Conveying perceptible virtual tactile maps with a minimalist sensory substitution device

Luca Brayda and Claudio Campus
Robotics, Brain and Cognitive Sciences dept.
Italian Institute of Technology
16163, Via Morego 30, Genoa, Italy
Email: {luca.brayda, claudio.campus}@iit.it

Abstract—Many sensory substitution interfaces for visually impaired subjects have been proposed, but very few are currently used, probably because of the lack of assessments from both sensory and cognitive standpoints. In this work we have assessed a minimalist sensory substitution device, as part of our DIGEYE system, able to elicit 3D cognitive maps from virtual tactile objects: first, we computed sensory thresholds allowing subjects to well discriminate height profiles with touch, then we assessed how two measures, one objective (Stimuli Rate) and one subjective (Perceived Levels of Difficulty), could be related to the complexity of different virtual objects. Trends indicate that our device is suitable within the psychophysical ranges, that both the rate at which tactile information is acquired and the perceived difficulty in processing this information seem to follow the complexity of explored virtual objects. This work is the base for further validations involving blind subjects.

I. INTRODUCTION

Visually impaired subjects feel, today, the large gap between technologies aimed at efficient navigation using vision and their accessibility. Commercially available touch screens cannot paradoxically be used without a visual trigger, because limited or no tactile feedback is given. The potential of touch-based interfaces is largely unexploited to help with visual impairments. Many systems [1], [2], [3] have been proposed so far which make use of feedback channels alternative to vision, exploiting the phenomenon known in neuroscience as sensory substitution [3]. They can be grouped in two sets: the first comprises those which convey direct spatial information, meaning with this an information that can be perceived in a tangible user space, with stimuli similar to real haptic feedback. They generally employ expensive or bulky apparatuses [4], [5], probably more suitable as accurate experimental setups, less as potential devices for the large public. The second set comprises those conveying indirect spatial information, by attempting to transform visual information in another domain (non verbal, e.g. vibrational, acoustic, or verbal), but with a coding scheme sometimes non intuitive, therefore requiring long periods of training [2]. In fact, the inherent difficulty for this set of devices is to find the trade off between richness of coded information and number of abstraction layers to understand it: namely that the coded spatial information overwhelms the user or, on the contrary, is too small and need to be integrated [6].

For example, a semantic sequence of words describing the shape of a building, its dimensions and locations relative to the user position is not direct and requires experience to be acquired, model-matched and kept in memory. Rather, the perception of its tactile representation on a map directly and quickly provides a similar amount of information and, importantly, mimics the perception of real objects with touch that blind subjects are so well trained at [7]. Since blind people, especially congenitally blind, have prior knowledge of objects and environments acquired by non-verbal sensory feedbacks, it is reasonable to assume that scaled tactile representations can provide less confusing information than verbal descriptions, just because the latter need to match with previous concepts that a blind person may not have, and which are part of modern (and extensive) orientation and mobility programs.

When a physical map is explored, a cognitive map is developed. Cognitive maps are defined as the internal representation of perceived environmental features or objects and the spatial relations among them [8]. Spatial representations are preserved in blind individuals [9] and tactile maps are de facto widespread and used, although obviously not available for every kind of object. They cannot be easily printed, they cannot be zoomed, updated or tailored to specific user needs. Virtual reality offers the way to solve this problem.

Some prototypes using virtual objects [4], [5] have already been proposed, with contradictory results [10] deriving in part from very few preliminary studies establishing the minimal information able to be conveyed by a touch-to-vision sensory substitution interface [11].

Studying the capability of the brain to integrate a minimal sensory information to build a map which is complex per se is important because it allows to assess the necessary and sufficient amount of sensory feedback to provide an efficient and comfortable coding through an (artificial) sensory substitution device. Too little information may make the device prohibitive to use, with long and frustrating training sessions. Conversely, a higher amount of information would decrease training time, but too much information may potentially cause overload and fatigue: a trade off is needed, better if measurable.

In this work we attempt to provide minimal tactile information with a device that, unlike previous similar prototypes [10], displays information by means of one single tactile equivalent of the pixel, called taxel [12]. Virtual objects are presented and freely, actively explored. We keep the task complexity as...
low as possible at the tactile level, in order to measure it at the cognitive level, all by providing no training.

Unlike previous work about sensory substitution interfaces, in this work the tactile interface is validated at the two primary stages of the construction of cognitive maps, the first being the tactile perception of information, the second being the mental mapping obtained by collecting spatially distributed noticeable information of variable complexity. Our device gives the third dimension via the tactile feedback, but the remaining two dimensions are built by the active movement of hand and arm, therefore via proprioceptive feedback. Our methodology consists, in practice, to assess and make sure that the third dimension is well perceived, so that the brain can be put in sufficient conditions to make mental interpolations and construct the explored object through an active searching process. Our measures are intended to be net of training and can be considered as sensory lower bounds.

This work answers to the following research questions:

- Is it possible to perceptually distinguish different levels (heights) with our device?
- If so, are there behavioral and cognitive cues possibly related to the amount of different three-dimensional haptic information?

We have already shown that it is possible with our device to mentally construct objects of different shape, although we have not directly measured the perceptual limits [13]. We also qualitatively assessed the strategies that subjects adopt when exploring an unknown object [14]. Here, to answer to the first research question we computed perceptual thresholds of the tactile feedback given by our device. To answer to the second research we made subjects explore perceptible objects having simple and abstract 3D shapes. We showed a possible link between an objective behavioral parameter, a subjective cognitive parameter and the complexity of virtual objects.

II. THE DIGEYE SYSTEM

The (DIGital EYE) DIGEYE system [13] comprises two parts: a TActile MOuse-shaped device (TAMO), coupled with a sensing tablet; a personal computer, where a software application presents a virtual object synchronized with the TAMO.

Figure 1 shows that given some height function, $H = f(X,Y)$, the TAMO is able to display height information, $H$, at any absolute position $(X,Y)$ of the 210×297-mm tablet. While subjects freely move the TAMO on the tablet, the software receives the current position and wirelessly sends back the value of $H$ to the controller inside the TAMO. The motor controller generates tactile feedback by activating a stepper motor, which raises a small lever proportionally to $H$. TAMO thereby generates a taxel for each pixel of the tablet, much like a tactile bas-relief representation. When the subjects explore the ground level, no tactile stimulation is provided by the lever. When the subjects cross a level other than ground, the lever rises to signal a virtual object edge. When the height of the object is constant, the height of the lever is also constant. When the height of the object increases/decreases, the height of the lever varies proportionally. Subjects actively seek spatial information by freely moving the TAMO device without any prior cue and attempting to build a cognitive map of the virtual object. More details are available in [14]. While using the tactile device, behavioral data (i.e., $X$, $Y$, and $H$ values) are recorded.

III. PSYCHOPHYSICAL EVALUATION

To understand how closely two heights $H$ can be perceived as different (our first research question) we performed a psychophysical evaluation.

A. Experimental setup

14 volunteers (7 males and 7 females) with age 32 ± 5.4 (range: 24-45 years) participated in the evaluation. All subjects were right handed. Subjects were blindfolded and kept their right hand on the TAMO device at rest, while passively receiving tactile stimulations. Subjects were comfortably seating at a desk, their elbow and wrist resting on the table top. They received instructions by a synthetic voice through a pair of headphones. To avoid possible acoustic influences due to the motor non-stationary noise, a background stationary pink noise was continuously played. In this setup the lever of the device could reach four possible configurations: 1) "Ground" level: the lever is hidden in the loophole of the TAMO (visible under the lever in the leftmost subfigure of Fig. 1), without stimulating the finger; 2) "Level $h \in H$" (with $h = (1,2,3)$) the lever raises from "Ground" level and reaches a height $H$ corresponding to a rotation of the lever of $h \ast i \ast R/B$, where
\( h \) 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 constants \( R = 180 \) and \( B = 255 \) are built-in the chosen motor, while \( i = 25 \) 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). It is then possible to rotate the lever by paces of a fraction of a degree, namely increasing vertical under the finger). It is then possible to rotate the lever experimenter to cover most of the angle range (i.e. testing the are built-in the chosen motor, while \( i \) of the stepper motor. The constants \( B \) range of degrees attainable by the motor and \( i \) steps of the motor for each height variation, \( h \) translates in increases of \( h \) is one of the possible virtual heights, \( i \) varying \( h \) is 100\% of the times indicated. When considering the whole data set, \( i \) translates in increases of \( \Delta = 0 \). As such, by varying \( h = (1, 2, 3) \) it is possible to raise the lever up to three, distinct, and possibly distinguishable, levels. Just Noticeable Differences (JND), in the \( i \) domain, are computed by means of the method of constant stimuli and a two Alternative Forces Choice protocol (2AFC), to establish the minimum \( i \) value statistically necessary to perceive two different stimuli, for each reference stimulus \( h \).

Fig. 2 shows the experimental protocol. Subjects compared two stimuli presented in random order, choosing which was the greater. The stimuli were a constant reference stimulus at a given \( h \) and a variable stimulus at \( h + i + \Delta \).

The reference stimuli was noticeable because it was kept higher than the absolute thresholds of elevation known for tactile map symbols [11]: we also verified in a pilot study with two subjects that stimulus at \( h + i \) is 100\% of the times distinguishable from the "Ground level". Eleven values of \( \Delta \) were chosen (one value for \( \Delta = 0 \) and ten in a range between 1 and 25) and logarithmically spaced to ensure an estimation of the psychometric function as reliable as possible. The spacing was not the same for each reference \( h \), as larger reference stimuli, as expected, required larger ranges of \( \Delta \). 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.

It was expected that, with increasing values of \( h \), i.e. with initial stimuli of increasing amplitude, the corresponding JND, expressed in \( i \) units, would also increase. If, in addition, the relation would be linear, then the stimulation of the TAMO device would likely respect the Weber law.

**B. Statistical analysis**

Psychometric functions were evaluated for each subject and for each of the three reference stimuli, using Bayesian Inference [15], using the psignifit package within Matlab\textsuperscript{®}. For each function, Just Noticeable Differences are retained as differential stimulus corresponding to 75\% proportion of correct guesses. The JNDS of all subjects, for each reference stimulus, are grouped in three distinct distributions, which are compared with ANOVA and post-hoc analyses.

**C. Results**

Fig. 3 shows one example of psychometric function for all 110 trials of a single subject: 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. 4 shows whisker plots representing the distribution of thresholds of the chosen three different initial height stimuli. The ANOVA shows that the distributions of JNDS are significantly different \( F(2, 39) = 5.08, P \cong 0.01 \). Post-hoc analysis revealed significantly greater JND for the highest reference stimulus (Level 3) compared with the smallest (Level 1) \( (t(13) = -3.5, P = 0.004, 95\% \ CI=[-6.2, -1.5]) \), while Level 2 compared with Level 1 reveals to be weakly significantly
should differ at least by one JND, to ensure that they can be statistically distinguished.

We observe that differential thresholds (JNDs) increase with increasing reference stimulus. Since the riser in-between steps is the same (i.e. the lever rises by \( \frac{R}{B} \) equally from Level 1 to Level 2, than from Level 2 and Level 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. These results suggest 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° (which corresponds to the worst case) to be distinguished at least 75% of the times.

IV. BEHAVIOURAL EVALUATION

Our psychophysical results help to define the geometry of perceptually different virtual objects. In fact, we are now allowed to display virtual objects of heights which differ at least by one JND, to ensure that they can be statistically distinguished.

We recall that the TAMO device is designed to provide minimal tactile information: the shape of the 3D object must be reconstructed by active exploration. Our objective being that of measuring how difficult it is to map objects with different number of levels, we kept this number low to minimize memory retention effects.

This issue can be solved by Miller’s method of absolute judgements [16], which measures how much information humans are able to receive by a generic device as the covariation between input (displayed information) and output (perceived information). The upper limit of such amount is the channel capacity (where the channel is in our case the haptic sense) and is measurable in bits. Since a meta-study on different modalities revealed an average channel capacity of 2.6 bits, with a standard deviation of 0.6 bits, in this work we assume that 2 bits of information can reliably be retained in memory, i.e. that subjects are able to discriminate four different heights.

Using the same notations and a similar set of stimuli than our psychophysical tests, we assumed that four levels \( (h = (1, 2, 3, 4)) \) equispaced by \( i = 25 \) should ensure perceptible levels and easy to remember: the difference between subsequent levels is therefore constant and almost double of the worst JND among the tested subjects in the psychophysical evaluation (see upper bound of the whisker plot number 3 in 4). This would ensure that the correct detection of the most similar tactile stimuli happens in a percentage statistically very close to 100% (guesses for \( i = 25 \) are well on the asymptote of Fig. 3.) Since three out of four stimuli are identical to those used in the psychophysical evaluation as references, the computed thresholds hold true also in this second setup.

In this behavioural evaluation, we let subjects construct a cognitive map of three different virtual objects, each one having a different number of height levels.

We hypothesized that objects with more levels would be perceived as more complex. Therefore, we analyzed possible relationship between the complexity (a generally wide concept, but restricted in this work to the number of levels) and two parameters: one behavioural, the Stimuli Rate (SR), and one cognitive, the Perceived Levels of Difficulty (PLD). SR is defined as the total amount of stimuli given by the up and down movements of the lever during the exploration of a single virtual object. We assumed that this objective behavioral parameter was linked to the way subjects explore the unknown object, namely the strategy chosen to map it. The strategy must necessarily depend on the kind and/or amount of tactile information that the subject is actively gathering. We hypothesized that the strategy, and thus the SR, was linked to the object complexity. PLD, instead is defined as a subjective voting measure, of range 1 to 10, given by subjects and increasing with the difficulty encountered in constructing a single object.

We hypothesized that also this parameter could be linked to object complexity, expecting increased votes with increased number of virtual levels.

A. Experimental setup

30 volunteers (15 males and 15 females) with age 35 ± 14 years (range: 20-60 years) participated to the study. The experimental protocol is represented in Fig. 5. Subjects subsequently explored three, gradually complexified virtual scenarios: one square parallelepiped, two and four levels zigzagrat with a 2 min pause between objects. Apart from the Ground level, the first zigzagrat ("object1") displays one level, i.e. \( h = 1 \); the second ("object2") two levels, i.e. \( h = (1, 2) \); the third ("object3") four levels, i.e. \( h = (1, 2, 3, 4) \). The aim was to construct a cognitive map of explored objects in a constrained amount of time. The coordinates of TAMO and the height of
Fig. 5. Experimental protocol of the behavioral evaluation. Subjects explored each object ten times, each time for 10s. Every trial started and stopped with two distinct sounds. Every trial was preceded and followed by 10s of rest. Finally, subjects assigned to the exploration of each object a subjective perceived level of difficulty (PLD) ranked on a scale from 1 to 10 (higher grades corresponded to higher perceived difficulty).

Fig. 6. Stimuli Rate (SR), expressed in number of up-down and down-up transitions, in function of the complexity of the virtual environment. The increasing trend is statistically significant.

Fig. 7. Perceived Levels of Difficulty (PLD) in function of the complexity of the virtual environment. The increasing trend is statistically significant.

1. Active explorations (Exp)
   obj 1  obj 2  obj 3

2. Cognitive load estimation
   Start sound 10s  Stop sound 10s
   Exp
   Rest
   Vote PLD
   time

0 20 40 60 80 100 120
SR
object 1 object 2 object 3

0 2 4 6 8 10
PLD
object 1 object 2 object 3

B. Statistical analysis

The effect of complexity was statistically evaluated for SR and PLD. We used two one-way repeated measure ANOVA post hoc analyses. Post-hoc comparisons were performed using Tukey HSD test with FDR correction. R software was used for all statistical analyses (http://www.r-project.org).

C. Results

Comparing different objects (with 1, 2 and 4 levels), ANOVA was significant both for SR, $F(2,87) = 62.3$, $P < 0.0001$ and for PLD, $F(2,87) = 16.6$, $P < 0.0001$. As for post hoc analyses both behavioral parameters showed significant trends from “object1” to “object3”, always with $P<0.01$. The trend was increasing both for SR (Fig. 6) and PLD (Fig. 7). Results show that the amount of complexity does have an effect on the Stimuli Rate, suggesting that subjects were acquiring more information, when encountering more levels.

Results also show that the amount of complexity does have an effect on the Perceived Levels of Difficulty, i.e. that more complex objects are harder to explore. SR and PLD are two a priori different measures. SR is subjective and it indirectly reflects sensory behavioral aspects. PLD is subjective and is more related to the amount of cognitive load required to build the map. However, both SR and PLD exhibit similar trends, in accordance with a global process of information acquisition and processing.

V. DISCUSSION

In this work we have shown that our tactile mouse-shaped device is able to provide distinguishable virtual height sets. In a first psychophysical experiment, we computed perceptual thresholds for various height sets and noticed that thresholds increased with the increasing reference stimuli of the lever, in degrees. This helped us to set a second experiment, where three virtual objects were shaped according to the computed thresholds and were displayed with the same device, this time in an active exploration task. One behavioural measure, Stimuli Rate and one cognitive measure, Perceived Levels of Difficulty, showed to be linked to the complexity of virtual objects, with similar trends. We can then preliminarily positively answer to both our research questions. Importantly, the increasing trend of the JND with respect to the amplitude of the reference stimuli indicates that measuring thresholds with degrees is consistent in this work: since the lever rotates, it stimulates the whole finger, both in vertical and horizontal direction. The elicited sensation is most likely a combination of two factors: tactile feedback (caused by the movement of the tip of the lever on the fingertip) and kinesthetic feedback (the joints of the finger are moved by the lever, especially for higher values of $i$). Their relative contribution is difficult to determine and is behind the scope of this paper. However, we speculate that a stimulation which physically changes the kind of feedback across its range may even increase the level of understanding the displayed information, although this should be proven. Since only three height values are displayed in
our psychophysical experiment, our interpretations are still preliminary. Further work, involving a modified experimental setup with more height levels, and some fitting measures, is required to clarify this aspect. The JNMDs computed in the first experiment did not require motion of hand, as required in the second experiment. While the hand motion might have an impact on thresholds, it may be that thresholds would in principle differ only by an offset. This offset could be due to the friction on the tablet - always occurring, no matter the presence or amount of the tactile stimulus -, which could disturb detection, therefore increasing thresholds. We attempted to minimize this possible effect using a low-friction tablet, covered by a built-in plastic layer. In the second setup, we deliberately chose a discrete range of stimuli in the third dimension not only for purely psychophysical reasons: it has been shown [17] that categorizing information helps in the identification of tangible pictures, with raised-line drawings, for both sighted and a group of blind subjects. Our device mimics this task, thus we expect similar trends when testing with blind subjects. This work does not treat data linked to success of the reconstruction process, which will be a matter of a future study. Future work will also include blind subjects, although it is generally accepted that blind subjects perform well within the thresholds of sighted subjects due to increased tactile acuity [7]. Importantly, building the TAMO device from scratch costs around 300$, which is a promising upper bound.

VI. CONCLUSION

We have assessed a minimalist sensory substitution device as part of our DIGEYE system. First, we computed thresholds allowing subjects to well discriminate tactile information in one dimension only. Then we assessed how difficult was for subjects to construct three-dimensional objects. In this case we used both an objective sensory measure and a subjective cognitive measure. Trends indicate that the complexity of virtual objects increases both the rate at which tactile information is acquired and the difficulty perceived by subjects in processing the acquired information. This work is the base for further validations involving blind subjects.

ACKNOWLEDGMENT

The authors would like to thank Monica Gori for suggestions about the psychophysical protocol, Bruno Bonino, Davide Dellepiane and Julien Jenven for support in maintaining the hardware setup, as well as Jesus Ortiz for software tuning. This research was founded by the Italian Institute of Technology.

REFERENCES