Multimodality in VR: A Survey

DANIEL MARTIN, SANDRA MALPICA, DIEGO GUTIERREZ, BELEN MASIA, and
ANA SERRANO, Universidad de Zaragoza, I3A, Spain

Virtual reality (VR) is rapidly growing, with the potential to change the way we create and consume content. In VR, users integrate multimodal sensory information they receive to create a unified perception of the virtual world. In this survey, we review the body of work addressing multimodality in VR and its role and benefits in user experience, together with different applications that leverage multimodality in many disciplines. These works thus encompass several fields of research and demonstrate that multimodality plays a fundamental role in VR, enhancing the experience, improving overall performance, and yielding unprecedented abilities in skill and knowledge transfer.

CCS Concepts: • **Human-centered computing → Virtual reality;**

Additional Key Words and Phrases: Virtual reality, immersive environments, multimodality

1 INTRODUCTION

Virtual Reality (VR) is inherently different from traditional media, since it introduces additional degrees of freedom, a wider field of view, more sophisticated sound spatialization, or even gives users control of the camera. VR immersive setups (such as **head-mounted displays (HMDs)** or CAVE-like systems) thus have the potential to change the way in which content is consumed, increasing realism, immersion, and engagement. This has impacted many application areas such as education and training [29], rehabilitation and neuroscience [183, 237], or virtual cinematography [194]. One of the key aspects of these systems lies in their ability to reproduce sensory information from different modalities (mainly visual and auditory, but also haptic, olfactory, gustatory, or proprioceptive), giving them an unprecedented potential.

Fig. 1. VR can be used to systematically analyze the interactions of multimodal information. In this example, the influence of auditory signals in the perception of visual motion is studied [115]. The authors found that different temporal synchronization profiles affected how the stimuli were perceived: When the visual (red balls moving) and auditory (an impact sound) stimuli were correctly synchronized, users perceived a unified event, in particular a collision between both balls.

Although visual stimuli tend to be the predominant source of information for humans [22, 208], additional sensory information helps to increase our understanding of the world. Our brain integrates different sources of sensory feedback, including both external stimuli (visual, auditory, or haptic information) and internal stimuli (vestibular or proprioceptive cues), thus creating a coherent, stable perception of objects, events, and oneself. The unified experience of the world as we perceive it therefore emerges from these multimodal cues [161, 196]. These different sources of information must be correctly synchronized to be perceived as belonging together [141, 157], and synchronization sensitivity varies depending on the context, task, and individual [45]. In general, different modalities will be perceived as coming from a single event or object as long as their temporal incongruency is shorter than their corresponding window of integration [115].

When exploring virtual environments, the presence of stimuli from multiple sources and senses (i.e., multimodality) and their potential overlaps (i.e., crossmodality) may also enhance the final experience [130]. Many works have described techniques to integrate some of these stimuli to produce more engaging VR experiences or to analyze the rich interplay of the different senses. For instance, leveraging the window of integration mentioned above may alleviate hardware limitations and lag time, producing the illusion of real-time performance; this is particularly useful when different modalities are reproduced at different refresh rates [30]. Moreover, VR is also inherently well suited to systematically study the integration process of multimodal stimuli [11] and analyze the complex interactions that occur when combining different stimuli [115] (see Figure 1).

In this survey, we provide an in-depth review of multimodality in VR. Sensory modalities include information from the five senses: visual for sight, auditory for hearing, olfactory for smell, gustatory for taste, and haptic and thermal for touch. Apart from the five senses, we also consider proprioception, which can be defined as the sense of self-movement and has been defined as the sixth sense [34, 223]. We synthesize the existing body of knowledge with a particular focus on the *interaction* between sensory modalities focusing on visual, auditory, haptic, and proprioceptive feedback; in addition, we offer an extensive overview of existing VR applications that directly take multimodality into account.

1.1 The five senses

The way we perceive the world is defined by the five senses: sight, hearing, smell, taste, and touch. Vision is the dominant sense when retrieving information of our surroundings [169]. We are capable of understanding complex scenes with varying visual patterns, we can detect moving objects in our peripheral view, and we are highly sensitive to light [220]. However, we tend to focus our visual attention in a narrow region of frontal space [208]. In that sense, we rely on hearing to retrieve information from unseen locations. Auditory stimuli can grab our attention irrespective of our orientation, and we are good at filtering out particular sounds in a noisy environment (e.g., the cocktail party phenomenon [9]). The sense of touch includes different aspects: haptic, kinaesthetic (related with proprioception), temperature, pressure, and pain sensing. Touch occurs across the whole body, although our hands are our primary interface for this sense. Finally, the senses of smell and taste are closely related. They are often linked to emotions or memories and can even trigger aversive reactions [133]. Most importantly, besides the particularities of each different sense, and as we will see through this review, our multiple senses influence each other.

1.2 Proprioception

Proprioception arises from static (position) and dynamic (motion) information [22]. It plays a key role in the concept of self, and has been more traditionally defined as "awareness of the spatial and mechanical status of the musculoskeletal framework" [224]. Proprioceptive information comes mainly from mechanosensory neurons next to muscles, tendons, and joints, although other senses can induce proprioceptive sensations as well. A well-known example are visual cues inducing the phantom limb illusion [162].

Proprioception plays an important role in VR as well. On the one hand, it helps provide the subjective sensation of *being there* [177, 206]. On the other hand, proprioception is tied to cybersickness, since simulator sickness is strongly related to the consistency between visual, vestibular, and proprioceptive information; significant conflicts between them could potentially lead to discomfort [100, 129].

1.3 Reproducing sensory modalities in VR

Important efforts have been made in VR so that all the modalities previously mentioned can be integrated. Visual and auditory feedback are the most commonly used, and almost all consumer-level devices integrate these modalities. There is currently a wide variety of manufacturers providing different HMD systems to enjoy VR at consumer level. Each of them offers devices with different capabilities and specifications at different costs. Table 1 compiles an overview of the most relevant devices currently in the market. An open issue in VR is latency [46]: Newer HMD models feature higher refresh rates, as well as significantly increased spatial resolution. Usually, those displays feature a **field-of-view (FoV)** slightly smaller than that allowed by human peripheral vision. However, new stereoscopic rendering techniques allow to present content in 3D, and therefore perception of materials, depth, or many other cues can be achieved through visual cues [7]. *Auditory* feedback, which is often integrated in the HMD as a built-in feature (Table 1), is generally enabled by speakers or headphones, and spatial audio rendering techniques also support our perception of space in virtual environments [6], even enhancing perceived visual properties [115]. *Haptic* feedback is still in an exploratory phase and can be achieved through a variety of sensors, including wearables [155, 179], physical accessories [213], ultrasounds [119], controller devices [214, 234], rotary components [84], or electric muscle stimulation [108]. Other resources such as fans, hear lamps, or even spray bottles have been used to provide additional tactile stimuli in VR [235]. Recently, advances in ultrasound sensor technology have resulted in the creation of a novel haptic

Table 1. Overview of Predominant Current HMD Devices

Manufacturer and model	Resolution per eye	Positional tracking	Max. refresh rate (Hz)	Field of view (degrees)	Display type	Integrated audio	Price
FOVE	1280 × 1440	Yes	70	100	OLED	No - Jack 3.5mm	$600
HP Reverb - Pro	2160 × 2160	Inside-out	90	114	LCD	Built-in headphones	$649
HTC VIVE Pro	1440 × 1600	Yes	90	110	AMOLED	Built-in headphones	$799
HTC VIVE Pro 2	2448 × 2448	Yes	120	120	LCD	Built-in headphones	$800
Oculus Rift S	1280 × 1440	Inside-out	80	110	LCD	In-line speakers	$399
Oculus Quest 2	1832 × 1920	Inside-out	120	104	LCD	Stereo speakers	$399
Samsung Odyssey	1440 × 1600	Inside-out	90	110	AMOLED	Built-in headphones	$500
PlayStation VR	960 × 1080	Outside-in	120	100	OLED	No - Jack 3.5mm	$299
Valve Index	1440 × 1600	Yes	144	130	LCD	No - Jack 3.5mm	$999
Varjo VR-3	2880 × 2720	Yes	90	115	uOLED	Off-ear speakers	$3195

For each of them, we include the resolution per eye, whether they provide positional tracking, their maximum refresh rate (in Hz), their field of view (FoV, in degrees), the type of display, and a current estimate of the final consumer price. The better specs (in terms of refresh rate and FoV) offered by Valve Index come at a higher cost, while other manufacturers opt for cheaper HMDs, potentially more affordable to consumers.

device that will allow for mid-air force around virtual objects and interactions [107] this device is about to reach the consumer market [47]. *Olfactory* stimuli can be provided through smell cartridges or specific hardware [69], and electric stimulation of taste buds has been tested to generate flavors [163]. How all these different feedback modalities can be integrated depends on the particular context, with some algorithms and techniques already proposed for the most common sensory combinations, including audiovisual or audiohaptic stimuli [113].

1.4   Related surveys

Our perception of the world depends on the integration of information from multiple senses. Several works have reviewed the influence of individual senses *in isolation*, including sight [167], sound [191], or touch [102], which are the three key modalities in VR (see Table 5). Several surveys exist focusing on particular aspects of VR. Rubio et al. [175] systematically reviewed advances in communication, interaction, and simulation in VR, pointing out that the key factors to generate appealing virtual experiences include a good interactivity, representation, gameplay, and narrative. The latter, narrative, has been explored in depth from a cinematic perspective [172], since content creators no longer have the same level of control over how the viewer attention is directed: To find new ways of guiding viewer's attention, the authors reviewed current attention techniques in virtual environments, either unimodal or multimodal, emphasizing how auditory cues can be critical, and the still-unexploited potential of haptic devices (see Section 8.3)

Other surveys target specific applications of VR, such as education or medicine [170]. For the specific case of clinical medicine, Li et al. [103] found that most of the works in the literature leverage the capabilities of haptic devices to simulate real, clinical tasks, in line with our insights (see Section 8.1). Freina et al. [55] reviewed works focused on using VR in education, concluding that it increases the learner's involvement and motivation, which are enhanced with multimodality (Section 8.2).

Some works are concerned particularly with cognitive aspects in multimodal environments. Hecht et al. [68] briefly studied how integrating multiple sources can increase presence, enhance attention, and improve response time. Koelewijn et al. [91] focused on surveying works related to low-level, audiovisual interactions, concluding that both multisensory integration and attentional processes take place and can interact at multiple stages in the brain. However, multisensory overload can sometimes lead to a preference on simpler environments if not handled correctly; we delve deeper into this in Section 3. Other surveys have studied multimodality in *traditional*

media, including cognition [209], interaction [83], human-computer interfaces [44, 64], or fusion and integration techniques [10]. Closer to the present work, Melo et al. [130] systematically studied the impact of multisensory stimuli in virtual experiences. Their study suggests that 85% of works tackling multimodality in VR report positive impacts, with only 1% of them reporting negative impacts. They also reported how multisensory experiences in VR are mainly applied in the health domain, science and engineering, teaching, or machinery; which is in line with our reports in Section 8. While they report that these are the more common applications among the 105 studies they surveyed, we provide a discussion on how multimodality is impacting each of these fields.

However, and different from all these works, in this survey we focus on the integration of multimodal information and the benefits and experiences that can be achieved that way, and we compile a large body of works studying not only those positive effects, but also their applicability into different disciplines.

1.5 The challenges of multimodality

One of the main challenges when considering a fully multimodal immersive experience is the gaps of empirical knowledge that exist in this field. As stated before, in this survey our main focus lies in the visual, auditory, haptic, and proprioceptive modalities. This is partly related to the fact that many modalities and their interactions remain unexplored, and there is still much to learn about them. Moreover, the available data on multimodality in VR (both referring to multimodal stimuli and to user data while experiencing multimodal VR) is scarce at best. It is also important to consider the window of integration. The necessity of synchronizing different modalities implies the need for real-time, high-fidelity computation. Hardware processing limitations might also imply a constraint in what multimodal techniques should be used in different scenarios. Moreover, not all VR headsets are equally prepared to support multimodality: Although most of them can give audiovisual feedback, proprioception and haptic feedback are sometimes limited, while olfactory and gustatory feedback are usually not found at all in consumer-level headsets. For example, most smartphone-based VR headsets do not include controllers, hindering the possibility of including haptic feedback. Many other basic VR systems are not able to track translations either (i.e., only have three degrees of rotatory freedom), which limits proprioceptive feedback. However, one of the most critical risks with multimodality is its definition itself. Although multimodality has the potential to improve user experience, increase immersion, or even improve performance in certain tasks, multimodal applications have to be very carefully designed, making sure that each modality has its function or purpose. This is not a trivial task, since the dimensionality of the problem grows with each added modality. Otherwise, additional modalities might distract and overload the final user, hampering the experience and diminishing user satisfaction.

2 SCOPE AND ORGANIZATION OF THIS SURVEY

In this survey, we provide an in-depth review of the most significant works devoted to explore the role and effects of multimodality in virtual reality. We gather knowledge about how multiple sensory modalities interact and affect the perception, the creation, and the interaction with the virtual experience.

The structure of this survey can be seen in Figure 2. Since our focus is not on any specific part of the VR pipeline, but rather on the *VR experience* for the user, we have identified several areas of the VR experience in which multimodality plays a key role. First, Section 3 is devoted to the realism of the VR experience, which is tied to immersion and the sense of presence that the user experiences. Second, Section 4 looks into how multimodality can affect the attentive process of the user in the virtual environment, determining how they explore the environment and what drives their attention within it. Third, Section 5 delves into works that demonstrate how multimodality

3/ Multimodality contribution to realism	4/ The effects of multimodality in users' attention
How does multimodality help improve environment fidelity and embodiment?	What draws users' attention? How can their attention be guided?

5/ Multimodality in users' performance	6/ Multimodal illusions in virtual reality
How can multimodality enhance users' performance and the outcome of virtual tasks?	How can multimodality exploit human mechanisms to enhance the experience?

7/ Navigation in virtual relality	8/ Applications
How does multimodality affect the way users navigate and move in immersive environments?	Research areas that have benefited from multimodality in virtual reality

Fig. 2. Structure of this state-of-the-art report. We divide it into different areas of the VR experience in which multimodality can play a key role: user immersion, presence, and realism of the experience (Section 3); user attention when exploring the virtual environment (Section 4); user performance when completing tasks (Section 5); multimodal perceptual illusions that can be leveraged in VR (Section 6); and navigational effects of multimodality in virtual environments (Section 7). Finally, we review different applications where multimodality has been shown to improve the end goal (Section 8) and finalize with a discussion on the need for multimodality and open avenues of research (Section 9).

can help the user in completing certain tasks, essentially improving user performance in the virtual environment.

Additionally, there are a number of works devoted to analyzing multimodal perceptual illusions and their perception in VR environments. These, which we compile in Section 6, can be leveraged by future techniques to improve any of the aforementioned areas of the VR experience. Some other works have tackled the problem of navigation in VR, which is another integral part of the virtual experience, and Section 7 encompasses them. A complete perspective of all these sections is also included in Table 5. Finally, we devote Section 8 to reviewing application areas that have benefited from the use of multimodal virtual experiences, and we conclude (Section 9) with a discussion of the potential of multimodality in VR and interesting avenues of future research.

3 THE EFFECTS OF MULTIMODALITY IN PERCEIVED REALISM

Perceived realism elicits realistic responses in immersive virtual environments [205], and is tied to the overall perception of the experience. There are two key factors that can lead to users responding in a realistic manner: the place illusion and the plausibility illusion [204]. The former, also called "presence," defines the sensation of "being there" and is dependent on sensorimotor information, while the latter refers to the illusion that the scenario that is apparently happening is actually taking place and is determined by the ability of the system to produce events that relate to the user, i.e., the overall credibility of the scenario being depicted in comparison with the user's expectations. Slater [204] argued that participants respond realistically to an immersive VR environment when these two factors are present. Similar observations were made in telepresence systems [211], where sensorially rich mediated environments were proved to actually elicit more realistic responses.

Increasing the feeling of presence can therefore enhance the experience by eliciting more realistic responses from the users, and as aforementioned, increasing the perceived realism has a positive impact in the feeling of presence. This actually depends on both the virtual environment where the user is placed and its own representation in there. As happens in the real world [59], all human modalities play a fundamental role and must be correctly integrated to construct a coherent notion of the both the virtual environment and the self. In this section, we will therefore

focus on how multimodal cues can affect perceived realism by affecting both the perception of the environment and the perception of the self.

3.1 Perception of the environment

The perceived realism of virtual environments is a key concern when designing virtual experiences, therefore many works have been devoted to investigate how multimodality and crossmodality can indeed help achieve sensorially rich experiences. While multimodality refers to the binding of different inputs from multiple sensory modalities, crossmodality involves interactions between different sensory modalities that influence the perception of one another [96, 210]. Chalmers et al. [26] discussed how crossmodal effects in human multisensory perception can be exploited to selectively deliver high-fidelity virtual environments; for instance, rendering with higher visual quality those items related to the current auditory information of the scene, allowing to reduce computational costs in unattended regions of the virtual environment. This work also reports that humans perceive sensory information with more or less attention, depending on the task they are executing (i.e., some tasks require more attention to particular types of stimuli) or on if they have already been preconditioned to that kind of virtual environment (e.g., they are used to it). Traditionally, sound has proven to facilitate visual perception, including enabling a better understanding of the environment, yielding a more comfortable experience, or even increasing performance of visual-related tasks [79, 198]. Seitz et al. [189] conducted a 10-day experiment where two groups of people were trained for auditory-visual motion-detection tasks, one with only visual, and the other with audiovisual stimuli. Although all of them improved their performance over time, those trained with multimodal stimuli showed significantly better performances. Various works have been thus devoted to this audiovisual integration: Morgado et al. [134] presented a system that generates ambisonic audio for 360° panoramas, so auditory information is represented in a spherical, smoother way (see Figure 3, left). Similarly, Huang et al. [77] proposed a system that automatically adds spatialized sounds to create more realistic environments (see Figure 3, right), validating by means of user studies the overall preference of this solution in terms of realism. Indeed, different soundscapes (a sound or combination of sounds created from an immersive environment) are able to increase the sense of presence in VR [190], and as Liao et al. [104] studied, combining visual and auditory zeitgebers (periodically occurring natural phenomena that act as cues in the regulation of biological rhythms), which act like synchronizers, could actually enhance presence, even influencing time perception. All these previous works suggest that using auditory information, either spatialized or not, enhances the realism of the experience, although some of them warn about the potential backfire of increasing the cognitive load, which can negatively impact users' confidence [86].

However, multimodal integration can also present some drawbacks: Akhtar and Falk [1] surveyed current audiovisual quality assessment and found that auditory information may cause discomfort and decrease the quality of the virtual experience [176]. To avoid negative effects during multimodal integration, different sensory cues should be not only realistic, but also coherent to the environment and between them.

Proprioception also plays an important role in eliciting realism, as it contributes to the feeling of the user being there. Although some works have demonstrated that some manipulations in virtual movement directions and distances can be unnoticeably performed (either by manipulating the environment itself through the game engine or by modifying the real-to-virtual mapping of users' movement) [97, 193], users tend to expect their virtual movements to match their real ones, to maintain a coherent experience. In this line, Mast and Oman [125] studied the so-called visual reorientation illusions: When the environment is rotated above a given noticeable threshold in any axis, users can perceive that the expected vertical axis does not match the virtual one, and conflicts

Fig. 3. Including correct and coherent auditory information in the virtual environment has been proved to increase realism and immersion. *Left*: A system that automatically generates ambisonic information that creates a smoother acoustic experience for the scene [134]. *Right*: A framework to include auditory information into 360° panoramas, depending on the elements in the scene [77]. In both cases, their validation experiments yield users' preference when auditory information is included and an overall increase in the perceived realism and immersion.

between visual and vestibular cues may arise, potentially causing motion sickness. Although the effect of this illusion is stronger for elder users [74], an incoherent spatial estimation in VR can potentially diminish the perceived realism.

Including additional modalities can also enhance environment realism. In particular, giving realistic feedback with respect to what users expect to happen actually increases plausibility. Normand et al. [149] showed that it is possible to induce a body distortion illusion by synchronous visual-tactile and visual-motor correlations (see Figure 4). In a similar fashion, Hoffman compared the realism of virtually touching an object with that of touching it physically at the same time [71], yielding a significant increase in perceived realism when the object was physically touched, too. Similar results were obtained with taste and olfactory cues [73]: They found a preference on smelling and physically biting a chocolate bar in contrast to only virtually biting it. The level of presence achieved depends on the different combinations of sensory feedback, and multi-sensory systems have been proved to be superior to traditional audio-visual virtual systems in terms of the sense of presence and user preference [86]. Similar conclusions were obtained by Hecht et al. [68], who reported that multimodality led to a faster start of the cognitive process, which ultimately contributed to an enhanced sense of presence. However, and even if the benefits of multimodal integration are widely known and shared between researchers and practitioners, there is still much to learn about the limits and drawbacks of multisensory integration, and studying up to what point multimodal interaction can be safely applied to increase perceived realism in different scenarios remains an interesting future avenue.

3.2 Perception of the self

Virtual experiences are designed for humans, and in many occasions, users are provided with a virtual representation of themselves. This is a very effective way of establishing their presence in the virtual environment, hence contributing to the place illusion [204]. This representation does not need to be visually realistic, but it has to be coherent enough with the users' actions and expectations to maintain the consistency of the experience. In the following, we review different works that have leveraged multimodality in virtual reality to achieve consciousness of the self and embodiment, and therefore to create realistic representations of the users.
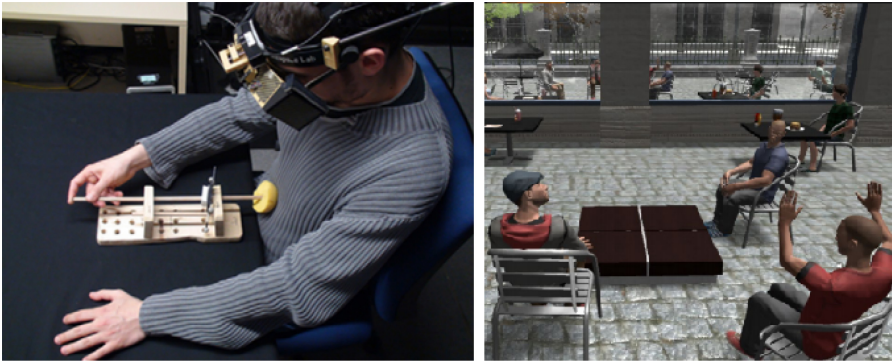
Fig. 4. *Left*: Synchronizing different modalities increases the feeling of presence and the perception of the self. Moreover, multimodality can even create a distortion of that perception: Normand et al. [149] presented a study where a body distortion illusion is achieved by synchronous visual-tactile and visual-motor correlations. *Right*: Some works have studied how different physical and behavioral factors can directly affect, and even manipulate, embodiment [140], and therefore, the perception of the self.

Having the feeling of being in control of oneself is possibly one of the main characteristics that VR offers [204]. The feeling of presence is possible without being in control; however, being able to control a virtual body highly increases this illusion [187]. The sense of embodiment gathers the feeling of owning, controlling, and being inside a body. As Kilteni et al. [89] reported, this depends on various subcomponents, namely, the sense of self-location (a determined volume in space where one feels to be located), the sense of agency (having the subjective experience of action, control, intention, motor selection, and the conscious experience of will), and the sense of body ownership (having one's self-attribution of a body, implying that the body is the source of the experienced sensations). Other factors like the proximity of virtual objects to the body also have an effect on the sense of embodiment [188]. All these concepts (such as presence or embodiment) are intrinsic characteristics that VR can achieve, and they yield the self-consciousness feeling that makes VR so different from other media.

Multimodality has been largely studied as a means to enhance those sensations. Particularly, presence is tied to the integration of multiple modalities, and many works have demonstrated how it is increased when multiple sensory information is combined [182], as opposed to unimodal (i.e., only visual) systems [86]. For instance, Gonçalves et al. [59] designed an experiment where three groups of people were exposed to virtual environments including different amount of modalities in the presented stimuli; they reported how users experiencing more modalities reported a higher involvement. Moreover, they remark the positive impact of including haptic feedback in an experience. Blanke et al. [16] discussed the relevance of a series of principles to achieve a correct sensation of bodily self-consciousness, requiring body-centered perception (hand, face, and trunk), and integrating proprioceptive, vestibular, and visual bodily inputs, along with spatio-temporal multisensory information. Sakhardande et al. [178] presented a systematic study to compare the effect of tactile, visual, visual-motor, and olfactory stimuli on body association in VR, with the latter having the strongest effect on body association. Similar insights were proposed by Pozeg et al. [159], who demonstrated the importance of first-person visual-spatial viewpoints for the integration of visual-tactile stimuli, in this case for the sense of leg ownership. The main factors to build embodiment and body-ownership in VR have been widely studied [123]. Spanlang et al. [207] presented technical guidelines to create a core virtual embodiment system, defining three key aspects: (i) a VR module to handle creation, management, and rendering of all virtual entities, (ii) a

head-tracking module to map real movements to the virtual environment, and (iii) a display module to present all that environment. However, designing experiences that are too realistic can have negative aspects and be a drawback in certain specific cases: For example, group pressure of alien virtual avatars can result in users performing potentially harmful actions towards others that they would not normally carry out [140].

The sense of moving (which depends on agency and body ownership, as previously mentioned) is also key to achieve self-consciousness. Kruijff et al. [93] presented a work showing that adding walking-related auditory, visual, and vibrotactile cues could all enhance the participants' sensation of self-motion and presence. Various works have been presented in this line, e.g., investigating the integration of tendon vibrations to give standing users the impression of physically walking [92]. Sometimes locomotion is not possible, and it has to be externally generated, e.g., by means of a virtual walking system for sitting observers using only passive sensations such as optic flow and foot vibrations [126]. However, these techniques are likely to creating the well-known self-motion illusion: Although users are not actually moving, their brain unconsciously assumes they are moving, and their body sometimes generates postural responses [42] to control their stability. Meyer et al. [131] studied the impact of having multimodal (visual, auditory, and haptic) anchor points in the virtual environment in users' postural sway. They report how incongruent cues diminish perceived realism. However, they also remark on the complexity of providing dynamic tactile signals in VR, which leaves an interesting research line in how to exploit tactile cues to increase presence. Some other works have also explored alternatives for cases when locomotion is not feasible, for instance proposing and evaluating a virtual walking system for sitting observers using only passive sensations such as optic flow and foot vibrations [126].

Other modalities may also play an important role in users' self-consciousness: Several works have shown that multimodality can dramatically increase the sense of presence [56], although confidence levels for certain tasks are higher in traditional (i.e., audio-visual) virtual environments, due to a higher cognitive load [86]. Besides additional modalities, other factors such as immersion and emotion have been analyzed and argued to have a clear impact on the sense of presence [14]. In particular, audiovisual content eliciting emotional responses (like sadness) can increase engagement and presence, somehow bypassing the immersive effects of specific displays.

As reported in some of the aforementioned works, multimodality presents some challenges and limitations: Gallace et al. [56] focused on the ones associated with the simultaneous stimulation of multiple senses, including the senses of touch, smell, taste, and even nioceptive (i.e., painful) sense, given the cognitive limitations in the human sensory perception bandwidth when users have to divide their attention between multiple sensory modalities. Moreover, situations where some modalities violate interpersonal space may also lead to diminishing presence and comfort [236]. Ultimately, achieving user's self-consciousness depends on finding the right balance between different multimodal cues and the users' comfort, confidence, and capacity to integrate them. Establishing guidelines towards this balance remains one of the most interesting avenues in multimodal interaction.

4   THE EFFECTS OF MULTIMODALITY IN USERS' ATTENTION

When users are exploring or interacting with a virtual environment, different elements or events can draw their attention. Visual attention influences the processing of visual information, since it induces gaze to be directed to the regions that are considered more interesting or relevant (salient regions). The saliency of different regions results from a combination of top-down attentional modulation mechanisms (task-based) and the bottom-up multisensory information these regions provide (feature-based), creating an integrated saliency map of the environment [221]. As discussed in the previous section, VR setups may produce realistic responses and interactions, which can be

Fig. 5. Saliency maps show the likelihood of users directing their attention to each part of the scene. Most of the current literature has been devoted to estimating saliency in unimodal, visual stimuli. This image shows the recent visual saliency estimation method proposed by Martin et al. [121] (Left: Input panorama. Right: Estimated saliency). It has been shown that each sensory modality has the potential of influencing users' attentional behavior, therefore, there is a need for further exploration of multimodal saliency in VR.

different from traditional media due to the differences in perceived realism and interaction methods. Therefore, some works have been devoted to understanding saliency and users' attention in VR, offering some key insights about head-gaze coordination and users' exploratory behavior in VR. For example, Sitzmann et al. [203] detected the *equator bias* when users are freely exploring omnistereo panoramas: They observed a bias towards gazing at the central latitude of the scene (equator bias), which often corresponds to the horizon plane.

Saliency has been widely studied, both inside and outside the VR field [239]. Users are more likely to turn their attention and interact with those regions of the scene that provide more sensory information. Therefore, knowing *a priori* which parts of the scene are more salient may help anticipating how users will behave. So far, most of the works on saliency in VR have been carried out following a unimodal perspective [121, 242]: Although several senses can be considered to create a saliency map, they all leverage users' head position and gaze orientation to create probabilistic maps indicating the chances of a user looking at each part of the virtual scene (see Figure 5). Based on the study of visual cues, various works have presented systems able to predict user's gaze, depending on the environment, and also on user's previous behavior [76, 120].

Multimodality in saliency estimation has been only tackled in traditional media: The integration of visual and auditory information for saliency prediction in videos has been widely explored [38, 132]. All these approaches work under the assumption of audiovisual correlation: Moving elements were the source of the auditory cues. In a different approach, Evangelopoulos et al. [48] proposed the addition of text information in form of subtitles when speech was present in the auditory stream. In their work, saliency was considered as a top-down process, since the interpretation of the subtitles, a complex cognitive task, can distract viewer's attention from other parts of the scene.

Multimodality in saliency prediction for VR still remains in early phases, and only very few works have been devoted to it. Chao et al. [27] proposed the first work that studies user behavior (including saliency corresponding to sound source locations, viewing navigation congruence between observers, and the distribution of gaze behavior) in virtual environments containing both visual and auditory cues (including both monaural and ambisonic sounds). However, there are still many open avenues for future research: Visual saliency and gaze prediction in VR is still in an early phase, and the effects of auditory cues in saliency in virtual scenarios remain to be further explored. Auditory cues in VR may produce more complex effects and interactions than in traditional scenarios, since sound sources are not always in the user's field of view, and there might be several competing audiovisual cues. Additionally, investigating how other senses interact and predominate in saliency and attention can be useful for many applications, specially for content creation. Furthermore, with the proliferation of data-driven methods, it is also crucial to elaborate

Fig. 6. Examples of visual guidance methods in VR, adapted from Rothe et al.'s review on users' guidance for cinematic content [172]. Three visual guidance techniques are presented in this image: Arrows pointing to regions of interest, picture-in-picture techniques that show information of rear regions, and the position of a point of interest marked with red bars. Most of these techniques are intrusive, hence they may break immersion and realism. With the addition of multimodal cues, guidance can be facilitated while maintaining a positive user experience.

datasets that encompass enough variety of multimodal stimuli to support the formulation of new multimodal attentional models.

Although there is still much to learn about how multimodal cues compete and alter users' behavior, it is well known that multimodality itself has consequences on how users behave in immersive environments [27, 222]. One of the main difficulties when designing and creating content for VR lies in the fact that users' typically have control over the camera, and therefore each user may pay attention to different regions of the scene and create a different experience [117, 194]. Therefore, it is usually hard to make assumptions about users' behavior and attention. To support the creation of engaging experiences that convey the creators' intentions, multimodality can be exploited so cues from different modalities can induce specific behaviors and even guide users' attention.

For the case of attention, understanding and guiding users attention in VR has been a hot topic during the last years. Various works have explored the use of visual guiding mechanisms, such as central arrows and peripheral flicker to guide attention in panoramic videos [186]. The recent work of Wallgrun et al. [231] compares different visual guiding mechanisms to guide attention in 360° environments [231]. Lin et al. [106] proposed a picture-in-picture method that includes insets of regions of interest that are not in the current field of view, so users are aware of all the elements that surround them. Inducing the users to direct their attention to a specific part of the scene has also been explored, for example, using focus assistance techniques [105], such as indicating the direction of the relevant part or automatically orienting the world so users do not miss that part of the experience. Following this line, Gugenheimer et al. [63] presented a motorized swivel chair to rotate users until they were focusing on the relevant part of the scene, while Nielsen et al. [142] forced virtual body orientation to guide users' attention to the most relevant region. Other techniques directly let the viewer press a button to immediately reorient the scene to the part containing the relevant information [151]. It is worth mentioning that these kinds of techniques have to be taken into consideration with caution, since they can cause dizziness or discomfort due to visual-vestibular conflicts. We refer the reader to Rothe et al.'s work [172] for a complete survey about guidance in VR (see Figure 6).

However, guidance techniques are not necessarily constrained to visual manipulations. Multimodality can be also exploited to guide, focus, and redirect attention in VR, in many cases achieving more subtle, less intrusive methods. This is important to maintain the users' experience, as intrusive methods can alter the sense of presence, immersion, or suspension of disbelief (the temporary acceptance as believable of events or places that would ordinarily be seen as incredible). As shown in previous sections, sound can help enhance the virtual experience. Besides, it can also be used to manipulate or guide users' attention. Rothe et al. [174] demonstrated that the attention of the viewer could be effectively directed by sound and movements and later [173] investigated and compared three methods for implicitly guiding attention: Lights, movements, and sounds, showing that sounds elicit users' exploratory behavior, while moving lights can also easily draw attention. Other works have explored various unobtrusive techniques combining auditory and visual information, showing that auditory cues indeed reinforce users' attention being drawn towards specific parts of the environment [21]. Similar insights were obtained by Masia et al. [124], who investigated the impact of directional sound during cinematic cuts in VR, finding that in the presence of directional sound cues, users converge much faster to the main action after a cut, even if the sound is misaligned with the region of interest. Given the importance of including sound in VR experiences, Bala et al. [12] presented a software for adding sound to panoramic videos and studied how sound helped people direct their attention. Later, they examined the use of sound spatialization for orientation purposes [13]. In particular, they found that full spatial manipulation of sound (e.g., locating music in a visual region of interest) helped guide attention. In a similar fashion, some works have studied how to design sound to influence attention in VR [180] and how decision-making processes are affected by auditory and visual cues of diegetic (i.e., sounds emanating for the virtual environment itself) and non-diegetic (i.e., sounds that do not originate from the virtual environment itself) origins [23]. However, non-diegetic cues need to be analyzed and presented carefully: The work by Peck et al. [153] showed that a distractor audio can be successful at fostering users' head rotations (and thus redirection); however, users considered this method as unnatural. It has been also suggested that too many sound sources in a VR cinematic video can produce clutter, and therefore hinder the identification of relevant sound sources in the movie [172]. How to use multimodality for guiding users' attention has many open possibilities for further investigation. In this context, establishing guidelines regarding which senses to use, how to combine them, and up to what extent each of them can surpass the others remains for now a complex, unresolved task.

5 MULTIMODALITY IN USERS' PERFORMANCE

Understanding how users perform different tasks in VR is key for developing better interfaces and experiences. Although in many cases task performance highly depends on the users' skills and experience, there are many scenarios where multimodality can play an important role in this aspect: By integrating multiple sensory information, we can mimic better the real world, and this can lead to higher performance in different scenarios, comparable to real life. Additionally, multimodal VR technologies are becoming a very powerful tool for training and education, especially in scenarios that can be expensive, or even dangerous, in real life. In those cases, multimodality can help complete some tasks in a shorter period or with a higher accuracy [68].

The effects of multimodality in task performance have been largely studied in traditional media. Lovelace et al. [109] demonstrated how the presence of a task-irrelevant light enhances the detectability of a brief, low-intensity sound. This behavior also holds in the inverse direction: Concurrent auditory stimuli could enhance the ability to detect brief visual events [146]. Therefore, integrating audiovisual cues may diminish the risk of users losing some relevant information. In a similar line, Van der Burg et al. [225] reported that a simple auditory *pip* drastically decreased

detection time for a synchronized visual stimuli. These effects are not only present in audiovisual stimuli: Tactile-visual interactions also affect search times for visual stimuli [226]. In most of these work, the experiments were carried out in laboratory conditions with simple stimuli, and therefore studying their applicability and limitations in more complex scenarios remains an interesting avenue. Furthermore, Maggioni et al. [114] studied the potential of smell for conveying and recalling information. They compared the effectiveness of visual and olfactory cues, and their combination in this task, and demonstrated that olfactory cues indeed improved users' confidence and performance. Therefore, the integration of multiple cues has been widely proved to be effective in terms of detectability and efficiency. As Hecht et al. [68] studied, this improvement in terms of performance also holds for multimodal VR: When there are multiple senses involved, users start their cognitive process faster, thus they can pay attention to more cues and details, resulting in a richer, more complete and coherent experience.

Performance in spatial tasks can greatly benefit from multimodality [5, 57]. Auditory cues are extremely useful in spatial tasks in VR, and therefore have been widely explored: The effect of sound beacons in navigation performance when no visual cues are available has been explored [230], with some works proving that navigation when no visual information is available is possible using only auditory cues [62]. Other works have exploited this, proposing a novel technique to visualize sounds, similar to how echolocation would work in animals, which improved the space perception in VR, thanks to the integration of auditory and visual information [171], or combining the spatial information contained in echoes to benefit visual tasks requiring spatial reasoning [57]. Other senses have also been explored with the goal of enhancing spatial search tasks: Ammi and Katz [5] proposed a method coupling auditory and haptic information to enhance spatial reasoning, and thus improving performance in search tasks. Direct interaction tasks can be also enhanced by multimodality: Auditory stimuli has been proved to facilitate touching a virtual object outside user's field of view, hence creating a more natural interaction [90]. Egocentric interaction is also likely to happen, and proprioception plays an important role on those cases. Poupyrev et al. [156] presented a formal study comparing virtual hand and virtual pointer as interaction metaphors, in object selection and positioning experiments, yielding that indeed both techniques were suitable for different interaction scenarios.

As aforementioned, when developing VR experiences requiring users' to complete some tasks, the integration of multiple modalities can increase their performance and spatial reasoning, leading to better, more consistent results. Furthermore, adding certain modalities (e.g., olfactory or haptic information) is not always easy, specially at consumer level. Enabling these modalities within current consumer-level devices (Table 1) remains a future avenue that would not only greatly benefit multimodality in terms of performance, but it would also improve the whole experience. In spite of that, in some cases, combining several modalities can lead to the opposite effect, suppressing or diminishing some abilities [116], hence special care must be paid when designing multimodal experiences (Section 1.5).

6   MULTIMODAL ILLUSIONS IN VR

Multimodality can be leveraged to trick the self-perception of the users or to alter how they perceive the world around them by means of facilitatory or inhibitory (suppressive) effects, which can have direct implications on how users behave in the virtual environment. Being able to manipulate the experience can be very useful in certain contexts and applications: For instance, sometimes it can be useful to guide the user towards a particular aspect of the virtual environment without disrupting the experience (e.g., in cinematography or videogames). A forced guidance could lead to reduce an immersion feeling or even rupture the suspension of disbelief. In other cases, physical space is constrained, and manipulating users' movement may allow to reduce the

necessary physical space to complete a task [193]. Manipulating the experience can be also useful to reduce simulator sickness; for instance, by means of manipulating camera control depending on some characteristics such as velocity, acceleration, or scene depth [75]. Although this can be done using a single modality, the use of multimodal cues can improve the effectiveness of these techniques.

Illusion refers to an incorrect perception or interpretation of a real, external stimulus. It can lead to interpreting reality in several ways. Any healthy person can experiment illusions without experiencing any pathological condition. However, not every person is affected in the same way by an illusion. Illusions can have physiological (e.g., an after-image caused by a strong light [82]) or cognitive (e.g., the Rubin vase [152]) components. They have been widely studied, as understanding illusions yields valuable information about the limitations of human senses and helps understanding the underlying neural mechanisms that help create the perception of the outside world. Moreover, illusions can allow to alter users responses to certain tasks, even increasing performance. For instance, Chauvel et al. [28] conducted some experiments where non-golfers practiced putting golf balls, some of them with manipulated holes to enhance their visual acuity. Those who trained under these conditions showed a more effective learning outcome and a better performance when trying in real-life scenarios. In this subsection, we will focus on multimodal illusions or effects, or how illusions in other senses can affect visual perception. For visual-only illusions, we refer the reader to *The Oxford Compendium of Visual Illusions* [199].

Multimodal illusions can be useful for boosting accuracy in certain tasks. For example, multi-sensory cues can improve depth perception when using handheld devices by simulating tactile responses when holding or interacting with a virtual object with a force feedback system [20, 216]. Using a small number of worn haptic devices, Glyn et al. [101] improved spatial awareness in virtual environments without the need of creating physical prototypes. Instead of applying contact (haptic feedback) at the exact physical point of the user's body that was touching a virtual object, they used a small, fixed set of haptic devices to convey the same information. Their work was based on the funnelling illusion [15], in which the perceived point of contact can be manipulated by adjusting relative intensities of adjacent tactile devices. Visuo-haptic illusions allow not only to better perceive the virtual space, but also to feel certain virtual object properties, like weight, that are not easy to simulate. Even further, these properties can be unnoticeably altered when combining multiple sensory information. Carlon [25] showed that users' perception of heaviness can be unnoticeably altered when manipulating their movements in a virtual environment.

The rubber hand illusion is an illusion where users are induced to feel like a rubber hand is part of their body. In VR, proprioceptive and haptic cues can lead to a similar feeling induced either for an arm [241] or for the whole body [154] (see Figure 7). Similarly, proprioception can also be altered by modifying the virtual avatar (i.e., distorting the position or length of the virtual arms and hands) while retaining body ownership, allowing users to explore a bigger area of the virtual environment with their body [52]. Regarding audiovisual illusions, the well-known McGurk effect has been replicated in VR. The McGurk effect happens when the audio of a syllable is paired with visual stimuli of a second syllable, raising the perception of a third, different syllable. This illusion has been used to study how audio spatialization affects speech perception, suggesting that sounds can be located at different positions and still create a correct speech experience [201, 202]. It was also found that the spatial mismatch does not affect immersion levels, suggesting that computational resources devoted to audio localization could be decreased without affecting the overall user experience. Another interesting audiovisual illusion that appears both in conventional media and VR is the ventriloquist effect, where auditory stimuli coming from a distant source seem to emerge from an actors' lips. The best located or dominant modality (usually vision) overrides the spatial information of the weak modality giving rise to the apparent
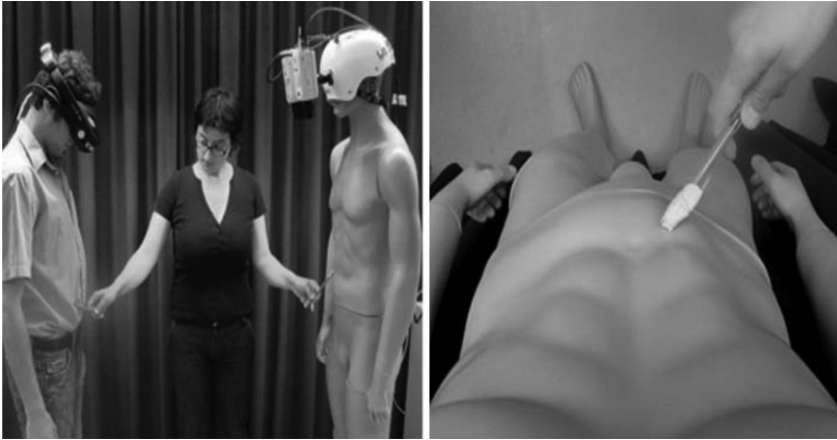
Fig. 7. Multimodality illusions can change the way users perceive both themselves and the environment. For instance, Petkova et al. [154] studied how proprioceptive and haptic cues could lead to body ownership illusions.

translation of sound to the location of the visual stimulus [2]. In this sense, auditory stimuli are affected by visual cues [184], with visual stimuli influencing the processing of binaural directional cues of sound localization. In a complementary way, auditory perception can also act as a support for visual perception, orienting users to regions of interest outside the field of view [95]. Not every audiovisual illusion has to do with speech. In the sound-induced flash illusion [197], a single flash paired with two brief sounds was perceived as two separate flashes. The reverse illusion also happened when two flashes were concurrent with a single beep, raising the perception of a single flash.

In addition to illusions, in which new stimuli are sometimes created, there is also the phenomenon of perceptual suppression, in which one stimulus is no longer (completely or partially) perceived due to an external circumstance. For example, visual suppression is often present in the human visual system. The human brain has evolved to discard visual information when needed to maintain a coherent and stable image of the surrounding environment. Two good examples of visual suppression are blinks and saccades [18, 215], which avoid the processing of blurry information without causing perceptual breaks. Perceptual suppression has been demonstrated and used both in conventional media and in VR [215], usually allowing for environmental changes without the user's awareness, which is useful in many applications, such as navigation in VR. It has also been studied how stimuli of a given modality can alter or suppress information of a different modality, usually visual. In particular, for traditional media, both auditory [70] or haptic [81] stimuli can suppress visual stimuli. Functional imaging studies [99] suggest that crossmodal suppression occurs at neural levels, involving sensory cortices of different modalities. Crossmodal suppression has not been widely studied in VR. However, a recent study [116] shows that auditory stimuli can degrade visual performance in VR using a specific spatiotemporal layout (see Figure 8). Nonetheless, there is still much to investigate about traditional illusions and whether they still hold in virtual environments. The interaction between senses, and in particular the predominance of some of them against the rest, may also diminish or enhance these phenomena, and therefore remains an interesting avenue for future work. We thus believe that a deeper study of crossmodal interactions, both facilitatory and inhibitory, could greatly benefit VR applications, as well as increase our knowledge on sensory perceptual processing in humans.
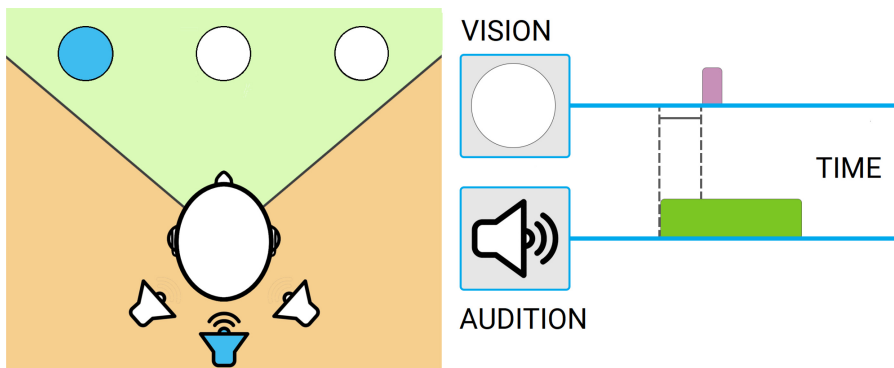
Fig. 8. Spatiotemporal layout of an auditory-triggered effect that degrades visual perception. Sound cues located outside the field of view are concurrent with the appearance of a visual target (inside the field of view), causing the visual target to be missed by participants even when they are directly fixating on it. Figure adapted from Malpica et al. [116].

7 MULTIMODALITY IN NAVIGATION

As discussed in previous sections, agency has an effect in the feeling of realism in a virtual experience, and it is achieved when users feel that their avatar responses are coherent with their real actions, which has a direct implication in body ownership (which also depends on other factors [232]). One characteristic that makes VR intrinsically different from any other traditional media, and that contributes to the users' feeling of control of themselves, is its ability to reproduce each movement of the user into the virtual world. Virtual environments naturally elicit exploration, which usually requires the user to move across the virtual environment. In many cases, movement is heavily constrained by the physical space available [193], and therefore a complete 1:1 reproduction of the movement is not feasible.

Enabling full locomotion in a VR application (i.e., allowing the user to freely move in the virtual space) would increase the possibilities of the virtual experience. However, designers and practitioners are aware of the limited size of physical spaces in which users can consume VR. **Redirected walking techniques (RDW)** emerged in the pursuit of alleviating this limitation: These techniques propose different ways to subtly or overtly manipulate either the user or the environment during locomotion to allow the exploration of virtual worlds larger than the available physical space. Nilsson et al. [144] presented an overview of research works in this field, since redirected walking was first practically demonstrated. Nevertheless, most of these works rely on visual manipulations: Some of them exploit only visual cues or mechanisms, such as saccades [215] or blinks [97] to perform inadvertent discrete manipulations, whereas others exploit continuous manipulations that remain unnoticed by users [127, 193]. However, these previous works do not exploit cues from other sensory modalities. As we have presented along this work, integrating multiple senses can take these kinds of techniques a step further.

Rhythmic auditory stimuli affects how we move [112], and auditory stimuli can be therefore used to actively manipulate our motion perception. Serafin et al. [192] described two psychophysical experiments showing that humans can unknowingly be virtually turned about 20% more or 12% less than their physical rotation by using auditory stimuli: With no visual information available, and with an alarm sound as the only informative cue, users could not reliably discriminate whether their physical rotations had been smaller or larger than the virtual ones. Nogalski and Fohl [147] presented a similar experiment, aiming for detection thresholds for acoustic redirected walking, in

this case by means of wave field synthesis: By designing a scenario surrounded by speakers, and with no visual information available, they demonstrated that some curvature gains can be applied when users walk towards, or turn away from, some sound source. Their work yielded similar rotation detection thresholds of ±20%, which is additionally in line with other works, proving the ability of acoustic signals to manipulate users' movements [49] and the potential benefits of using auditory stimuli in complex navigational tasks. Later, Rewkowski et al. [165] confirmed that RDW with auditory distractors can be safely used in complex navigational tasks such as crossing streets and avoiding obstacles. Nilsson et al. [145] revealed similar detection thresholds for conditions involving moving or static correlated audio-visual stimuli. Additionally, Nogalski and Fohl [148] summarized how users' behavior significantly varies between audio-visual and auditory only stimuli, with the latter yielding more pronounced and less constant curvatures than with audio-visual information.

Other sensory modalities can be used to manipulate users' virtual movement, improve agency and therefore lead to a more natural navigation. Hayashi et al. [67] presented a technique that allows to manipulate the mapping of the user's physical jumping distance and direction. Jumping is an action strongly correlated to proprioception, but it is usually unfeasible due to the available physical space. Manipulating the virtual distance when jumping can allow users to physically jump even when space is constrained, hence proprioceptive cues and realism can be maintained in the experience. Campos et al. [24] also introduced an integration of visual and proprioceptive cues for travelled distance perception, demonstrating that body-based cues contributed to walked distance estimation, attributable to vestibular inputs. Matsumoto et al. [128] presented a combination of redirected walking techniques with visuo-haptic interaction and a path planning algorithm. Haptic feedback directly applied to feet can also influence audiovisual self-motion illusions [143]. Exogenous cues (i.e., any external information coming from the environment) can also play a role in these kinds of manipulations. Feng et al. [50] examined the effects, influence, and interactions of multi-sensory cues during non-fatiguing walking, including movement directional wind, footstep vibrations, and footstep sounds, yielding results that evidenced the improvement on user experience and realism when these cues were available.

In some cases, motion is not possible at all, hence it is necessary to generate an external, visual motion. This self-motion illusion is commonly known as vection and sometimes leads to some postural responses (pursuing a correct vestibular and proprioceptive integration of information). It has been demonstrated that auditory cues increase vection strength in comparison with purely visual cues [88], and that moving sounds enhance circular vection [168]. Moreover, vection may also depend on the environment itself: Meyer et al. [131] explored which factors actually modulate those postural responses and showed that real and virtual foreground objects serve as static visual, auditory, and haptic reference points. Some of the experiments in these works were carried out under rigidly controlled setups and in laboratory conditions, and therefore may not apply to free viewing or other complex conditions. Exploring the effectiveness (or degrading effects) of these insights in more complex scenarios can be an interesting future avenue for research. Finally, the effects of other senses besides auditory and haptics in navigation remain unexplored.

8 APPLICATIONS

We have reviewed different aspects of multimodality in VR, as well as crossmodal interactions between the different sensory modalities, together with different achievable effects. A summary of all those benefits that multimodality can lead to in VR can be seen in Table 5. Different disciplines have leveraged these benefits to enhance different VR applications, showing that multimodality can indeed deliver more realistic and immersive VR experiences. While the application scenarios
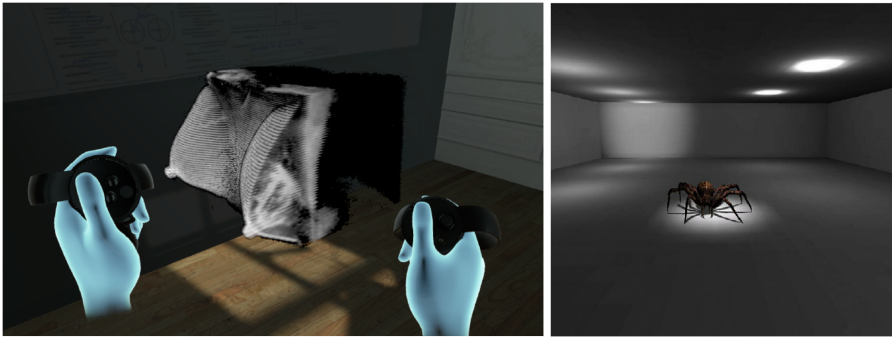
Fig. 9. Two representative examples of different applications of multimodality in medicine. *Left:* Data visualization and manipulation frameworks [160] are important in medical and surgical education and training, and multimodality may enhance the realism and immersion, thus achieving better learning transfer. *Right:* Multimodal VR is a key tool for phobia treatments, since it is able to create realistic environments that face users against their fears, without actually exposing them [200].

range across many disciplines, here, we focus on applications of multimodality to three areas where VR has made a critical impact: medicine, training and education, and entertainment.

8.1 Medicine

The potential uses of VR for medical applications have been studied for decades, and research on this field has evolved as the virtual technologies have done so. Satava et al. [217] presented a review about how VR has become an integral technology of medicine both for professionals and patients alike: from medical image visualization and preoperative planning to teaching and simulation, including teleinterventions and rehabilitation. Other works focused more deeply on the use of VR in the areas of surgical planning, interoperative navigation, and surgical simulations [170, 185]. The possibility of virtualizing a real human body previously scanned and watching it from a far more realistic, immersive perspective than through a conventional display is of great use for health professionals. This has been possible, to a large extent, due to the increasingly photorealistic representation of the anatomy (both in terms of physical tissue properties and of physiological parameters) that virtual environments are achieving.

One of the most pervasive applications of VR in medicine is training, since it can provide a realistic environment for training without the risks of its real counterpart. The enhanced realism and immersion that multimodality provides can lead to improved training and education. Lu et al. [110] presented an audio-visual platform for medical education purposes. One step further, multimodal setups including haptics have been proposed for medical surgery training, where the realism of the feedback significantly improved the learning effect for both virtual [78] and augmented [66] reality interfaces. In the area of medical visualization, Prange et al. [160] also exploited virtual environments and presented a multimodal medical 3D image system where users could walk freely inside a room and interact with the system by means of speech and manipulate patients' information with gestures (see Figure 9). Multimodal VR applications are, however, not constrained to medical training and visualization areas: Psychological research relying on VR has also experienced an unprecedented growth, as Wilson and Soranzo reviewed [237], emphasizing both the advantages (e.g., greater control over stimulus presentation, safe exposure to adverse conditions) and challenges (e.g., VR-induced side effects) of VR in this area. Similarly, Bohil et al. studied the latest advances in VR technology and their applications to neuroscience research [17], highlighting

its high compatibility with medical imaging technologies (such as functional magnetic resonance imaging—fMRI), which allow for a high degree of ecological validity and control over the therapeutic experience.

Other areas that have leveraged the benefits of multimodality are rehabilitative medicine and psychiatry, where significant progress has been made. Psychiatric therapies can benefit from multimodality, since different aspects of behavioral syndromes can be extensively analyzed in virtual environments: Given the suitability of VR to manipulate the virtual world and control certain tasks, it has proven to be a fitting paradigm to treat diseases such as OCD [33] or Parkinson's disease [32]. Phobia treatment is one of the main areas leveraging the benefits of multimodal environments: The realism that multimodality offers over visual-only VR experiences enhances these experiences and increases the effectiveness of the treatment. In addition, VR allows exposing patients to their fears in a safe and highly controlled way, minimizing any potential risks of exposure therapy. Shiban et al. [200] studied the effect of multiple context exposure on renewal in spider phobia (see Figure 9), suggesting that exposure in multiple contexts improves the generalizability of exposure to a new context, therefore helping patients to reduce the chances of future relapses. The work of Hoffman et al. [72] went a step further: They explored not only whether VR exposure therapy reduces fear of spiders, but also concluded that giving patients the illusion of physically touching the virtual spider increases treatment effectiveness. Muhlberger et al. [137] studied the effect of VR in the treatment of fear of flying, exploiting not only visual and acoustic cues, but also proprioceptive information, since motion simulation may increase realism and help induce fear. Later, they studied the long-term effect of the exposure treatment [136], proving its efficacy in treating the fear of flying. The effect of auditory feedback has been studied in other domains, such as the particular case of agoraphobic patients [228], where multimodality increases patients' immersion feeling, hence facilitating emotional responses. However, those techniques should be applied with caution, since large exposures to VR scenarios may hinder patients' ability to distinguish between the real and the virtual world [80], leading to the disorder known as **Chronic Alternate-World Disorder (CAWD)**.

Rehabilitation has also leveraged advances in VR, yielding impressive results. Sano et al. [183] demonstrated that phantom limb pain (which is the sensation of an amputated limb still attached) was reliably reduced when multimodal sensory feedback was included in the VR therapy of patients with brachial plexus avulsion or arm amputation. Fordell et al. [54] presented a treatment method for chronic neglect, where a VR forced feedback interface provided sensorimotor activation in the contra-lesional arm, which, combined with visual scanning training, yielded improvements in activities of daily life that require spatial attention and an improvement in transfer to real life. Moreover, spatialized sound was also beneficial to improve rehabilitation of postural control dysfunction [233].

Table 2 compiles examples leveraging multimodal VR for medical applications. As for the future, VR has the potential to serve medicine even in extreme situations. Virtual care has become an option to foster personalized connections between doctors and patients when in-person appointments are not possible, continuously adapting to the realities of the COVID-19 pandemic [118].

8.2 Education and training

Training and education are areas in which VR holds great promise and in which it has already begun to show its capabilities: Jensen and Konradsen presented a review on the use of VR headsets for training and education and showed that in many cases, better learning transfer can be achieved in this medium compared to traditional media [85].

In education, VR has been widely studied as a new paradigm for teaching: Designing ad hoc environments helps create adequate scenarios for each learning purpose, hence facilitating the

Table 2. Example Works of Different Medical Applications where Multimodality Plays an Important Role

Application	Example work	Additional involved senses (other than vision)				Brief description
		Audition	Proprioception	Haptics	Other	
Rehabilitation	Fordell et al. [54]	✗	✓	✓	✗	Chronic neglect treatment, with a force feedback interface.
	Sano et al. [183]	✓	✓	✓	✗	Multimodal sensory feedback to reduce phantom limb.
Phobia treatment	Viaud et al. [228]	✓	✗	✗	✗	Effects of auditory feedback in agoraphobic patients.
	Mülberger et al. [136, 137]	✓	✓	✗	✗	Multimodality short- and long-term effects on fear of flight.
	Hoffman et al. [72]	✓	✗	✓	✗	Illusions of touching to reduce fear of spiders.
OCD therapy	Cipresso et al. [33]	✗	✓	✗	✗	Different instructions to analyze behavioral syndromes.
Medical data visualization	Prange et al. [160]	✓	✓	✓	✗	Visualize and manipulate patients' medical data in 3D.
Surgery training	Hutchins et al. [78]	✓	✗	✓	✗	Medical training simulator with haptic feedback.
	Harders et al. [66]	✗	✗	✓	✗	Medical training simulator with AR features.
Medical education	Lu et al. [110]	✓	✗	✗	✗	Virtual platform to educate on medicine.

transfer of knowledge to real-life scenarios. Stojvsic et al. [212] reviewed the literature on VR applications in education and conducted a small study showing that teachers perceived benefits in introducing immersive technologies, since students were more motivated and immersed in the topic of interest. At the same time, childhood education processes have been shown to be improved by leveraging multimodality in virtual environments by means of human-computer interaction methods including feedback and interaction from multiple modalities [31] or somatic interaction (hand gestures and body movements) [4, 51].

Many frameworks regarding VR in teaching and education have been studied and evaluated, demonstrating that using virtual manipulatives (i.e., virtual interaction paradigms) that provide multimodal interactions actually yields richer perceptual experiences than classical methodologies in the cases of e.g., mathematics learning [150] or chemistry education [3]. In the case of the latter, a virtual multimodal laboratory was designed, where the user could perform chemistry experiments like in the real world, through a 3D interaction interface with also audio-visual feedback, which indeed improved the learning capabilities of students. Similarly, Tang et al. [219] introduced an immersive multimodal virtual environment supporting interactions with 3D deformable models through haptic devices, where not only gestures were replicated but also touching forces were correctly simulated, hence generating realistic scenarios. One step further, Richard et al. [166] surveyed existing works including haptic or olfactory feedback in the field of education and described a simulation VR platform that provides haptic, olfactory, and auditory feedback, which they tested in various teaching scenarios, demonstrating they affected student engagement and learning positively, and obtaining similar insights as other reviews in educational scenarios, such as in **STEAM (science, technology, engineering, arts, and mathematics)** classrooms [218].

It is worth mentioning that multimodality can also help alleviate sensory impairments, since environments can be designed to maximize the use of the non-affected senses. Following this idea, Yu and Brewster [240] studied the strengths of a multimodal interface (i.e., with speech interactions) against traditional tactile diagrams in conveying information to visually impaired and blind people, showing benefits of this approach in terms of the accuracy obtained by users.

One widespread technique to enhance learning leverages the so-called *serious games*, which enable learning by means of interactive, yet enriching video-games. Checca and Bustillo [29] reviewed the use of immersive VR serious games in this context and their positive effect on learning processes and transfer. Multimodal VR can actually benefit the learning process of these learning-based serious games [41], since multisensory feedback can enhance many of the cognitive processes involved. Covaci et al. [36] presented a multisensory educational game to investigate how

olfactory stimuli could contribute to users' learning experience: It made the experience more enjoyable, but also led to an improvement in users' performance and overall learning.

As aforementioned, multimodal VR has potential in the transfer of knowledge. Given this, it is well suited for simulating and training complex and usually expensive real-life skills requiring high cognitive loads. Gopher [60] highlighted how virtual multimodal training conditions give better results when compared with traditional training conditions in many domains, including sports, rehabilitation, industry, or surgery; with the latter being the core of Van der Meijden et al.'s work [227], which reviewed the use of haptic feedback for surgery training, concluding how the addition of this information yields positive assessments in the majority of the cases and even reduces surgical errors. Transferring learning from training simulators to real-life situations is one of the most relevant parts of the learning process, and multimodality has been proved to enhance it [98].

In the manufacturing industry, many processes require learning specific skills, and multimodal virtual environments can offer new ways of training. Some works have studied the usability of VR for a manufacturing application, such as the assembly of components into a final product, where proprioception and haptic manipulation was required [19]. Other works have proposed a virtual system dedicated to train workers in the use and programming of milling machines, offering visual, audio, and haptic (force) feedback [37], also replacing the use of conventional mechanical milling machines. Since fine motor skills can be transferred to the performance of manual tasks, other studies have analyzed the effectiveness of virtual training in the specific case of industry in contrast to real-life training [158]. At the end, the aforementioned works on virtual skill training agree that virtual training could replace real training, since learning is correctly transferred, and the virtual counterparts are usually less expensive and time-consuming.

VR is also extremely helpful for assembly and maintenance processes (e.g., virtual prototyping [40]), since it provides a cheap method to directly inspect, interact with, and modify 3D prototypes without the need of a physical industrial manufacturing process [195]. In this context, haptic feedback might be crucial to provide feedback in assembly simulations [65].

Other complex tasks can also benefit from multimodal virtual training. MacDonald et al. [111] focused on the air traffic control problem and evaluated the relevant aspects of the auditory modality to improve the detection of sonic warnings, including the best design patterns to maximize performance, signal positioning, and optimal distances on the interaural axis, depending on the sound amplitudes. Real-time acoustic spatialized simulation can be also used in architecture when designing acoustic isolation or studying how sound will be propagated through an indoor environment [229].

All the works mentioned in this section concluded that multimodality offered higher user engagement than unimodal or traditional environments, leading to a better experience and learning transfer. Training in virtual environments has proven to be useful, especially in contexts that are hard or expensive to replicate in real life. However, while VR training is effective, the lack of a particular modality (e.g., haptic feedback when learning to manipulate pumps [238]) could diminish the effectiveness of VR with regard to traditional *hands-on* experiences. Hence, it is important to include all the useful sensory information that is needed for each particular experience and make it as realistic and reliable as possible. A list of some representative applications of multimodal VR in training and education can be found in Table 3.

8.3 Entertainment

Entertainment is undergoing an important revolution with the re-emergence of VR as a new medium: As VR devices become more affordable, their use at consumer level is rapidly increasing. Leisure by means of VR videogames, cinematography, or narrative experiences is becoming

Table 3. Example Works of Different Education and Training Applications where Multimodality Plays an Important Role

| Application	Example work	Additional involved senses (other than vision)				Brief description
		Audition	Proprioception	Haptics	Other	
Education	Christopoulos and Gaitatzes [31]	✓	✓	✗	✗	Children education on history.
	Alves et al. [4]	✗	✓	✗	✗	Serious games for children education on history.
	Ali et al. [3]	✓	✗	✗	✗	Children education on chemistry.
	Tang et al. [219]	✗	✗	✓	✗	Education on deformable materials.
	Lu et al. [110]	✓	✗	✗	✗	Education on medicine.
	Richard et al. [166]	✓	✗	✓	Olfactory	Education on physics.
Accessibility in education	Yu and Brewster [240]	✓	✗	✓	✗	Accessibility for blind people.
Serious games	Deng et al. [41]	✗	✗	✓	✗	Review on multimodality for serious games.
Skill training	Gopher [60]	✓	✗	✓	✗	Review on multimodality for skill training.
	Boud et al. [19]	✗	✗	✓	✗	Skill training for industrial processes.
	Crison et al. [37]	✓	✗	✓	✗	Skill training for industrial processes.
	Ha et al. [65]	✗	✗	✓	✗	Skill training for virtual prototyping.
	MacDonald et al. [111]	✓	✗	✗	✗	Skill training for air traffic control.
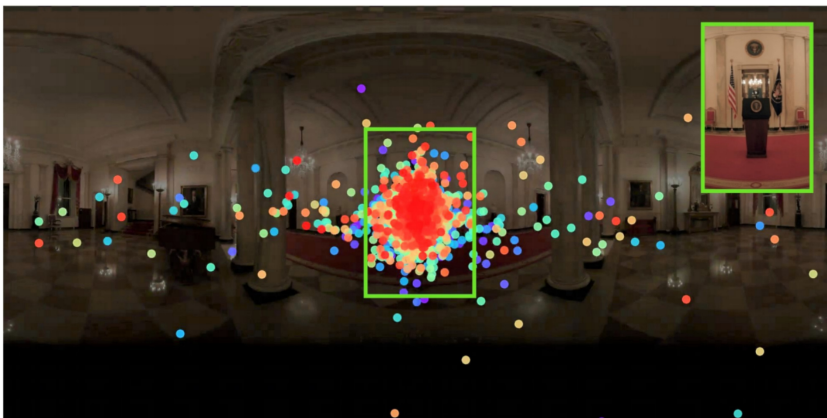
Fig. 10. Representative image of the work of Marañes et al. [117], where they analyze users' gaze behavior during visualization of VR cinematic content. One of the key open problems in VR is the generation of engaging virtual experiences that meet users' expectations. To that end, it is necessary to understand users' behavior in such virtual experiences.

increasingly common, and creating realistic, engaging experiences is the main goal of content creators. Multimodality can be instrumental in improving both realism and engagement.

Videogames allow users to interact with a virtual environment, controlling characters or avatars that respond based on their actions. Traditional videogames have leveraged narrative characteristics to connect with the player, to immerse them in the virtual world, so the experience feels more engaging. With the appearance of VR, immersive games are evolving: Higher realism and stronger feelings of presence and agency can potentially be achieved now with this technology.

Nesbitt and Hoskens [139] hypothesized that integrating information from different senses could assist players in their performance. They evaluated visual, auditory, and haptic

information combinations, and although no significant performance improvement was achieved, players reported improved immersion, confidence, and satisfaction in the multisensory cases. Since haptic devices may enhance the experience, some works have devoted in developing different toolkits to offer these interactions in VR (e.g., vibrotactile interactions [122]), while other works have exploited somatic interactions, including not only haptic but whole proprioceptive cues. Alves et al. [4] studied user experience in games that included hand gestures and body movements, identifying problems and potential uses of gestural interaction devices in an integrated manner. Many narrative experiences may require the user to have the feeling of walking, and it may be one of the hardest scenarios to get a realistic response, since multiple sensory information is combined. In this scope, some works investigated the addition of multisensory walking-related cues in locomotion [93], showing that adding auditory cues (i.e., footstep sounds), visual cues (i.e., head motions during walking), and vibrotactile cues (under participants' feet) could all enhance participants' sensation of self-motion (vection) and presence. Sometimes, full locomotion is not permitted, however, realism can still be achieved: Colley et al. [35] went a step further in exploiting body proprioception, presenting a work that proposed using an HMD in skiing and snowboarding training while the user was on a real slope, so proprioceptive cues were completely realistic.

Although many of the current VR videogames exploit audiovisual and somatic cues (which are the easiest to provide with current technology), some have tried to work with additional cues. As in previously mentioned learning processes, some works have explored the use of olfactory cues to investigate how enabling olfaction can contribute to users' learning performance, engagement, and quality of experience [36], although this modality still remains in an early exploratory phase.

In a similar manner, gustatory cues have been studied in several works. Arnold et al. [8] presented a game involving eating real food to survive, which, combined with the capture and reproduction of chewing sounds, increased the realism of the experience. Following this line, Mueller et al. [135] highlighted the potential technologies and designs to support eating as a form of play.

Multisensory feedback can enhance many of the high and complex cognitive processes involved in VR [41]. Particularly, multimodality can trigger different emotional responses in immersive games: Kruijff et al. [94] investigated those emotional effects and proposed guidelines that can be applied to reproduce diverse emotional responses in multimodal games.

Within the wide area of entertainment, cinematographic and narrative experiences in VR have been emerging during the past years. As explored in Section 4, guiding users' attention is especially challenging in virtual environments, where users cannot see the whole scenario at once. Although some traditional continuity editing rules may still apply [194], and visual cuts may impact users' behavior [117], the presence of directional sounds can also influence how users explore immersive environments [124], thus special attention must be paid to sound design when considering narrative experiences in VR [61]. To explore how different cues may actually define how users drive their attention in cinematic VR, Rothe et al. [173] investigated implicitly guiding the attention of the viewer by means of lights, movements, and sounds, integrating auditory and visual modalities, while Ranasinghe et al. [164] proposed adding olfactory and haptic (thermal and wind) stimuli to virtual narrative experiences to achieve enhanced sensory engagement. A compilation of representative applications of multimodal VR in entertainment can be found in Table 4.

9 CONCLUSIONS

Virtual reality can dramatically change the way we create and consume content in many aspects of our everyday life, including entertainment, training, design and manufacturing, communication, or advertising. In the past years, it has been rapidly growing and evolving as a field, with the thrust of impressive technical innovations in both acquisition and visualization hardware and software. However, if this new medium is going to succeed, it will be based on its ability to create *compelling*

Table 4. Example Works of Different Applications in Entertainment Where Multimodality Plays
an Important Role

Application	Example work	Additional involved senses (other than vision)				Brief description
		Audition	Proprioception	Haptics	Other	
Videogames	Nesbitt et al. [139]	✓	✗	✓	✗	Multimodality to assist players' performance.
	Martinez et al. [122]	✗	✗	✓	✗	Vibrotactile toolkit for immersive videogames.
	Alves et al. [4]	✗	✓	✗	✗	Serious games for children education on history.
Physical activity simulation	Kruijff et al. [93]	✓	✓	✓	✗	Walking simulation for leisure applications.
	Colley et al. [35]	✗	✓	✗	✗	Proprioceptive cues to simulate skiing.
Cognitive and emotional effects	Kruijff et al. [94]	✓	✗	✓	Olfactory	Study of emotional responses in virtual experiences.
	Deng et al. [41]	✓	✗	✓	✗	Cognitive load and processes in serious games.
Narrative experiences	Rothe et al. [173, 174]	✓	✗	✗	✗	Attention guidance in narrative experiences.
	Ranasinghe et al. [164]	✓	✗	✓	Olfactory	Enhancing engagement in narrative experiences.

user experiences. The interaction between different sensory modalities (such as the five senses, or proprioception) has always been of interest to content creators, but in a VR setting, in which the user is immersed in an alternative reality, the importance of multimodal sensory input plays a more relevant role, since the feedback either from any modality, or from the combination of multiple of them, affects the final experience. In fact, it becomes both a possible liability, if not handled properly, and a potential strength, that if adequately leveraged can boost realism, help direct user attention, or improve user performance. Throughout this survey, we have summarized not only the main lines of research in these areas, but also outlined relevant insights for future directions in each of them.

While making use of multimodal setups can provide benefits to the experience, it also increases costs and complexity. From the point of view of the hardware, however, audiovisual integration is almost always present in current systems (see Table 1), and this is also the case for proprioception (except for smartphone-based and related headsets). Most controllers also include some kind of haptic feedback, although in this case it is quite simple and rudimentary, with ample room for improvement and sophistication in consumer-level systems. While research-level technology in haptics is quite advanced, transforming it into consumer-level solutions has been and still is a challenge, due to systems complexity, durability, or cost. We are currently witnessing the first attempts at providing more sophisticated haptic interactions with simpler, consumer-level hardware by using ultrasound; and certainly more advances are to be expected in this area, given the importance of haptics to the multimodal experience. Taste and smell are almost untapped in terms of hardware. Unlike the case of touch, where haptics is abundantly explored at research level, these senses are in their infancy from a research standpoint as well. Thus, special effort should be made towards developing hardware that is able to simulate compelling stimuli for these underexplored senses. From the point of view of the software, inclusion of multimodal input increases the bandwidth and computational resources needed, both current stumbling blocks of VR experiences, particularly collaborative ones. Thus, compression techniques and computational optimizations (both hardware- and software-based) are two of the most active areas of research in VR that would also help an increased use of multimodal input. At the same time, works have shown that multimodal input can help maintain realism and immersion with lower quality visual input, so it can also be an advantage in these areas. Additionally, even if it implies an increase in cost and complexity, and depending on the final application scenario, these increased costs may still be more than

Table 5. Illustrative Examples of Different Effects That Can Be Achieved through Multimodality in VR

Effect	How	Sensory modalities (other than visual)				Related works
		Auditory	Haptic	Proprioc.	Other	
Better perception and understanding of the environment (Section 3)	Integrating the visual appearance of an object or scene in the VE with a realistic and coherent sound of the same object	✓	✗	✗	✗	[79, 115]
	Correctly spatializing audio in a 360° environment	✓	✗	✗	✗	[58, 62, 77, 134, 230]
	Enhancing realism by touching a physical object and a virtual object at the same time	✗	✓	✓	Olfactory, gustatory	[71, 73]
	Proprioceptive cues (e.g., a hand instead of a pointer) are suitable for interaction scenarios	✗	✗	✓	✗	[156]
Increase in spatial awareness (Section 3, 7)	Training users to navigate a VE using auditory reflections and reverberation	✓	✗	✗	✗	[6]
	Creating sensory illusions with worn haptic devices that map different points in the VE to the body	✗	✓	✓	✗	[101]
	Being able to control a virtual body	✗	✗	✓	✗	[187]
Increase in the feeling of presence, realism, or immersion (Section 3)	Body-centered perception (hand, face, and trunk) achieved via spatio-temporal multisensory information integration within peripersonal space	✓	✓	✓	Olfactory	[16, 154, 178]
	Achieving partial body ownership considering first person visuo-spatial viewpoint and anatomical similarity	✗	✓	✓	✗	[159, 241]
	Integrating as much multisensory information as possible (spatialized sound, vibrations, wind, real objects, physical movement, scent) coherently with the VE	✓	✓	✓	Olfactory	[59, 68, 86]
	Use of soundscapes or zeitgebers in the virtual scenes	✓	✗	✗	✗	[104, 190]
Self-motion (also related to presence) (Section 7)	Adding walking-related multimodal cues, integrating vibrations in standing or even sitting users	✓	✓	✓	✗	[92, 93, 126, 131]
Modification of the saliency of the environment (Section 4)	Users tend to pay attention to regions with more sensory information (mostly explored in traditional media).	✓	✗	✗	✗	[38, 39, 43, 87, 132, 138]
	Auditory and visual stimuli are spatio-temporally correlated to increase saliency	✓	✗	✗	✗	[27]
Guidance or direction of attention (Section 4)	Using diegetic sound and moving cues to trigger exploratory behavior	✓	✗	✗	✗	[23, 153, 173, 174, 180]
	Spatialized auditory stimuli can also direct attention to specific parts of the environment, including those outside the field of view	✓	✗	✗	✗	[12, 13, 21, 95]
Improvement of user performance (Section 5)	Simultaneous stimuli of irrelevant sensory modalities increase detection of target stimulus and can also decrease search times or increase memory retention	✓	✓	✗	Olfactory	[109, 114, 146, 226]
	Auditory stimuli improve spatial awareness and thus reduce search times and improve spatial reasoning	✓	✗	✗	✗	[57, 62, 171, 230]
	Cognitive process starts faster in the presence of multimodality, allowing users to pay attention to more cues and details	✓	✓	✗	✗	[68]
	Combination of audio-haptic tempos to convey spatial information and relevant stimuli increase performance in search tasks	✓	✓	✗	✗	[5]
	Auditory stimuli facilitate touch of virtual objects outside the field of view	✓	✓	✗	✗	[90]
	Improvement of depth perception via simulated touch with a force-feedback system	✗	✓	✗	✗	[20, 216]
Simulation of physical properties (Section 3, 5)	Increasing the space that the user can explore with their body modifying the virtual avatar	✗	✗	✓	✗	[52, 53]
	Weight or touch simulation by haptic feedback	✗	✓	✓	✗	[25, 181]

advantageous if the alternative is setting up a similar, real scenario, in, e.g., emergency or medical training.

The inherent increased complexity resulting from the interaction between sources also poses a challenge for researchers in this area. We have reviewed a number of studies analyzing the interaction of *two* sensory modalities. Most of them were based on constrained experiments under laboratory conditions. However, the final goal of VR is to be present at consumer level, where more complex phenomena and interactions are likely to happen. Thus, lifting constraints on the experimental conditions, and exploring to what extent the insights found generalize and hold in free-viewing scenarios with more confounding factors, remains a critical avenue of future work, which undoubtedly needs to be built on the findings from controlled, constrained experiments. Works exploring three or more modalities are more rare. The integration of input from multiple senses has been an open area of research for over a century, partly because of the curse of dimensionality into which one runs when tackling this problem: The size of the parameter space grows exponentially and soon becomes intractable. Even when the data was available, deriving models to explain it has been a challenge, and analytical models often were not able to explain phenomena outside the particular scenario and parameter space explored, partly because of their lack of generality, partly because the type of data gathered can be very sensitive to the particular experimental setup. Current data-driven approaches certainly provide a new tool to address the problem, and some works have already started to rely on them, as is the case with audiovisual attention modeling. For this to be a solid path forward, however, we need public, carefully crafted datasets that can be used by the community and in benchmarks, and we need reproducible experimental setups. Incidentally, VR is in itself a great experimental scenario for reproducibility, as opposed to physical, real-world setups.

Being aware of how the different sensory inputs interact thus helps researchers and practitioners in the field in two ways: In a first level, it aids them to create believable, successful experiences with the limited hardware and software resources available to them. At the next level, they can leverage the way the different sensory inputs will interact to overcome some of the limitations imposed by the hardware and software available, and even to improve the *design* of such hardware and software. As multimodal interactions become known and well understood, they can then be leveraged for algorithm design, content generation, or even hardware development, essentially contributing to create better virtual experiences for users and helping unleash the true potential of this medium.

REFERENCES

[1] Z. Akhtar and T. H. Falk. 2017. Audio-visual multimedia quality assessment: A comprehensive survey. *IEEE Access* 5 (2017), 21090–21117.

[2] D. Alais and D. Burr. 2004. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 3 (2004).

[3] Numan Ali, Sehat Ullah, Ihsan Rabbi, and Aftab Alam. 2014. The effect of multimodal virtual chemistry laboratory on students' learning improvement. In *International Conference on Augmented and Virtual Reality*. Springer, 65–76.

[4] Luis Miguel Alves Fernandes et al. 2016. Exploring educational immersive videogames: An empirical study with a 3D multimodal interaction prototype. *Behav. Inf. Technol.* 35, 11 (2016), 907–918.

[5] Mehdi Ammi and Brian F. G. Katz. 2014. Intermodal audio-haptic metaphor: Improvement of target search in abstract environments. *Int. J. Hum.-comput. Interact.* 30, 11 (2014), 921–933.

[6] A. Andreasen, M. Geronazzo, N. Nilsson, J. Zovnercuka, K. Konovalov, and S. Serafin. 2019. Auditory feedback for navigation with echoes in virtual environments: Training procedure and orientation strategies. *IEEE Trans. Visualiz. Comput. Graph.* 25, 5 (2019), 1876–1886.

[7] Claudia Armbrüster, Marc Wolter, Torsten Kuhlen, Will Spijkers, and Bruno Fimm. 2008. Depth perception in virtual reality: Distance estimations in peri-and extrapersonal space. *Cyberpsychol. Behav.* 11, 1 (2008), 9–15.

[8] Peter Arnold. 2017. You better eat to survive! Exploring edible interactions in a virtual reality game. In *CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 206–209.

[9] Barry Arons. 1992. A review of the cocktail party effect. *J. Amer. Voice I/O Societ.* 12, 7 (1992).

[10] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, and Mohan S. Kankanhalli. 2010. Multimodal fusion for multimedia analysis: A survey. *Multim. Syst.* 16, 6 (2010), 345–379.

[11] Hudson Diggs Bailey, Aidan B. Mullaney, Kyla D. Gibney, and Leslie Dowell Kwakye. 2018. Audiovisual integration varies with target and environment richness in immersive virtual reality. *Multisens. Res.* 31, 7 (2018).

[12] Paulo Bala, Raul Masu, Valentina Nisi, and Nuno Nunes. 2018. Cue control: Interactive sound spatialization for 360° videos. In *International Conference on Interactive Digital Storytelling*. Springer, 333–337.

[13] Paulo Bala, Raul Masu, Valentina Nisi, and Nuno Nunes. 2019. "When the elephant trumps"—A comparative study on spatial audio for orientation in 360º videos. In *Conference on Human Factors in Computing Systems*. 1–13.

[14] Rosa María Baños, Cristina Botella, Mariano Alcañiz, Víctor Liaño, Belén Guerrero, and Beatriz Rey. 2004. Immersion and emotion: Their impact on the sense of presence. *Cyberpsychol. Behav.* 7, 6 (2004), 734–741.

[15] A. Barghout, J. Cha, A. El Saddik, J. Kammerl, and E. Steinbach. 2009. Spatial resolution of vibrotactile perception on the human forearm when exploiting funneling illusion. In *IEEE International Workshop on Haptic Audiovisual Environments and Games*.

[16] Olaf Blanke, Mel Slater, and Andrea Serino. 2015. Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron* 88, 1 (2015), 145–166.

[17] Corey J. Bohil, Bradly Alicea, and Frank A. Biocca. 2011. Virtual reality in neuroscience research and therapy. *Nat. Rev. Neurosci.* 12, 12 (2011), 752–762.

[18] Benjamin Bolte and Markus Lappe. 2015. Subliminal reorientation and repositioning in immersive virtual environments using saccadic suppression. *IEEE Trans. Visualiz. Comput. Graph.* 21, 4 (2015), 545–552.

[19] A. C. Boud, Chris Baber, and S. J. Steiner. 2000. Virtual reality: A tool for assembly? *Pres.: Teleop. Virt. Environ.* 9, 5 (2000), 486–496.

[20] Laroussi Bouguila, Masahiro Ishii, and Makoto Sato. 2000. Effect of coupling haptics and stereopsis on depth perception in virtual environment. In *1st Workshop on Haptic Human Computer Interaction*. 54–62.

[21] A. Brown, A. Sheikh, M. Evans, and Z. Watson. 2016. Directing attention in 360-degree video. In *IBC Conference*.

[22] E. Burns, S. Razzaque, A. T. Panter, M. C. Whitton, M. R. McCallus, and F. P. Brooks. 2005. The hand is slower than the eye: A quantitative exploration of visual dominance over proprioception. In *IEEE Virtual Reality Conference*. 3–10.

[23] Anıl Çamcı. 2019. Exploring the effects of diegetic and non-diegetic audiovisual cues on decision-making in virtual reality. In *16th Sound and Music Computing Conference*. 28–31.

[24] Jennifer L. Campos, John S. Butler, and Heinrich H. Bülthoff. 2012. Multisensory integration in the estimation of walked distances. *Experim. Brain Res.* 218, 4 (2012), 551–565.

[25] Alexsis Danae Carlon. 2018. *Virtual Reality's Utility for Examining the Multimodal Perception of Heaviness*. Ph.D. Dissertation. California State University, Fresno.

[26] Alan Chalmers, Kurt Debattista, and Belma Ramic-Brkic. 2009. Towards high-fidelity multi-sensory virtual environments. *Vis. Comput.* 25, 12 (2009), 1101.

[27] F. Chao, C. Ozcinar, C. Wang, E. Zerman, L. Zhang, W. Hamidouche, O. Deforges, and A. Smolic. 2020. Audio-visual perception of omnidirectional video for virtual reality applications. In *IEEE International Conference on Multimedia & Expo Workshops*.

[28] Guillaume Chauvel, Gabriele Wulf, and François Maquestiaux. 2015. Visual illusions can facilitate sport skill learning. *Psychon. Bull. Rev.* 22, 3 (2015), 717–721.

[29] David Checa and Andres Bustillo. 2019. A review of immersive virtual reality serious games to enhance learning and training. *Multim. Tools Applic.* 79, 9 (2019), 1–27.

[30] Haonan Cheng and Shiguang Liu. 2019. Haptic force guided sound synthesis in multisensory virtual reality (VR) simulation for rigid-fluid interaction. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE.

[31] Dimitrios Christopoulos and Athanasios Gaitatzes. 2009. Multimodal interfaces for educational virtual environments. In *13th Panhellenic Conference on Informatics*. IEEE, 197–201.

[32] P. Cipresso, G. Albani et al. 2014. Virtual multiple errands test (VMET): A virtual reality-based tool to detect early executive functions deficit in Parkinson's disease. *Front. Behav. Neurosci.* 8 (2014), 405.

[33] P. Cipresso, F. La Paglia, C. La Cascia, G. Riva, G. Albani, and D. La Barbera. 2013. Break in volition: A virtual reality study in patients with obsessive-compulsive disorder. *Experim. Brain Res.* 229, 3 (2013), 443–449.

[34] Jonathan Cole and Barbara Montero. 2007. Affective proprioception. *Janus Head* 9, 2 (2007), 299–317.

[35] Ashley Colley, Jani Väyrynen, and Jonna Häkkilä. 2015. Skiing in a blended virtuality: An in-the-wild experiment. In *19th International Academic Mindtrek Conference*. 89–91.

[36] A. Covaci, G. Ghinea, C. Lin, S. Huang, and J. Shih. 2018. Multisensory games-based learning-lessons learnt from olfactory enhancement of a digital board game. *Multim. Tools Applic.* 77, 16 (2018).

[37] F. Crison, A. Lecuyer, D. d'Huart, J. Burkhardt, G. Michel, and J. Dautin. 2005. Virtual technical trainer: Learning how to use milling machines with multi-sensory feedback in virtual reality. In *IEEE Virtual Reality Conference*. 139–145.

[38] Bert De Coensel and Dick Botteldooren. 2010. A model of saliency-based auditory attention to environmental sound. In *20th International Congress on Acoustics (ICA'10)*. 1–8.

[39] Bert De Coensel, Dick Botteldooren, Birgitta Berglund, and Mats E. Nilsson. 2009. A computational model for auditory saliency of environmental sound. *J. Acoust. Societ. Amer.* 125, 4 (2009), 2528–2528.

[40] Antonino Gomes De Sa and Gabriel Zachmann. 1999. Virtual reality as a tool for verification of assembly and maintenance processes. *Comput. Graph.* 23, 3 (1999), 389–403.

[41] Shujie Deng, Julie A. Kirkby, Jian Chang, and Jian J. Zhang. 2014. Multimodality with eye tracking and haptics: A new horizon for serious games? *Int. J. Ser. Games* 1, 4 (2014), 17–34.

[42] Johannes Dichgans and Thomas Brandt. 1978. Visual-vestibular interaction: Effects on self-motion perception and postural control. In *Perception*. Springer, 755–804.

[43] Varinthira Duangudom and David V. Anderson. 2007. Using auditory saliency to understand complex auditory scenes. In *15th European Signal Processing Conference*. IEEE, 1206–1210.

[44] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. 2009. Multimodal interfaces: A survey of principles, models and frameworks. In *Human Machine Interaction*. Springer, 3–26.

[45] Ragnhild Eg and Dawn M. Behne. 2015. Perceived synchrony for realistic and dynamic audiovisual events. *Front. Psychol.* 6 (2015), 736.

[46] Mohammed S. Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. 2018. Toward low-latency and ultra-reliable virtual reality. *IEEE Netw.* 32, 2 (2018), 78–84.

[47] Emerge. 2021. Bringing touch and emotion to virtual experiences. Retrieved from https://emerge.io/.

[48] G. Evangelopoulos, A. Zlatintsi, A. Potamianos, P. Maragos, K. Rapantzikos, G. Skoumas, and Y. Avrithis. 2013. Multimodal saliency and fusion for movie summarization based on aural, visual, and textual attention. *IEEE Trans. Multim.* 15, 7 (2013), 1553–1568.

[49] Tobias Feigl, Eliise Kõre, Christopher Mutschler, and Michael Philippsen. 2017. Acoustical manipulation for redirected walking. In *23rd ACM Symposium on Virtual Reality Software and Technology*. 1–2.

[50] Mi Feng, Arindam Dey, and Robert W. Lindeman. 2016. The effect of multi-sensory cues on performance and experience during walking in immersive virtual environments. In *IEEE Virtual Reality Conference (VR'16)*. IEEE, 173–174.

[51] Luís Fernandes et al. 2015. Bringing user experience empirical data to gesture-control and somatic interaction in virtual reality videogames: An exploratory study with a multimodal interaction prototype. In *Conferência Ciências E Tecnologias Da Interação (SciTecIn'15)*.

[52] Tiare Feuchtner and Jörg Müller. 2017. Extending the body for interaction with reality. In *Conference on Human Factors in Computing Systems*. 5145–5157.

[53] Tiare Feuchtner and Jörg Müller. 2018. Ownershift: Facilitating overhead interaction in virtual reality with an ownership-preserving hand space shift. In *31st ACM Symposium on User Interface Software and Technology*.

[54] Helena Fordell, Kenneth Bodin, Anders Eklund, and Jan Malm. 2016. RehAtt-Scanning training for neglect enhanced by multi-sensory stimulation in virtual reality. *Topics Stroke Rehabil.* 23, 3 (2016), 191–199.

[55] Laura Freina and Michela Ott. 2015. A literature review on immersive virtual reality in education: State of the art and perspectives. In *International Scientific Conference on Elearning and Software for Education*. 10–1007.

[56] Alberto Gallace, Mary K. Ngo, John Sulaitis, and Charles Spence. 2012. Multisensory presence in virtual reality: Possibilities & limitations. In *Multiple Sensorial Media Advances and Applications*. IGI Global, 1–38.

[57] Ruohan Gao, Changan Chen, Ziad Al-Halah, Carl Schissler, and Kristen Grauman. 2020. VisualEchoes: Spatial image representation learning through echolocation. *arXiv preprint arXiv:2005.01616* (2020).

[58] R. Gao and K. Grauman. 2019. 2.5D visual sound. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[59] Guilherme Gonçalves, Miguel Melo, José Vasconcelos-Raposo, and Maximino Bessa. 2019. Impact of different sensory stimuli on presence in credible virtual environments. *IEEE Trans. Visualiz. Comput. Graph.* 26, 11 (2019).

[60] Daniel Gopher. 2012. Skill training in multimodal virtual environments. *Work* 41, Supplement 1 (2012), 2284–2287.

[61] M. Gospodarek, A. Genovese, D. Dembeck, C. Brenner, Ag. Roginska, and K. Perlin. 2019. Sound design and reproduction techniques for co-located narrative VR experiences. In *Audio Engineering Society Convention*.

[62] Matti Groehn, Tapio Lokki, Lauri Savioja, and Tapio Takala. 2001. Some aspects of role of audio in immersive visualization. In *Visual Data Exploration and Analysis VIII*, Vol. 4302. International Society for Optics and Photonics.

[63] J. Gugenheimer, D. Wolf, G. Haas, S. Krebs, and E. Rukzio. 2016. SwiVRChair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. In *Conference on Human Factors in Computer Systems*.

[64] Hayrettin Gürkök and Anton Nijholt. 2012. Brain–computer interfaces for multimodal interaction: A survey and principles. *Int. J. Hum.-comput. Interact.* 28, 5 (2012), 292–307.

[65] Sungdo Ha, Laehyun Kim, Sehyung Park, Cha-Soo Jun, and H. Rho. 2009. Virtual prototyping enhanced by a haptic interface. *CIRP Ann.* 58, 1 (2009), 135–138.

[66] Matthias Harders, Gerald Bianchi, and Benjamin Knoerlein. 2007. Multimodal augmented reality in medicine. In *International Conference on Universal Access in Human-computer Interaction*. Springer, 652–658.

[67] O. Hayashi, K. Fujita, K. Takashima, R. Lindernan, and Y. Kitarnura. 2019. Redirected jumping: Imperceptibly manip-
      ulating jump motions in virtual reality. In *IEEE Conference on Virtual Reality and 3D User Interfaces*.
[68] David Hecht, Miriam Reiner, and Gad Halevy. 2006. Multimodal virtual environments: Response times, attention,
      and presence. *Pres.: Teleop. Virt. Environ.* 15, 5 (2006), 515–523.
[69] Nicolas S. Herrera and Ryan P. McMahan. 2014. Development of a simple and low-cost olfactory display for immer-
      sive media experiences. In *ACM International Workshop on Immersive Media Experiences*. 1–6.
[70] Souta Hidaka and Masakazu Ide. 2015. Sound can suppress visual perception. *Sci. Rep.* 5, 1 (2015), 1–9.
[71] Hunter G. Hoffman. 1998. Physically touching virtual objects using tactile augmentation enhances the realism of
      virtual environments. In *IEEE Virtual Reality Annual International Symposium*. IEEE, 59–63.
[72] Hunter G. Hoffman, Azucena Garcia-Palacios, Albert Carlin, Thomas A. Furness Iii, and Cristina Botella-Arbona.
      2003. Interfaces that heal: Coupling real and virtual objects to treat spider phobia. *Int. J. HCI* 16, 2 (2003).
[73] Hunter G. Hoffman, Ari Hollander, Konrad Schroder, Scott Rousseau, and Tom Furness. 1998. Physically touching
      and tasting virtual objects enhances the realism of virtual experiences. *Virt. Real.* 3, 4 (1998), 226–234.
[74] I. P. Howard, H. L. Jenkin, and G. Hu. 2000. Visually-induced reorientation illusions as a function of age. *Aviat., Space,
      Environ. Med.* 71, 9 Suppl (2000), A87–91.
[75] Ping Hu, Qi Sun, Piotr Didyk, Li-Yi Wei, and Arie E. Kaufman. 2019. Reducing simulator sickness with perceptual
      camera control. *ACM Trans. Graph.* 38, 6 (2019), 1–12.
[76] Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. 2020. DGaze: CNN-based
      gaze prediction in dynamic scenes. *IEEE Trans. Visualiz. Comput. Graph.* 26, 5 (2020), 1902–1911.
[77] Haikun Huang, Michael Solah, Dingzeyu Li, and Lap-Fai Yu. 2019. Audible panorama: Automatic spatial audio gen-
      eration for panorama imagery. In *CHI Conference on Human Factors in Computing Systems*.
[78] Matthew Hutchins, Matt Adcock, Duncan Stevenson, Chris Gunn, and Alexander Krumpholz. 2005. The design of
      perceptual representations for practical networked multimodal virtual training environments. In *11th International
      Conference on Human-Computer Interaction*.
[79] T. Iachini, L. Maffei, F. Ruotolo, V. Senese, G. Ruggiero, M. Masullo, and N. Alekseeva. 2012. Multisensory assessment
      of acoustic comfort aboard metros: A virtual reality study. *Appl. Cog. Psychol.* 26, 5 (2012).
[80] Atsushi Ichimura, Isao Nakajima, and Hiroshi Juzoji. 2001. Investigation and analysis of a reported incident resulting
      in an actual airline hijacking due to a fanatical and engrossed VR state. *CyberPsychol. Behav.* 4, 3 (2001).
[81] Masakazu Ide and Souta Hidaka. 2013. Tactile stimulation can suppress visual perception. *Sci. Rep.* 3 (2013).
[82] Hiroyuki Ito. 2012. Cortical shape adaptation transforms a circle into a hexagon: A novel afterimage illusion. *Psychol.
      Sci.* 23, 2 (2012), 126–132.
[83] Alejandro Jaimes and Nicu Sebe. 2007. Multimodal human–computer interaction: A survey. *Comput. Vis. Image Un-
      derst.* 108, 1-2 (2007), 116–134.
[84] Seungwoo Je, Myung Jin Kim, Woojin Lee, Byungjoo Lee, Xing-Dong Yang, Pedro Lopes, and Andrea Bianchi. 2019.
      Aero-plane: A handheld force-feedback device that renders weight motion illusion on a virtual 2D plane. In *ACM
      Symposium on User Interface Software and Technology*. 763–775.
[85] Lasse Jensen and Flemming Konradsen. 2018. A review of the use of virtual reality head-mounted displays in educa-
      tion and training. *Educ. Inf. Technol.* 23, 4 (2018), 1515–1529.
[86] S. Jung, A. Wood, S. Hoermann, P. Abhayawardhana, and R. Lindeman. 2020. The impact of multi-sensory stimuli on
      confidence levels for perceptual-cognitive tasks in VR. In *IEEE Conference on Virtual Reality and 3D User Interfaces*.
[87] Christoph Kayser, Christopher I. Petkov, Michael Lippert, and Nikos K. Logothetis. 2005. Mechanisms for allocating
      auditory attention: An auditory saliency map. *Curr. Biol.* 15, 21 (2005), 1943–1947.
[88] Behrang Keshavarz, Lawrence J. Hettinger, Daniel Vena, and Jennifer L. Campos. 2014. Combined effects of auditory
      and visual cues on the perception of vection. *Experim. Brain Res.* 232, 3 (2014), 827–836.
[89] Konstantina Kilteni, Raphaela Groten, and Mel Slater. 2012. The sense of embodiment in virtual reality. *Pres.: Teleop.
      Virt. Environ.* 21, 4 (2012), 373–387.
[90] Zentaro Kimura and Mie Sato. 2020. Auditory stimulation on touching a virtual object outside a user's field of view.
      In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW'20)*. IEEE.
[91] Thomas Koelewijn, Adelbert Bronkhorst, and Jan Theeuwes. 2010. Attention and the multiple stages of multisensory
      integration: A review of audiovisual studies. *Acta Psychol.* 134, 3 (2010), 372–384.
[92] Kean Kouakoua, Cyril Duclos, Rachid Aissaoui, Sylvie Nadeau, and David R. Labbe. 2020. Rhythmic proprioceptive
      stimulation improves embodiment in a walking avatar when added to visual stimulation. In *IEEE Conference on
      Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW'20)*. IEEE, 573–574.
[93] E. Kruijff, A. Marquardt, C. Trepkowski, R. Lindeman, A. Hinkenjann, J. Maiero, and B. Riecke. 2016. On your feet!
      Enhancing vection in leaning-based interfaces through multisensory stimuli. In *Symposium on Spatial User Interac-
      tions*.

[94] E. Kruijff, A. Marquardt, C. Trepkowski, J. Schild, and A. Hinkenjann. 2015. Enhancing user engagement in immersive games through multisensory cues. In *International Conference on Games and Virtual Worlds for Serious Applications*.

[95] Mikko Kytö, Kenta Kusumoto, and Pirkko Oittinen. 2015. The ventriloquist effect in augmented reality. In *IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 49–53.

[96] Christophe Lalanne and Jean Lorenceau. 2004. Crossmodal integration for perception and action. *J. Physiol.-Paris* 98, 1–3 (2004), 265–279.

[97] E. Langbehn, F. Steinicke, M. Lappe, G. Welch, and G. Bruder. 2018. In the blink of an eye: Leveraging blink-induced suppression for imperceptible position and orientation redirection in virtual reality. *ACM Trans. Graph.* 37, 4 (2018).

[98] C. Lathan, M. Tracey, M. Sebrechts, D. Clawson, and G. Higgins. 2002. Using virtual environments as training simulators: Measuring transfer. *Handb. Virt. Environ.: Des., Implement. Applic.* (2002).

[99] Paul J. Laurienti, Jonathan H. Burdette, Mark T. Wallace, Yi-Fen Yen, Aaron S. Field, and Barry E. Stein. 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. *J. Cog. Neurosci.* 14, 3 (2002).

[100] Joseph J. LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* 32, 1 (2000).

[101] G. Lawson, T. Roper, and C. Abdullah. 2016. Multimodal "sensory illusions" for improving spatial awareness in virtual environments. In *European Conference on Cognitive Ergonomics*.

[102] Stephen D. Laycock and A. M. Day. 2007. A survey of haptic rendering techniques. In *Computer Graphics Forum*, Vol. 26. Wiley Online Library, 50–65.

[103] Lan Li, Fei Yu, Dongquan Shi, Jianping Shi, Zongjun Tian, Jiquan Yang, Xingsong Wang, and Qing Jiang. 2017. Application of virtual reality technology in clinical medicine. *Amer. J. Translat. Res.* 9, 9 (2017).

[104] Haodong Liao, Ning Xie, Huiyuan Li, Yuhang Li, Jianping Su, Feng Jiang, Weipeng Huang, and Heng Tao Shen. 2020. Data-driven spatio-temporal analysis via multi-modal Zeitgebers and cognitive load in VR. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*. IEEE, 473–482.

[105] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360 video. In *CHI Conference on Human Factors in Computing Systems*. 2535–2545.

[106] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. 2017. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. In *30th Annual ACM Symposium on User Interface Software and Technology*. 255–265.

[107] Benjamin Long, Sue Ann Seah, Tom Carter, and Sriram Subramanian. 2014. Rendering volumetric haptic shapes in mid-air using ultrasound. *ACM Trans. Graph.* 33, 6 (2014), 1–10.

[108] P. Lopes, S. You, A. Ion, and P. Baudisch. 2018. Adding force feedback to mixed reality experiences and games using electrical muscle stimulation. In *Conference on Human Factors in Computing Systems*. 1–13.

[109] C. Lovelace, B. Stein, and M. Wallace. 2003. An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cog. Brain Res.* 17, 2 (2003), 447–453.

[110] Jianfeng Lu, Li Li, and Goh Poh Sun. 2010. A multimodal virtual anatomy e-learning tool for medical education. In *International Conference on Technologies for E-Learning and Digital Entertainment*. Springer, 278–287.

[111] Justin A. MacDonald, J. D. Balakrishnan, Michael D. Orosz, and Walter J. Karplus. 2002. Intelligibility of speech in a virtual 3-D environment. *Hum. Fact.* 44, 2 (2002), 272–286.

[112] Justyna Maculewicz, Niels Christian Nilsson, and Stefania Serafin. 2016. An investigation of the effect of immersive visual and auditory feedback on rhythmic walking interaction. In *Audio Mostly Conference*. 194–201.

[113] Eduardo Magalhães, João Jacob, Niels Nilsson, Rolf Nordahl, and Gilberto Bernardes. 2020. Physics-based concatenative sound synthesis of photogrammetric models for aural and haptic feedback in virtual environments. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW'20)*. 376–379.

[114] Emanuela Maggioni, Robert Cobden, Dmitrijs Dmitrenko, and Marianna Obrist. 2018. Smell-O-Message: Integration of olfactory notifications into a messaging application to improve users' performance. In *20th ACM International Conference on Multimodal Interaction*. 45–54.

[115] S. Malpica, A. Serrano, M. Allue, M. G. Bedia, and B. Masia. 2020. Crossmodal perception in virtual reality. *Multim. Tools Applic.* 79, 5 (2020), 3311–3331.

[116] Sandra Malpica, Ana Serrano, Diego Gutierrez, and Belen Masia. 2020. Auditory stimuli degrade visual performance in virtual reality. *Sci. Rep.* 10 (2020).

[117] Carlos Marañes, Diego Gutierrez, and Ana Serrano. 2020. Exploring the impact of 360° movie cuts in users' attention. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*. IEEE, 73–82.

[118] Donna Marbury. 2021. What Does the Future Hold for AR and VR in Healthcare? Retrieved from https://healthtechmagazine.net/article/2020/11/what-does-future-hold-ar-and-vr-healthcare.

[119] Maud Marchal, Gerard Gallagher, Anatole Lécuyer, and Claudio Pacchierotti. 2020. Can stiffness sensations be rendered in virtual reality using mid-air ultrasound haptic technologies? In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 297–306.

[120] Daniel Martin, Ana Serrano, Alexander W. Bergman, Gordon Wetzstein, and Belen Masia. 2022. ScanGAN360: A generative model of realistic scanpaths for 360° images. *IEEE Trans. Visualiz. Comput. Graph.* 1 (2022).

[121] Daniel Martin, Ana Serrano, and Belen Masia. 2020. Panoramic convolutions for 360° single-image saliency prediction. In *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*.

[122] J. Martínez, A. García, M. Oliver, J. Molina, and P. González. 2014. Vitaki: A vibrotactile prototyping toolkit for virtual reality and video games. *Int. J. Hum.-comput. Interact.* 30, 11 (2014).

[123] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7 (2013), 83.

[124] B. Masia, J. Camon, D. Gutierrez, and A. Serrano. 2021. Influence of directional sound cues on users exploration across 360 movie cuts. *IEEE Comput. Graph. Applic.* 41, 4 (2021), 1–1.

[125] Fred W. Mast and Charles M. Oman. 2004. Top-down processing and visual reorientation illusions in a virtual reality environment. *Swiss J. Psychol./Schweizerische Zeitschrift für Psychologie* 63, 3 (2004), 143.

[126] Y. Matsuda, J. Nakamura, T. Amemiya, Y. Ikei, and M. Kitazaki. 2020. Perception of walking self-body avatar enhances virtual-walking sensation. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*.

[127] K. Matsumoto, E. Langbehn, T. Narumi, and F. Steinicke. 2020. Detection thresholds for vertical gains in VR and drone-based telepresence systems. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*. 101–107.

[128] K. Matsumoto, T. Narumi, Y. Ban, Y. Yanase, T. Tanikawa, and M. Hirose. 2019. Unlimited corridor: A visuo-haptic redirection system. In *International Conference on Virtual-reality Continuum and Its Applications in Industry*. 1–9.

[129] Mark McGill, Alexander Ng, and Stephen Brewster. 2017. I am the passenger: How visual motion cues can influence sickness for in-car VR. In *Proceedings of the Conference on Human Factors in Computing Systems*. 5655–5668.

[130] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa. 2020. Do multisensory stimuli benefit the virtual reality experience? A systematic review. *IEEE Trans. Visualiz. Comput. Graph.* (2020).

[131] Georg F. Meyer, Fei Shao, Mark D. White, Carl Hopkins, and Antony J. Robotham. 2013. Modulation of visually evoked postural responses by contextual visual, haptic and auditory information: A "virtual reality check." *PloS One* 8, 6 (2013).

[132] X. Min, G. Zhai, J. Zhou, X. Zhang, X. Yang, and X. Guan. 2020. A multimodal saliency model for videos with high audio-visual correspondence. *IEEE Trans. Image Process.* 29 (2020), 3805–3819.

[133] María Isabel Miranda. 2012. Taste and odor recognition memory: The emotional flavor of life. *Rev. Neurosci.* 23, 5-6 (2012), 481–499.

[134] Pedro Morgado, Nuno Nvasconcelos, Timothy Langlois, and Oliver Wang. 2018. Self-supervised generation of spatial audio for 360 video. In *Conference on Advances in Neural Information Processing Systems*. 362–372.

[135] F. Mueller, T. Kari, R. Khot, Z. Li, Y. Wang, Y. Mehta, and P. Arnold. 2018. Towards experiencing eating as a form of play. In *Symposium on Computer-human Interaction in Play Companion Extended Abstracts*.

[136] A. Mühlberger, A. Weik, P. Pauli, and G. Wiedemann. 2006. One-session virtual reality exposure treatment for fear of flying: 1-year follow-up and graduation flight accompaniment effects. *Psychother. Res.* 16, 1 (2006).

[137] A. Muhlberger, Georg Wiedemann, and Paul Pauli. 2003. Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychother. Res.* 13, 3 (2003), 323–336.

[138] Jiro Nakajima, Akihiro Sugimoto, and Kazuhiko Kawamoto. 2013. Incorporating audio signals into constructing a visual saliency map. In *Pacific-Rim Symposium on Image and Video Technology*. Springer, 468–480.

[139] Keith V. Nesbitt and Ian Hoskens. 2008. Multi-sensory game interface improves player satisfaction but not performance. In *9th Conference on Australasian User Interface*. 13–18.

[140] Solène Neyret, Xavi Navarro, Alejandro Beacco, Ramon Oliva, Pierre Bourdin, Jose Valenzuela, Itxaso Barberia, and Mel Slater. 2020. An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later Milgram obedience scenario. *Sci. Rep.* 10, 1 (2020), 1–18.

[141] Aaron R. Nidiffer, Adele Diederich, Ramnarayan Ramachandran, and Mark T. Wallace. 2018. Multisensory perception reflects individual differences in processing temporal correlations. *Sci. Rep.* 8, 1 (2018), 1–15.

[142] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. In *22nd ACM Conference on Virtual Reality Software and Technology*. 229–232.

[143] Niels C. Nilsson, Rolf Nordahl, Erik Sikström, Luca Turchet, and Stefania Serafin. 2012. Haptically induced illusory self-motion and the influence of context of motion. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 349–360.

[144] Niels Christian Nilsson, Tabitha Peck, Gerd Bruder, Eri Hodgson, Stefania Serafin, Mary Whitton, Frank Steinicke, and Evan Suma Rosenberg. 2018. 15 years of research on redirected walking in immersive virtual environments. *IEEE Comput. Graph. Applic.* 38, 2 (2018), 44–56.

[145] Niels Christian Nilsson, Evan Suma, Rolf Nordahl, Mark Bolas, and Stefania Serafin. 2016. Estimation of detection thresholds for audiovisual rotation gains. In *IEEE Virtual Reality Conference (VR'16)*. IEEE, 241–242.

[146] Toemme Noesselt, Daniel Bergmann, Maria Hake, Hans-Jochen Heinze, and Robert Fendrich. 2008. Sound increases the saliency of visual events. *Brain Res.* 1220 (2008), 157–163.

[147] M. Nogalski and W. Fohl. 2016. Acoustic redirected walking with auditory cues by means of wave field synthesis. In *IEEE Virtual Reality Conference.*

[148] Malte Nogalski and Wolfgang Fohl. 2017. Curvature gains in redirected walking: A closer look. In *IEEE Virtual Reality Conference (VR'17).* IEEE, 267–268.

[149] Jean-Marie Normand, Elias Giannopoulos, Bernhard Spanlang, and Mel Slater. 2011. Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PloS One* 6, 1 (2011).

[150] Seungoh Paek. 2012. *The Impact of Multimodal Virtual Manipulatives on Young Children's Mathematics Learning.* Ph.D. Dissertation. Teachers College, Columbia University.

[151] Amy Pavel, Björn Hartmann, and Maneesh Agrawala. 2017. Shot orientation controls for interactive cinematography with 360 video. In *30th Annual ACM Symposium on User Interface Software and Technology.*

[152] Nicholas Peatfield, Nadia Mueller, Phillipp Ruhnau, and Nathan Weisz. 2015. Rubin-vase illusion perception is predicted by prestimulus activity and connectivity. *J. Vis.* 15, 12 (2015), 429–429.

[153] Tabitha C. Peck, Henry Fuchs, and Mary C. Whitton. 2009. Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Trans. Visualiz. Comput. Graph.* 15, 3 (2009).

[154] V. Petkova and H. Ehrsson. 2008. If I were you: Perceptual illusion of body swapping. *PloS One* 3, 12 (2008).

[155] Evan Pezent, Marcia K. O'Malley, Ali Israr, Majed Samad, Shea Robinson, Priyanshu Agarwal, Hrvoje Benko, and Nicholas Colonnese. 2020. Explorations of wrist haptic feedback for AR/VR interactions with Tasbi. In *Extended Abstracts of the Conference on Human Factors in Computing Systems.* 1–4.

[156] Ivan Poupyrev, Tadao Ichikawa, Suzanne Weghorst, and Mark Billinghurst. 1998. Egocentric object manipulation in virtual environments: Empirical evaluation of interaction techniques. In *Computer Graphics Forum*, Vol. 17. Blackwell Publishers Ltd.

[157] Albert R. Powers Iii, Andrea Hillock-Dunn, and Mark T. Wallace. 2016. Generalization of multisensory perceptual learning. *Sci. Rep.* 6 (2016), 23374.

[158] Matthieu Poyade. 2013. Motor skill training using virtual reality and haptic interaction–A case study in industrial maintenance. *MÁLAGA* (2013). Universidad de Malaga.

[159] Polona Pozeg, Giulia Galli, and Olaf Blanke. 2015. Those are your legs: The effect of visuo-spatial viewpoint on visuo-tactile integration and body ownership. *Front. Psychol.* 6 (2015), 1749.

[160] Alexander Prange, Michael Barz, and Daniel Sonntag. 2018. Medical 3D images in multimodal virtual reality. In *23rd International Conference on Intelligent User Interfaces Companion.* 1–2.

[161] Jesse J. Prinz. 2006. Is the mind really modular? In *Contemporary Debates in Cognitive Science.* Citeseer 22–36.

[162] Vilayanur S. Ramachandran and Diane Rogers-Ramachandran. 1996. Synaesthesia in phantom limbs induced with mirrors. *Proc. Roy. Societ. Lond. Series B: Biol. Sci.* 263, 1369 (1996), 377–386.

[163] Nimesha Ranasinghe, Adrian Cheok, Ryohei Nakatsu, and Ellen Yi-Luen Do. 2013. Simulating the sensation of taste for immersive experiences. In *ACM International Workshop on Immersive Media Experiences.*

[164] N. Ranasinghe, P. Jain, N. Thi Ngoc Tram, K. Koh, D. Tolley, S. Karwita, L. Lien-Ya, Y. Liangkun, K. Shamaiah, C. Eason Wai Tung et al. 2018. Season traveller: Multisensory narration for enhancing the virtual reality experience. In *Conference on Human Factors in Computing Systems.* 1–13.

[165] N. Rewkowski, A. Rungta, M. Whitton, and M. Lin. 2019. Evaluating the effectiveness of redirected walking with auditory distractors for navigation in virtual environments. In *IEEE Conference on Virtual Reality and 3D User Interfaces.*

[166] Emmanuelle Richard, Angèle Tijou, and Paul Richard. 2006. Multi-modal virtual environments for education: From illusion to immersion. In *International Conference on Technologies for E-learning and Digital Entertainment.* Springer.

[167] Christian Richardt, James Tompkin, and Gordon Wetzstein. 2020. Capture, reconstruction, and representation of the visual real world for virtual reality. In *Real VR–Immersive Digital Reality.* Springer, 3–32.

[168] Bernhard E. Riecke, Aleksander Väljamäe, and Jörg Schulte-Pelkum. 2009. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Trans. Appl. Percept.* 6, 2 (2009).

[169] Irvin Rock and Jack Victor. 1964. Vision and touch: An experimentally created conflict between the two senses. *Science* 143, 3606 (1964), 594–596.

[170] Joseph M. Rosen, Hooman Soltanian, Richard J. Redett, and Donald R Laub. 1996. Evolution of virtual reality [medicine]. *IEEE Eng. Med. Biol. Mag.* 15, 2 (1996), 16–22.

[171] Amalie Rosenkvist, David Sebastian Eriksen, Jeppe Koehlert, Miicha Valimaa, Mikkel Brogaard Vittrup, Anastasia Andreasen, and George Palamas. 2019. Hearing with eyes in virtual reality. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR'19).* IEEE, 1349–1350.

[172] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. 2019. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimod. Technol. Interact.* 3, 1 (2019), 19.

[173] Sylvia Rothe and Heinrich Hußmann. 2018. Guiding the viewer in cinematic virtual reality by diegetic cues. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 101–117.

[174] Sylvia Rothe, Heinrich Hußmann, and Mathias Allary. 2017. Diegetic cues for guiding the viewer in cinematic virtual reality. In *23rd ACM Symposium on Virtual Reality Software and Technology*. 1–2.

[175] J. Rubio-Tamayo, M. Gertrudix Barrio, and F. García García. 2017. Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. *Multimod. Technol. Interact.* 1, 4 (2017).

[176] Francesco Ruotolo, Luigi Maffei, Maria Di Gabriele, Tina Iachini, Massimiliano Masullo, Gennaro Ruggiero, and Vincenzo Paolo Senese. 2013. Immersive virtual reality and environmental noise assessment: An innovative audio–visual approach. *Environ. Impact Assess. Rev.* 41 (2013), 10–20.

[177] Wallace Sadowski and Kay Stanney. 2002. Presence in virtual environments. In *Human Factors and Ergonomics. Handbook of Virtual Environments: Design, Implementation, and Applications*. Lawrence Erlbaum Associates Publishers 791–806.

[178] P. Sakhardande, A. Murugan, and J. Pillai. 2020. Exploring effect of different external stimuli on body association in VR. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW'20)*. 689–690.

[179] S. Salazar, C. Pacchierotti, X. de Tinguy, A. Maciel, and M. Marchal. 2020. Altering the stiffness, friction, and shape perception of tangible objects in virtual reality using wearable haptics. *IEEE Trans. Haptics* 13, 1 (2020), 167–174.

[180] Inês Salselas, Rui Penha, and Gilberto Bernardes. 2020. Sound design inducing attention in the context of audiovisual immersive environments. *Person. Ubiq. Comput.* 25, 4 (2020), 1–12.

[181] M. Samad, E. Gatti, A. Hermes, H. Benko, and C. Parise. 2019. Pseudo-haptic weight: Changing the perceived weight of virtual objects by manipulating control-display ratio. In *Conference on Human Factors in Computing Systems*.

[182] Maria V. Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nat. Rev. Neurosci.* 6, 4 (2005), 332–339.

[183] Yuko Sano, Akimichi Ichinose, Naoki Wake, Michihiro Osumi, Masahiko Sumitani, Shin-ichiro Kumagaya, and Yasuo Kuniyoshi. 2015. Reliability of phantom pain relief in neurorehabilitation using a multimodal virtual reality system. In *Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'15)*.

[184] Ludivine Sarlat, Olivier Warusfel, and Isabelle Viaud-Delmon. 2006. Ventriloquism aftereffects occur in the rear hemisphere. *Neurosci. Lett.* 404, 3 (2006), 324–329.

[185] R. Satava and S. Jones. 1998. Current and future applications of virtual reality for medicine. *Proc. IEEE* 86, 3 (1998).

[186] A. Schmitz, A. MacQuarrie, S. Julier et al. 2020. Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. In *IEEE Conference on Virtual Reality and 3D User Interfaces*.

[187] Martijn J. Schuemie, Peter Van Der Straaten, Merel Krijn, and Charles A. P. G. Van Der Mast. 2001. Research on presence in virtual reality: A survey. *CyberPsychol. Behav.* 4, 2 (2001), 183–201.

[188] Sofia Seinfeld, Tiare Feuchtner, Johannes Pinzek, and Jörg Müller. 2020. Impact of information placement and user representations in VR on performance and embodiment. *arXiv preprint arXiv:2002.12007*.

[189] Aaron R. Seitz, Robyn Kim, and Ladan Shams. 2006. Sound facilitates visual learning. *Curr. Biol.* 16, 14 (2006).

[190] G. Serafin and S. Serafin. 2004. Sound design to enhance presence in photorealistic virtual reality. Georgia Institute of Technology.

[191] Stefania Serafin, Michele Geronazzo, Cumhur Erkut, Niels C. Nilsson, and Rolf Nordahl. 2018. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE Comput. Graph. Applic.* 38, 2 (2018).

[192] Stefania Serafin, Niels C. Nilsson, Erik Sikstrom, Amalia De Goetzen, and Rolf Nordahl. 2013. Estimation of detection thresholds for acoustic based redirected walking techniques. In *IEEE Virtual Reality Conference (VR'13)*. IEEE, 161–162.

[193] Ana Serrano, Daniel Martin, Diego Gutierrez, Karol Myszkowski, and Belen Masia. 2020. Imperceptible manipulation of lateral camera motion for improved virtual reality applications. *ACM Trans. Graph.* 39, 6 (2020).

[194] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. 2017. Movie editing and cognitive event segmentation in virtual reality video. *ACM Trans. Graph.* 36, 4 (2017).

[195] Abhishek Seth, Judy M. Vance, and James H. Oliver. 2011. Virtual reality for assembly methods prototyping: A review. *Virt. Real.* 15, 1 (2011), 5–20.

[196] Ladan Shams and Robyn Kim. 2010. Crossmodal influences on visual perception. *Phys. Life Rev.* 7, 3 (2010).

[197] L. Shams, W. Ma, and U. Beierholm. 2005. Sound-induced flash illusion as an optimal percept. *Neuroreport* 16, 17 (2005).

[198] Ladan Shams and Aaron R. Seitz. 2008. Benefits of multisensory learning. *Trends Cog. Sci.* 12, 11 (2008).

[199] Arthur G. Shapiro and Dejan Todorovic. 2016. *The Oxford Compendium of Visual Illusions*. Oxford University Press.

[200] Youssef Shiban, Paul Pauli, and Andreas Mühlberger. 2013. Effect of multiple context exposure on renewal in spider phobia. *Behav. Res. Therap.* 51, 2 (2013), 68–74.

[201] A. Siddig, A. Ragano, H. Jahromi, and A. Hines. 2019. Fusion confusion: Exploring ambisonic spatial localisation for audio-visual immersion using the McGurk effect. In *ACM Workshop on Immersive Mixed and Virtual Environment Systems*.

[202] A. Siddig, P. Sun, M. Parker, and A. Hines. 2019. Perception deception: Audio-visual mismatch in virtual reality using the McGurk effect. In *AICS*. 176–187.

[203] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2017. How do people explore virtual environments? *IEEE Trans. Visualiz. Comput. Graph.* 24, 4 (2017).

[204] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. Roy. Societ. B: Biol. Sci.* 364, 1535 (2009), 3549–3557.

[205] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE Comput. Graph. Applic.* 29, 3 (2009), 76–84.

[206] Mel Slater and Martin Usoh. 1993. Presence in immersive virtual environments. In *IEEE Virtual Reality Annual International Symposium*. IEEE, 90–96.

[207] Bernhard Spanlang. 2014. How to build an embodiment lab: Achieving body representation illusions in virtual reality. *Front. Robot. AI* 1 (2014), 9.

[208] Charles Spence, Jae Lee, and Nathan Van der Stoep. 2017. Responding to sounds from unseen locations: Crossmodal attentional orienting in response to sounds presented from the rear. *Eur. J. Neurosci.* 51, 5 (2017).

[209] Charles Spence and Cesare Parise. 2010. Prior-entry: A review. *Conscious. Cog.* 19, 1 (2010), 364–379.

[210] Charles Spence, Daniel Senkowski, and Brigitte Röder. 2009. Crossmodal processing. *Experimental Brain Research* 198, 2 (2009), 107–111.

[211] Jonathan Steuer. 1992. Defining virtual reality: Dimensions determining telepresence. *J. Comm.* 42, 4 (1992).

[212] Ivan Stojšić, Anđelija Ivkov-Džigurski, and Olja Maričić. 2019. Virtual reality as a learning tool: How and where to start with immersive teaching. In *Didactics of Smart Pedagogy*. Springer, 353–369.

[213] P. Strandholt, O. Dogaru, N. Nilsson et al. 2020. Knock on wood: Combining redirected touching and physical props for tool-based interaction in virtual reality. In *Conference on Human Factors in Computing Systems*.

[214] E. Strasnick, C. Holz, E. Ofek, M. Sinclair, and H. Benko. 2018. Haptic links: Bimanual haptics for virtual reality using variable stiffness actuation. In *Conference on Human Factors in Computing Systems*.

[215] Q. Sun, A. Patney, L. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman. 2018. Towards virtual reality infinite walking: Dynamic saccadic redirection. *ACM Trans. Graph.* 37, 4 (2018), 67.

[216] David Swapp, Vijay Pawar, and Céline Loscos. 2006. Interaction with co-located haptic feedback in virtual reality. *Virt. Real.* 10, 1 (2006), 24–30.

[217] G. Szàkely and R. Satava. 1999. Virtual reality in medicine. *BMJ* 319, 7220 (1999), 1305.

[218] Johann Taljaard. 2016. A review of multi-sensory technologies in a science, technology, engineering, arts and mathematics (STEAM) classroom. *J. Learn. Des.* 9, 2 (2016), 46–55.

[219] Ziying Tang, Anant Patel, Xiaohu Guo, and Balakrishnan Prabhakaran. 2010. A multimodal virtual environment for interacting with 3D deformable models. In *ACM International Conference on Multimedia*.

[220] Jonathan N. Tinsley, Maxim I. Molodtsov, Robert Prevedel, David Wartmann, Jofre Espigulé-Pons, Mattias Lauwers, and Alipasha Vaziri. 2016. Direct detection of a single photon by humans. *Nat. Commun.* 7, 1 (2016), 1–9.

[221] S. Treue. 2003. Visual attention: The where, what, how and why of saliency. *Curr. Opin. Neurobiol.* 13, 4 (2003).

[222] Manos Tsakiris. 2017. The multisensory basis of the self: From body to identity to others. *Quart. J. Experim. Psychol.* 70, 4 (2017), 597–609.

[223] John C. Tuthill and Eiman Azim. 2018. Proprioception. *Curr. Biol.* 28, 5 (2018), R194–R203.

[224] I. Valori, P. McKenna-Plumley, R. Bayramova, Claudio Zandonella C., G. Altoè, and T. Farroni. 2020. Proprioceptive accuracy in immersive virtual reality: A developmental perspective. *PloS One* 15, 1 (2020).

[225] E. Van der Burg, C. N. L. Olivers, A. Bronkhorst, and J. Theeuwes. 2008. Pip and pop: Nonspatial auditory signals improve spatial visual search. *J. Experim. Psychol.: Hum. Percept. Perform.* 34, 5 (2008).

[226] Erik Van der Burg, Christian N. L. Olivers, Adelbert W. Bronkhorst, and Jan Theeuwes. 2009. Poke and pop: Tactile–visual synchrony increases visual saliency. *Neurosci. Lett.* 450, 1 (2009), 60–64.

[227] Olivier A. J. Van der Meijden and Marlies P. Schijven. 2009. The value of haptic feedback in conventional and robot-assisted minimal invasive surgery and virtual reality training: A current review. *Surgic. Endosc.* 23, 6 (2009).

[228] Isabelle Viaud-Delmon, Olivier Warusfel, Angeline Seguelas, Emmanuel Rio, and Roland Jouvent. 2006. High sensitivity to multisensory conflicts in agoraphobia exhibited by virtual reality. *Eur. Psychiat.* 21, 7 (2006).

[229] Michael Vorländer, Dirk Schröder, Sönke Pelzer, and Frank Wefers. 2015. Virtual reality for architectural acoustics. *J. Build. Perform. Simul.* 8, 1 (2015), 15–25.

[230] Bruce N. Walker and Jeff Lindsay. 2003. Effect of beacon sounds on navigation performance in a virtual reality environment. Georgia Institute of Technology.

[231] J. Wallgrün, M. Bagher, P. Sajjadi, and A. Klippel. 2020. A comparison of visual attention guiding approaches for 360° image-based VR tours. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*. IEEE, 83–91.

[232] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Trans. Visualiz. Comput. Graph.* 24, 4 (2018).

[233] Zhu Wang, Anat Lubetzky, Marta Gospodarek, Makan TaghaviDilamani, and Ken Perlin. 2019. Virtual environments for rehabilitation of postural control dysfunction. *arXiv preprint arXiv:1902.10223*.

[234] E. Whitmire, H. Benko, C. Holz, E. Ofek, and M. Sinclair. 2018. Haptic revolver: Touch, shear, texture, and shape rendering on a reconfigurable virtual reality controller. In *Proc. of the Conf. on Human Factors in Computing Systems*.

[235] A. Wilberz, D. Leschtschow, C. Trepkowski, J. Maiero, E. Kruijff, and B. Riecke. 2020. FaceHaptics: Robot arm based versatile facial haptics for immersive environments. In *Conference on Human Factors in Computing Systems*.

[236] Laurie M. Wilcox, Robert S. Allison, Samuel Elfassy, and Cynthia Grelik. 2006. Personal space in virtual reality. *ACM Trans. Percept.* 3, 4 (2006), 412–428.

[237] Christopher J. Wilson and Alessandro Soranzo. 2015. The use of virtual reality in psychology: A case study in visual perception. *Comput. Math. Meth. Med.* (2015).

[238] F. Winther, L. Ravindran, Kasper P. Svendsen, and T. Feuchtner. 2020. Design and evaluation of a VR training simulation for pump maintenance. In *Extended Abstracts of the Conference on Human Factors in Computing Systems*. 1–8.

[239] Mai Xu, Chen Li, Shanyi Zhang, and Patrick Le Callet. 2020. State-of-the-art in 360 video/image processing: Perception, assessment and compression. *IEEE J. Select. Topics Sig. Process.* 14, 1 (2020), 5–26.

[240] Wai Yu and Stephen Brewster. 2002. Multimodal virtual reality versus printed medium in visualization for blind people. In *International ACM Conference on Assistive Technologies*. 57–64.

[241] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality? In *IEEE Virtual Reality Conference (VR'10)*. IEEE, 95–102.

[242] Y. Zhu, G. Zhai, and X. Min. 2018. The prediction of head and eye movement for 360 degree images. *Sig. Process.: Image Commun.* 69 (2018), 15–25.