Tactylus, a Pen-Input Device exploring Audiotactile Sensory Binding

Ernst Kruijff
Fraunhofer Institute for Intelligent Analysis and Information Systems, Institute for Computer Graphics Graz University of Technology

ernst@icg.tu-graz.ac.at

Gerold Wesche
Fraunhofer Institute for Intelligent Analysis and Information Systems Sankt Augustin, Germany

wesche@iais.fraunhofer.de

Kai Riege
Fraunhofer Institute for Intelligent Analysis and Information Systems Sankt Augustin, Germany

riege@iais.fraunhofer.de

Gernot Goebbels
fleXilution GmbH
Köln, Germany

gernot.goebbels@flexilution.de

Martijn Kunstman
XIOX
Culemborg, Netherlands

kunstman@xiox.nl

Dieter Schmalstieg
Institute for Computer Graphics Graz University of Technology Graz, Austria

schmalstieg@icg.tu-graz.ac.at

ABSTRACT

Recent studies have shown that through a careful combination of multiple sensory channels, so called multisensory binding effects can be achieved that can be beneficial for collision detection and texture recognition feedback. During the design of a new pen-input device called Tactylus, specific focus was put on exploring multisensory effects of audiotactile cues to create a new, but effective way to interact in virtual environments with the purpose to overcome several of the problems noticed in current devices.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**] Multimedia Information Systems — *artificial, augmented, and virtual realities*; H.5.2 [**Information Interfaces and Presentation**] User Interfaces — *Haptic I/O*; I.3.6 [**Computer Graphics**] Methodology and Techniques – *Interaction Techniques*; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism—*Virtual Reality*

General Terms

Measurement, Performance, Design, Experimentation, Human Factors

Keywords

3D User Interfaces, audiotactile feedback, sensory substitution

1. INTRODUCTION

Nowadays, many virtual reality (VR) setups make use of some

kind of pen-input device to control a VR application. When observing frequently used pen-input devices, especially in more complex task environments some ergonomic problems can be noticed, next to control and feedback issues. A pen-input device is generally regarded as a time-multiplexed device. Multiple tasks are performed in a serial manner, thereby resulting in a composite flow of action. However, the control possibilities do not always match the complexity of an application, regularly leading to the usage of a system control technique to change the mode of interaction, or the usage of a second device. The pen-input devices we evaluated (from Polhemus and Intersense) are not very ergonomic when analyzing grip and balance issues. Using these devices, manipulation tasks are not always easy to perform, and might benefit from more advanced feedback mechanisms besides visual and self-maintained feedback, especially in visually challenging (occluded) tasks. Finally, we noticed an increased demand on pen-like devices to support integration in hybrid interaction setups. In order to potentially solve these problems, several issues where addressed during the design of a new input device, called the Tactylus. We created an ergonomically improved form, limiting fatigue problems, and increased the amount and diversity of buttons to improve the quality of system control, including the support for hybrid interaction. Finally, we searched for apt feedback solutions to increase performance, resulting in the application of sensory substitution methods [1] by using audiotactile cues to simulate haptic feedback. A key issue was to analyze multisensory effects of closely coupled visual, auditory and tactile cues. Recently, some research (section 2) has been focusing on the effects of a process called multisensory binding. The usage of congruent signals can lead to a better perception of roughness of textures, up to perceptual modulation that may serve as a collision detector [2, 3]. The objective when designing the Tactylus was to take advantage of multi-sensory binding, finding out how it would affect performance and perception of certain events. The research on the Tactylus continues work on hybrid interfaces by coupling desktop and immersive interaction [4, 5], and vibrotaction feedback methods [6, 7]. The specific integration of vibrotactile and pen-like

interface devices is still rather rare. Some undocumented examples exist, but only a few have been published. Directly related developments include the Haptic Stylus [8], which emulates the feel of buttons, the SenStylus [9], which integrates vibrotactile feedback in a pen-like device.

2. MULTI-SENSORY BINDING

The majority of approaches dealing with multimodal feedback regard output modalities as separate entities. More recently, this idea is shifting towards what is called multi-sensory binding [10] or cross-modal integration [2]. This integration refers to activities in the brain that affect the perception of feedback obtained from multiple sensory channels. Based on the plasticity of the brain to associate sensory information, modalities interact among each other, through which perceptual events can be integrated [11]. Three different situations can be identified: cross-modal bias, enrichment and transfer. During cross-modal bias, stimuli from two or more sensory systems can differ and can affect each other leading to modified or even incorrect perception. Enrichment refers to the strengthening of perception by adding one or more sensory channels, whereas transfer deals with sensory potentials that trigger potential in another sensory channel, thereby not necessary biasing perception. One of the keys to integration is the weighting of sensor potentials, dealing with the domination of one sensory channel over another one. Generally, visual alters the other modalities, though it has been shown that sound alters the temporal aspects of vision and, based on the modality appropriateness theory, may also alter other aspects of vision [1]. This refers to task-specific characters of actions: the modality which is most appropriate for the task at hand will most likely dominate over any other sensory channel. The effects of multi-sensory binding can be well used to develop sensory substitution systems, in which sensory potential and resulting perception lie closely together. An example which has greatly inspired the design of the Tactylus is the integration of visual, tactile and sound stimuli to provide correct collision information by Shimojo and Shams [2], suitable for many engineering environments. Shimojo and Shams showed that in an ambiguous motion display, crossing objects could be perceived as bouncing objects, once an additional stimulus was added (visual flash, sound or vibration) with high enough synchronicity with the visual event. Ultimately, vibrotactile stimulation provides directional cues, but this would require multiple vibrotactors, which is hardly possibly in small devices. By integrating spatial audio cues and vibrotactility, this problem can potentially be solved.

The second issue addressed by the binding of visual, vibrotactile and auditory stimuli is the perception of surfaces. Lederman et al. report on the strength of vibrotactilty for exploring textures, even though the resulting psycho-physiological function differs from using the bare finger [12]. As observed by Weisenberger and Poling., audio can contribute to the perception of surfaces, like material properties and texture roughness [3]. Observing both haptic and auditory cues, they found out that when perceiving textures, the weighting is predominantly haptic (62%), though auditory cues play an important role too (38%). They further noticed that auditory cues are not particularly salient upon initial experience – the cues need to be learned. This may be in line with Lederman et al's study, which reported that sound may actually slow down texture recognition. An important factor when dealing with adding auditory cues is the level of correspondence – binding of different stimuli can lead to a different perception of

textures as compared to when only visual information is provided. Hence, it seems that haptic and audio cues can either bias or disambiguate visual information. In any case, auditory cues need to make use of suitable sound models [13] to afford apt interpretation.

Figure 1: The Tactylus using a power grip

3. DEVICE DESIGN

In order to develop the Tactylus, we performed a detailed analysis of control-body linkages and specific control tasks associated with the usage of pen-input devices, to come to a balanced device with a good grip. Using a user group (9 male and female users), we found that users applied two kinds of power grips and one precision grip. Hereby, a different interplay between foremost the thumb and the pointing finger can be noticed. A closer analysis of available pen-input devices showed that most devices are off-balanced due to cables pulling out of the back, without having the possibility to truly counterbalance due to limited grip. In longer sessions this often leads to re-grasping of the device to relieve muscle tension in the arm and the hand. Whereas fatigue is not a big problem in sweeping tasks (high speed, low accuracy tasks such as quickly moving an object), it becomes an increasing problem with fine grain tasks like fine object placement.

The control-body linkage and identified control task parameters lead to specific design variables. The grip of the new device should work for both power and precision grip, and have an ambidextrous form, allowing left and right-handed usage. The thumb should have a good grip to counterbalance the weight of the device allowing the musculature of the pointing finger to relieve in precision grip modus. Additionally, the main body of the stylus should be large enough to hold the needed electronics of the device, including 2 buttons and a wheel, tracking sensor, vibrotactor, and pen-tip. With the user group, ergonomics of different pen-forms made out of foam were tested to come to a final polyamide-based solution, produced by selective laser sintering. The electronics were mostly placed in the belly of the device at the front, in order to counterbalance the cables. The tip inserted in the front of the device can be replaced to hold reflective markers for optical tracking. Of particular interest is the way tactile feedback is provided. The vibrotactor (1.5VDC, 60mA pager motor) is connected to an amplifier and a rectifier, which generates a pulsating current making the vibrotactor rotate at the sound frequencies put on a specific audio channel. Hence, it can vibrate at the same frequency as the audio heard through the speakers – the user is literally feeling the audio files. We took this approach since the recorded audio files show particular roughness details when observing the waveform patterns. Audio recorded of

collision (friction) between an object and rough textures shows patterns with quickly changing and highly differing frequency peaks, whereas collisions with smooth objects show a rather flat waveform. We believed that this is especially good for providing texture information via sensory substitution. An additional advantage of using audio-based control of the vibrotactor is the potential high-synchronicity between audio and vibrotactility. Even though the adaptivity speed of a vibrotactor is slightly lower than of an audio speaker, the delay can be neglected.

4. EXPERIMENTS

Two experiments were prepared that focused at collision detection and texture recognition tasks, specifically dealing with analyzing the effects of multi-sensory binding, both in direction of either biased or disambiguated perception. 11 subjects (8 male and 3 female novice (5) and more experienced users, aged 24 – 34) participated in the first experiment that focused on collision detection. The quality of multi-sensory feedback was tested on object placement ("key-lock") tasks with different levels of complexity. Every user performed 4 placement tasks, in half in clearly visual manner, in half where the placement area was visually occluded. The users performed the same placement task under different feedback conditions, namely purely visual, visual and audio feedback, or visual and audiotactile feedback, thus resulting in a total of 12 placement tasks per subject. Cues included both collision and friction audiotactile cues. Performing the more complex key-lock placement under the visually occluded setting could be rated as moderately complex. Also, especially in the visually occluded placement task, visual information was highly ambiguous, since it was rather hard to see if an object would collide or not, due to absence of visual cues such as highlighting or shadows. We hypothesized that users would rate the audiotactile as most appropriate, especially in more complex task situations. In the second experiment, dealing with texture recognition, 10 subjects (7 male, 3 female novice (4) and more experienced users, aged 24 – 49) took part, of which 8 users also took part in the first experiment. The texture recognition experiment primarily focused on biasing or disambiguation effects of visual information for recognizing different levels of roughness of textures. Subjects were presented with 15 different combinations of visual, tactile and auditory cues, to test binding effects between different modalities. The combinations represented 5 different stages of roughness. For this purpose, 5 different textures (size 512 x 512 pixels) were selected with different visual roughness. Sounds (with a continuous roughness scaling) were generated by recording collision between real materials (metal) and synthesizing them. The vibrotactor would vibrate according to the wave pattern of the selected audio file. We hypothesized that by differing combinations of visuals, audio and vibrotactile feedback, audio and vibrotaction could alter the visual perception. Before the experiment, users were shown 2 reference textures in order for the user to correctly understand the continuous roughness scaling. Users were not informed about the methods behind the test (i.e., the differing of feedback combinations). After the experiments, users were interviewed and asked to fill out a questionnaire with 9 questions, using a 7 point Likert scale rating. The experiment was performed at a back projection display, the TwoView display, driven by a dual-pipe PC with 2,4Ghz, and a NVIDIA Quadro 4400 graphics board. Audio was provided by a pair of loudspeakers, whereas the ART tracking system was used to optically track the Tactylus.

5. RESULTS AND DISCUSSION

Analyzing the results of the experiments, some statements can be made. Due to the rather non homogenous user group, these statements need to be further tested but provide very interesting indications.

Audiotactile cues can enhance collision detection
The object placement experiment evaluating collision perception seemed to confirm the stated hypothesis: users preferred using audiotactile feedback to perform object placement to more correctly interpret collision between objects. In the less complex object placement tasks, users rated audio-only feedback as being good (average (avg) 5.00, standard deviation (stdev) 1.41), but most users clearly preferred the combination of audio and tactile feedback (avg 5.73, stdev 1.27). For the more complex tasks (occluded placement), users rated the feedback slightly lower, both in case of audio (avg 4.64, stdev 1.29) and audiotactile feedback (avg 5.45, stdev 1.13). The lower rating can be explained by the higher complexity of the task: even though most people (78%) found the audiotactile feedback more appropriate in the more complex than in the easier situations, the total amount of feedback is less, since visual feedback is highly reduced through occlusion in the complex scenarios. The majority of users stated they could perform their actions more precisely with audiotactile feedback (avg. 5.46, stdev 1.23), in which 63% claimed the precision increase to be very good (rating between 6 and 7).

Audiotactile cues can disambiguate visual information
Most users seemed to be able to disambiguate the visual information well in the visually occluded task by using the audiotactile feedback. The accuracy of spatial audio was rated less well and highly diverse: about half of the people could well make use of the directional information, the others had problems with it (avg. 3.45, stdev 2.02). None of the users reported on problems with binding the audio and tactile information (avg 5.18, stdev 1.25). Hence, it seems that audiotactile feedback can effectively be used to illicit collision information to users, and can substantially aid in especially visually complex situations. It should be stated that this could be increased by adding visual cues such as highlighting or shadows. Nonetheless, especially in visually occluded situations as tested in this experiment, such cues become increasingly less effective.

Audiotactile cues can enhance texture perception
The texture roughness recognition experiment turned out to be a hard and complex task. 67% of the users stated that they could correctly and well interpret the textures using visual, audio and tactile information, thereby supporting the results reported in [12]. 2 users found it extremely hard, though. When observing the interpretation of the first 5 textures (represented by the "correct" combination of visual, auditory and tactile information), interpretation offsets were rather small, indicating that users were rather precise in texture roughness recognition. The average offset was extremely low at rough textures, whereas with light textures at most around half a scale in average. Strangely enough, some offset was seen with the interpretation of the smoothest texture, though shown as reference texture at the start of the test – the roughest texture (also shown as reference), was interpreted flawlessly afterwards. Some learning effects could be noticed of users who also expressed to be more visually oriented: based on visual properties only, the recognition rate increased. As soon as

the combination of visual, auditory and tactile information would get biased, some interpretation changes can be noticed. Based on the noted texture interpretations, a first important notification is that as long as the change in auditory and / or vibrotactile information is small, users seem to interpret the texture visually, or at least remember the original combination. Thus, changing on or two scales up or down with auditory or vibrotactile feedback does not necessary bias visual perception. This was especially true by biasing visually smooth textures with rough sound and vibration feedback. In a single case, similar visual and auditory texture of a scale 4 texture information was biased by extremely light (scale 1) vibration: with around half the users, this lead to interpreting the texture as being smoother, but interestingly enough, the other half interpreted this texture as being rougher than before. Currently, we do not have any explanation for this.

Audio can alter vision

A clear offset was noticed when the (previously flawless interpreted) roughest texture was shown with level 5 vibration, but with level 3 audio roughness feedback. 89% now claimed that the texture was at least one level smoother (avg. 3.90), which is a rather clear proof that audio can alter visual perception. The interviews with the subjects made some things more clear which could not be clearly derived from the numerical analysis. First, most users reported they were sometimes annoyed by the feedback they got, supporting the hypothesis that people do not mainly base their perception on visual information but tend to be able to get biased by other sensorial information. Most users noted that they would first take audio into account (avg 4.80 suitability / increase of interpretation over visual only, 1.75 stdev), whereas vibration only played a less important role (avg 3.70 suitability for interpretation, with 2.01 stdev). These results were biased by a few users rating extremely low. Most users also reported on the lower usability of vibration for texture recognition: some noticed that the difference between vibrations was not good enough, some others found it good enough but just did not put too much focus on it. Interestingly enough, about half of the subjects stated that audio did disambiguate the visual information: the textures were purposely represented in a flat way (put on a plain), as such that users could clearly notice roughness differences, but not the height differences. Overall, our test results seem to imply that audio is more important than vibrotactile information for texture recognition, which contradicts with findings from Lederman [12].

6. CONCLUSION

In this article, we presented a new pen input device which provided successful indications for both the usability and complexity of binding audiotactile cues. Audio seems to have a larger impact on perception of collisions and textures than vibration, even though this could be leveled out when discrimination between levels of vibration is higher. The results support previous findings in the direction of multi-sensory processing, and provide some ideas on how both its strength and weaknesses can be applied to increase interaction performance.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Kaczmarek, K., et al., Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomechanical Engineering, 38, 1, 1991.*

[2] Shimojo, S. and Shams, L. Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology, 11, 2001, 505-509.*

[3] Weisenberger, J. and Poling, G. Multisensory Roughness Perception of Virtual Surfaces: Effects of Correlated Cues. In *Proceedings of the 12th International Symposium on Haptic Interfaces for Virtual Environments and Teleoperators Systems (HAPTICS'04).* 2004.

[4] Szalavari, Z. and Gervautz, M. The Personal Interaction Panel - a Two-Handed Interface for Augmented Reality. Computer Graphics Forum, 16, 3, 1997, 335-346.

[5] Watsen, K., Darken, R. and Capps, M. A Handheld Computer as an Interaction Device to a Virtual Environment. In *Proceedings of the Third Immersive Projection Technology Workshop.* 1999.

[6] Cheng, L.-T., R. Kazman, and Robinson, J. Vibrotactile feedback in delicate virtual reality operations. In *Proceedings of the Fourth ACM International Conference on Multimedia.* 1996.

[7] Lindeman, R., et al. The Design and Deployment of a Wearable Vibrotactile Feedback System. In *Proceedings of the 8th IEEE International Symposium on Wearable Computers.* 2004.

[8] MERL, Haptic Stylus project summary, available at: http://www.merl.com/projects/hapticstylus/. 2004.

[9] Fiorentino, M., Monno, G. and Uva, A. The Senstylus: a novel rumble-feedback pen device for CAD application in Virtual Reality. In *Proceedings of the 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005 (WSCG 2005).* 2005.

[10] Spence, C. and Squire, S. Multisensory Integration: Maintaining the Perception of Synchrony. *Current Biology, 13, 2003, 519-521.*

[11] Pai, D. Multisensory Interaction: Real and Virtual. In *Proceedings of the International Symposium on Robotics Research.* 2003.

[12] Lederman, S., et al. Relative Performance using Haptic and/or Touch-Produced Auditory Cues in a Remote Absolute Texture Identification Task. In *Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (IEEE HAPTICS'03).* 2004.

[13] Guest, S., et al., Audiotactile interactions in roughness perception. Experimental Brain Research, 146, 2002, 161-171.