stimulus). $\alpha_i$ was considered to reflect the objective difficulty (according
4 Results

4.1 Just noticeable differences (JND)

Fig. 4 shows one example of psychometric function for all 110 trials of a single subject and one reference value: the chosen range of deltas covers all the guessing range, from the chance level to the asymptote where all guesses are correct. The JND is also indicated. When considering the whole data set, Fig. 5 shows box plots representing the distribution of thresholds in function of the three reference stimuli. JND distributions computed over 14 subjects resulted not gaussian (Lilliefors test gives \( D = 0.27, p = 0.02 \)). Therefore, instead of the ANOVA, we used the Kruskal-Wallis test and found that the reference stimuli has an effect \( (\chi^2(13) = 8.66, p = 0.01) \). Post-hoc analysis with Wilcoxon test revealed significantly greater JND for the highest reference stimulus \( (\alpha_3) \) compared with the smallest \( (\alpha_1) \). As well for the \( \alpha_2 \) compared with \( \alpha_3 \). We observe that differential thresholds (JNDs) increase with increasing reference stimuli. Since the riser in-between steps is the same (i.e. the lever rises by \( i = R/B = 18^\circ \) equally from \( \alpha_2 \) to \( \alpha_3 \), than from \( \alpha_2 \) and \( \alpha_3 \)), JNDs shows a quasi-linear behaviour and the Weber fraction can be assumed approximately constant.

Importantly, there are no JNDS below 1, meaning that the precision of the device is sufficient for this task. This results allow to establish the haptic resolution of the third dimension of any virtual object: it suggests that, if one wants to display with TAMO a virtual object composed by different levels, then the two closest levels should differ at least by 4.3\(^\circ\) (which corresponds to the worst case) to be distinguished at least 75% of the times.

No significant effect was found for \( O_h \) \( (F(2,33) = 0.89, p = 0.42) \).

4.2 Perceived performances (PPE)

Figure 6 shows that PPE decreases with increasing reference stimulus. The ANOVA shows that the distributions of PPEs are significantly different \( F(2,33) = 3.47, p < 0.05 \). Specifically, post-hoc with t-test show that \( \alpha_1 \) and \( \alpha_3 \) are significantly different \( t(11) = 2.67, p = 0.02, 95\% CI=[0.24,2.48], \) starred link between the boxes in Figure 6) as well as \( \alpha_1 \) and \( \alpha_2 \) \( t(11) = 2.59, p = 0.03, 95\% CI=[0.12,1.51], \) while this does not hold for \( \alpha_2 \) compared with \( \alpha_3 \). The degrees of freedom of these tests are different than those exhibited by the JND because the subjective measures of two subjects were not available. These results indicates that, as the amplitude of the reference stimulus grows, subjects tend to judge their performance as lower and lower. Note that JND and PPE exhibit similar post-hoc associations.

Figure 8 shows a linear fit between JND and all PPE for all subjects and all references (every subject has one blue dot for \( \alpha_1 \), one green for \( \alpha_2 \) and one red for \( \alpha_3 \)). A negative significant correlation between PPE and JND appears \( (df = 34, R^2 = 0.15, p = 0.03) \). This result entails that - with this device - are able to predict their subjective performance, independently of its quality.

No significant effect was found for \( O_h \) \( (F(2,33) = 2.14, p = 0.13) \).

4.3 Perceived levels of difficulty (PLD)

PLD increases with increasing reference stimulus \( h \). Figure 7 shows the distributions of discriminative perceived levels of difficulty (PLD) in function of the three different reference stimuli. The distributions of PLD does not exhibit Gaussian profile as confirmed by Lilliefors test \( (D = 0.26, p = 0.02) \). The Kruskal-Wallis test showed that the PLD distributions were significantly different \( (\chi^2(11) = 6.06, p < 0.05) \). Post-hoc analysis with Wilcoxon test revealed significantly greater PLD for the highest reference stimulus \( (\alpha_3) \) compared with the smallest \( (\alpha_1) \) \( (W = 108, p = 0.01) \), starred black bar link connecting the boxes in Figure 7). Moreover, we found a negative correlation between PPE and PLD \( (df = 34, R^2 = 0.152, p = 0.02) \), see Figure 9.

These results indicates that subjects were able to correctly evaluate the objective difficulty of the stimuli.
Perceived fatigue (FAT)

FAT is affected by the temporal order of the stimuli. Figure 11 shows FAT in function of the reference stimuli, while Figure 10 shows FAT in function of the session order. In practice, while the first figure shows FAT in function of the reference stimulus, while Figure 11 shows FAT in function of the amplitude of the reference stimulus. Subjective difficulty increases with objective difficulty. Significantly different conditions are joined with a starred link (\( p < 0.05 \)).

No correlation was found between PLD and JND (\( df = 34, R^2 = 0.02, p = 0.33 \)). No significant effect was found for \( \alpha_h \) (\( F(2, 33) = 0.32, p = 0.73 \)).

4.4 Perceived fatigue (FAT)

FAT is not affected by \( h \), but it only increases with \( \alpha_h \) (i.e. for later presented \( h \)). Figure 10 shows the distributions of FAT in function of the reference stimuli, while Figure 11 shows FAT in function of the session order. In practice, while the first figure takes into account the latin square design, the second does not. The ANOVA shows that the FAT is not affected by the reference stimulus \( h \) (\( F(2, 33) = 0.04, p = 0.96 \)). Unlike, JND, PLD and PPE, the FAT is the only variable to be affected by the temporal order of the stimuli \( \alpha_h \) (\( F(2, 33) = 7.44, p < 0.005 \)). Post hoc indicate significant differences among all pairs (\( \alpha_1 \) and \( \alpha_2 \) : \( t(11) = -3.64, p = 0.004, 95\%CI = [-2.04, -0.5] \); \( \alpha_2 \) and \( \alpha_3 \) : \( t(11) = -4.35, p = 0.001, 95\%CI = [-3.08, -1.01] \); \( \alpha_2 \) and \( \alpha_3 \) : \( t(11) = -3.72, p = 0.003, 95\%CI = [-1.22, -0.31] \)). These results show that there is no significant link between reference stimuli of different heights and the self-estimated level of fatigue. Rather, the level of fatigue appears to be linked only to the time of the experiment, since clearly at the third session (i.e. after about one hour of continuous stimulation) subjects reach a higher level of fatigue. Moreover, FAT didn’t show any correlation with JND (\( df = 34, R^2 = 0.01, p = 0.53 \)), as with PLD (\( df = 34, R^2 = 0.07, p = 0.09 \)) and with PPE (\( df = 34, R^2 = 0.04, p = 0.20 \)). Then, observed PLD, PPE and JND, are not likely to be affected by the level of fatigue.

5 DISCUSSION

To the best of our knowledge, this is the first study attempting to correlate psychophysical objective with subjective parameters.

The methodology that we propose in this study adds to classic psychophysics a peculiar relation with subjective measures. Importantly, our approach can be applied to other devices without loss of generality. In particular, for sensory-deprived subjects, especially within a rehabilitation framework, it is important to do well and to be aware of it. This means knowing that self evaluation is correct, i.e. that it reflects a real haptic percept. We believe that self-confidence and acceptance of novel technology also depends on this aspect.

In this work we have shown that our tactile mouse-shaped device (TAMO) is able to provide distinguishable virtual height sets. We investigated a haptic sensitivity task from both a performance and difficulty perspective, considering both the objective and the subjective domain (see Table 1). As a measure of objective performance, we computed the perceptual threshold (JND) for each objective difficulty level (\( h \)). Moreover, for each \( h \) we collected three subjective variables: perceived level of difficulty (PLD), perceived performance (PPE) and perceived level of fatigue (FAT).

Our results answer to both research questions:

1. What is the effect of a haptic task difficulty on performance?

Difficulty seems to worsen performance (vertical arrows in
1. Perceived performance in function of Just Noticeable Difference. The α corresponding to three reference stimuli are plotted together with different colors (in blue α₁, in green α₂, in red α₃). PPE and JND are negatively correlated: subjective self-evaluated performance predicts objective performance.

2. Perceived performance in function of perceived levels of difficulty. The α corresponding three reference stimuli are plotted together with different colors (in blue α₁, in green α₂, in red α₃). PPE and PLD are negatively correlated: perceived difficulty affects self-evaluated performance.

3. Perceived levels of fatigue in function of the amplitude of the reference stimulus. FAT is the only variable which is not affected by h.

4. Perceived levels of fatigue in function of the order in which the reference stimuli were presented. FAT is the only variable affected by Oh. Significantly different conditions are joined with a starred link ($p < 0.05$).

Fig. 12). We found that increasing objective difficulty entails worse objective performance. The same occurs in the subjective domain. Then, it is fundamental to consider both subjective and objective aspects when a device is evaluated.

2. Can subjects reliably evaluate themselves while using the TAMO device in a haptic discrimination task? Yes, because subjective evaluation seems to closely reflect objective condition (horizontal arrows in Fig. 12). In fact, we found that objective difficulty and performance were associated to the corresponding subjective counterparts. In particular, subjectively perceived performance predicts objective sensitivity. Then, TAMO seems to be a "What You Touch Is What You Get" interface.

Importantly, we found that while h modulates PLD, JND and PPE, it does not modulate FAT.

Specifically, increasing h increases PLD. This suggests that subjects judged the task as more difficult when it actually (objectively) was. Therefore, subjects seem to be able to evaluate the difficulty of their own task.

Similarly, higher h corresponds to higher JND (i.e. lower sensitivity). This result indicates that subjects performed worse when the task was more difficult. Then, objective performance seems to be affected - as expected - by objective difficulty. The trend qualitatively follows the Weber’s law. However, to draw more robust conclusions about this point more h values will be evaluated in future experiments. An increasing JND was also reported with other haptic feedback devices [1] and was necessary to establish a suitable resolution for our device: the higher the JND, the smaller the number of different haptic heights that can be displayed.

Interestingly, increasing h decreases PPE. This result suggests that subjects also thought to perform worse when the task was more difficult. Thus, subjectively perceived performance is reliably affected by objective difficulty of the task. Considering this aspect to ensure a sufficient self-confidence of users is crucial for the success of any device aimed at rehabilitation.

As expected, PPE is negatively correlated with PLD, suggesting that, from a purely subjective standpoint, subjects felt to perform worse when they estimated that the task was harder to perform. Even if both PPE and PLD are subjective variables, they reflect different (reliably complementary) aspects of the task. Moreover, this link mirrors the one between the two corresponding objective counterparts (h and JND), as depicted in Fig 12. Therefore, the task difficulty seems to have, regardless of the domain in which it
is evaluated, an important influence on the final performance. This result could appear stronger if more values of $h$ were available, as it is planned to be done in future studies.

Importantly, PPE is negatively correlated also with JND. This association suggests that subjects though to perform worse when haptic sensitivity was actually worse. The result is not obvious: it could have happened that subjects were aware of their performance only when it was good, but in our case this happened also for bad performance. Subjective and objective performance are therefore tightly linked.

Interestingly, there was no correlation between PLD and JND. This confirms that the relationship between objective performance and subjective perceived difficulty is not trivial. Generally speaking, both performance and difficulty aspects should be considered, as it is done in current evaluation questionnaires [6].

Results obtained so far may have been biased by uncontrolled factors. The Latin square design attempts to minimize learning effects. Furthermore, the fatigue may increase with the experimental time and could also affect the perception reference stimuli. However, there is no link between reference stimuli of different heights and the self-estimated level of fatigue. All variables except FAT are affected by $h$. Instead, FAT is the only variable influenced by the order in which $h$ were presented ($O_h$). Moreover, fatigue didn’t correlate with JND, PLD or PPE. All these results suggest that no measure in this work is affected by FAT.

Our tests were performed without any preliminary training. Therefore the TAMO device allows subjects to immediately well-evaluate their performance: this is important both to improve our device and more generally to design haptic interfaces which consistently reflect users’ expectations. Subjects may realize by themselves how to tune up or down the haptic height resolution, without necessarily being assisted by a software - or a specialist - evaluating their performance. While this needs to be verified with visually impaired users, it may increase a lot the accessibility of such human-machine interface.

6 Conclusion

Subjects must be aware of their own performance while interacting with a novel device. In this study we have shown that a perceptual task with a haptic device should be evaluated at the same time by a performance and difficulty perspective, jointly using subjective and objective measures. This new approach can ensure the development of “What You Touch Is What You Get” haptic interfaces which can be practically self-managed by end-users.

Acknowledgements

The authors would like to thank Julien Jenvrin and Stefano Saliceti for designing and assembling the electronics and the shell of the TAMO device, as well as Jesus Ortiz for software tuning. Thanks to Bruno Bonino and Davide Dellepiane for support in maintaining the hardware setup. Special thanks to all our volunteers. This research was funded by the Italian Institute of Technology.

References


