CMMR²⁰¹³ Marseille

und, Music & Motion October 15 -18



Proceedings

10th International Symposium on Computer Music Multidisciplinary Research

www.cmmr2013.cnrs-mrs.fr

Influence of Auditory Pitch on Haptic Estimation of Spatial Height

Michele Geronazzo¹, Federico Avanzini¹ and Massimo Grassi²

¹ Dept. of Information Engineering, University of Padova ² Dept. of General Psychology, University of Padova {geronazzo, avanzini}@dei.unipd.it, massimo.grassi@unipd.it

Abstract. This paper presents an experiment aimed at assessing the influence of auditory feedback on haptic estimation of size. Experimental subjects were instructed to explore a virtual 3D object (a stair-step) with a haptic device, and to return a verbal estimate of the step riser height. Haptic exploration was accompanied with a real-time generated sinusoid whose pitch varied as a function of the interaction point's height within two different ranges. Experimental results show that the haptic estimation is robust and accurate regardless the frequency range of the accompanying sound.

Keywords: Multimodal virtual environment, haptic virtual objects, height estimation, auditory pitch

1 Introduction

The traditional approach to perception investigates one sense at time [5]. This approach is useful to understand how single senses work, but it does not take into account that perception is intimately a multimodal process: sensations come all simultaneously, so that while touching an object we perceive its size with our eyes and hands.

Research in multimodal perception provides the ground for the design of multimodal interfaces and virtual environments. It has been long recognised that properly designed and synchronized haptic and auditory displays can provide greater immersion in a virtual environment than a high-fidelity visual display alone [6, 13]. Audio-haptic interaction is particularly interesting for applications involving interaction with virtual objects and environments [1], including music performance and interaction with musical instruments [8]. Moreover, haptic and auditory modalities can be exploited to design interfaces for non-sighted users, e.g. to render spatial information in non-visual exploration of maps [2].

One of the main contributions of the multimodal approach to perception is that flows of information carried by the senses interact and modulate each other. The amount of this intermodulation depends on the type of information and on the ability of each sense to estimate certain information [4]. As an example, vision estimates geometrical properties (e.g., size) better than audition. Therefore, audition has only a limited possibility to modulate size-information carried by



Fig. 1: Experiment manager station (a) and subject performing a trial (b).

eyes. Similarly, audition outperforms vision in temporal tasks, and vision has limited possibilities to modulate temporal information carried by our ears [12].

Modulation across senses can also occur when the information carried by two senses is semantically coherent. As an example, musical pitch is often classified as "high" and "low", i.e., with intimately spatial terms. Rusconi *et al.* [9, 10] showed that this spatial connotation interacts with motor action so that, when we are asked to respond quickly whether a pitch is high or low in comparison to a reference, we are faster if the response is coherent with the spatial position of the response key (e.g., the response is "high" and the response key is in the upper part of the keyboard), rather than viceversa.

Empirical results on this type of interaction have led researchers to hypothesize that the representations of heterogeneous continua share a common nucleus. As an example, according to the ATOM's theory [15] the representation of space, time and number is processed by a common mechanism. Other authors, in contrast, suggest that some representations (i.e., time and numbers) are spatially mapped [3, 7]. At any rate, the work reported in [9, 10] shows that musical pitch interacts with non-auditory continua such as motor space. Here, we investigated weather a tone's pitch (i.e., a stimulus subtly subtending a spatial representation) can influence a robust perception such as the haptic estimate of the size of an object in absence of vision. In the experiment described in Sec. 2, blindfoldedsubjects explored a virtual 3D object by means of a haptic device and and had to return a verbal estimate of object's height. Haptic exploration was accompanied with a continuous sound, a sinusoid whose pitch varied as a function of the interaction point's height within two ranges. Experimental results, discussed in Sec. 3 show that the information carried by the frequency sweep (i.e., larger sweep, larger tonal space, therefore larger "space") can modulate only to a limited extent a robust and inherently spatial perception such as the haptic perception.

2 The experiment

Apparatus. The experiment was carried out in a silent booth. The experimental setup is depicted in Fig. 1. A computer acts as the control unit and is con-



Fig. 2: A simplified scheme of the experimental setup.

nected to a Sensable PHANTOM Desktop haptic device³ and to a Motu 896mk3 sound card, which transmits the acoustic feedback to two loudspeakers (Genelec 8030A). The graphical user interface is implemented in MATLAB. The developed software setup is outlined in Fig. 2. In order to build a extensible interactive multimodal virtual environment, the setup integrates H3DAPI⁴ (an open source platform using OpenGL and $X3D^5$ with haptics in one unified scene graph) and Pure Data⁶ (an open source real-time environment for audio processing). Communication is managed through Open Sound Control (OSC).

Stimuli. Test subjects are asked to haptically estimate the heights $h_i = 2 \cdot i$ ($i = 1 \dots 5$) of five virtual stair-steps. To guarantee a sufficient workspace, all the stair-steps span a $22 \times 22 = 484$ cm² horizontal square area (see Fig. 3) The step riser lies in the yz-plane of the virtual scene and in the midsagittal plane related to subject posture. The upper tread of the stair-step lies at the left or right side of the yz-plane, for a right-handed or a left-handed subject respectively. Normalized static and dynamic friction coefficients are set to 0.1 and 0.4 respectively. The normalized stiffness is set to 1 in order to render an impenetrable virtual stair-step without causing the device to become unstable.

Upon collision of the cursor with the stair-step, an auditory feedback is produced. The y coordinate of the haptic interaction point (HIP) is mapped to the frequency of a sine wave. Therefore, continuous interaction with the surface produces a dynamic sine sweep. The mapping is defined through three parameters: f_{min} , the lowest frequency associated to ground level; Δf , the frequency range spanned above f_{min} (thus the maximum frequency associated with the upper tread is $f_{max} = f_{min} + \Delta f$); the mapping strategy, i.e. the function that maps HIP to the frequency domain within the prescribed range.

In this work we choose $f_{min} = 200$ Hz and a linear mapping strategy. Thus, for the height h_i the fundamental frequency f of the sine sweep is:

$$f = f_{min} + \Delta f \frac{y}{h_i}.$$
 (1)

 $^{^{3}}$ A 6-DOF position input/3-DOF force output device with a stylus grip [11].

⁴ http://www.h3dapi.org/

⁵ http://www.web3d.org/x3d/

⁶ http://puredata.info/



Fig. 3: A schematic representation of the experiment spatial arrangement.

The only varying parameter is Δf , which takes the two values f_{min} and $3f_{min}$. These result in f_{max} values that are one and two octaves above f_{min} , respectively.

Along the x HIP coordinate, a simple linear panning approach spatially places the auditory feedback between two loudspeakers in order to improve spatial orientation in the virtual scene [14] and the localization of the step riser. The gains for the left and right channel are $G_{l,r} = \frac{1}{2}(1 \pm P)$, where $P \in [-1, +1]$ corresponds to the horizontal panning position, i.e. $P = \pm 1$ at the left/right loudspeaker positions, and P = 0 at the step riser. Levels are set so as to produce 70 dB SPL at the approximate listener head position.

The choice of sine waves is coherent with our initial research question of how pitch alone interacts with haptic size estimation.

Procedure. Participants are informed about the use of the stylus for exploring objects, and no indication about their shape is provided. They are led to believe that they will explore real, physical objects, and no mention of the haptic device is made. They are blindfolded before entering the silent booth and guided to the experiment-ready position (see Fig. 3).

The 10 stimuli (5 heights $\times 2 \Delta f$ values) are presented with 4 repetitions. The order of the 40 trials is randomized. Participants are instructed to be accurate to 1/10th cm in their verbal height estimations and to explore the whole objects surface (including step risers). They are allowed to interact with each stimulus for 10 seconds before answering. No feedback concerning the accuracy of their responses is given. The role of the auditory feedback is not explained or commented. At the end of the trials participants are guided out of the silent booth and asked to answer a questionnaire.

Subjects. A total of twenty subjects 12 males and 8 females, aged between 20 and 30 (mean = 23, SD = 2.92), caucasian, 18 right-handed and 2 left-handed, took part to the experiment. All participants self-reported normal hearing and no impairment in limb movements. They took, on average, about 35 minutes to complete the experiment. They were students and apprentices of the University of Padova and had no knowledge nor experience of haptic force-feedback devices.



Fig. 4: Mean height estimations (a) and mean relative error of height (b) \pm standard deviation, as a function of the real stair-steps height, for the two sinesweep frequency ranges across 20 subjects. Perfect estimations (black solid lines) lie on the bisector (a) and on the x axis (b).

3 Results and Discussion

Subject estimates were averaged separately for stair-step height and frequency range (see Fig. 4a). Two (frequency ranges) by five (stair-step heights) two ways analysis of variance (ANOVA) on the resulting values revealed that subjects' estimates increase as a function of the stair-step F(4, 74) = 66.00, p < .001.

Figure 4a shows that absolute height estimates were accurate on average. A slight tendency to underestimate the stair-step height was observed for most subjects in all conditions. More importantly, subjects produced larger estimates when the stair-step was accompanied by a larger frequency range sweep. From Fig. 4a it can be seen that, for each height, averaged estimates are larger when the frequency range spans two octaves. However, the ANOVA showed that this result was not statistically significant: F(1, 19) = 1.79, p = .19.

In order to asses how estimates changed as a function of the height of the stair-step the subjective estimates were transformed in percentages of underor over-estimation of the stair-step (see Fig. 4b) and the two-ways ANOVA was recalculated. From Fig. 4b, the effect of the auditory feedback can be appreciated more clearly. The subject's percent error did not change as a function of the stair-step size, F(4, 76) = 0.57, p > .05. However, also in this case the ANOVA confirmed the non significant effect of sound: F(1, 19) = 1.58, p = .22.

Although current results are not statistically significant, we propose a further analysis. For each subject, the mean value among equal-height repetitions defines the *personal perceived audio-haptic height*. The sum of trials that exhibit an estimated height greater than the corresponding reference forms the *personal over-estimation data pool* from which we can identify two sub-groups with respect to their frequency ranges. Figure 5a depicts the relative increment in cardinality (%) of $\Delta f = 2 - oct$ sub-groups related to $\Delta f = 1 - oct$ sub-groups



Fig. 5: (a) Increment in cardinality (%) of $\Delta f = 2 - oct$ sub-group related to $\Delta f = 1 - oct$ sub-group per stair-step height across 20 subjects. (b) Subjects' answers to Q2 (black solid line: mean; black dotted line: standard deviation), and Q3 (blue solid line: mean; blue dotted line: standard deviation.

in dependence of haptic stair-step heights. The resulting curve increases monotonically, especially in 4-8 cm range encouraging further studies, particularly in haptic heights greater than 8 cm where a detectable audio effect occurs.

Finally, three questions from the post-experimental questionnaire are reported and the corresponding answers by subjects are discussed:

- **Q1:** indicate if the object(s) was (were) real;
- Q2: evaluate to which extent the haptic feedback helps your estimates;

Q3: evaluate to which extent the auditory feedback helps your estimates.

The first question is binary evaluated (yes or no). Questions Q2, Q3 are evaluated on a visual analogue scale (VAS) [0 = not at all, 10 = very much].

Interestingly, answers to Q1 reveal that the majority of subjects indicate the stair-steps as real (yes: 12; no: 6; do not know: 2), confirming the high degree of realism in the haptic simulation achievable by the PHANTOM device. Answers to for Q2 and Q3 are summarized in Fig. 5b. It can be seen that most subjects considered the haptic feedback to be very helpful. On the contrary, subjects on average reported the auditory feedback to be substantially less helpful, although the answers to Q3 exhibit a larger variability. This again supports the experimental results, which show to some extent an effect of the auditory feedback.

In fact, by looking at individual subject performances, it can be seen that the frequency range of the sine-sweep has some effect also on subjects who reported no influence by the auditory feedback (e.g., Subject 8 in Fig. 6a). Furthermore, subjects who reported to equally take into account both modalities (e.g. Subject 7 in Fig. 6b) are clearly influenced by the frequency range.



Fig. 6: Mean relative error \pm standard deviation as a function of stair-steps height, for the two sine-sweep frequency range. (a) Subject 8; (b) Subject 7.

4 Conclusions and ongoing work

In this paper we presented a pilot experiment aimed at studying the effects of auditory feedback (and particularly pitch) on the haptic estimation of object size (and particularly height). Our initial results demonstrate the presence of such an effect. More precisely, when pitch is varied along height using a larger scaling factor, subjects' estimates of height become both larger and less reliable.

According to current data, the effect is not statistically significant. This result may suggest that haptic information related to object size is extremely robust and that audition has only a limited possibility to modulate size information acquired haptically. This view is also supported by the overall accuracy of subjects in their absolute estimations of step heights.

However, the experiment presented in this paper was conceived as a pilot experiment from which more extensive tests can be designed. Therefore, firm conclusions can be drawn only upon these tests being completed. In particular, upcoming experiments will also measure subjects' performance without auditory feedback, in order to compare unimodal and bimodal conditions.

Moreover, other kinds of auditory feedback will be explored (e.g., using loudness, or spectral centroid, or other parameters as indicators of step heights). The goal in this case is to assess whether certain parameters are semantically more coherent than others with the concept of height. Although such a semantic link has been proven to exist for pitch [9, 10], other parameters (e.g., loudness) may provide an even stronger link.

Different experimental set-up will also be explored. In particular, headphones will be used instead of loudspeakers, in order to minimize audibility of mechanical sounds coming from the PHANTOM device, and to provide better spatial co-location of PHANTOM stylus and feedback sound.

Acknowledgments. We acknowledge the support of Alberto Ziliotto for software development, Laura Anolfo and Alice Cazzolato for data collection.

Bibliography

- Avanzini, F., Crosato, P.: Integrating physically-based sound models in a multimodal rendering architecture. Comp. Anim. Virtual Worlds 17(3-4), 411–419 (July 2006)
- [2] Campus, C., Brayda, L., Carli, F.D., Chellali, R., Famà, F., Bruzzo, C., Lucagrossi, L., Rodriguez, G.: Tactile exploration of virtual objects for blind and sighted people: the role of beta 1 EEG band in sensory substitution and supra-modal mental mapping. J. Neurophysiol. 107(10), 2713–2729 (2011)
- [3] Dehaene, S., Bossini, S., Giraux, P.: The mental representation of parity and number magnitude. Journal of Experimental Psychology: General 122(3), 371 (1993)
- [4] Ernst, M.O., Banks, M.S.: Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 429–433 (2002)
- [5] Fechner, G.T.: Elemente der Psychophysik. Breitkopf & Härtel, Leipzig (1889)
- [6] Hahn, J.K., Fouad, H., Gritz, L., Lee, J.W.: Integrating sounds in virtual environments. Presence: Teleoperators and Virtual Environment 7(1), 67–77 (Feb 1998)
- [7] Ishihara, M., Keller, P.E., Rossetti, Y., Prinz, W.: Horizontal spatial representations of time: Evidence for the stearc effect. Cortex 44(4), 454–461 (2008)
- [8] Miranda, E.R., Wanderley, M.: New Digital Musical Instruments: Control And Interaction Beyond the Keyboard (Computer Music and Digital Audio Series). A-R Editions, Inc., 1st edn. (2006)
- [9] Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C., Butterworth, B.: The mental space of pitch height. Annals of the New York Academy of Sciences 1060(1), 195–197 (2005)
- [10] Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C., Butterworth, B.: Spatial representation of pitch height: the smarc effect. Cognition 99(2), 113–129 (2006)
- [11] SensAble: Phantom desktop technical specification (2013), http://www.sensable.com
- [12] Shams, L., Kamitani, Y., Shimojo, S.: Visual illusion induced by sound. Cognitive Brain Research 14(1), 147–152 (June 2002)
- [13] Srinivasan, M.A., Basdogan, C.: Haptics in virtual environments: taxonomy, research status, and challenges. Comput. & Graphics 21(4), 393–404 (1997)
- [14] Stamm, M., Altinsoy, M., Merchel, S.: Influence of the auditory localization direction on the haptic estimation of virtual length. In: Cooper, E., Kryssanov, V., Ogawa, H., Brewster, S. (eds.) Haptic and Audio Interaction Design, Lecture Notes in Computer Science, vol. 6851, pp. 101–109. Springer Berlin / Heidelberg (2011)
- [15] Walsh, V.: A theory of magnitude: common cortical metrics of time, space and quantity. Trends in cognitive sciences 7(11), 483–488 (2003)