

Influence of auditory and visual feedback for perceiving walking over bumps and holes in desktop VR

Luca Turchet* AAU Copenhagen, Denmark	Maud Marchal† INSA/INRIA Rennes, France	Anatole Lécuyer‡ INRIA Rennes, France	Stefania Serafin§ AAU Copenhagen, Denmark	Rolf Nordahl¶ AAU Copenhagen, Denmark
---	---	---	---	---

Abstract

In this paper, we present an experiment whose goal is to investigate the role of sound and vision in the recognition of different surface profiles in a walking scenario. Fifteen subjects participated to two within-subjects experiments where they were asked to interact with a desktop system simulating bumps, holes and flat surfaces by means of audio, visual and audio-visual cues. Results of the first experiment show that participants are able to successfully identify the surface profiles provided through the proposed audio-visual techniques. Results of a second experiment in which conflictual audio-visual stimuli were presented, reveal that for some of the proposed visual effects the visual feedback is dominant on the auditory one, while for the others the role of dominance is inverted.

CR Categories: I.3.5 [Computer Graphics]: Physics Based Modeling—; I.6.5 [Simulation and Modeling]: Model Development— [H.5.5]: Information Interfaces and Presentation (e.g. HCI)—Sound and Music Computing.

Keywords: Bump and hole, pseudo-haptic, design, audio-visual.

1 Introduction

In most current virtual reality setups, users are restricted to walk on a flat surface. Only few locomotion interfaces are able to render uneven grounds, but they have the disadvantages of being costly and cumbersome [Hollerbach et al. 2003; Iwata et al. 2001]. Recently, research has shown that it is possible to simulate the act of walking on unflat surfaces only using visual cues [Marchal et al. 2010]. These results are a development of previous research on pseudo-haptic simulation [Lécuyer et al. 2004]. The main idea of the research described in [Lécuyer et al. 2004] was to investigate whether it was possible to simulate a bump or a hole by only using visual feedback.

In the sound synthesis community, all previous research on walking sounds has focused on flat surfaces [Cook 2002; Fontana and Bresin 2003; Farnell 2007; Miner and Caudell 2005]. Recently, a sound synthesis engine able to simulate footsteps sounds on aggregate and solid surfaces has been presented [Nordahl et al. 2010; Turchet et al. 2010]. Such sound engine is based on physical models synthesis and it works both offline and in realtime. In all the different surfaces simulated, the system energy parameter of the

engine is controlled by a ground reaction force (GRF), i.e., the reaction force supplied by the ground at every step. During the offline working of the engine, such force is estimated from recordings on real footsteps sounds.

In [Serafin et al. 2010], researchers used such engine in order to run a preliminary experiment whose goal was to assess the role of temporal aspects in sonically simulating the act of walking on a bump or a hole. In particular, they investigated whether the timing between heel and toe and the timing between footsteps affected perception of walking on unflat surfaces. Results showed that it is possible to simulate a bump or a hole only by using temporal information.

In this paper, we are interested in understanding whether the addition of sounds to the visual feedback enhances the pseudo-haptic feedback for the simulation of unflat surfaces. Our hypothesis is that sound plays an important role in the recognition of the surface profile and that it can be dominant on the visual modality according to the visual stimulus presented.

2 Visual and auditory simulation

2.1 Visual feedback

The visual techniques used to simulate the act of walking over bumps and holes were the same proposed in [Marchal et al. 2010]: a straightforward modification of the camera’s height (H), a modification of the camera’s navigation velocity (V), a modification of the camera’s orientation (O), and the combination of the three effects (HOV). Our simulations used a known mathematical profile, the Gaussian one, which was used for the simulations of both holes and bumps.

Since the velocity technique was used for the generation of the auditory feedback we briefly recall it here. The velocity effect is based on the variation of the camera velocity. Thus, the camera velocity is decreased when the user is going up and increased when the user is going down. We used a different algorithm for the ascending and descending cases. The algorithm computes the ratio R_{Velocity} applied between the real user velocity and the virtual camera velocity. The camera velocity is then modified following the equation:

$$\text{Velocity}^t = \text{Velocity}^{t-1} \cdot R_{\text{Velocity}}^t \quad (1)$$

- Ascending case:

$$R_{\text{Velocity}}^t = \exp(-R_{\text{AscendingV}} \cdot \alpha^t) \quad (2)$$

where α is the tangent angle of the Gaussian curve and $R_{\text{AscendingV}}$ is a constant.

- Descending case:

This algorithm is designed to give a run up for a while after the bump or at the beginning of the hole. At time t , the ratio is updated regarding the difference between the user height in scene at times $t - 1$ and t :

$$R_{\text{Velocity}}^t = R_{\text{Velocity}}^{t-1} + \Delta_{\text{Height}} \cdot R_{\text{DescendingV}} \quad (3)$$

where the ratio $R_{\text{DescendingV}}$ is a constant. When the subject reaches the end of the descent, his speed is at a maximum.

*e-mail:tur@media.aau.dk

†e-mail:maud.marchal@irisa.fr

‡e-mail:anatole.lecuyer@irisa.fr

§e-mail:sts@media.aau.dk

¶e-mail:rn@media.aau.dk

If he is walking in a hole, then he starts to go up and his speed value will be given by the ascending algorithm. If the subject is on a bump, he will reach the plane ground after the bump. His speed ratio R_{Velocity} will start decreasing at 0.1 unit per second, until another bump/hole is reached or the ratio is back to normal.

2.2 Auditory feedback

As concerns the sounds, the goal was to add an auditory feedback consistent with the images provided at the screen, but since we were interested just in a proof of concept no interaction between the visual and sound engines was created.

The technique adopted to render bumps and holes at auditory level has been the placement of footsteps sounds at different temporal intervals, taking as reference the work of the visual engine set with the velocity effect. In order to create a coherent mapping between the space travelled by the visual rendering along the surface and the temporal distances between steps, we chose to divide the surface profile in equal parts covering the same amount of space, and to place in those points a step (see Figure 1). Subsequently the temporal intervals to reach each point located on the curve have been calculated according to the equations utilized for the velocity effect by the visual engine (see section 2.1). Precisely, such calculations have been performed using the same parameters defining the Gaussian curve (height and sigma), but modifying the values of the R factors regulating the lows for the ascending and descending case because of the necessity of synchronizing properly the two systems.

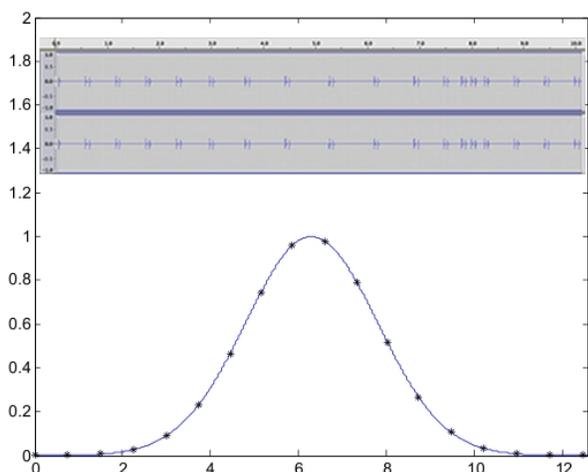


Figure 1: Gaussian profile used for a bump simulation with the indication of the points in which the footstep sounds occur, and stereo waveform of the corresponding sound file.

The input files for the sound synthesis engine (see section 1) resulting from such calculations, have been generated by means of MATLAB placing at different temporal patterns a single footstep sound (the same footstep was used to create the different stimuli presented to the subjects). The use of the same footstep sound was justified by the fact that we did not want other factors, such as changes in amplitude, to affect the results of the experiment. The footstep sound used was a recording of a real footstep on concrete. Such sound was chosen among those available in the Hollywood Edge sound effects library.¹

For the purpose of this experiment, two types of surfaces, gravel and wood, were chosen. The reason for choosing two materials

was to assess whether the surface type affected the quality of the results.

3 Experiment

We conducted two within-subjects experiments using three different kinds of stimuli: visual only, audio only, and audio visual.

1. Experiment 1: recognition of bumps, holes and flat surfaces by means of visual, audio and (coherent) audio-visual stimuli.
2. Experiment 2: recognition of bumps and holes in presence of incoherent audio-visual stimuli.

The images provided during the experiments were generated by the visual engine, while the sounds consisted of footsteps sounds generated by the offline use of the sound synthesis engine.

The goal of the first experiment was to investigate the ability of subjects to recognize the different walking surface profiles they were exposed to. One of our hypotheses was that the recognition would have improved using the bimodal stimuli rather than the single modalities alone.

The goal of the second experiment was to determine the role of dominance of the two modalities involved, in presence of incoherent stimuli. Our hypothesis was that the visual modality would have been dominant on the auditory modality for the H, O and HOV effects, while for the V effect the role of dominance would have been held by the auditory modality.

3.1 Participants

The experiments were performed by 15 participants, 12 men and 3 women, aged between 22 and 29 (mean=26.2, standard deviation=2.1662). All participants had normal or corrected vision and reported normal hearing conditions. All participants were naive with respect to the experimental setup and to the purpose of the experiment. The participants took on average about 40 minutes to complete the two experiments and to fill the questionnaire.

3.2 Task

During experiment 1 participants were given the list of three different surfaces (bump, hole, flat), presented as forced alternate choice, while during experiment 2 the possible choices were only two, i.e., bump or hole, in order to understand which modality was dominant. The task consisted of recognizing to which surface the walk corresponded after the presentation of the stimulus. Each stimulus was presented only one time before giving the answer, and when moving to the next stimulus participants could not change the answer to the previous stimuli.

After both experiments, a preference questionnaire was proposed in which participants had to grade from 1 (low appreciation) to 7 (high appreciation) three further effects (a bump provided at audio, visual and audio-visual level) according to 3 subjective criteria: efficiency of the simulation, realism of the simulation, and global appreciation of the effect. In the questionnaire we did not mention the words audio, visual or audio-visual, we simply called them condition 1, 2 and 3, in order to leave to the participants the task of creating the proper mapping.

At the end of the questionnaire, participants were also given the opportunity to leave an open comment on their experience interacting with the system.

3.3 Setup

All experiments were carried out in an acoustically isolated laboratory where the setup for the experiment was installed. It con-

¹www.hollywoodedge.com/

sisted of a simple graphical user interface with which the participants were asked to interact. The interaction consisted of pressing keys on the keyboard in order to start each trial and to give the corresponding answer.

The virtual environment was a simple corridor with given dimensions (height=3.0m, length=12.6m, width=2.0m). There was a part in the center of the corridor where the height could be modified during the experiments: the user could experience either on a bump, a hole or a plane. To symbolize this variable part of the corridor, a transparent cube was represented on the ground with a height of 0.5m and a surface of 6.6mx2m, as illustrated in Figure 2. The variable height of the ground was not visible in order to exclude visual cues from the scene.

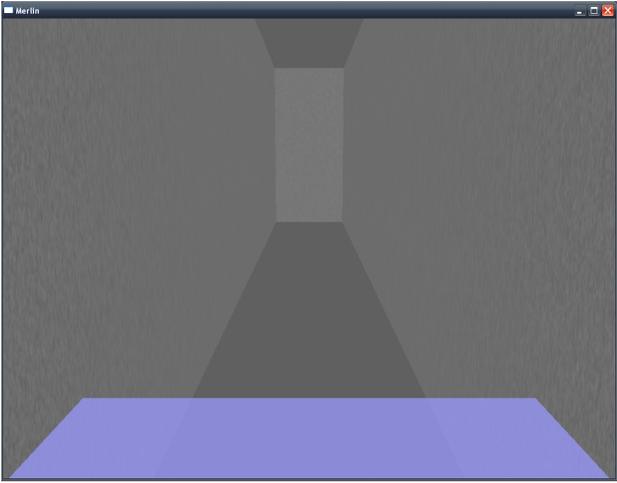


Figure 2: A screenshot at the beginning of a trial.

The computer utilized for the experiment had clock speed equal to 3.40 GHz, 2 GB of memory, it run the Windows XP operative system, and it was equipped with the SigmaTel Sound Card. Sounds were provided to the user by means of a set of headphones.² During the experiments subjects were sitting on a chair, listening to the sounds through headphones and interacting with the interface.

3.4 Experimental plan

Participants were exposed to 105 trials in experiment 1, where 15 stimuli were repeated seven times. The stimuli consisted of bumps, holes and flat surfaces rendered by means of sounds (footsteps on two types of materials, gravel and wood), images (using the HOV technique), and the combinations of images and sounds.

During experiment 2 participants were exposed to 48 trials, where 8 audio-visual incoherent stimuli were repeated six times. In detail, only bumps and holes were rendered: when a bump was presented at visual level a hole was provided at auditory level, and vice versa. The techniques used to render the bumps and holes at visual level were Height, Orientation, Velocity and HOV, while the sounds provided were the same presented in experiment 1 using wood as material.

The temporal durations of the trials were 10.5 seconds for bumps, 9 seconds for holes, and 9.5 seconds for the flat surfaces.

4 Results

For each participant, the percentage of correct answers was estimated for the different experimental conditions. Results are shown

²Sennheiser HD 600, <http://www.sennheiser.com>

in tables 1, 2.

Concerning experiment 1, the first noticeable element emerging from results is the high percentages of correct answers for all the three modalities. The presentation of flat surfaces led to better performances rather than bumps and holes. The addition of audio stimuli did not produce higher percentages for the recognition of the surfaces rather than the visual only modality (which is already very high), and an in depth analysis performed by means of the Kruskal-Wallis test revealed that all the differences are not statistically significant.

Moreover no substantial differences have been found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel.

		Bump	Hole	Flat	%Correct answers
Visual	Bump	98	7	0	93.33
	Hole	7	98	0	93.33
	Flat	0	0	105	100
Audio 1	Bump	96	8	1	91.42
	Hole	8	97	0	92.38
	Flat	6	0	99	94.28
Audio 2	Bump	96	5	4	91.42
	Hole	4	98	3	93.33
	Flat	3	1	101	96.19
Audio-Visual 1	Bump	98	7	0	93.33
	Hole	8	97	0	92.38
	Flat	0	0	105	100
Audio-Visual 2	Bump	99	5	1	94.28
	Hole	8	97	0	92.38
	Flat	1	0	104	99.04

Table 1: Confusion matrix illustrating the results for experiment 1. Legend: Visual: the HOV effect; Audio 1: sounds of footsteps on gravel; Audio 2: sounds of footsteps on wood; Audio-Visual 1: sounds of footsteps on gravel plus the HOV effect; Audio-Visual 2: sounds of footsteps on wood plus the HOV effect.

As regards experiment 2, results show clearly that in presence of audio-visual conflicts audio is dominated by vision when H and O effects are presented. Conversely, vision is dominated by audio when V and HOV effects are presented. In particular the highest role of dominance has been found for audio stimuli respect visual stimuli provided by means of the Velocity effect.

	% Visual dominance	% Audio dominance
Bump H	85.55	14.45
Hole H	81.11	18.89
Bump O	72.22	27.78
Hole O	71.11	28.89
Bump V	8.88	91.12
Hole V	3.33	96.67
Bump HOV	30	70
Hole HOV	26.66	73.34

Table 2: Table illustrating results for experiment 2. Legend: H: the Height effect; O: the Orientation effect; V: the Velocity effect; HOV: the combination of the three effects.

Table 3 shows the results concerning the grades obtained by the three different modalities for each of the subjective criteria. Results clearly show better evaluations in the audio-visual modality (condition 3) rather than the audio (condition 1) and visual (condition 2) modalities alone. In particular an in depth analysis by means of t-test reveals that such differences are statistically significant for the efficiency criterium for condition 3 respect to condi-

tion 1 ($p=0.01644$), for the realism criterium for condition 3 respect condition 2 ($p=0.02156$), and for the appreciation criterium for condition 3 respect to condition 1 ($p=0.02047$) and 2 ($p=0.006438$). No significant differences in the criteria investigated have been found between condition 1 and 2.

	Audio		Visual		Audio-Visual	
	μ	σ	μ	σ	μ	σ
Efficiency	4.5333	1.7265	5.3333	1.3452	5.8667	0.9904
Realism	4.5333	1.5523	4.1333	1.5976	5.5333	1.5523
Appreciation	4.4667	1.7674	4.4667	1.3020	5.8667	1.3020

Table 3: Table illustrating the answers of the questionnaire (average scores from a seven-point Likert scale, and relative standard deviation).

5 General discussion

Results of experiment 1 show that the simulated surface profiles were recognized with high percentages in all the three modalities, and this is a confirm of the success of the techniques proposed in [Marchal et al. 2010] and [Serafin et al. 2010]. Although the addition of audio stimuli did not produce significant higher percentages rather than the visual only modality, the subjective questionnaire revealed, for the criteria investigated, a clear preference of the bimodal stimuli respect to the stimuli presented in the single modalities. In addition no significant differences were found between stimuli involving footsteps sounds on wood and those involving footsteps sounds on gravel and this is a further confirm of the results illustrated in [Serafin et al. 2010].

In case of conflictual audio-visual stimuli the relationships of dominance between modalities turned out to be very clear, according to the visual effect adopted. Our hypotheses were confirmed for all the effects with the exception of HOV for which, surprisingly, the auditory modality is dominant on the visual one. Presumably an explanation lies in the fact that the V effect was present in the HOV condition, and it is well known in the field of multimodal perception that the temporal resolution of the auditory system is better than the one of the visual system [Stein and Meredith 1993]. Our results confirm such behaviour of the perceptive system.

What emerges from an in depth analysis of results of experiment 2 is that participants on average were consistent in their choices: indeed in case of conflicts, they tended to choose always the same modality as guide (according to the visual effect presented). Some of the subjects reported in their comments that in experiment 2 they noticed conflicts between what they heard and what they saw at screen. In particular one of them explicitly reported that in case of conflicts he was driven in his decision by the visual stimulus since not seeing the feet of the person walking in the animation, he imagined that the person would have been able to place the steps in an arbitrary way.

6 Conclusion and future work

In this paper, we described an experiment whose goal is to assess whether the combination of auditory and visual information affects the perception of walking on a hole or a bump using a desktop system. Results show that the surface profiles simulated by means of audio, visual and audio-visual stimuli were accurately recognized in all the three modalities, but the subjective questionnaire revealed a clear preference of the bimodal stimuli.

As concerns the auditory stimuli, they were created only varying temporal parameters of footsteps, such as the distance between steps. Other important aspects, such as amplitude and spectral information was not modified. These parameters are object of future investigations. The results presented in this paper have interesting

applications in the field of navigations in virtual environments and computer games, where more realistic auditory feedback can enhance the realism of the simulated experience. In future works we plan to integrate the visual system with the sound synthesis engine in order to have an audio-visual interactive simulation of the proposed surface profiles, both with a desktop system and an immersive VR configuration with HMD.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme under FET-Open grant agreement 222107 NIW - Natural Interactive Walking.³ This publication is supported by COST Action IC0601.

References

- COOK, P. 2002. Modeling Bill's Gait: Analysis and Parametric Synthesis of Walking Sounds. *Proceedings of the AES 22nd International Conference on Virtual, Synthetic, and Entertainment Audio*, 73–78.
- FARNELL, A. 2007. Marching onwards: procedural synthetic footsteps for video games and animation. *Proceedings of the Pure Data Convention*.
- FONTANA, F., AND BRESIN, R. 2003. Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. *Proc. Colloquium on Musical Informatics*, 109–114.
- HOLLERBACH, J., CHECCACCI, D., NOMA, H., YANAGIDA, Y., AND TETSUTANI, N. 2003. Simulating side slopes on locomotion interfaces using torso forces. In *Haptic Symposium*, Cite-seer, 91–98.
- IWATA, H., YANO, H., AND NAKAIZUMI, F. 2001. Gait master: A versatile locomotion interface for uneven virtual terrain. *vr*, 131.
- LÉCUYER, A., BURKHARDT, J., AND ETIENNE, L. 2004. Feeling bumps and holes without a haptic interface: the perception of pseudo-haptic textures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 246.
- MARCHAL, M., LÉCUYER, A., CIRIO, G., BONNET, L., AND EMILY, M. 2010. Walking Up and Down in Immersive Virtual Worlds: Novel Interactive Techniques Based on Visual Feedback. In *Proceedings of IEEE Symposium on 3D User Interface*.
- MINER, N., AND CAUDELL, T. 2005. Using wavelets to synthesize stochastic-based sounds for immersive virtual environments. *ACM Transactions on Applied Perception* 2, 4, 521–528.
- NORDAHL, R., SERAFIN, S., AND TURCHET, L. 2010. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *Proceedings of IEEE Virtual Reality*.
- SERAFIN, S., TURCHET, L., AND NORDAHL, R. 2010. Do you hear a bump or a hole? An experiment on temporal aspects in footsteps recognition. In *Proceedings of Digital Audio Effects Conference*.
- STEIN, B. E., AND MEREDITH, M. A. 1993. *The Merging of the Senses*, 1st ed. MIT Press.
- TURCHET, L., SERAFIN, S., DIMITROV, S., AND NORDAHL, R. 2010. Physically based sound synthesis and control of footsteps sounds. In *Proceedings of Digital Audio Effects Conference*.

³www.niwproject.eu