Multimedia Tools and Applications (2020) 79:3311–3331
https://doi.org/10.1007/s11042-019-7331-z

Crossmodal perception in virtual reality

S. Malpica[1] · A. Serrano[1] · M. Allue[1] · M. G. Bedia[1] · B. Masia[1]

Received: 1 December 2017 / Revised: 4 November 2018 / Accepted: 4 February 2019 /
Published online: 26 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
With the proliferation of low-cost, consumer level, head-mounted displays (HMDs) we are witnessing a reappearance of virtual reality. However, there are still important stumbling blocks that hinder the achievable visual quality of the results. Knowledge of human perception in virtual environments can help overcome these limitations. In this work, within the much-studied area of perception in virtual environments, we look into the less explored area of crossmodal perception, that is, the interaction of different senses when perceiving the environment. In particular, we look at the influence of sound on visual perception in a virtual reality scenario. First, we assert the existence of a crossmodal visuo-auditory effect in a VR scenario through two experiments, and find that, similar to what has been reported in conventional displays, our visual perception is affected by auditory stimuli in a VR setup. The crossmodal effect in VR is, however, lower than that present in a conventional display counterpart. Having asserted the effect, a third experiment looks at visuo-auditory crossmodality in the context of material appearance perception. We test different rendering qualities, together with the presence of sound, for a series of materials. The goal of the third experiment is twofold: testing whether known interactions in traditional displays hold in VR, and finding insights that can have practical applications in VR content generation (e.g., by reducing rendering costs).

Keywords 3D graphics and realism · Virtual reality · Crossmodal perception · Material perception

1 Introduction

During the last years, we are witnessing a reappearance of virtual reality (VR). New applications are developed every day, going far beyond entertainment and gaming, and including

✉ S. Malpica
smalpica@unizar.es

A. Serrano
anase@unizar.es

[1] Universidad de Zaragoza, I3A, Zaragoza, Spain

advertising [58], virtual tourism [19], prototyping [51], medicine [27], scientific visualization [26], or education [52], to name a few. There are still important stumbling blocks that hinder the development of more applications and reduce the visual quality of the results; examples include limited spatial resolution, chromatic aberrations, tracking issues, limited processing capability leading to lag, subsequent motion sickness, or content generation [62]. A relevant area which has received quite some interest but remains full of unanswered questions and open problems is how our perception is modified or altered when immersed in a virtual environment. Knowledge of human perception in virtual environments can help overcome the aforementioned current limitations. In the past, perception has been leveraged in many computer graphics-related areas such as rendering [41], material modeling and acquisition [57], or display [31]; a good review of applied perception in graphics can be found in the course by McNamara and colleagues [34].

In this paper, within the much-studied area of perception in virtual environments, we chose to look into the less explored area of crossmodal perception in HMDs, that is, the interaction of different senses when perceiving a virtual environment through a headset. HMDs are different from traditional displays in that they provide a more realistic and immersive experience, as well as introducing additional degrees of freedom (the user now controls the camera), spatialized sound, increased field of view, and more visual cues (e.g., motion parallax). Specifically, we looked at the *influence of sound on visual perception in a virtual reality scenario*.

Crossmodal perception, and in particular the interaction between visual and auditory stimuli, has been studied before in real scenes and on conventional displays. The crossmodal effect between these two sensory inputs has been assessed and documented in different works [49, 53, 56], which state, among other conclusions, that the presence of sound can alter the visual perception.

This paper is an extension of our previous work [1], where we replicated a well-known crossmodal perception experiment [49]. We found that crossmodal interaction was indeed present in VR, and that its effects persisted even in the presence of more complex stimuli. These experiments are described in Section 3. We further extend this initial work by, once we have asserted the presence of a visual-auditory crossmodal effect, analyzing the effects of sound in the visual perception of materials, in order to find practical applications for VR. This new experiment is described in Section 4 and constitutes the main contribution of the present work. Generating content for VR headsets requires rendering complex scenes in real time, at high resolution and at, ideally, at least 60 fps, which comes at a large computational cost, specially if the aim is to obtain a realistic appearance. Different works have investigated how visual perception is affected in VR, partly with the aim of reducing this rendering cost [5, 38]; conversely, other works have analyzed the effect of sound in material perception, but not in an immersive environment [6, 30]. In this work we have taken the first steps towards analyzing the influence of a visual-auditory effect on material perception in VR (Section 4), providing insights that can be used in the future to reduce computational costs, or improve the quality when rendering complex appearances. In particular, the research questions we investigate in this paper are the following:

- Manifestation of the crossmodal effect in VR environments with increasing complexity.
- Influence of crossmodal interactions in material perception in immersible VR environments.

2 Related work

2.1 Crossmodal interactions

Nowadays, a popular view in neuroscience holds that the human brain is structured into a large number of areas in which information is highly separated [13]. This perspective assumes that mental processes such as perception -but also emotions or intentions- are limited to neural processes inside the brain and confined to particular areas. In the same way, it is often assumed that inputs coming from different perceptual modalities are processed in the brain independently and in different brain regions [47].

However, the feeling of unified perceptions of objects and events is an ordinary experience. It suggests that information from different sensory modalities must somehow be bounded together in the brain in order to represent a single object or event [39]. This assumption is cornerstone in most recent alternative neurodynamic views (as for example, bodily and sensorimotor approaches) in order to propose solid explanatory alternatives to traditional and internalist perspectives of brain organization [60, 65]. In these alternative approaches, multisensory perception processes and different sensory modalities are understood as closely related through flexible integrations of the dynamics of brain by means of the emergence of transient assemblies of neural synchronization when a unified perception arises [28]. Thus, a complete understanding of perception would require to know the different ways in which one sense modality is able to impact another, creating crossmodal illusions [53]. If we understood the interactions among perceptual modalities, we could shed light on the true mechanisms that support perceptual processes.

It is worth highlighting that, until very recently, the neural principles of multisensory integration and crossmodal illusions have remained unexplored. The modular view of the brain has been so strong with respect to the visual stimuli that it has been considered in the past as independent from other modalities. However, in recent years the interest in understanding crossmodal phenomena and illusions has increased substantially [56]. Some of the deeper studies are those involved in alterations between auditory and visual senses. The best known example amongst these is the *ventriloquism effect* which refers to the perception of speech sounds as coming from a different direction than its real source, forced by the influence of visual stimuli from an apparent speaker [20]. Another well-known example is *the McGurk effect* [33] where lip movements of a subject are integrated with different but similar speech sounds.

In this work we first investigate the effect of auditory spatial information on the perception of moving visual stimuli. We focus on the case of *motion perception* because previous studies have suggested that there should exist common neural substrates between the visual and auditory modalities [54]. The work is inspired in a classical experiment developed in the 90s where sound influenced ambiguous visual motion perception as proposed by [49]. The authors found that when two objects -in a virtual and ambiguous simulation- moving along crossing trajectories reached the same position and then moved apart, they would be sometimes perceived by participants in the study as if moving on a constant trajectory and crossing. However, in other cases, participants reported that the objects reversed their direction as they would do following a collision. Sekuler et al. [49] discovered that this ambiguity was solved when a sound emerged at the moment of coincidence of the objects, as this would show that the sensory information perceived in one modality (audition) could modulate the perception of events occurring in another modality (visual motion perception). Although the crossmodal effect reported by Sekuler and collaborators was accused of simply showing a cognitive limit rather than a genuine crossmodal perceptual effect, the authors opened the debate regarding the perceptual nature of many other crossmodal illusions between visual and auditory stimuli. For instance, the effect known as sound-induced flash illusion [54, 55] showed how the perception of a brief visual stimuli could be altered by concurrent brief sounds. When a single flash of light was showed together with two beeps, the perception changed from a single flash to two flashes. The reverse illusion could also occur when two flashes were accompanied by a single beep (which would be then perceived as a single flash). Auditive clues have also shown to affect object recognition when added to visual information as Suied et al. [59] show in their work.

Regarding crossmodal interactions in VR environments, several works have used a crossmodal effect to modify the user's visual perception. For example, Nilsson et al. [36] explore redirection techniques for virtual walking with audiovisual stimuli and Maculewicz et al. [29] explore the influence of sound in walking interactions. Crossmodal interactions with binaural sound have also been used in VR to reduce the time to complete a given search task [22] and to compensate for distance compression [11]. Binaural sound has been used in AR to enhance the presence of a virtual object by producing virtual sound effects [3]. Also, moving sounds have been used to induce the sensation of circular [42] and linear [61] vection in VR. Visuo-haptic interactions have also been used in redirected walking techniques in Matsumoto et al.'s "unlimited corridor" experiment [32]. Lately, crossmodal visuo-haptic applications are gaining more attention as haptic devices get more accurate and reliable, such is the case of virtual body ownership illusions [25]. Crossmodal interactions can also play a role in intangible cultural heritage (ICH) modelling [10, 40], for example the project i-Treasures [9] relies on sensorimotor learning through an interactive 3D environment to contribute to the transmission of cultural expression.

2.2 Crossmodal material perception

The majority of works in material perception deal with the unimodal case of visual-only material representations, trying to understand how humans perceive the reflections of light in material surfaces. The influence of shape in material perception is studied by Vangorp et al. [64]. In addition, Vangorp [63] also studies visual material perception in realistic computer graphics. Material classification in visual and semantic domains was investigated by Fleming et al. [12]. Other works in material perception study sound-only representations. For example, Klatzky et al. [24] analyze the relation between material perception and contact sounds. Avanzini and Rocchesso [2] and Giordano and McAdams [16] use contact sounds to classify different materials. Grassi [17] analyzes the influence of contact sounds in the perceived size of an object. Here we focus, however, in the multimodal case.

Several works assert that material perception in humans is multimodal by nature. The use of different modalities interplays in an unknown way to give us more information. Among them, the most used combination in computer science is the association of vision and sound, of which we include here some examples. Mishra et al. [35] show the influence of audio in color perception. Taking one step further, Fujisaki et al. [14] studied the audiovisual information integration in the perception of materials. Later, they also studied [15] if a common subjective classification could be found in the perceived properties of wood regarding audio, visual and touch information. Grelaud et al. [18] take advantage of crossmodal perception to improve audiovisual rendering for games, showing that the object's impact sound and its quality affects the perceived visual quality of the material. Following a similar reasoning, Waltl et al. [66] improve the immersive sensation of a virtual environment through different sensory effects. Finally, Rojas et al. use different sound cues to modify the perceived visual quality on various works [43–46].

The two closest works to our own are the work of Bonneel et al. [6], and the work of Martin et al. [30]. Bonneel et al. [6] combined and analyzed levels of detail in audiovisual rendering. They designed a study in which subjects compared the similarity to a reference of sequences rendered with different auditory and visual levels of detail. The results of their study show that high quality sound improves the perceived similarity of a lower-quality visual approximation to the reference. Martin et al. [30] performed two experiments. In the first experiment the users were presented a full collection of materials in different presentations (visual, auditory and audiovisual) and were asked to rate different attributes. As a point of reference, subjects also performed all ratings on physical material samples. A key result of the experiment was that auditory cues strongly benefit the perception of certain qualities that are of a tactile nature (like hard/soft, rough/smooth). A follow-up experiment demonstrated that, to a certain extent, audio cues can also be transferred to other materials, exaggerating or attenuating some of their perceived qualities. Both works hint at the unified and integrated nature of perceptual constructs, and how no particular modality of sensorial perception can be characterized entirely in isolation from the others. In this work we look at these interactions in a virtual environment seen through a HMD; it is the first time, to our knowledge, that these experiments are performed within a VR scenario.

3 Crossmodal interaction

We have first performed two experiments in order to determine how much an immersive environment interferes with the crossmodal interaction between the visual and auditive systems. Our experiments are based in the work of Sekuler et al. [49], where they explore the perceptual consequences of sound altering visual motion perception. In their experiments, they showed two identical disks that moved steadily towards each other, coincided, and then continued in the same direction. This scenario is consistent with two different interpretations: either the two spheres did not collide and continued in their original directions (they *streamed*), or they collided and *bounced*, changing their traveling direction. The goal of the experiment is to analyze whether a sound at the moment of the impact can affect the interpretation of the scenario.

We build upon Sekuler et al.'s work, and extend their experiment to virtual reality, aiming to explore the consequences on crossmodal interactions of introducing the user inside a more realistic and complex environment presented with a *head mounted display* (HMD).

3.1 Experiment 1

Goal We first reproduce the experiment described in Sekuler et al.'s work both in a regular screen and in a HMD (*Oculus Rift DK2*). The goal of this experiment was to test whether the effect of sound altering visual motion perception as reported in the experiments carried out by Sekuler et al. is also observed when reproduced in a virtual environment with an HMD.

Stimuli The visual stimuli were rendered with *Unity*. They consisted of two spheres with radius *0.5 degrees*, placed over a white plane. The material of the spheres was brown and very diffuse to avoid introducing additional visual cues. The two spheres were initially separated by a distance of *4.2 degrees*, and moved towards each other at a constant speed of *6 degrees per second*. After they coincided, they continued moving without changing their original direction. We show in Fig. 1 the initial layout of the scene. In this scenario we presented three different visual conditions: the spheres moved continuously, paused one frame

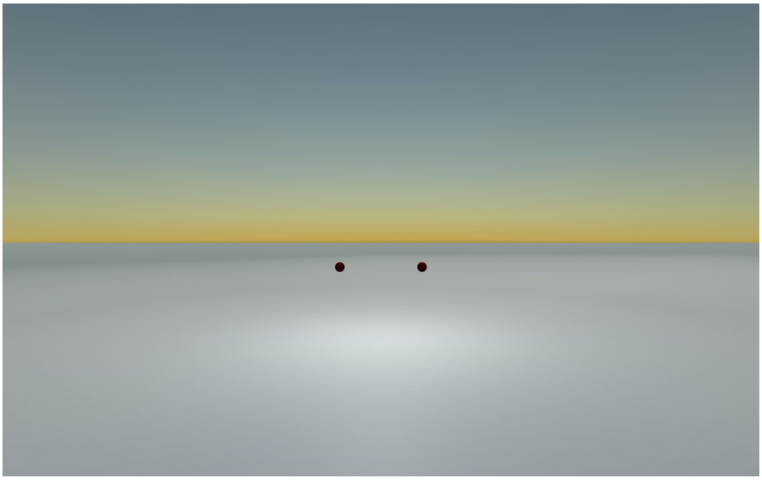


Fig. 1 Initial layout of the scene for Experiment 1

at the point of their coincidence, or paused two frames at the point of their coincidence.[1] These three visual conditions were presented together with one of the four following auditory conditions: no sound, accompanied by a brief click sound (frequency of *2000 Hz*, duration of *3 milliseconds*) triggered *150 milliseconds* before or after the coincidence, or accompanied by a brief click sound at the point of coincidence.

Participants Thirteen participants took part in the experiment (three female, ten male), with ages ranging from 18 to 28 years. All the participants volunteered to perform our experiments, and they were not aware of the purpose of each experiment. They were requested to fill a questionnaire about visual health, and we conducted a stereoscopic vision test to discard those participants with defective depth perception. They all had normal or corrected-to-normal vision.

Procedure During the experiment we presented a total of twelve different conditions to each participant, three visual (continuous movement, pause one or two frames at the coincidence) and four auditory (no sound, sound at, before, or after the coincidence). Each of these conditions was presented *ten* times, making a total of 120 trials that appeared in a random order. We performed two blocks of the same experiment ordered randomly: one displayed on a regular screen (*Acer AL2216W TFT 22"*), and the other one displayed on an HMD (*Oculus Rift DK2*).

Before the HMD block, the lenses of the *Oculus Rift DK2* were adjusted to the participant eyes. We additionally introduced a training session before this block, where we showed two spheres at different depths and the participant had to choose which one was closer. We presented ten trials of the training with spheres at random depths. With this training the user gets used to the device, setup, and answering procedure.

[1] The original experiment [49] reported frames in a regular analog screen whose typical framerate is 25 frames per second. Since the framerate of our screen and the HMD (*Oculus Rift*) were very different, we adjusted the pause to last 1/25 seconds. Therefore, throughout the paper the terminology is as follows: one frame is equivalent to 1/25 seconds, and two frames are equivalent to 2/25 seconds.

Table 1 Results (*F-test* and *significance*) of the analysis of the data with repeated measures ANOVA for Experiment 1

	F	Sig.
Sound vs percent. bounce	83.664	0.000
Pause vs percent. bounce	63.528	0.000
Display vs percent. bounce	13.176	0.000

We test the influence of three factors in the perceived percentages of bounce responses

We guided the participants through the test by showing several slides with descriptions of each phase of the experiment. After each trial, a slide was displayed with the question *“Did the spheres bounce or stream?”*, and a visual aid indicating the participant to answer with a mouse click (right or left).

Analysis and results We use repeated measures ANOVA to test the influence of each of the conditions independently in the observed responses. For every participant, we take into account the answer (*bounce* or *stream*) in each of the ten trials. We need the repeated measures scheme because we measure the same independent variables (e.g., frames paused) under different conditions performed by the same subjects. We fix a significance value (p-value) of 0.05 in all the tests, and in those cases in which results from Mauchly’s test of sphericity indicate that variances and covariances are not uniform, we report the results with the corresponding correction applied to the degrees of freedom (Greenhouse Geisser correction [7]). Prior to the analysis, we perform outlier rejection as detailed in the Appendix. We have three factors or variables of infuence: (i) the overall influence of the display (2D scene presented on a *screen*, or 3D environment presented on an *HMD*); (ii) the influence of the *sound* when the spheres collide; and (iii) the influence of the length of the *pause* at the point of coincidence between the spheres. Results are presented in Table 1.

We can conclude that all three factors have a significant effect in the percentage of bounce responses, since all the p-values are below 0.05. We show in Fig. 2 the mean percentages of bounce responses for the tested factors (error bars represent the standard error of the mean). We observe that the percentage of bounce responses decreases when using the HMD display. However, the main findings of Sekuler et al.’s work hold: a

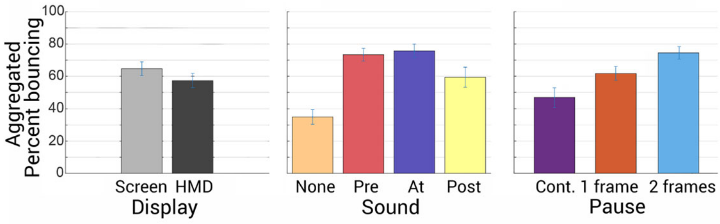


Fig. 2 Aggregated percentages of bounce responses and corresponding error bars (standard error of the mean) for the Experiment 1. From left to right: Percentages for two display conditions (screen or HMD), percentages for four auditory conditions (no sound, sound at, before, or after the moment of coincidence of the spheres), and percentages for three visual conditions (continuous movement, pause one, or two frames at the point of coincidence of the spheres)

sound at the moment of coincidence, and a pause of two frames at the point of coincidence promote the perception of bouncing. We believe that the decrease in perceived bouncing in the tests with the HMD comes from the increase in the amount of visual cues due to the stereoscopic view. Sound promotes perception of bouncing when compared with the absence of sound; however, it has significantly less effect when reproduced after the point of coincidence. Still, there is a high tolerance for asynchrony between the sound and the visual input: even when the sound is delayed, the percentage of bounce responses increases. Also, as reported previously by Sekuler and others [4, 48, 49], the overall percentage of bounce responses increases with the duration of the pause.

3.2 Experiment 2

Goal The goal of this experiment was to test whether a more complex scene could influence the crossmodal effect of sound altering visual motion perception. In order to do this, we increase the realism of the scene in three different ways (we term them three *blocks*) while keeping the proportions between distances and speed of the spheres of the original experiment.

Stimuli The visual stimuli were rendered once again with *Unity*. We designed a new scene where the spheres are placed on a white table, inside a furnished room, and with a more realistic illumination. With respect to the first experiment we also increased the size of the spheres to *1 degree* of radius, and the distance between them to *8.4 degrees*, to make them more visible. A screenshot of the initial layout of the scene for the first block of the experiment is shown in Fig. 3, left. For the second block of the experiment, starting from the scene in the first block, we additionally introduced two more visual cues to the spheres. First, we increased the glossiness of the material of the spheres, and second, we slightly lifted the spheres over the table in order to have more visible shadows (see Fig. 3 middle). Finally, for the third block of the experiment, starting from the scene in the first block, we also rotated the plane of the collision between the spheres. We show a screenshot of the initial layout for this block in Fig. 3 right.

Participants Twenty seven participants took part in the experiment (two female, twenty-five male) with ages ranging from 19 to 32 years. As in the previous experiment, participants volunteered and took a questionnaire about visual health, and a stereoscopic depth test to assure that they all had correct depth vision. They all had normal or corrected-to-normal vision.

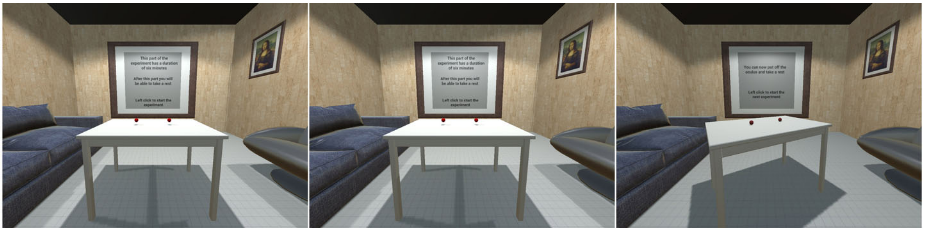

This part of the experiment has a duration of six minutes
After this part you will be able to take a rest
Left click to start the experiment
This part of the experiment has a duration of six minutes
After this part you will be able to take a rest
Left click to start the experiment
You can now put off the oculus and take a rest
Left click to start the next experiment

Fig. 3 Initial layout of the scene for the three different blocks in Experiment 2. Left: increased radius of the spheres (block 1), middle: increased radius of the spheres and additional visual cues (block 2), and right: increased radius of the spheres and rotated plane of the collision (block 3)

Table 2 Results (*F-test* and *significance*) of the analysis of the data with repeated measures ANOVA for Experiment 2

	F	Sig.
Sound vs percent. bounce	124.137	0.000
Pause vs percent. bounce	845.386	0.000
Scene vs percent. bounce	0.022	0.979

We test the influence of three factors in the perceived percentages of bounces

Procedure During the experiment we presented a total of six different conditions, two visual (continuous movement, pause two frames at the coincidence), and three auditory (no sound, *click sound* at, or after the coincidence). Based on the results of the first experiment we removed the visual condition with a pause of one frame because the percentage of bouncing perceived was similar to the one perceived with the pause of two frames, and the auditory condition corresponding to the sound before the coincidence, also because of its similarity with the sound after the coincidence. Each of these conditions was presented *ten* times, making a total of 60 trials that appeared in a random order. All the blocks of the experiment were presented in the *HMD*, and each participant performed three randomly ordered blocks that corresponded to the three scenes described in the *Stimuli* section, totalling 180 trials per subject. Before starting the test, the participants performed the same training described in Experiment 1.

Finally, in this experiment the slides with instructions about the test were shown on a frame on the back of the room striving to preserve as much as possible the realism of the environment.

Analysis and results Again, we wanted to test three factors: the influence of each of the three scenes (three blocks), the influence of the *sound* when the spheres collide, and the influence of the *pause* at the point of coincidence between the spheres. Similarly to Experiment 1, we perform a repeated measures ANOVA; results are presented in Table 2. In Fig. 4 we show the mean percentages of bounce responses for the tested factors, and the associated error bars representing the standard error of the mean.

The analysis with the ANOVA reveals that, as before, there is a significant effect of the *sound*, and the *pause* in the perceived percentage of bounces. However, the p-value for the

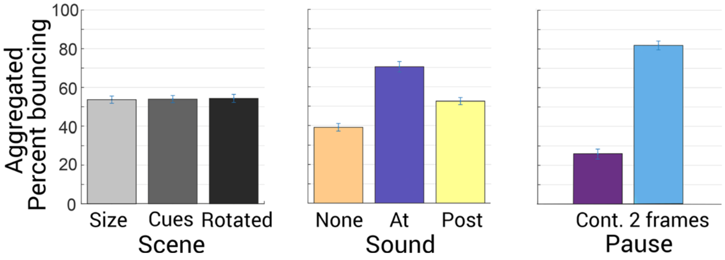


Fig. 4 Aggregated percentages and error bars (standard error of the mean) for the Experiment 2. From left to right: Percentages for the three different scenes or blocks (increase in the size of the spheres, additional visual cues in the spheres, or rotated plane of the movement); percentages for three auditory conditions (no sound, sound at, or after the moment of coincidence of the spheres); and percentages for two visual conditions (continuous movement, or pause two frames at the point of coincidence of the spheres)

test with different scenes is very high, therefore we cannot draw any significant conclusion about the relationship between the three different scenes and the observed percentage of bouncing. When comparing Experiments 1 and 2 we can see that even when increasing the level of realism of the scene, the crossmodal effect of the sound altering the perceived motion still holds, although there is a general shift downwards of the percentage of bounce responses which can be observed by comparing the corresponding percentages of Figs. 2 and 4. This shift downwards is possibly due to the presence of additional cues; however the high p-value of the scene factor, further indicates that there is no significant difference on the effect on crossmodal interaction between the three scenes (blocks) tested (i.e., no cue has proven to be significantly stronger or weaker in the detection of bouncing).

4 Crossmodal material perception

Once we've proven that crossmodal interactions hold in VR we aim to analyze whether these interactions influence material perception. Our goal is twofold: we want to increase once more the stimuli complexity (not just a single sound with equal spheres, but different sounds paired with different visual stimuli), as well as determine if the presence of sound could help improving the immersion experience in VR environments, or even reducing its rendering costs. We have performed an experiment in order to determine how much the perception of material appearance is affected in virtual environments when a crossmodal interaction (visual and auditory stimuli) is presented in comparison with unimodal stimuli (only visual stimuli).

4.1 Experiment 3: description

We use a HMD to determine if the presence of a collision sound can alter the perceived appearance of a material in a virtual environment. We presented different materials and asked the participants to rate a set of perceptual attributes. This attributes included low-level perceptual traits (*soft/hard*, *glossy/matte*, and *rough/smooth*), and high-level descriptors of appearance (*realistic*, *metallic-like*, *plastic-like*, *fabric-like*, and *ceramic-like*). We chose these attributes because they are discriminatory [50], and they have also been used previously for assessing the interactions of sound and visual stimuli [30]. The participants wore isolating headphones (Vic Firth SIH1) during the experiment and they provided answers to the rating questions with a Xbox controller.

Stimuli The visual stimuli were rendered in *Unity* with the default material model (*GGX*). In the visual-only stimuli, they consisted on a sphere placed in front of the camera. In the audiovisual stimuli, the same sphere was presented, but this time with a wooden drumstick hitting it periodically from behind. Figure 5 shows an example of an audiovisual stimulus. The auditory stimuli were recorded mono sounds from the MIT hit sounds dataset [37], that were synchronised to play when the drumstick hits the sphere (in the MIT hit sounds database, it is also a wooden drumstick that is used to produce the sounds). We virtually placed sound sources in the 3D scene, effectively spatializing the mono sound regarding the participant and the sphere's relative position. Note that this is different from using stereo sound tracks, since participants actually perceive a 3D audio effect (i.e., they perceive effects such as head-shadowing). The same sound was always presented for the same material, regardless of its rendering quality. We used four different materials for the sphere. The materials were modeled in Unity and chosen to cover a range of material categories,

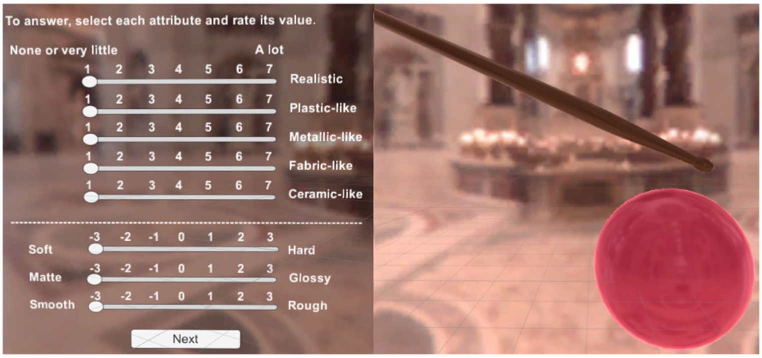

To answer, select each attribute and rate its value.
None or very little
A lot
1 2 3 4 5 6 7 Realistic
1 2 3 4 5 6 7 Plastic-like
1 2 3 4 5 6 7 Metallic-like
1 2 3 4 5 6 7 Fabric-like
1 2 3 4 5 6 7 Ceramic-like
Soft -3 -2 -1 0 1 2 3 Hard
Matte -3 -2 -1 0 1 2 3 Glossy
Smooth -3 -2 -1 0 1 2 3 Rough
Next

Fig. 5 Left: The panel with the attributes that the participants had to rate. With the controller's joystick they could set the rating value and move between the attributes and the "next" button. Right: Presentation of a stimulus in the scene, showing both a sample sphere and the wooden drumstick

which are chosen based on the types of materials present in the MERL database. In particular, we have: *metal*, *fabric*, *plastic*, and a *phenolic* material, (a specular material typically used as coating and to which we associated a ceramic-like sound). Each of the materials was presented twice: one with Unity's light-probe default rendering illumination quality (high resolution, 128 samples) and another with a reduced quality (low resolution, 32 samples). Figure 6 shows these eight combinations. The illumination in all cases was the environment map *St. Peters*, from the Light Probe Image Gallery [8], since real-world illumination, and that environment map in particular, facilitates material discrimination tasks [12].

Participants Thirteen new participants took part in the experiment (two female, eleven male), with ages ranging from 19 to 29 years. They all had normal or corrected-to-normal vision. Similarly to the two previous experiments, all participants took part in a questionnaire and a stereoscopic depth test.

Procedure During the experiment we presented a total of 24 different stimuli to the participants (4 (materials) × 2 (quality levels) × 2 (modalities) + 3 (control materials) × 2 (modalities) + 2 (training stimuli)). Each of the stimuli was shown once. First, a brief explanation of the procedure and the attributes to be used was made. Then, the participants

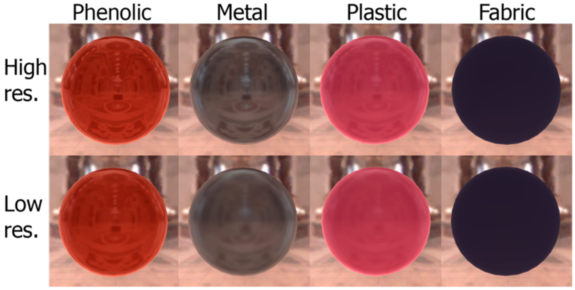


Fig. 6 Each column shows one of the four possible materials used in the experiment. From left to right: Phenolic, metal, plastic, and fabric. Each row shows the material on high resolution (top) and low resolution (bottom)

Table 3 Conditions in our experiment

	Low res.	High res.
Visual only	C_0	C_1
Audiovisual	C_2	C_3

underwent a training with two different stimuli to make sure they understood what they were being asked to do and to learn how the controller worked. This training helped the user to get used to the device, setup, and answering procedure.

The experiment was divided in *two* different blocks, with a total of four conditions (see Table 3): visual-only stimuli ($\{C_0, C_1\}$ for the low and high quality rendering, respectively) and audiovisual stimuli ($\{C_2, C_3\}$ corresponding to the low and high quality rendering, respectively). The order of these two blocks was randomized: Half the participants started with visual-only stimuli and the other half with audiovisual stimuli. Each of the blocks had 11 different stimuli (the four materials were presented in low and high quality, and there were 2 control materials). The presentation order of the stimuli within a block was also randomized, although ensuring that two qualities of the same material did not appear successively. To the left of the stimuli, a panel with the questions of the experiment was presented (Fig. 5, left). Each stimulus, together with the questions, was displayed for 60 seconds. At the end of the 60 seconds, only the questions panel remained. A counter showing the remaining time before the stimulus disappeared was also displayed to make the user aware of the remaining time. Each question pertained to an attribute and a 7-point scale was used to provide the rating answer.

If the participant had rated all the attributes before the 60 seconds had passed, she could move forward to the next stimulus. Between each pair of stimuli, a gray screen with a red cube appeared so that the participants could take a rest if needed before continuing the experiment. The next stimulus appeared when the participants aligned a visual target with the red cube; in this way we also ensured that they were all looking at the same point of the scene when each stimulus is first presented.

The following subsection describes the analysis performed on the gathered rating data, and the insights drawn from it.

4.2 Experiment 3: analysis and results

For the analysis we first performed outlier rejection by using our control materials: subjects were discarded when they did not provide a reasonable answer to the attribute *glossiness* in

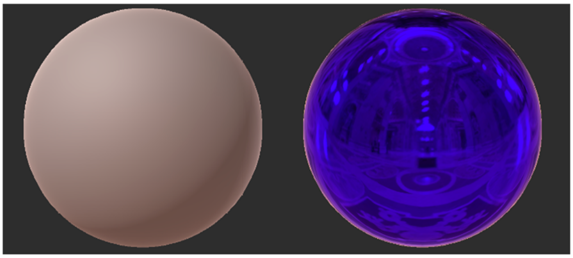


Fig. 7 Control materials used to discard outliers. We discarded a subject if her rating for the attribute *glossiness* was above 2 for a very diffuse material (*left*), or below 6 for a very specular material (*right*), on a 7-point scale

Table 4 Summary of the results (significance) of the analysis of the data with *Wilcoxon Signed-Rank* tests for Experiment 3

	Mat.	Att.	Cond.	Sig.
Influence of resolution	Metallic	Glossy	$C_0 < C_1$	0.041
Influence of sound	Metallic	Plastic	$C_0 > C_2$	0.041
	Metallic	Metallic	$C_0 < C_2$	0.048
	Phenolic	Plastic	$C_0 > C_2$	0.027
			$C_1 > C_3$	0.017
	Phenolic	Ceramic	$C_0 < C_2$	0.027
			$C_1 < C_3$	0.017

We compare the mean value of the *attribute* assigned to the *material* for the specified *conditions*

our control materials (see Fig. 7). We discarded *two* subjects with this procedure, leaving a total of *eleven* users to analyze. We tested our data for normality using the *Shaphiro-Wilk* test, which is well suited for small samples. The ratings for all our attributes did not present a normal distribution ($p < 0.05$), we therefore turned to non-parametric methods to carry out the analysis of our four conditions. For each material and for each attribute we perform pairwise comparisons between the four conditions ($\{C_0, C_1, C_2, C_3\}$) by using the *Wilcoxon Signed-Rank* test. This test is a nonparametric equivalent to the dependent *t-test*, and can be used to investigate changes in ratings when subjects are presented to several conditions. Following Kerr and Pellacini [23] we consider significant p-values below 0.1, which indicates a 90% confidence that the means of the two different conditions differ. Our main insights are summarized in Table 4, and described in detail in the following.

Influence of resolution The resolution of the light-probe plays an important role in the perceived *glossiness* of the material, as can be seen in Fig. 8. This resolution affects the specular reflections (see Fig. 6), therefore it is particularly noticeable in very specular materials, i.e., there is a significant difference between the high and low resolution stimuli for the *metallic* material while for the *fabric* material this difference is barely noticeable. We found a significant interaction in the *metallic* material between the *resolution* and the perceived

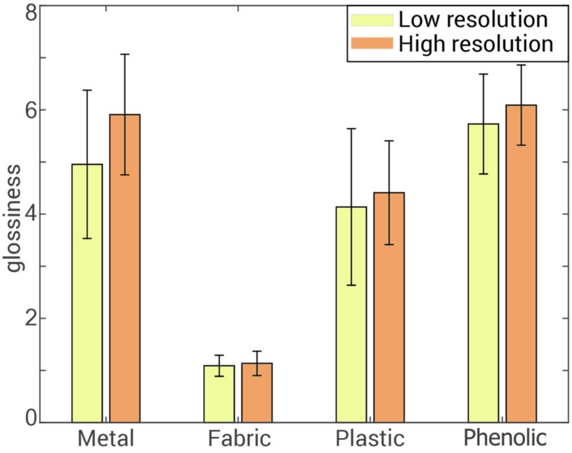


Fig. 8 Mean ratings for the *glossy* attribute when the user is presented with the low resolution (yellow) and the high resolution visual stimuli (orange) for our four materials analysed. Errorbars show ±1 SEM. There is a trend indicating that the perceived glossiness increases in the high resolution stimuli

glossiness ($p = 0.041$ for $\{C_0, C_1\}$). The same trend can be observed for the conditions $\{C_2, C_3\}$. For the other three materials, interestingly, we observe no significant difference in the perception of glossiness regardless of resolution. These findings could be useful to save rendering costs by adjusting the resolution of light-probes according to the material, since the resolution of the light-probe has little effect in the perception of diffuse materials.

Influence of sound We have found several interactions describing a significant effect of the presence of sound in the ratings for the high-level attributes. For the *metallic* material the ratings for the *plastic* attribute are significantly lower when the stimuli is presented together with sound ($p = 0.041$ for $\{C_0, C_2\}$). Conversely, the ratings for the *metallic* attribute are significantly higher ($p = 0.048$ for $\{C_0, C_2\}$). This effect is significant when we compare the low resolution conditions $\{C_0, C_2\}$, but not when we compare the high resolution conditions $\{C_1, C_3\}$. We believe this can be due to the high resolution visual stimuli better conveying the visual traits of the material; this undermines the effect of the auditory stimuli, since the user recognizes the material well enough just with the visual stimuli. This suggests that the effect of sound in material identification tasks may be more relevant when the visual stimuli has a low quality. For the *phenolic* material the mean of the *plastic* attribute significantly decreases when the user is presented with the multimodal stimuli. In this case,

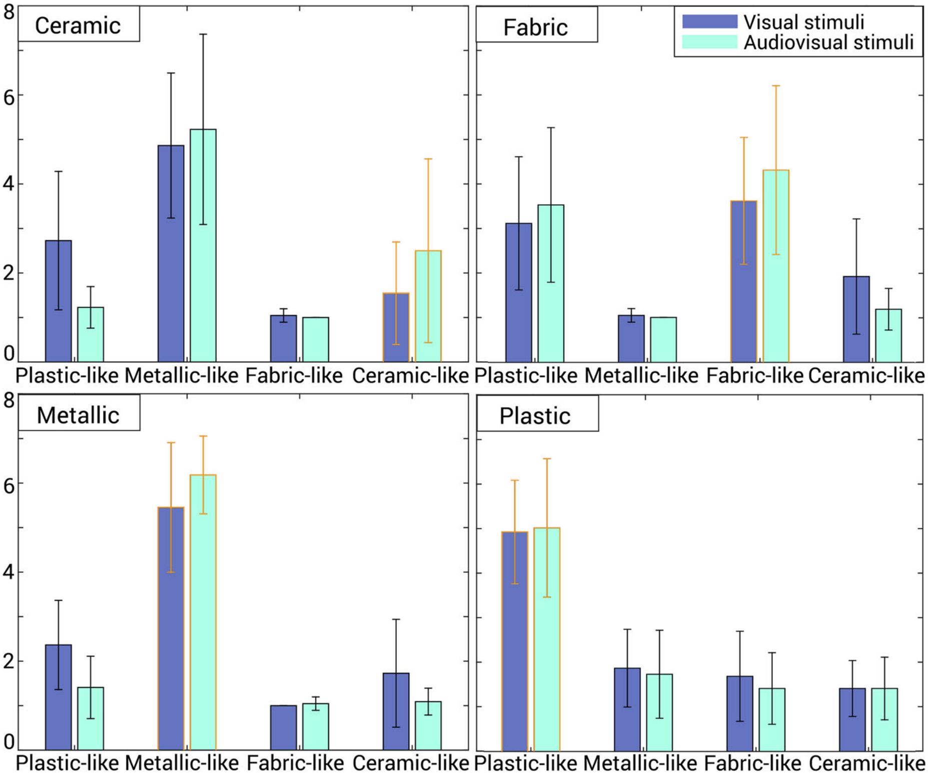


Fig. 9 Mean ratings for the high-level attributes when the user is presented with the visual only stimuli (blue) and the audiovisual stimuli (green) for our four materials. Errorbars show ±1 SEM. For every material, there is an increase in the mean rating of its corresponding attribute (marked by an orange outline) when the visual stimuli is accompanied by sound

the effect is noticeable both for the low resolution ($p = 0.027$ for $\{C_0,\ C_2\}$) and high resolution ($p = 0.017$ for $\{C_1,\ C_3\}$) conditions. For this same material, the mean of the *ceramic* attribute increases ($p = 0.078$ for $\{C_0,\ C_2\}$ and $p = 0.077$ for $\{C_1,\ C_3\}$), which indicates that the sound effectively helps the users identifying the material. We did not find significant interactions for the *fabric* and the *plastic* materials, however, a similar trend can be seen in Fig. 9: for every material there is an increase in the mean rating of its corresponding attribute (bars outlined in orange in Fig. 9) when the user is presented with the audiovisual stimuli. These findings agree with those of Giordano and McAdams [16], which supported that impact sounds were good descriptors for material identification tasks, and they suggest that the sound also benefits material discrimination tasks in VR, particularly when such materials are not easily recognizable only by its visual traits.. Our findings indicate that a high resolution is required for material identification when its representation consists on visual stimuli only, however if additional auditory stimuli are introduced, the resolution could be lowered while keeping the perceived appearance, thus saving rendering costs.

5 Conclusions and discussion

In this paper, we have performed an exploration of crossmodal perception in virtual reality scenarios. We have studied the influence of auditory signals in the perception of visual motion. To do so, we first replicated an existing experiment which demonstrated the existence of a crossmodal interaction between both senses with simple stimuli on a 2D conventional display. We were able to successfully replicate it, obtaining the same trends in the results, and then extended it to virtual reality with a HMD. We found that the same trends hold on a HMD (i.e., the factors explored had the same influence on the crossmodal effect), but that there is a reduction in the crossmodal effect. This reduction essentially means that there is a shift in the results towards a better accuracy of subjects in performing the tasks assigned in the HMD setup. This can be due to the presence of additional cues, in particular depth cues including binocular disparity and possibly motion parallax. A similar conclusion can be drawn in our second experiment: We repeated the first experiment (only on the HMD), with new subjects, and with more complex stimuli (we had three different variations of the initial stimulus) to see whether the effect would still hold with more realistic scenery. We observed a further reduction of the crossmodal effect (subjects were better at detecting the correct behavior of the stimuli), which we hypothesize is due to the presence of additional cues, in this case pictorial cues (shading, perspective, texture).

We then move on to the particular case of material appearance perception, with the aim of laying the foundation for future practical applications. When analyzing crossmodal effects in a VR setup, we have observed that findings previously reported for conventional displays hold: the presence of sound improves material recognition. We have also included two different rendering qualities for the material, and observed two main findings: First, that the influence of the rendering quality on the perception of low-level attributes such as glossiness varies between material categories. Second, that the effect of sound in the recognition of materials is more relevant for the low quality-rendering case than for the high quality one.

In summary, regarding the research questions posed in Section 1, we can conclude that:

- The crossmodal effect holds in VR environments, even when increasing the complexity of scenes.
- Crossmodal interactions influence the perception of material traits in VR environments. More research is necessary to be able to quantify this effect and further understand it.

As in all studies of similar nature, some of our findings may not generalize to conditions outside our study. We have focused on simple sounds and scenes with a controlled increase of complexity. This allows us to isolate the effects of each condition, and perform a systematic analysis. We believe these are just a few steps in the exploration of crossmodal perception in virtual reality. In the future, we would like to expand these experiments by including other potentially influencing factors or effects, and by further increasing the complexity of the stimuli. An interesting avenue for future research would be to use different sound types and qualities in addition to the rendering qualities. In the area of material perception, we hope this work serves as the foundation for future explorations. Here we have employed representative materials of four main categories, future works should further delve into the problem, analyzing a larger variety of materials, especially among specular ones where there is more to be gained from exploitation of this crossmodal interaction. This could result in the development of cuantitative prediction models to enable further practical applications of crossmodal perception in VR environments.

Appendix

Data processing in Experiments 1 and 2 We first processed the collected data by rejecting those users with stereo vision problems. In order to do this, we discarded a user if during the training the percentage of successful answers was equal or under 70%. We further processed the data by rejecting outliers. To do this, we first calculated for each participant and for each of the twelve conditions the percentage of *bouncing* answers over the ten trials. Then we used the first and third quartiles (Q_1 and Q_3), and the interquartile difference (Q_d) to find outliers for each condition [21]. We discarded a condition if it fulfilled any of the following inequalities:

$$\begin{array}{c} condition < (Q_1 - K_d * Q_d) \\ condition > (Q_3 + K_d * Q_d) \end{array} \qquad (1)$$

with $Q_d = Q_3 - Q_1$ and $K_d = 1.5$. Additionally, if a participant was marked as an outlier for more than one condition, all the answers of the participant were discarded.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Allue M, Serrano A, Bedia MG, Masia B (2016) Crossmodal perception in immersive environments. In: Spanish computer graphics conference (CEIG)
2. Avanzini F, Rocchesso D (2001) Controlling material properties in physical models of sounding objects. In: ICMC
3. Baughman AK, McCrory NA, Pandey D, Pandey R Augmented reality enabled response modification, Feb. 13 2018. US Patent 9,891,884
4. Bertenthal BI, Banton T, Bradbury A (1993) Directional bias in the perception of translating patterns. Perception 22(2):193–207
5. Billger M, d'Elia S Color appearance in virtual reality: a comparison between a full-scale room and a virtual reality simulation. In: 9th Congress of the international color association (2002), International Society for Optics and Photonics, pp 122–126
6. Bonneel N, Suied C, Viaud-Delmon I, Drettakis G (2010) Bimodal perception of audio-visual material properties for virtual environments. ACM Trans Appl Percept 7(1):1–16

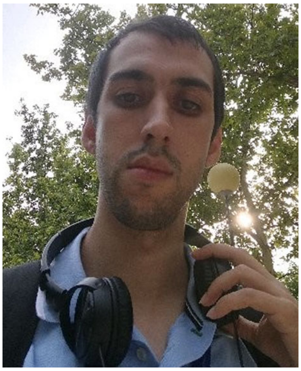
7. Cunningham D, Wallraven C (2011) Experimental design: from user studies to psychophysics, 1st edn. A. K Peters, Ltd., Natick
8. Debevec P (1998) Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Proceedings of the 25th annual conference on computer graphics and interactive techniques. ACM, pp 189–198
9. Dimitropoulos K, Manitsaris S, Tsalakanidou F, Nikolopoulos S, Denby B, Al Kork S, Crevier-Buchman L, Pillot-Loiseau C, Adda-Decker M, Dupont S et al (2014) Capturing the intangible an introduction to the i-treasures project. In: 2014 International conference on computer vision theory and applications (VISAPP), vol 2. IEEE, pp 773–781
10. Doulamis N, Doulamis A, Ioannidis C, Klein M, Ioannides M (2017) Modelling of static and moving objects: digitizing tangible and intangible cultural heritage. In: Mixed reality and gamification for cultural heritage. Springer, pp 567–589
11. Finnegan DJ, O'Neill E, Proulx MJ (2016) Compensating for distance compression in audiovisual virtual environments using incongruence. In: Proceedings of the 2016 CHI conference on human factors in computing systems, CHI '16. ACM, New York, pp 200–212
12. Fleming RW, Dror RO, Adelson EH (2003) Real-world illumination and the perception of surface reflectance properties. J Vis 3(5):3–3
13. Fodor J (2000) The mind doesn't work that way: the scope and limits of computational psychology. MIT Press, Cambridge
14. Fujisaki W, Goda N, Motoyoshi I, Komatsu H, Nishida S (2014) Audiovisual integration in the human perception of materials. J Vis 14(4):12–12
15. Fujisaki W, Tokita M, Kariya K (2015) Perception of the material properties of wood based on vision, audition, and touch. Vis Res 109:185–200
16. Giordano BL, McAdams S (2006) Material identification of real impact sounds: effects of size variation in steel, glass, wood, and plexiglass plates. J Acoust Soc Am 119(2):1171–1181
17. Grassi M (2005) Do we hear size or sound? Balls dropped on plates. Percep Psychophys 67(2):274–284
18. Grelaud D, Bonneel N, Wimmer M, Asselot M, Drettakis G (2009) Efficient and practical audio-visual rendering for games using crossmodal perception. In: Proceedings of the 2009 symposium on interactive 3D graphics and games, I3D '09. ACM, New York, pp 177–182
19. Guttentag DA (2010) Virtual reality: applications and implications for tourism. Tour Manage 31(5):637–651
20. Hairston DW, Hodges DA, Burdette JH, Wallace MT (2006) Auditory enhancement of visual temporal order judgment. NeuroReport 17(8):791–5
21. Hoaglin DC, Iglewicz B (1987) Fine-tunning some resistant rules for outlier labeling. J Am Stat Assoc 82(400):1147–1149
22. Hoeg ER, Gerry LJ, Thomsen L, Nilsson NC, Serafin S (2017) Binaural sound reduces reaction time in a virtual reality search task. In: 2017 IEEE 3rd VR workshop on sonic interactions for virtual environments (SIVE), pp 1–4
23. Kerr WB, Pellacini F (2010) Toward evaluating material design interface paradigms for novice users. In: ACM SIGGRAPH 2010 Papers, ACM, pp 35:1–35:10
24. Klatzky RL, Pai DK, Krotkov EP (2000) Perception of material from contact sounds. Presence: Teleoperators Virt Environ 9(4):399–410
25. Kokkinara E, Slater M (2014) Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. Perception 43(1):43–58
26. Koutek CDM, Koutek M Scientific visualization in virtual reality: interaction techniques and application development
27. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS (2009) Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. Bmj 338:b1802
28. Le Van Quyen M (2011) The brainweb of cross-scale interactions. New Ideas Psychol 29:57–63
29. Maculewicz J, Nilsson NC, Serafin S (2016) An investigation of the effect of immersive visual and auditory feedback on rhythmic walking interaction. In: Proceedings of the audio mostly 2016, AM '16. ACM, New York, pp 194–201
30. Martín R, Iseringhausen J, Weinmann M, Hullin MB Multimodal perception of material properties. In: Proceedings of the ACM SIGGRAPH symposium on applied perception (2015). ACM, pp 33–40
31. Masia B, Wetzstein G, Didyk P, Gutierrez D (2013) A survey on computational displays: pushing the boundaries of optics, computation, and perception. Comput Graph 37(8):1012–1038
32. Matsumoto K, Ban Y, Narumi T, Yanase Y, Tanikawa T, Hirose M (2016) Unlimited corridor: redirected walking techniques using visuo haptic interaction. In: ACM SIGGRAPH 2016 emerging technologies, SIGGRAPH '16. ACM, New York, pp 20:1–20:2

33. McGurk HMJ (1976) Hearing lips and seeing voices. Nature 264:746–8
34. McNamara A, Mania K, Gutierrez D (2011) Perception in graphics, visualization, virtual environments and animation. SIGGRAPH Asia Courses
35. Mishra J, Martinez A, Hillyard SA (2013) Audition influences color processing in the sound-induced visual flash illusion. Vis res 93:74–79
36. Nilsson NC, Suma E, Nordahl R, Bolas M, Serafin S (2016) Estimation of detection thresholds for audiovisual rotation gains. In: 2016 IEEE virtual reality (VR) pp 241–242
37. Owens A, Isola P, McDermott J, Torralba A, Adelson EH, Freeman WT (2015) Visually indicated sounds. arXiv:abs/1512.08512
38. Patney A, Salvi M, Kim J, Kaplanyan A, Wyman C, Benty N, Luebke D, Lefohn A (2016) Towards foveated rendering for gaze-tracked virtual reality. ACM Trans Graph (TOG) 35(6):179
39. Prinz J (2006) Is the mind really modular? In: Stainton RJ (ed) Contemporary debates in cognitive science. Contemporary debates in philosophy. Blackwell Publishing, Malden
40. Rallis I, Georgoulas I, Doulamis N, Voulodimos A, Terzopoulos P (2017) Extraction of key postures from 3d human motion data for choreography summarization. In: 2017 9th International conference on virtual worlds and games for serious applications (VS-Games). IEEE, pp 94–101
41. Ramanarayanan G, Ferwerda J, Walter B, Bala K (2007) Visual equivalence: towards a new standard for image fidelity. ACM Trans Graph 26:3
42. Riecke BE, Väljamäe A, Schulte-Pelkum J (2009) Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. ACM Trans Appl Percept 6(2):7:1–7:27
43. Rojas D, Kapralos B, Cristancho S, Collins K, Hogue A, Conati C, Dubrowski A (2012) Developing effective serious games: the effect of background sound on visual fidelity perception with varying texture resolution. In: MMVR, pp 386–392
44. Rojas D, Kapralos B, Hogue A, Collins K, Nacke L, Cristancho S, Conati C, Dubrowski A (2013) The effect of sound on visual fidelity perception in stereoscopic 3-d. IEEE Trans Cybern 43(6):1572–1583
45. Rojas D, Kapralos B, Collins K, Dubrowski A (2014) The effect of contextual sound cues on visual fidelity perception. Stud Health Technol Inf 196:346–352
46. Rojas D, Cowan B, Kapralos B, Colllins K, Dubrowski A (2015) The effect of sound on visual realism perception and task completion time in a cel-shaded serious gaming virtual environment. In: 2015 Seventh international workshop on quality of multimedia experience (QoMEX). IEEE, pp 1–6
47. Samuels R (2000) Massively modular minds: evolutionary psychology and cognitive architecture. In: Carruthers P, Chamberlain A (eds) Evolution and the human mind. Cambridge University Press, Cambridge
48. Sekuler R, Sekuler A, Brackett T (1995) When visual objects collide: repulsion and streaming. Investig Ophthalmol Vis Sci 36:50
49. Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385(6614):308
50. Serrano A, Gutierrez D, Myszkowski K, Seidel H-P, Masia B (2016) An intuitive control space for material appearance. ACM Trans Graph (SIGGRAPH ASIA) 2016(35):6
51. Seth A, Vance JM, Oliver JH (2011) Virtual reality for assembly methods prototyping: a review. Virt Real 15(1):5–20
52. Seymour NE, Gallagher AG, Roman SA, O'brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 236(4):458
53. Shams LKR (2010) Crossmodal influences on visual perception. Physics of Life Reviews
54. Shams L, Kamitani YSS (2000) What you see is what you hear? Nature 408:788
55. Shams L, Kamitani Y, Shimojo S (2002) Visual illusion induced by sound. Cogn Brain Res 14:147–152
56. Shimojo S, Scheier C, Nijhawan R, Shams L, Kamitani Y, Watanabe K (2001) Beyond perceptual modality: auditory effects on visual perception. Acoust Sci Technol 22(2):61–67
57. Sillion FX, Rushmeier H, Dorsey J (2008) Digital modeling of material appearance. Morgan Kaufmann/Elsevier
58. Suh K-S, Lee YE (2005) The effects of virtual reality on consumer learning: an empirical investigation. Mis Q, 673–697
59. Suied C, Bonneel N, Viaud-Delmon I (2008) Integration of auditory and visual information in the recognition of realistic objects. Exp Brain Res 194(1):91
60. Tononi G, Edelman GM (1998) Consciousness and complexity. Science 282:1846–1851
61. Väljamäe A, Larsson P, Västfjäll D, Kleiner M (2008) Sound representing self-motion in virtual environments enhances linear vection. Presence: Teleoper Virt Environ 17(1):43–56
62. Van Krevelen D, Poelman R (2010) A survey of augmented reality technologies, applications and limitations. Int J Virt Real 9(2):1

63. Vangorp P (2009) Human visual perception of materials in realistic computer graphics. PhD thesis, Department of Computer Science, KU Leuven Celestijnenlaan 200A, 3001 Heverlee. Belgium
64. Vangorp P, Laurijssen J, Dutré P (2007) The influence of shape on the perception of material reflectance. ACM Trans Graph 26:3
65. Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2:229–239
66. Waltl M, Timmerer C, Hellwagner H (2010) Improving the quality of multimedia experience through sensory effects. Quality of Multimedia Experience (QoMEX)

S. Malpica is a Computer Engineer and a PhD student now at Universidad de Zaragoza (Spain). She is currently working in the Graphics and Imaging Lab group at the same university, and her research interests include material appearance and perception, either in conventional displays or in Virtual Reality.

A. Serrano is an Electrical Engineer and a PhD student at Universidad de Zaragoza (Spain) under the supervision of Prof. Diego Gutierrez and Prof. Belen Masia. Her research interests lie in the fields of Image Processing and Computational Photography. Currently her main interests include material appearance, and visual attention and behavior in virtual reality. Her work has been published in several venues, including Transactions on Graphics, and Computer Graphics Forum.



M. Allue is a computer engineer graduated by Zaragoza's University. His interests include Virtual Reality (VR) and Augmented Reality (AR).

M. G. Bedia (group leader of Isaac Lab) is an Associate Professor of Computer Science at the Department of Computer Science and Systems Engineering (University of Zaragoza, Spain). He holds a BSc in Physics, an MSc in Technological Innovation management and a PhD in Computer Science and Artificial Intelligence, all from the University of Salamanca (Spain). He has worked as a research fellow in the field of artificial cognitive systems in the Department of Computer Science at the University of Salamanca, the Planning and Learning Group at the University Carlos III of Madrid, and the Multidisciplinary Institute at the University Complutense of Madrid. He has also been a visiting postdoctoral researcher at the Institute of Perception, Action and Behavior (Edinburgh, UK) and the Centre for Computational Neuroscience and Robotics at the University of Sussex (Brighton, UK). His areas of interest are mainly focused to the development of dynamical models of cognitive and embodied systems, the convergence between dynamical systems and information theory and the application of evolutionary and A-life techniques. He was co-founder of the "Spanish Network of Cognitive Sciences," thematic network established to promote and coordinate research in Cognitive Systems with emphasis on the relationships between scientific and educational policies, and the Spanish university system.

B. Masia received the Ph.D. degree in computer science from Universidad de Zaragoza, Zaragoza, Spain, in 2013. She has been a Postdoctoral Researcher at the Max Planck Institute for Informatics in Saarbruecken, Germany, and a Visiting Researcher at the Massachusetts Institute of Technology. She is currently an Assistant Professor in the Department of Computer Science, Universidad de Zaragoza, and a member of the I3A Research Institute. Her research in computational imaging, displays, and applied perception has been published in top venues such as ACM Transactions on Graphics. She is an Associate Editor of ACM Transactions on Applied Perception, and of Computers & Graphics. She has also served in numerous technical papers program committees, including ACM SIGGRAPH Asia, Eurographics, and Pacific Graphics. She has received a number of awards, including the Eurographics Young Researcher Award in 2017, a Eurographics PhD Award in 2015, she was selected as one of the top ten Innovators Under 35 in Spain by MIT Technology Review in 2014, and has also received a NVIDIA Graduate Fellowship in 2012. She is a Eurographics Young Fellow.