Combining Multi-Sensory Stimuli in Virtual Worlds – A Progress Report

Julia Fröhlich and Ipke Wachsmuth

AI & VR Lab, Faculty of Technology, Bielefeld University
Universitätsstraße 25, 33615 Bielefeld, Germany
{jfroehli,ipke}@techfak.uni-bielefeld.de

Abstract. In order to make a significant step towards more realistic virtual experiences, we created a multi-sensory stimuli display for a CAVE-like environment. It comprises graphics, sound, tactile feedback, wind and warmth. In the present report we discuss the possibilities and constraints tied to such an enhancement. To use a multi-modal display in a proper way, many considerations have to be taken into account. This includes safety requirements, hardware devices and software integration. For each stimulus different possibilities are reviewed with regard to their assets and drawbacks. Eventually the resulting setup realized in our lab is described – to our knowledge one of the most comprehensive systems. Technical evaluations as well as user studies accompanied the development and gave hints with respect to necessities and chances.

Keywords: Multi-Sensory Stimuli, Wind, Warmth, Presence, Virtual Reality

1 Introduction

Surprisingly today's virtual reality setups fall short on combining multi-sensory stimuli, even though virtual reality visionaries already included such. For instance Morton Heilig's Sensorama Simulator in 1962 [12] presented multi-sensory feedback like sound, wind, force feedback and smell. In 1992 David Zeltzer presented the AIP cube – aiming to provide dimensions for the comparison of virtual realities [26]. As seen in Figure 1 it consists of the three components: presence, interaction and autonomy. Zelzer introduced the presence axis as *"a rough lumped measure of the number and fidelity of available sensory input and output channels"* ([26] p.128).

In [11] we presented a user study on combining modalities in virtual worlds. Results suggested that there might be an uncanny valley effect, when only one additional cue is presented. We concluded that it is worthwhile to add as many modalities as possible which is supported by other evaluations (e.g. [6]). This has two effects: (1) the virtual world gets more believable the more out-put is generated and (2) the user is distracted from the real world if more senses are stimulated virtually.

This paper will introduce a modern approach that includes a set of multi-modal stimuli as well as the integration of autonomy. It consists of a three-sided CAVE-like environment and is enhanced with 3D-sound, tactile feedback, wind and warmth.
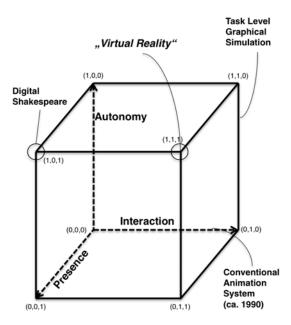

Fig. 1. The AIP-cube redrawn from [26]

2   Related Work

Many systems have been developed which add one modality to a visual virtual environment. This has mostly been **auditive** (e.g. [16], [19]) or **haptic** enrichment (e.g. [3], [23]). Some efforts including **olfactory** stimuli can be found, either presenting the scents via a wearable device (e.g. [24], [13]) or using global displays, which fill a whole room with odor. The main Problem with global approaches is fast neutralization. One System minimizing this problem is introduced by Yanagida et al. [25].

The integration of **wind** is realized either with a stationary system (constructed around the interaction space), or with a local system which is directly mounted onto the user. A real scooter is used for navigation inside the VR-Scooter-project [5] and generates wind in front of the user as well as tactile feedback with vibrotactile sensors. The authors state that perceived wind improved both, task performance and the subjective user experience. A disadvantage of this system is the inability to provide directional wind. Moon and Kim

[22] introduced a setup which consists of 20 fans arranged in three levels around the user. Their evaluation showed that users cannot distinguish between winds which come from neighboring fans, if the angle between is less than 45 degrees. Another system was presented by Cadin et al. [4], consisting of eight CPU fans mounted circularly around the user's head. In order to evaluate the difference between stationary and local systems, Lehmann et al. conducted a user study [20]. The addition of wind significantly improved participants presence and the stationary system was preferred.

There is not much research describing the role of **warmth** in VR. Dionisio published an evaluation of possible hardware devices for warmth sensations in 1996, concluding that fans, infrared lamps and Peltier-elements are the best option for temperature sensations [7]. An exemplary setup was introduced as a "virtual hell" [8], with each three fans and infrared lamps arranged around the user as well as Peltier-elements attached directly on the skin. Only very few systems combine those stimuli in order to create a multi-modal sensation system. For example Dinh et al. [6] implemented a virtual office enriched with visual, audio, wind, warmth and olfactory cues. The setup was constructed to fit the requirements of the user study, but not suited to easily implement another multi-modal virtual environment.

In the sections to follow we describe a progress report on the development of a hardware and software setup designed to create multi-modal virtual worlds with minimum effort.

3   Software Architecture

The biggest problem with designing a multi-modal virtual world is that it is such a time consuming task. Often this effort does not seem to be worthwhile with regard to the benefits like improved immersion. One approach to simplify the creation of virtual worlds is to semantically enrich virtual objects. This concept has proven to be efficient in creating Intelligent Virtual Environments (IVEs) [21]. But until now this idea has mostly been used to store additional knowledge about the graphical representation only.

We have developed a framework in which objects, enriched with information about their multi-modal properties, are being processed to generate enriched virtual worlds. This is done with an automatic processing of the 3D-scene and therefore minimizes the effort needed to develop such multi-modal virtual worlds. The semantic annotation of objects works as a type declaration, and all further information like possible sound and haptic information are drawn from a database. This is employed in a 3-step process: parsing the file, comparing semantic annotation with a database, generating a new file which includes multi-modal properties. Our framework is based on InstantReality, a consistent, platform independent framework for fast and efficient application development for Virtual and Augmented Reality [9]. It employs the X3D standard for the description of 3D-models which uses the XML format. Therefore files have to be stored in the X3D-format, in which objects can be annotated through the `<metadata>` tag.

This knowledge is not held in a separate knowledge base, but anchored at the object itself. Figure 2 shows the process on the example of sound integration. An existing X3D file, with semantic annotation inside the `<metadata>`-tag is parsed. Database entries to the corresponding semantic annotation are searched and added to the scenegraph. Consequently the additional time to create such multi-modal worlds is minimized.
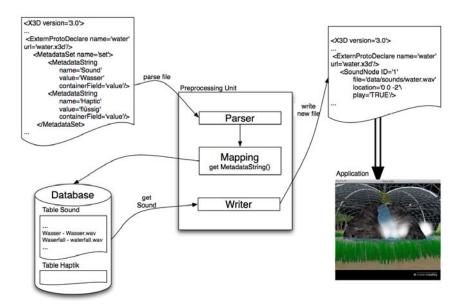

Fig. 2. Concept of enriched virtual objects to create automated multi-modal feedback.

4   Hardware Setup

The original hardware setup installed 2003 consists of a three sided CAVE-like environment and an optical tracking-system (ART) with 11 infrared cameras. A collection of different interaction and navigation techniques is available. Eight speakers are installed in the corners of the interaction area, and subsonic sound devices underneath the floor. but not yet enabled with spatial sound. Figure 3 gives an overview.

During the last four years a system was constructed, enriching the original setup with improved auditory output, natural hand interaction now including tactile feedback and physical reaction, a wind setup and a warmth installation. Each step is subsequently described, giving advice on how to integrate such output modalities in CAVE-like environments. Technical and empirical evaluations accompanied the development process, to ensure a satisfactory system. In the
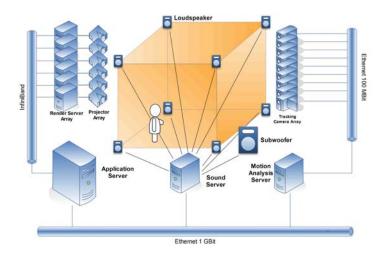
Fig. 3. Technical Setup without multi-sensory integration.

beginning, some requirements were imposed – some tied to the existing hardware setup, others to enable transferability to other VR-systems:

1. Ensure safety of developers and users
2. Ensure safety of existing hardware setup
3. Real-time requirements have to be met, tantamount to fast reaction-time of hardware devices and computation of output
4. Ability of continuous hardware adjustment is preferable
5. Added hardware must be low-cost
6. Existing computational capabilities should be kept, therefore computational cost for additions should be minimized

4.1   Sound

Spatial sound is realized with the existing eight speakers (one in each corner) and two subwoofers underneath the floor. Because the integration of a realistic auditory rendering system (e.g. wave-field synthesis) did not meet *requirements 3, 5 and 6*, a vector based amplitude pannig (VBAP) approach was included. Even though two drawbacks exist (sweet-spots and the inability to simulate realistic output for sound sources close to the user's position), it is a good method with regard to computational and hardware costs and creates quite believable sound fields. For each sound the three speakers closest to the source are chosen, and the volume is regulated with regard to the distance, where doubling the distance results in a decrease of 6 dB [2]. A list with all sounds and activation levels of all speakers is passed on to the FMOD sound library which handles

digital mastering [1]. The FMOD-library, which is free to use for non-commercial purposes, ensures real-time adjustments.

Our concept divides auditory output into three different types [10]. *Ambient sounds* represent a base level of output which is more or less constant over a larger region of a scene. As long as the user is within the defined area, ambient sound will be played without direction and steady volume. Only one ambient sound can be played at a time, thus all not applying ambient sounds are faded out if the user is in a subarea of the world . In addition, this concept allows for the definition of environmental properties which influence the audio rendering (e.g. adding reverberation), to fit the environment, such as an outside scenario, a dripstone cave or a concert hall. *Static sounds* are directly coupled to an object. They are adjusted in volume and direction with regard to their position relative to the user. *Event sounds* are only triggered when the related event occurs (e.g. a ball hitting the floor).

4.2   Tactile feedback

Tactile feedback is accomplished by ART fingertracking devices which track the thumb, the index finger and the middle finger. At the tip of each finger three wires made of memory metal are attached which create a vibration, serving as tactile feedback when heated in short intervals.

To enable natural interaction the ODE physic engine in conjunction with a gesture-interaction are embedded. Collision detection allows for an appropriate visual reaction as well as generation of event sounds as mentioned above.

Technical evaluation showed a satisfactory result regarding the natural interaction, as long as a solid gesture training is performed in the beginning. The tactile feedback however was not perceived well enough in all cases – the vibration seems to be not strong enough to be noted by every participant. Further investigation of appropriate thresholds and feasible alternative hardware devices will be performed in future work. In general physical interaction is regarded essential to employ realistic feedback to user actions, but since physical calculation is costly the number of possible collisions is limited to approximately 40.

4.3   Wind

Wind effects are accomplished by eight low-cost controllable fans by ADDA Corp., which are located at the upper bound of the projection area. Since the user can rotate freely while being in the virtual world it was necessary to place them around the interaction space. In consideration of available space, costs, as well as fine-grained adaptation of wind direction we chose a setup in which the fans are mounted evenly distributed on a nearly circular arrangement [22]. The fans were chosen with a special focus on being as silent as possible and for this reason are driven with 115 Volt instead of 230 Volt. They have a diameter of 25.4 cm and a wind performance of roughly 13 m^3/min. To control all fans continuously, they are connected to a MultiDim MKIII Dimmerpack by ShowTec and driven via the DMX-Protocol.

Similar to the sound generation, wind sources are divided into three types [15]. *Directional wind* simulates the overall wind effect from one far away direction. *Spot wind* is coupled with a virtual object (e.g. a virtual fan) and therefore has a position and only influences a small region in front of it. Further, *event wind* was added in order to simulate sudden wind events. Due to reaction time, very short events, as the drift of a closing door, cannot be simulated, while the perception of a passing train enriched with wind worked quite well. The first system did not perform occlusion detection which was considered a major drawback during a pilot study and therefore was included later on.

To determine important thresholds, an exploratory study with N = 9 test subjects was conducted to measure accuracy of wind direction estimation, reaction times to recognize the effect and needed activation thresholds. Results show that the wind direction is detected with an accuracy of M = 24.13 degrees (SD = 30.25). It took participants about 3 seconds (M=3.1s SD=0.29s) to notice the wind stimulus, whereas the disappearance of wind was noticed much faster (M=1.3s, SD=0.85s). Activation levels needed to be about 50% in order to be sufficient. Detailed results are given in [14].

4.4 Warmth

To enable thermal stimuli as well, a heat display was added, consisting of six 250 Watt infrared lamps, which are mounted above the interaction area. Infrared lamps met all necessary requirements – a detailed discussion about the hardware considerations can be found in [15]. The height can be regulated to ensure users' safety and enable equivalent experience among different sized people. Maximum heat in front of the lamp is $100°C$, while temperature sensors are used to monitor the whole system and ensure safety requirements. Again, a differentiation between different heat sources is implemented. *Directional warmth* can be used to model effects like the heat of the sun, whereas *point warmth* is used to model smaller heat sources (e.g. a fireplace).

Since our tracking system uses infrared light, a major concern was whether the infrared lamps could disturb the performance of the user-tracking. Tracking coverage in the interaction space was tested and no influence was measurable. Although it must be suspected that in a different setting (e.g. with fewer cameras or more infrared lamps), the lamps may affect the tracking. The same test procedure as for wind suggests an activation threshold of 57% (SD = 21%). Mean time for perceiving the appearance of warmth was 2.5 seconds, whereas the disappearance was noticed faster (M = 2.0s).

5 Further Modalities

The enrichment with odor was not yet approached, because hardware availability is limited and therefore costly. Additionally our global setup should be preferably extended with a global scent system (e.g. [25]) which are not that common.

Referring back to the AIP-cube (cf. Figure 1), this system accomplished a great step towards point $(0, 1, 1)$, but in order to create an even better virtual experience, the autonomy axis has to be considered as well. It is defined as *"a qualitative measure of the ability of a computational model to act and react to simulated events and stimuli"* ([26] p.127). A lot of research is done in the field of virtual agents, which strives for this level of autonomy, but integration into virtual worlds where the virtual human interacts with the user and acts autonomously, is even less common in VR applications than multi-sensory stimuli. Therefore the integration of a virtual human introduces yet another level of possible presence enhancement. To make a step toward point $(1, 1, 1)$ the embodied conversational agent MAX was included in our virtual setup. MAX is an autonomous agent developed in the AI & VR Lab at Bielefeld University and for example is employed as a museum guide in Germanys biggest computer museum, the Heinz Nixdorf MuseumsForum in Paderborn [17]. The agent has the ability to generate speech and gestures naturally and in some scenarios acts as a virtual interaction partner [18]. A virtual lab setup was designed, where MAX acts as a virtual guide, including some interaction possibilities like bowling as well as some chit chatting. He shows people around, acts autonomously, and reacts to user input. The environment is enriched with sound, wind, warmth and tactile feedback as well. Figure 4 gives a glance at the setup.

Fig. 4. MAX (added as autonomy factor to a multi-modal virtual environment) in the virtual lab on a hoover disc, to take visitors along.

6   Conclusion

The resulting system combines visual with auditory, tactile, wind and warmth stimuli (cf. Figure 5). Due to the earlier described software-architecture is it easy to include in existing or new virtual reality projects. To our knowledge this is a VR-setup with one of the most extensive multi-sensory displays. Since low cost (hardware and computational wise) was one main criteria, the system may be added to another CAVE-based VR-system with little effort. In addition a major user study to examine the impact on perceived presence were performed, with promising results [11]. The addition of as many output modalities as possible seems to be worthwhile, especially if hardware costs are kept to a minimum as well as the effort needed to integrate the output modalities. In future work, the effects of multi-sensory stimuli will be investigated further, with an emphasis on cross-modal dependencies that might occur.

An evaluation of the integration of an autonomy factor as proposed by Zeltzer is desirable. Within the proposed lab setup, test subjects are guided either by a real or virtual human. Further investigation of the impact on perceived presence with regard to the absence or presence of an autonomous agent, is assumed to help with the design of more comprehensive virtual reality systems.
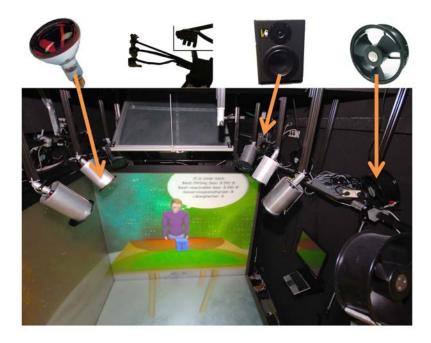

Fig. 5. Hardware devices used with exemplary position in the overall setup.

Acknowledgements

References

1. fmod - interactive audio middleware. http://www.fmod.org, 2010.
2. D. R. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.
3. G. Burdea, G. C. Burdea, and C. Burdea. *Force and touch feedback for virtual reality*. Wiley New York, 1996.
4. S. Cardin, D. Thalmann, and F. Vexo. Head Mounted Wind. In *Computer Animation and Social Agents*, pages 101–108, 2007.
5. L. Deligiannidis and R. J. K. Jacob. The VR Scooter: Wind and Tactile Feedback Improve User Performance. In *3DUI'06*, pages 143–150, 2006.
6. H. Q. Dinh, N. Walker, C. Song, A. Kobayashi, and L. F. Hodges. Evaluating the Importance of Multi-sensory Input on Memory and the Sense of Presence in Virtual Environments. In *Proceedings of the IEEE Virtual Reality*, VR '99, pages 222–228, Washington, DC, USA, 1999. IEEE Computer Society.
7. J. Dionisio. Temperature feedback in Virtual Environments. In *Imaging Sciences and Display Technologies*, pages 233–243, 1996.
8. J. Dionisio. Virtual Hell: A Trip Through the Flames. *IEEE Comput. Graph. Appl.*, 17(3):11–14, May 1997.
9. D. Fellner, J. Behr, and U. Bockholt. Instantreality - a framework for industrial augmented and virtual reality applications. In *The 2nd Sino-German Workshop Virtual Reality & Augmented Reality in Industry*.
10. J. Fröhlich and I. Wachsmuth. A phong-based concept for 3D-audio generation. In *Smart Graphics*, pages 184–187. Springer, 2011.
11. J. Fröhlich and I. Wachsmuth. The Visual, the Auditory and the Haptic – A User Study on Combining Modalities in Virtual Worlds. In *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pages 159–168. Springer, 2013.
12. M. L. Heilig. Sensorama simulator, u.s.patent no.3050870, August 1962.
13. K. Hirota, Y. Ito, T. Amemiya, and Y. Ikei. Presentation of Odor in Multi-Sensory Theater. pages 372–379, 2013.
14. F. Hülsmann, J. Fröhlich, N. Mattar, and I. Wachsmuth. Wind and Warmth in Virtual Reality: Implementation and Evaluation (in press). VRIC '14: Proceedings of the Virtual Reality International Conference: Laval Virtual. ACM, 2014.
15. F. Hülsmann, N. Mattar, J. Fröhlich, and I. Wachsmuth. Wind and Warmth in Virtual Reality – Requirements and Chances. In *Virtuelle und Erweiterte Realität: 10. Workshop der GI-Fachgruppe VR/AR*, pages 133–144, 2013.
16. S. Husung, R. Mikalauskas, C. Weber, and T. Kästner. Modelling of sound propagation of technical systems for real-time VR-applications. *MECHANIKA*, (4):33–37, 2010.

17. S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth. A conversational agent as museum guide - Design and evaluation of a real-world application. Intelligent virtual agents: Proceedings 5th International Conference, pages 329–343. Springer, 2005.
18. S. Kopp, B. Jung, N. Leßmann, and I. Wachsmuth. Max - A multimodal assistant in virtual reality construction. *KI - Künstliche Intelligenz*, 4(03):11–17, 2003.
19. P. Larsson, D. Västfjäll, and M. Kleiner. Ecological acoustics and the multi-modal perception of rooms: real and unreal experiences of auditory-visual virtual environments. In *Proc. of the Conf. on Auditory Display*, pages 245–249, 2001.
20. A. Lehmann, C. Geiger, B. Wöldecke, and S. J. Poster: Design and Evaluation of 3D Content with Wind Output. In *IEEE Symposium on 3D User Interfaces*, pages 152–152, 2009.
21. M. Luck and R. Aylett. Applying Artificial Intelligence to Virtual Reality: Intelligent Virtual Environments. *Applied Artificial Intelligence*, 14(1):3–32, 2000.
22. T. Moon and G. J. Kim. Design and Evaluation of a Wind Display for Virtual Reality. In *VRST '04 Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pages 122 – 128, 2004.
23. B. Weber, M. Sagardia, T. Hulin, and C. Preusche. Visual, Vibrotactile, and Force Feedback of Collisions in Virtual Environments: Effects on Performance, Mental Workload and Spatial Orientation. In *Virtual Augmented and Mixed Reality. Designing and Developing Augmented and Virtual Environments*, pages 241–250. Springer, 2013.
24. T. Yamada, S. Yokoyama, T. Tanikawa, K. Hirota, and M. Hirose. Wearable Olfactory Display: Using Odor in Outdoor Environment. In *VR*, pages 199–206, 2006.
25. Y. Yanagida, S. Kawato, H. Noma, A. Tomono, and N. Tetsutani. Projection-Based Olfactory Display with Nose Tracking. In *VR*, pages 43–50, 2004.
26. D. Zeltzer. Autonomy, interaction, and presence. *Presence: Teleoper. Virtual Environ.*, 1(1):127–132, Jan. 1992.