Understanding spatial, semantic and temporal influences on audiovisual distance compression in virtual environments

Daniel J. Finnegan[1]

School of Computer Science and Informatics, Cardiff University

Karin Petrini, Michael Proulx

Department of Psychology, University of Bath

Eamonn O'Neill

Department of Computer Science, University of Bath

Abstract

Perception of distance in virtual reality (VR) is compressed; that is, objects and the distance between them and the observer are consistently perceived as closer than intended by the designers of the VR environment. Although well documented, this phenomenon is still not fully understood or defined with respect to the factors influencing such compression. Studies on distance compression typically factor auditory or visual stimuli *individually*, critically neglecting to study how such stimuli may interact. They also tend to focus on simple static environments involving simple objects that don't move. VR can be – and at its best should be – a multisensory experience, involving not only vision but also auditory and potentially other senses. We report a study encompassing 2 experiments exploring spatial, semantic, and temporal factors of congruency in environments–environments where visual and audio cues do not correlate one-to-one as they would in a physical environment–on distance compression. Results suggest no impact of semantic association, yet significant effects for temporal and spatial congruence. We discuss the impact of our findings on virtual environment design and implementation.

Keywords: distance perception, distance compression, perceptual theory, virtual reality, head mounted display, spatial audio

[1] For correspondence: finnegand@cardiff.ac.uk

1. Introduction

Egocentric distance perception is defined as the perception of distance between one's self and a target, from the perspective of one's self [1]. The ability to perceive and process distance information is particularly important when VR is used to simulate real world scenarios in which an action must be done quickly and accurately, e.g. reaching for an object; jumping over an obstacle or across a gap; moving to a target. A wide range of applications from virtual museum tours to VR-enabled remote medical surgery require a perception of space closely resembling that of the real world [2, 3, 4, 5, 6].

Virtual reality (VR) provides the means to completely control the visual field of view, and thus influence visual perception by manipulating the visual cues available to the observer. Many studies have shown that distance judgements are underestimated compared to real world judgements [7, 8, 9, 10]. This underestimation, commonly referred to as distance compression, is profound; for a given context and task, participants making distance judgements in the virtual world underestimate distances compared to when they make the same judgements in the real world. Distance compression in VR has been studied extensively in both audio and visual domains [11, 12, 13, 14]. There are many different factors involved in VR that have been associated with distance compression (weight and inertia, movement and optical flow, graphics fidelity, measurement method etc), although an exhaustive list has not been established. Table 1 presents some of the most widely researched visual based factors and relevant papers. We have categorised the various established factors as sensory, physiological, cognitive, and external or environmental related, based on the keywords of relevant articles and the IVs manipulated in the reported studies. As shown in Table 1, prior work in VR has emphasized visual cues to distance perception. As VR is at it's best a multisensory experience, one must consider work on understanding other senses and their role in distance judgements. In this paper, we focus our attention on audio and how auditory cues interact with visual cues. Auditory distance cues are classified as static and dynamic [29, 30]. Static cues typically represent properties of the audio signal itself, either in isolation or combined in various ways. They do not require motion of the sound source or of the listener. Dynamic cues relate to how the sound signal changes over time. As an emanating sound source moves through space, the signal we hear is modified by the source position, and the context of the sound (i.e. room dampening, sound medium,

Table 1: Key factors known to impact distance perception in virtual environments. Note these studies listed are mostly related to distance compression as measured using visual stimuli.

Factor	Category	Notes	Key References
Inertia	External Environment	Fake helmet used which replicated moment of inertia	[15][a]
Angular Declination & Field of View	External Environment	Artificially manipulated eye height; perception based on reflections of objects in the environment	[16, 17, 18, 19, 20, 21][abc]
Familiarity	Cognitive	Adaptation to the environment over time; trial with feedback, then without feedback	[22, 23][a]
Inter-pupillary Distance (IPD)	Physiological	Based on individual distance between pupils; measured for each participant; compared against average IPD value	[1, 24][a]
Sense of Presence	Cognitive	Compared real world to series of virtual models; only one model actually genuine; abstract, non-photorealistic	[25, 26][a]
Use of Visual Blur	Sensory/Cognitive	Rendered scene with aperture blur; compared algorithm's prediction against human perception in a psychophysical experiment	[27][c]
Measurement Method	Cognitive	Compared measurement protocols; observed varying compression rates; evidence for top-down vs bottom-up influences	[9, 28][a]

[a] Head Mounted Display (HMD)

[b] LSID/CAVE system

[c] Other technology

occluding objects). As one moves one's head, one can also change the signal perceived at each ear–so called binaural cues–, in turn integrating over space to create dynamic cues. Table 2 presents the core known audio cues which people consider when making distance judgements.

In audiovisual environments, observers must process a number of auditory and visual distance cues in concert when making distance judgements. When audiovisual cues are both present, they may be integrated at a low or high level depending on how they relate to one another [31]. Stimuli that occur spatially and/or temporally together may be integrated at a low level depending on the strength of the cross-modal correspondences between them. For example, speech is a well recognized signal with particular expected attributes. As the sound originates from the speaker's mouth, one would expect to see the mouth moving, with the sound and movement of the mouth in close temporal synchrony. As sound travels slower than light, only when an auditory and visual stimulus overlap within a certain time window will they be integrated together to inform distance judgements. However, this is just one half of the equation. The other involves the time taken to *process* stimuli by the observer. Although light travels faster than sound, the retina in the human eye takes longer to transduce light signals into something the human brain can process compared to the basilar membrane in the human ear [32]. Thus, the human sensory system has evolved to increase the time window for audio and visual stimuli to be integrated, accommodating for the delay in transducing light signals by extending the period of neural activation in the brain [33]. Though auditory and visual cues both contribute to distance perception as separate individual events, when cues appear within a given time window, subject to individual differences across the population [34], they are likely integrated into a single audiovisual event as the observer perceives them as coming from the same source. Given that perceived delays between the visual and auditory information can affect perceived distance and audiovisual localisation (e.g., [35]) and in turns the resulting compression in VR (by increasing or decreasing the compression) it is essential to examine the effect of audiovisual asynchrony on distance perception in VR environments to better inform developers.

High level integration processes are more complex, with stimuli integrated based on known semantic associations [36]. Visual stimuli may be expected to exhibit certain noises, for example a kettle is expected to exhibit sounds when boiling. Semantic congruence, where the sound from an audiovisual stimulus is as anticipated or known in advance due to prior exposure or learned association,

Table 2: Table of auditory cues grouped by range and classified type. Note that categories and ranges are not mutually exclusive; some cues have been identified in multiple categories and ranges.

Cue	Category	Range	References
Intensity	Static	Personal, Peripersonal, Vista	[6, 40]
Monaural Amplitude Modulation	Static	Peripersonal	[29, 41]
Direct-to-Reverberant Energy Ratio	Static	Peripersonal	[42, 43]
Spectral Content	Static	Vista, Dynamic	[44, 45]
Binaural Level Difference	Static, Dynamic	Personal	[29, 46]
Acoustic Tau	Dynamic	Peripersonal, Vista	[47, 48]

leads to more immersive environments [37], and is known to elicit faster and more accurate perception judgements [38, 39]. However, it is unclear how distance perception, and in turn distance compression, is influenced by the semantic relationships between what we see and hear in virtual environments. Understanding how sights and sounds are integrated together when making distance judgements, at both a high and low level, has direct implications for software systems that render virtual environments by synchronizing visual and auditory stimuli. Similarly, for systems rendering experiences which aim to immerse people in realistic enviroments, we must understand the role semantic association plays. In this paper we deploy psychophysical methods to assess the impact of semantic and temporal incongruence on human perception, and investigate how this impact propogates to distance compression in virtual environments. That is, we examine how the matching and mistmatching of visual and auditory information at a low sensory level as well as at a higher semantic level affects distance compression in VR.

Previous work has shown how changing the position of auditory and visual cues can positively impact distance perception by reducing distance compression [49]. When audio and visual cues are presented to an observer from different source positions, we call this *spatial incongruence.* In effect, rendering visual and auditory stimuli spatially incongruent to one another in an audiovisual virtual environment changes an observer's perceived distance to the audiovisual object or stimulus, thus resulting in a reduction in distance compression. In further understanding how semantic and temporal incongruency impact distance perception, we are motivated by the following research questions:

1. **RQ 1:** How does the semantic association between auditory and visual distance cues impact distance compression in virtual environments?
2. **RQ 2:** How does latency between auditory and visual distance cues impact distance compression in virtual environments?

2. Materials and Methods

We conducted 2 experiments, each applying a 3x2 design involving 3 independent variables. We created an abstract environment centered around a single audiovisual stimulus: a high quality 3D model of a dog accompanied by a prerecorded vocalization of either a dog's bark or a cat's meow. These stimuli were chosen as they are well known stimuli used in similar experiments investigating semantic associations in cross-modal and audiovisual research [50, 51]. Our environment is carefully designed to be highly controlled rather than natural; this is intentional, as we want to manipulate and measure the interaction between stimuli without surrounding the participant in a naturalistic environment and thus affording extra distance cues (e.g., familiarity, size and shape constancy (See Figure 1)). This is analogous to how illusions are used as a mechanism to study human perception: by identifying the limits of perception through careful and controlled manipulation of stimuli [52, 53]. To address RQ 1 we manipulate the semantic relationship between the sound heard and what the observer saw in the virtual environment. When the 3D model was accompanied by the sound of a dog's bark, we classify this as a *congruent* semantic association. When the 3D model was accompanied by the sound of the cat's meow, we classify this as an *incongruent* semantic association. Note how the semantic relationship is not characterized by the direction of congruity: stimuli are either semantically congruent or not.

In contrast to the semantic relationship, the temporal relationship is characterized by its directionality. Temporal congruence means a sound cue arrives at the same time as a visual cue. Conversely, temporal incongruence means a sound cue arrives at a different time as a visual cue. For ease of reading, when a sound cue arrives *after* a visual cue we call this positive temporal incongruence. When a sound cue arrives *before* a visual cue we call this negative temporal incongruence. Temporal incongruency may effect whether two stimuli are perceived as coming from the same source or not. Detecting temporal incongruency–a prerequisite to perceiving a relationship between a leading visual stimulus and a sound–is known to be difficult for times at and below 300ms e.g., [54]. Thus, one would expect a positive temporal incongruence (i.e. latency in sound) at or above a ms threshold (e.g., $\sim$ 300ms) to still result in both stimuli perceived as coming from the same source, but in turn would impact distance judgements as the source is perceived as further away due to the latency in sound [55]. Consider the analogy of thunder and lightning, where a visual stimulus leads a proceeding sound. As the delay in sound (thunder) increases, so does the perceived distance of the source.

In effect, positive temporal incongruence is expected to reduce distance compression in a virtual environment. On the other hand, a negative temporal incongruence of the same extent, i.e., $\sim$ 300ms is expected to possibly not result in a relationship perceived between the stimuli and thus have less impact on the distance judgement. To address RQ 2, we manipulated the temporal congruence between the visual and auditory components of the audiovisual stimuli across two primary conditions: in one condition the sound appeared 300ms before the dog appeared (negative temporal incongruence), in the other the sound appeared 300ms after the dog appeared (positive temporal incongruence). A note on parameters: although the speed of sound is constant in air (343 meters/second), and thus slower than light, we are less interested in the laws of physics and more interested in human perception. There is evidence to support the claim that distance perception in virtual environments is compressed compared to the real world for distances less than 10 meters [56]. Therefore, we are exploring how introducing artificial delays between the onset of auditory and visual stimuli, at the limits of what is perceived as causally related, influences distance compression. One must consider the problem from the participant's perspective: when presented with audiovisual stimuli, either of the sound, or the visual, or both may be taken as a reference point upon whcih to base their judgement. However, we are not testing audio visual integration: rather we are assessing the effect of perceived audio visual synchrony on spatial compression. Thus, we used stimuli that have been shown in the past to be effective when examining audio visual synchrony perception [50, 51, 57].

To assess the impact of semantic association and latency, we measured the just noticeable difference (jnd) thresholds of distance judgements between pairs of stimuli. For example, if one stimuli is 5 meters from the observer and another is 5.5 meters, but both are perceived as equidistant, then the jnd for this pair of stimuli is 0.5 meters. A lower jnd indicates that an individual is

Table 3: Factors manipulated in Experiments 1 & 2

Factor	Level 1	Level 2
Experiment 1		
Spatial	*Congruent:* Auditory and visual components in alignment	*Incongruent:* Auditory and visual components out of alignment
Semantic	*Congruent:* Dog model and dog sound	*Incongruent:* Dog model and cat sound
Noise	Gaussian blur at minimum level	Gaussian blur at maximum level
Experiment 2		
Spatial	*Congruent:* Auditory and visual components in alignment	*Incongruent* Auditory and visual components out of alignment
Temporal	*Positive Incongruence:* Sound onset 300ms before visual stimulus	*Negative Incongruence:* Sound onset 300ms after visual stimulus
Noise	Gaussian blur at minimum level	Gaussian blur at maximum level

better at distinguishing between two stimuli and thus will make correct distance judgements. For our study we formulated the following hypotheses, addressed over two experiments:

H1: Semantic congruence will positively influence distance compression, with reduced perceptual just noticeable difference (jnd) thresholds when the sound matches the visual stimulus compared to when they mismatch (Experiment 1).

H2: Temporal incongruence will positively influence distance compression, with reduced perceptual jnd thresholds when the onset of auditory stimulation is after the onset of visual stimulation compared to before (Experiment 2).

Table 3 shows a summary of factors manipulated in Experiments 1 and 2: spatial congruence, temporal incongruence, and visual noise. Ethical approval for this study was granted by the XXXX (REDACTED FOR ANONYMITY). In both experiments, the virtual environment was created using the Unity 3D game engine (from here on referred to as Unity) and presented to participants using the HTC Vive head mounted display (from here on referred to as HMD). Spatial congruence was manipulated by changing the position of the renderered 3D model and the sound source using the incongruence function introduced by Finnegan et al. [49]. The auditory stimulus was rendered binaurally using a custom built software renderer in Unity. The software renders 3D binaural audio (spatial) over consumer grade headphones. Auditory stimuli are attenuated with respect to distance using the same algorithm from [49] and which we include in the manuscript. We note that with this software perceived sound source does indeed change; see [58] for details.

When presented with conflicting sensory information, observers will bias towards the source exhibiting the highest signal-to-noise (SNR) ratio [59]. When making distance judgements, one will integrate the visual and auditory cues together based on their SNR. For example, in a dark environment one would place more weight on auditory cues which are unaffected by lighting. Likewise, in a loud environment one would place more weight on visual cues. Thus a high SNR in one modality creates a bias when making distance judgements with audio and visual cues. We introduced the blur to control for this bias i.e., participants completed trials in conditions which varied the visual SNR, while studying the incongruency effect and making our results comparable with previous work [49]. We implemented a gaussian noise shader applied as a post-process to the rendered video output in Unity, blurring the environment displayed through the head mounted display. The blur was parametrized by the size of the blur radius. We manipulated visual blur on two levels: medium blur with a radius of 5 pixels, and full blur with a radius of 10 pixels. Figure 1 Panels A shows the visual stimulus without blur applied. Panels B and C show the visual stimuli presented to participants for each blur level.

2.1. Design & Procedure

Ten participants took part in both experiments, with the order of the conditions counterbalanced. They were instructed to place the HMD over their head and adjust it so that the contents were clear to see. Next the experimenter placed a pair of Sennheiser HD 200 headphones over their ears. A sound test was performed to ensure that the headphones were working. After this setup was complete, the experimenter ran a custom computer program to generate the experimental conditions in a random order for each participant. Participants were directed to keep their head perfectly still, facing in a forward direction. Before beginning the experiment, the experimenter passed a keyboard to the participant and placed their index fingers over the A and L keys.

Each trial followed a 2-Alternative Forced Choice (2AFC) protocol. The 2AFC is a specification of a general forced n-choice protocol. N-choice paradigms are extremely popular in psychophysics due to their controlled nature and the

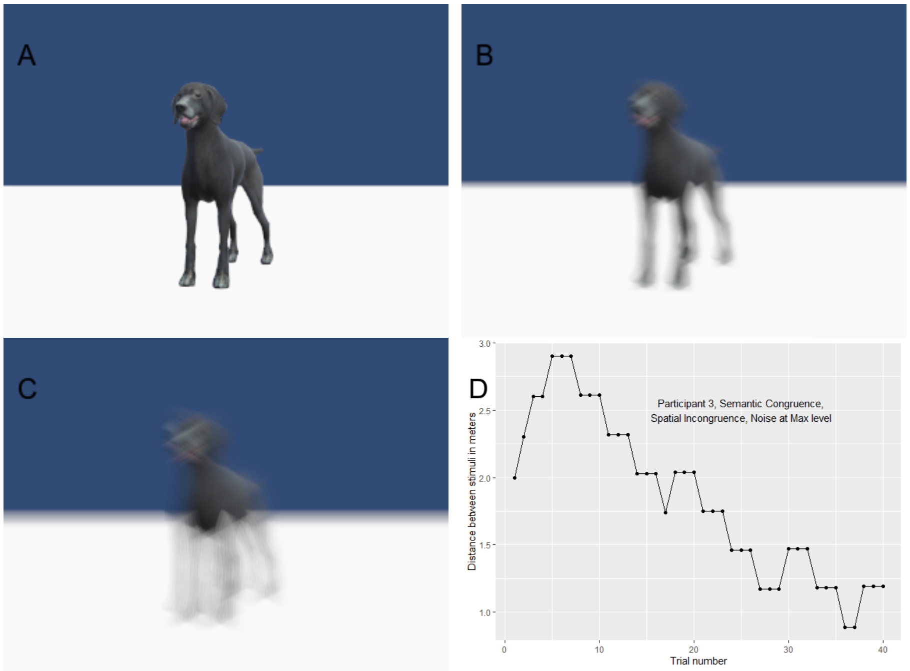

Figure 1: Panels A, B, and C illustrate the visual stimulus with no blur, min blur, and max blur respectively applied to the imagery inside the VR environment and adapted for for publication here. Panel D shows a representative staircase of the experiment trials to visualize our experiment procedure.

ability to account for the guessing as participants must choose at least one stimulus from the set of n. In HCI, they have been applied in a wide variety of contexts, from preference ratings for auditory attributes [60] to redirected walking for VR [61]. The audiovisual stimulus consisted of a high poly model of a dog with either the sound of a dog barking (semantic congruence) or the sound of a cat (semantic incongruence). Participants were presented with 1 audiovisual stimulus, proceeded by a 300ms pause, then followed by another audiovisual stimulus. The stimuli differed in congruence: one of the stimuli was presented congruently and the other incongruently, and participants had to choose which one they perceived as appearing closer to them. Participants input their response using the keyboard. They pressed the A key if they perceived the first stimulus as being closer, and the L key if they perceived the second stimulus as being closer to them. After inputting their response, the system presented the next trial.

Trials followed a staircase pattern where the spatial distance between each stmulus presented in a trial was either increased or decreased depending on the outcome of the preceding trial. Panel D in Figure 1 shows an example staircase for one condition in both experiments. The trials begin by setting the distance between stimuli at an initial starting point. Staircases followed a 3-up, 1-down procedure: if participants made 3 correct choices, the distance between stimuli decreased. If they made a single incorrect choice then the distance decreased. Each trial presented 2 choices to the participant (2-AFC); one rendered the dog and played the sound from the same distance. The other rendered the dog and played the sound according to the current experimental condition: if in the congruent condition, the dog was rendered and sound played from the same distance computed using the current staircase parameter value. If in the incongruent condition, the dog was rendered, and sound played from the distance computed using the incongruent function from [49] as follows: $\hat{y} = k\phi^{\alpha}$ where $\hat{y}$ is the distance to render the sound, ϕ is the distance of the dog, α and k are 2.22 and 0.61 respectively [62].

As trials progress, staircase procedures converge on the threshold where participants perceive stimuli as being similar. The jnd was measured by averaging the number of reversals in the staircase. By way of a working example, consider the following: For each participant, the experiment followed the psychophysical 3-up 1-down staircase procedure. Consider the following incongruent example:

1. The staircase contrast parameter is currently at 1 meter, the delta-up

factor is set at 0.2 meters, and the delta-down parameter is set at 0.3 meters.

2. First choice renders the dog and sound at 6 meters from the participant.
3. Second choice renders the dog at a position of 5 meters but the sound at a position of 5.6 meters.
4. Participants must then pick which choice they perceived as appearing closer to them.
5. If they choose correctly, the contrast parameter is updated based on the delta-up parameter becoming 0.7 meters. If they choose incorrectly it is set to 1.2 meters.
6. The participant chooses correctly. When the next trial begins, the first choice renders the dog at 4.8 meters with sound at 5.3 meters, and the second choice renders the dog at 6 meters with the sound at 6 meters.
7. So on and so forth.

One of the choices always rendered the dog at 6 meters (the standard stimulus). The order of the choices i.e., whether choice 1 was the standard or not, was randomized between trials. In total, we collected 40 trials across 10 participants across 8 experiment conditions, totalling 3200 data points. Data from 3 of our participants was discarded due to no reversals made during the staircase procedure. After computing thresholds, we had 8 data points per participant for a total of 56 data points for analysis.

3. Results

Null hypothesis significance testing (NHST) is subject to limitations when interpreting results, particularly with respect to measuring support for the null hypothesis (H0) [63]. To quantify the likelihood that our results are explained by H0, we report Bayes Factors (BF^{01}) for all null results using the method by Faulkenberry [63]. For significant results we report generalized η^2 effect sizes as they are not influenced by study design and are therefore comparable across designs [64]. For Experiment 1, results show a significant main effect of spatial congruence on the threshold, $F(1,6) = 7.08, p = .04, \eta^2 = 0.08$. There was no effect for semantic association ($F(1,6) = 0.06, p = .82$). Bayes Factor indicates weak support for the H0: the data are 1.3 times more likely explained by H0 ($BF_{01} = 1.32$). There was also no effect for noise ($F(1,6) = 0.00, p = .98$), with Bayes Factor again indicating weak support for H0 ($BF_{01} = 1.32$).

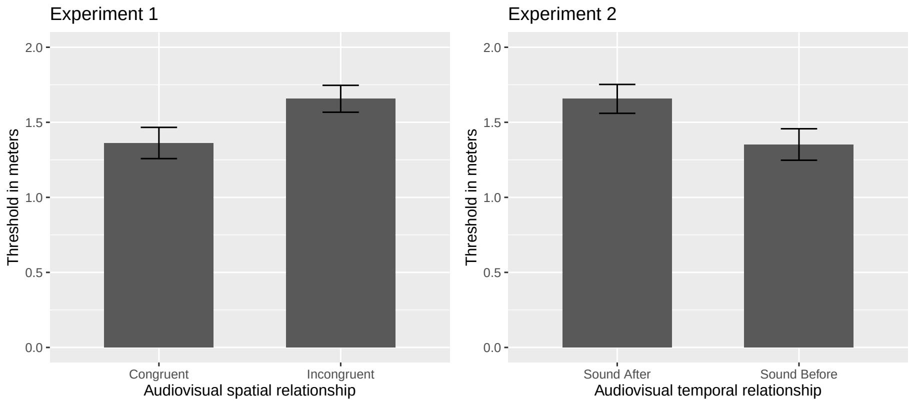

Figure 2: Results from both experiments in our study. The left hand side plots the mean threshold for spatial congruence in Experiment 1. The right hand side plots the mean threshold for temporal congruence in Experiment 2. Error bars represent standard error.

For Experiment 2, there was a significant main effect of temporal incongruency on the threshold, $F(1, 6) = 18.07, p = .005, \eta^2 = 0.08$. The threshold was smaller–distance compression was reduced–when sound appeared *before* the dog compared to after the dog. There was no statistically significant effect for spatial congruence ($F(1, 6) = 0.45, p = .53$). Bayes Factor indicates the data are 1.3 times more likely explained by H0 ($BF_{01} = 1.33$) indicating weak support for the null hypothesis. There was also no effect for noise ($F(1, 6) = 0.13, p = .74$) with Bayes Factor indicating the data are 1.3 times more likely explained by H0 ($BF_{01} = 1.32$). Figure 2 shows mean plots for the effects of spatial congruence in Experiment 1 and Temporal congruence in Experiment 2.

4. Discussion

From previous work involving spatial incongruence [49, 58], it was expected that spatially incongruent audiovisual stimuli would reduce distance compression in virtual environments. When auditory and visual cues are intentionally misaligned with respect to their distance from the observer, the combined effect results in participants performing better and more precise in distance discrimination tasks. However, previous work has focused on basic auditory and visual stimuli which drive little understanding into distance perception in more complex environments featuring rich distance cues, and offer little in terms of practical advice for constructing software tools for producing virtual environments where compresssion is limited. In Experiment 1, the congruent conditions resulted in a lower jnd compared to the incongruent conditions, meaning distance compression was reduced in the congruent condition, *not* the incongruent conditions. This result stands in contrast to previous work [49] where spatially incongruent stimuli were found to reduce distance compression in virtual environments. It is noteworthy that while Experiment 1 resulted in a significant difference between levels of the spatial congruence factor, this result was not replicated in Experiment 2. Our experiments' procedure differs from previous work in 2 distinct ways: first, in our analysis we compute the threshold of subjective equality as a function of staircase reversals in the 2AFC task detailed in Section 2.1. Secondly, while the study by Finnegan et al. used an abstract environment with a simplified audiovisual stimulus–a green cube and a pink noise burst–, our visual stimulus consisted of a high poly model of a dog. It could be that the change to more naturalistic stimuli–in this instance a high poly model of a dog and realistic vocalizations–resulted in the effect's reversal. If this is true, then what characteristics of the abstract stimuli would lead to a positive effect of incongruence? Spatial congruence typically results in better performance for a range of localization tasks [65], however, there is little precedent in the context of distance compression. Much of the literature on distance perception and compression in virtual reality focuses on unimodal environments, with a bias towards visual-only environments.

Prior work has shown a positive influence of semantic congruence on multisensory integration for determining action, for example feature discrimination tasks [36, 38]; so why not distance discrimination? A key contributing factor to multisensory integration is the unity assumption which states that multisensory stimuli sharing certain physical properties like temporal and/or spatial correspondence are grouped together as deriving from one object [39]. Thus under the unity assumption, semantic congruence was expected to have a positive effect on participants integrating what they saw and what they heard as the same object, in our case the dog model and the sound of the bark. However, we found semantic congruence to have no distinguishable impact on distance compression (H1). One reason could be that the dog mouth did not move and so the spatial incongruency did not work as well. By not having that cue to bind the stimuli, participants may not have been affected by semantic information much as the sound could have been perceived as coming from a different source in both incongruent and congruent sematic conditions. Future studies should consider this. Semantic congruence was restricted to a universally recognized pairing of audiovisual stimuli–the sound of a dog barking and a visual 3D model of a dog. We did not consider potential semantic relationships in the eye of the observer. For example, our study did not investigate any perceived anxiety or fear of the dog on the part of the observer. As fear has been shown to be a factor in distance perception [66, 67], high level semantic relationships, based on personal experience and subsequent association between the observer and the stimulus, remain to be explored.

H2 predicted that distance compression would be reduced when the sound appeared after the visual stimulus. Though temporal incongruence was statistically significant, the threshold was lower for the sound leading condition than for the visual leading condition, meaning distance compression was reduced when the auditory cue preceded the visual cue. In the context of distance perception, we expected that stimuli where the sound was heard after the visual stimulus would be perceived as further away because of the natural association between the spatial displacement of sound and the time it takes to reach our ears. Following the analogy of thunder and lightning, where thunder is heard after seeing a flash of lightning, these results are in line with previous work showing an effect of temporal misalignment on distance perception [68]. Similar results are observed in localization tasks, where training with spatio-temporally aligned stimuli has a positive impact on audio-only localization tasks [65]. Unexpectedly, observers were more precise in correctly judging which audiovisual stimuli was closer to them when sound *preceded* visual stimulation (inverse of H2). As thresholds were lower, distance compression was reduced when the sound came before the visual stimulus. There is evidence suggesting that temporal binding is impacted by the order in which crossmodal sensory stimuli are perceived by the observer [69]. Perceptually participants are better at detecting asynchrony for auditory first conditions than visual first conditions [54]. This may explain our results: participants in our study may have perceived the visual and auditory stimuli as not coming from the same source in the visual leading trials, and thus focused solely on the visual stimulus when making their distance judgements.

Many distance perception studies have compared different measurement methods, for example timed imagined walking & blind walking [70], and verbal estimation. Previous work has applied techniques involving absolute distance judgments, however, experiments involving verbal estimates of absolute distance judgments have shown a cognitive bias in participants' concepts of different metrics [9]. Our method of a 2-AFC task resembles the perceptual matching technique [71]. To mitigate the bias and interpret our findings considering previous work using non-conventional distance perception metrics [49], in our experiment we used a discrimination task. We argue this is appropriate as we are interested in how multimodal stimuli impacts distance perception, a prerequisite to studying distance compression. Considering the problem at such a low level, though forsaking ecological validity, the problem becomes one of binary outcome i.e., participants were tasked with choosing the stimulus pair–congruent or incongruent–which they perceived as closest to them. This is sufficient to address our research questions; however, future work should investigate a range of distances to validate the method further. Additionally, although a standard technique in the field of psychophysics, further work is required to compare the sensitivity and precision of our method with more common methods in distance perception research.

Finally, we note our study considered semantic and temporal congruence in combination with spatial congruence, yet did not consider an interaction between semantic and temporal congruence. Our methodology already involved a large number of conditions and many trials; we were concerned with exponentially growing numbers of trials and the impact this may have on participants e.g., fatigue. Considering our analogy of thunder and lightning, one does not perceive them as one object *per se*, but cognitively one knows they are, so one can use that information to figure out the distance. Future work should consider how semantic and temporal congruence interact with one another.

4.1. Implications for Future Virtual Environments

One application of our findings is in virtual reality training environments such as driving simulators, teleoperation and scenarios where distance perception is important. Prior research has demonstrated distance compression in virtual driving simulators [72, 73]. Given that the context of driving may present a combination of auditory and visual information, prolonged training with such systems that are not explicitly designed to account for cross-modal congruence may subject trainees to compressing their perceived distances, with potentially dangerous carry over to the real world. Virtual environment designers and developers should carefully consider the semantic and temporal relationships between audio and visual cues when creating their environments, paying particular attention to dynamic environments where sound cues are leading visual cues. This problem is particularly exacerbated by the added factor of motion in the display. For example, motion in a virtual environment is subject to a variety of visual filters, e.g. motion blur, and auditory filters, e.g. reverberance, to create a degree of realism in the training simulator. While blur has been shown to impact perceived depth [27] in static environments and where observers do not move, its impact on distance perception in environments where observers move has been contested [74]. In the context of our findings, blur coupled with temporally incongruent sources which may also move could result in worse distance compression.

Recent work has recommended a familiarization phase [75] to reduce distance compression when using head mounted display hardware. Yet this approach may not account for *what* is in the environment, rather than how it is displayed, and how the content of the environment may impact perceived distance. Familiarization with using a head mounted display is not the same as familarization with an environment displayed through a head mounted display. As simulations become more complex, increasingly exploit high fidelity multisensory environments using personalized 3D audio [76], and are used for training, it will become increasingly important to understand how various factors influence distance perception so that these simulations may accurately reflect their corresponding real world scenarios. As our findings demonstrate reduced distance compression with spatial congruence, while previous work demonstrates reduced distance compression with spatial incongruence [49], we must consider that this is probably due to differences in complexity and ecological level of the stimuli used. If the factors underlying multisensory integration of various stimuli, and how they interact, are better understood then we may create virtual reality experiences which mitigate the need for extensive familiarization and improve simulated to real world transitions.

Although visual cues receive a larger weighting in multisensory integration due to their tendency to be less noisy [59], our results suggest that, at least with respect to temporal appearance, auditory information may act as an ‘anchor’ for distance perception, establishing an initial estimate of how far away something is. For VR applications, these findings emphasize the importance of auditory cues in 3D spatial diegetic user interfaces, i.e interfaces which appear within the environment rather than overlayed on the screen. Designing a rich, multisensory environment taking our results into account, for example rendering sounds temporally incongruent to visual cues, may result in experiences which reduce distance compression on the part of users, making virtual reality a more robust tool for simulators, remote performance of tasks and other applications that rely on accurate and precise distance perception.

Ultimately, our results indicate how VR experience designers must take care when moving beyond visual only experiences which incorporate multisensory stimulation. Cross-modal correspondences between stimuli are not restricted to just visual and auditory information. As studies in HCI increasingly explore human behaviour in VR involving tactile, gustatory, and even olfactory stimulation [77], understanding fundamental relationships between all 5 senses and how they combine to form our perception and in turn understanding of reality, with implications for sense of presence and immersion, is critical to the design of better VR experiences.

5. Conclusion

Distance perception in virtual environments is a challenge involving many factors from top-down cognitive factors to bottom-up features of distance cues related to the stimuli perceived in the environment. Our focus is on audivisual virtual environments, and how audio and visual cues together result in varying distance judgements. We ran 2 experiments to study the impact of semantic association and latency between auditory and visual stimuli on distance perception. Previous work has assumed semantic congruence by the use of stimuli that have explicit, natural associations. Our study explicitly tested this assumption, with results showing no effect of semantic congruence on distance compression in a distance discrimination task. However, this semantic relationship was based on low level cues: future work should investigate semantic associations between stimuli which are individual or personal to the observer by exposing participants to stimuli beforehand or using digitized personal objects of participants. The impact of negative temporal incongruence i.e. latency, interacts with the spatial congruity between the sound and the visual object such that a sound delay when audiovisual stimuli were spatially incongruent resulted in reduced distance compression. Future work will investigate whether the effect of latency translates to dynamic environments with reverberation, if the impact of spatial and temporal incongruence on distance perception is consistent across more realistic environments, and how these factors may interact with semantic relationships between stimuli in densely populated environments. Future work should also investigate if these effects are observed in different distance perception tasks to further clarify their impact on egocentric distance perception and compression generally.

Considerable previous work shows how congruent presentation of multisensory stimuli results in more efficient processing of those stimuli, e.g. [31, 78]. Given that virtual reality applications are moving towards multisensory experiences for a wide range of applications, it is imperative to investigate how various multisensory cues can be combined, and how they in turn will influence features of the user experience such as distance perception. Future work will build a model which predicts how audio and visual cues are integrated together to inform distance judgements. This includes high level factors such as semantic relationships between multisensory stimuli. Taking into account the findings reported here, such a model may prove fruitful for practitioners and could be implemented as a software tool for aiding the construction of virtual environments which minimize distance compression, making such environments better suited to a wide range of applications that require accurate and precise spatial and distance perception.

Acknowledgments

Daniel J. Finnegan thanks the School of Computer Science at Cardiff University for supporting his research. Eamonn O'Neill's research was partly funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1. Michael Proulx's research was partly funded by EPSRC grant EP/J017205/1, Design Patterns for Inclusive Collaboration. Karin Petrini's research was supported by the University of Bath's Alumni fund. We thank Tayfun Essenkaya, Crescent Jicol, and Soumya Chinnachamy Barathi for their insightful comments and discussions.

References

[1] R. S. Renner, B. M. Velichkovsky, J. R. Helmert, The perception of egocentric distances in virtual environments - A review, ACM Computing Surveys 46 (2) (2013) 1–40. doi:10.1145/2543581.2543590.
URL http://dl.acm.org/citation.cfm?id=2543581.2543590

[2] B. Hervy, F. Laroche, J.-L. Kerouanton, A. Bernard, C. Courtin, L. D'haene, B. Guillet, A. Waels, Augmented historical scale model for museums, in: Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14, ACM Press, New York, New York, USA, 2014, pp. 1–4. doi:10.1145/2617841.2617843.

URL http://dl.acm.org/citation.cfm?doid=2617841.2617843

[3] R. S. Renner, E. Steindecker, M. Müller, B. M. Velichkovsky, R. Stelzer, S. Pannasch, J. R. Helmert, The Influence of the Stereo Base on Blind and Sighted Reaches in a Virtual Environment, ACM Transactions on Applied Perception 12 (2) (2015) 1–18. doi:10.1145/2724716.
URL http://dl.acm.org/citation.cfm?doid=2746686.2724716

[4] F. O. Matu, M. Thøgersen, B. Galsgaard, M. M. Jensen, M. Kraus, Stereoscopic augmented reality system for supervised training on minimal invasive surgery robots, in: Proceedings of the 2014 Virtual Reality International Conference on - VRIC '14, ACM Press, New York, New York, USA, 2014, pp. 1–4. doi:10.1145/2617841.2620722.
URL http://dl.acm.org/citation.cfm?doid=2617841.2620722

[5] G. Parseihian, C. Jouffrais, B. F. G. Katz, Reaching nearby sources: comparison between real and virtual sound and visual targets, Frontiers in Neuroscience 8 (September) (2014) 1–13. doi:10.3389/fnins.2014.00269.

[6] H. Wu, D. H. Ashmead, B. Bodenheimer, Using immersive virtual reality to evaluate pedestrian street crossing decisions at a roundabout, Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization - APGV '09 1 (212) (2009) 35. doi:10.1145/1620993.1621001.
URL http://portal.acm.org/citation.cfm?doid=1620993.1621001

[7] J. M. Loomis, J. A. Da Silva, N. Fujita, S. S. Fukusima, Visual space perception and visually directed action., Journal of Experimental Psychology: Human Perception and Performance 18 (4) (1992) 906–921. doi:10.1037/0096-1523.18.4.906.

[8] S. A. Kuhl, W. B. Thompson, S. H. Creem-Regehr, Minification influences spatial judgments in virtual environments, Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization - APGV '06 1 (212) (2006) 15. doi:10.1145/1140491.1140494.
URL http://portal.acm.org/citation.cfm?doid=1140491.1140494

[9] J. M. Loomis, J. W. Philbeck, Measuring spatial perception with spatial updating and action, in: M. B. Roerta L. Klatzky Brian MacWhinney (Ed.), Embodiment, Ego-Space, and Action, Carnegie Mellon Symposia on Cognition, Taylor and Francis, 2008, pp. 1–43.
URL `http://www.tandf.net/books/details/9780805862881/`

[10] K. M. Rand, M. R. Tarampi, S. H. Creem-Regehr, W. B. Thompson, The Importance of a Visual Horizon for Distance Judgments under Severely Degraded Vision, Perception 40 (2) (2012) 143–154.

[11] S. H. Creem-Regehr, J. K. Stefanucci, W. B. Thompson, Perceiving Absolute Scale in Virtual Environments: How Theory and Application Have Mutually Informed the Role of Body-Based Perception, in: The Pschology of Learning and Motivation, Vol. 62, 2015, pp. 195–224. `doi:10.1016/bs.plm.2014.09.006`.
URL `http://linkinghub.elsevier.com/retrieve/pii/S0079742114000073`

[12] S. A. Kuhl, W. B. Thompson, S. H. Creem-Regehr, HMD calibration and its effects on distance judgments, ACM Transactions on Applied Perception 6 (3) (2009) 1–20. `doi:10.1145/1577755.1577762`.
URL `http://portal.acm.org/citation.cfm?doid=1577755.1577762`

[13] D. Waller, A. R. Richardson, Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information., Journal of Experimental Psychology: Applied 14 (1) (2008) 61–72. `doi:10.1037/1076-898X.14.1.61`.
URL `http://doi.apa.org/getdoi.cfm?doi=10.1037/1076-898X.14.1.61`

[14] M. Paquier, N. Côté, F. Devillers, V. Koehl, Interaction between auditory and visual perceptions on distance estimations in a virtual environment, Applied Acoustics 105 (2016) 186–199. `doi:10.1016/j.apacoust.2015.12.014`.
URL `http://linkinghub.elsevier.com/retrieve/pii/S0003682X15003680`

[15] P. Willemsen, A. A. Gooch, W. B. Thompson, S. H. Creem-Regehr, Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments,

Presence: Teleoperators and Virtual Environments 17 (1) (2008) 91–101. doi:10.1162/pres.17.1.91.

[16] B. Williams, T. Rasor, G. Narasimham, Distance perception in virtual environments, in: Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization - APGV '09, ACM Press, New York, New York, USA, 2009, p. 7. doi:10.1145/1620993.1620995.
URL http://portal.acm.org/citation.cfm?doid=1620993.1620995

[17] R. Messing, F. H. Durgin, Distance Perception and the Visual Horizon in Head-Mounted Displays, ACM Transactions on Applied Perception 2 (3) (2005) 234–250. doi:10.1145/1077399.1077403.
URL http://portal.acm.org/citation.cfm?doid=1077399.1077403

[18] R. Toth, J. Hasselgren, T. Akenine-Möller, Perception of highlight disparity at a distance in consumer head-mounted displays, in: Proceedings of the 7th Conference on High-Performance Graphics - HPG '15, HPG '15, ACM Press, New York, New York, USA, 2015, pp. 61–66. doi:10.1145/2790060.2790062.
URL http://dx.doi.org/10.1145/2790060.2790062http://dl.acm.org/citation.cfm?doid=2790060.2790062http://doi.acm.org/10.1145/2790060.2790062

[19] J. G. P. Corujeira, I. Oakley, Stereoscopic Egocentric Distance Perception: The Impact of Eye Height and Display Devices, SAP '13 Proceedings of the ACM Symposium on Applied Perception.

[20] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, W. B. Thompson, The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments, Perception 34 (2) (2005) 191–204. doi:10.1068/p5144.
URL http://pec.sagepub.com/lookup/doi/10.1068/p5144

[21] I. V. Piryankova, S. de La Rosa, U. Kloos, H. H. Bülthoff, B. J. Mohler, Egocentric distance perception in large screen immersive displays, Displays 34 (2) (2013) 153–164. doi:10.1016/j.displa.2013.01.001.
URL http://dx.doi.org/10.1016/j.displa.2013.01.001

[22] T. Dat Nguyen, C. J. Ziemer, T. Grechkin, B. Chihak, J. M. Plumert, J. F. Cremer, J. K. Kearney, Effects of Scale Change on Distance Perception in Virtual Environments, ACM Trans. Appl. Percept. Article 8 (26). doi:10.1145/2043603.2043608.
URL http://dl.acm.org/citation.cfm?doid=2043603.2043608

[23] A. R. Richardson, D. Waller, Interaction With an Immersive Virtual Environment Corrects Users' Distance Estimates, Human Factors: The Journal of the Human Factors and Ergonomics Society 49 (3) (2007) 507–517. doi:10.1518/001872007x200139.
URL http://hfs.sagepub.com/content/49/3/507.abstract

[24] S. A. Kuhl, S. H. Creem-Regehr, W. B. Thompson, Individual differences in accuracy of blind walking to targets on the floor, Journal of Vision 6 (6) (2006) 726–726. doi:10.1167/6.6.726.
URL http://jov.arvojournals.org/Article.aspx?doi=10.1167/6.6.726

[25] B. Ries, V. Interrante, M. Kaeding, L. Anderson, The effect of self-embodiment on distance perception in immersive virtual environments, in: Proceedings of the 2008 ACM symposium on Virtual reality software and technology - VRST '08, ACM Press, New York, New York, USA, 2008, p. 167. doi:10.1145/1450579.1450614.
URL http://portal.acm.org/citation.cfm?doid=1450579.1450614

[26] L. Phillips, B. Ries, V. Interrante, M. Kaeding, L. Anderson, Distance perception in NPR immersive virtual environments, revisited, in: Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization - APGV '09, ACM Press, New York, New York, USA, 2009, p. 11. doi:10.1145/1620993.1620996.
URL http://portal.acm.org/citation.cfm?doid=1620993.1620996

[27] R. T. Held, E. A. Cooper, J. F. O'Brien, M. S. Banks, Using blur to affect perceived distance and size, ACM Transactions on Graphics 29 (2) (2010) 1–16. doi:10.1145/1731047.1731057.
URL http://portal.acm.org/citation.cfm?doid=1731047.1731057

[28] C. S. Sahm, S. H. Creem-Regehr, W. B. Thompson, P. Willemsen, Throwing versus walking as indicators of distance perception in similar real and virtual environments, ACM Transactions on Applied Perception 2 (1) (2005) 35–45. doi:10.1145/1048687.1048690.

[29] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, S. Pardhan, Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss, Attention, Perception, & Psychophysicsdoi:10.3758/s13414-015-1015-1.
URL http://link.springer.com/10.3758/s13414-015-1015-1

[30] P. Zahorik, D. S. Brungart, A. W. Bronkhorst, Auditory Distance Perception in Humans: A Summary of Past and Present Research, Acta Acustica united with Acustica 91 (3) (2005) 409–420.
URL http://www.ingentaconnect.com/content/dav/aaua/2005/00000091/00000003/art00003

[31] C. Spence, Audiovisual multisensory integration, Acoustical Science and Technology 28 (2) (2007) 61–70. doi:10.1250/ast.28.61.

[32] D. W. Massaro, M. M. Cohen, P. M. T. Smeele, Perception of asynchronous and conflicting visual and auditory speech, The Journal of the Acoustical Society of America 100 (3) (1996) 1777–1786, publisher: Acoustical Society of America. doi:10.1121/1.417342.
URL https://asa.scitation.org/doi/10.1121/1.417342

[33] C. Spence, S. Squire, Multisensory Integration: Maintaining the Perception of Synchrony, Current Biology 13 (13) (2003) R519–R521. doi:10.1016/S0960-9822(03)00445-7.
URL https://www.sciencedirect.com/science/article/pii/S0960982203004457

[34] B. Conrey, D. B. Pisoni, Auditory-visual speech perception and synchrony detection for speech and nonspeech signals, The Journal of the Acoustical Society of America 119 (6) (2006) 4065–4073, publisher: Acoustical Society of America. doi:10.1121/1.2195091.
URL https://asa.scitation.org/doi/10.1121/1.2195091

[35] D. Alais, D. Burr, The Ventriloquist Effect Results from Near-Optimal Bimodal Integration, Current Biology 14 (3) (2004) 257–262. doi:10.

1016/j.cub.2004.01.029.
URL https://www.sciencedirect.com/science/article/pii/S0960982204000430

[36] S. N. Macdonald, E. D. Richards, G. Desmarais, Impact of semantic similarity in novel associations: Direct and indirect routes to action, Attention, Perception, & Psychophysics 78 (1) (2016) 37–43. doi:10.3758/s13414-015-1041-z.
URL https://doi.org/10.3758/s13414-015-1041-z

[37] L. Turchet, Designing presence for real locomotion in immersive virtual environments: an affordance-based experiential approach, Virtual Reality 19 (3) (2015) 277–290. doi:10.1007/s10055-015-0267-3.
URL https://doi.org/10.1007/s10055-015-0267-3

[38] P. J. Laurienti, R. A. Kraft, J. A. Maldjian, J. H. Burdette, M. T. Wallace, Semantic congruence is a critical factor in multisensory behavioral performance, Experimental Brain Research 158 (4) (2004) 405–414. doi:10.1007/s00221-004-1913-2.
URL https://doi.org/10.1007/s00221-004-1913-2

[39] Y.-C. Chen, C. Spence, Assessing the Role of the Unity Assumption on Multisensory Integration: A Review, Frontiers in Psychology 8. doi:10.3389/fpsyg.2017.00445.
URL https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00445/full

[40] D. H. Mershon, E. L. King, Intensity and reverberation as factors in the auditory perception of egocentric distance, Perception & Psychophysics 18 (6) (1975) 409–415. doi:10.3758/BF03204113.
URL http://link.springer.com/article/10.3758/BF03204113%5Cnhttp://www.springerlink.com/index/10.3758/BF03204113

[41] D. O. Kim, P. Zahorik, L. H. Carney, B. B. Bishop, S. Kuwada, Auditory Distance Coding in Rabbit Midbrain Neurons and Human Perception: Monaural Amplitude Modulation Depth as a Cue, Journal of Neuroscience 35 (13) (2015) 5360–5372. doi:10.1523/JNEUROSCI.3798-14.2015.
URL http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.3798-14.2015

[42] A. Rungta, S. Rust, N. Morales, R. Klatzky, M. Lin, D. Manocha, Psychoacoustic Characterization of Propagation Effects in Virtual Environments, ACM Transactions on Applied Perception 13 (4) (2016) 1–18. doi:10.1145/2947508.
URL http://dl.acm.org/citation.cfm?doid=2974016.2947508

[43] P. Zahorik, Assessing auditory distance perception using virtual acoustics., The Journal of the Acoustical Society of America 111 (4) (2002) 1832–1846. doi:10.1121/1.1458027.

[44] S.-w. Jeon, Y.-c. Park, D. H. Youn, Auditory Distance Rendering Based on ICPD Control for Stereophonic 3D Audio System, IEEE Signal Processing Letters 22 (5) (2015) 529–533. doi:10.1109/LSP.2014.2363455.
URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6926755

[45] S. Werner, J. Liebetrau, Effects of shaping of binaural room impulse responses on localization, in: 2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX), IEEE, 2013, pp. 88–93. doi:10.1109/QoMEX.2013.6603216.
URL http://ieeexplore.ieee.org/document/6603216/

[46] S. Spagnol, E. Tavazzi, F. Avanzine, Relative auditory distance discrimination with virtual nearby sound sources, in: 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, 2015, pp. 1–6.
URL http://www.soundofvision.net/relative-auditory-distance-discrimination-with-virtual-nearby-sound-sources/

[47] M. S. Gordon, F. A. Russo, E. MacDonald, Spectral information for detection of acoustic time to arrival, Attention, Perception, & Psychophysics 75 (4) (2013) 738–750. doi:10.3758/s13414-013-0424-2.
URL http://link.springer.com/10.3758/s13414-013-0424-2

[48] J. M. Speigle, J. M. Loomis, Auditory distance perception by translating observers, in: Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium, IEEE Comput. Soc. Press, 1993, pp. 92–99. doi:10.1109/VRAIS.1993.378257.
URL http://ieeexplore.ieee.org/document/378257/

[49] D. J. Finnegan, E. O'Neill, M. Proulx, Compensating for Distance Compression in Audiovisual Virtual Environments Using Incongruence, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, San Jose, USA, 2016, pp. 200–212. doi:10.1145/2858036.2858065.

[50] Y.-C. Chen, C. Spence, When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures, Cognition 114 (3) (2010) 389–404. doi:10.1016/j.cognition.2009.10.012.
URL https://www.sciencedirect.com/science/article/pii/S0010027709002649

[51] S. Mastroberardino, V. Santangelo, E. Macaluso, Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks, Frontiers in Integrative Neuroscience 9 (2015) 45. doi:10.3389/fnint.2015.00045.

[52] C.-C. Carbon, Understanding human perception by human-made illusions, Frontiers in Human Neuroscience 8 (2014) 566. doi:10.3389/fnhum.2014.00566.
URL https://www.frontiersin.org/article/10.3389/fnhum.2014.00566

[53] A. Salagean, J. Hadnett-Hunter, D. J. Finnegan, A. A. de Sousa, M. Proulx, A Virtual Reality Application of the Rubber Hand Illusion Induced by Ultrasonic Mid-Air Haptic Stimulation, ACM Transactions on Applied Perception (2021) 1–15.

[54] S. A. Love, K. Petrini, A. Cheng, F. Pollick, A Psychophysical Investigation of Differences between Synchrony and Temporal Order Judgments, PloS onedoi:10.1371/journal.pone.0054798.

[55] Y. Sugita, Y. Suzuki, Implicit estimation of sound-arrival time, Nature 421 (6926) (2003) 911–911, number: 6926 Publisher: Nature Publishing Group. doi:10.1038/421911a.
URL https://www.nature.com/articles/421911a

[56] A. Clement, G. A. Radvansky, J. R. Brockmole, Compression of environmental representations following interactions with objects, Attention,

Perception, & Psychophysics 79 (8) (2017) 2460–2466. doi:10.3758/s13414-017-1401-y.
URL https://doi.org/10.3758/s13414-017-1401-y

[57] N. Russo, L. Mottron, J. A. Burack, B. Jemel, Parameters of semantic multisensory integration depend on timing and modality order among people on the autism spectrum: Evidence from event-related potentials, Neuropsychologia 50 (9) (2012) 2131–2141. doi:10.1016/j.neuropsychologia.2012.05.003.
URL https://www.sciencedirect.com/science/article/pii/S0028393212001947

[58] D. J. Finnegan, E. O'Neill, M. J. Proulx, An approach to reducing distance compression in audiovisual virtual environments, in: 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), 2017, pp. 1–6. doi:10.1109/SIVE.2017.7901607.

[59] M. O. Ernst, M. S. Banks, Humans integrate visual and haptic information in a statistically optimal fashion, Nature 415 (6870) (2002) 429–433. doi:10.1038/415429a.
URL http://www.nature.com/doifinder/10.1038/415429a

[60] S. Choisel, F. Wickelmaier, Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference, The Journal of the Acoustical Society of America 121 (1) (2007) 388–400. doi:10.1121/1.2385043.
URL http://asa.scitation.org/doi/10.1121/1.2385043

[61] G. Bruder, F. A. Sanz, A.-H. Olivier, A. Lecuyer, Distance estimation in large immersive projection systems, revisited, in: 2015 IEEE Virtual Reality (VR), IEEE, Arles, France, 2015, pp. 27–32. doi:10.1109/VR.2015.7223320.
URL http://ieeexplore.ieee.org/document/7223320/

[62] P. W. Anderson, P. Zahorik, Auditory/visual distance estimation: accuracy and variability, Frontiers in Psychology 5 (2014) 1–11. doi:10.3389/fpsyg.2014.01097.
URL http://www.frontiersin.org/Auditory_Cognitive_Neuroscience/10.3389/fpsyg.2014.01097/abstract

[63] T. J. Faulkenberry, Computing Bayes factors to measure evidence from experiments: An extension of the BIC approximation, Biometrical Letters 55 (1) (2018) 31–43, arXiv: 1803.00360. doi:10.2478/bile-2018-0003.
URL http://arxiv.org/abs/1803.00360

[64] R. Bakeman, Recommended effect size statistics for repeated measures designs, Behavior Research Methods 37 (3) (2005) 379–384. doi:10.3758/BF03192707.
URL https://doi.org/10.3758/BF03192707

[65] C. C. Berger, M. Gonzalez-Franco, A. Tajadura-Jiménez, D. Florencio, Z. Zhang, Generic HRTFs May be Good Enough in Virtual Reality. Improving Source Localization through Cross-Modal Plasticity, Frontiers in Neuroscience 12. doi:10.3389/fnins.2018.00021.
URL https://www.frontiersin.org/articles/10.3389/fnins.2018.00021/full

[66] K. T. Gagnon, M. N. Geuss, J. K. Stefanucci, Fear influences perceived reaching to targets in audition, but not vision, Evolution and Human Behavior 34 (1) (2013) 49–54. doi:10.1016/j.evolhumbehav.2012.09.002.
URL http://linkinghub.elsevier.com/retrieve/pii/S1090513812000918

[67] J. K. Stefanucci, K. T. Gagnon, C. L. Tompkins, K. E. Bullock, Plunging into the pool of death: Imagining a dangerous outcome influences distance perception, Perception 41 (1) (2012) 1–11. doi:10.1068/p7131.
URL http://pec.sagepub.com/lookup/doi/10.1068/p7131

[68] P. Jaekl, J. Seidlitz, L. R. Harris, D. Tadin, Audiovisual Delay as a Novel Cue to Visual Distance, Plos One 10 (10) (2015) e0141125. doi:10.1371/journal.pone.0141125.
URL http://dx.plos.org/10.1371/journal.pone.0141125

[69] M. Zampini, D. I. Shore, C. Spence, Audiovisual temporal order judgments, Experimental Brain Research 152 (2) (2003) 198–210. doi:10.1007/s00221-003-1536-z.
URL https://doi.org/10.1007/s00221-003-1536-z

[70] T. Y. Grechkin, T. D. Nguyen, J. M. Plumert, J. F. Cremer, J. K. Kearney, How does presentation method and measurement protocol affect distance

estimation in real and virtual environments?, ACM Transactions on Applied Perception 7 (4) (2010) 1–18. doi:10.1145/1823738.1823744.
URL http://portal.acm.org/citation.cfm?doid=1823738.1823744

[71] J. E. Swan, M. A. Livingston, H. S. Smallman, D. Brown, Y. Baillot, J. L. Gabbard, D. Hix, A perceptual matching technique for depth judgments in optical, see-through augmented reality, Proceedings - IEEE Virtual Reality 2006 (2006) 3. doi:10.1109/VR.2006.13.

[72] B. Baumberger, M. Flückiger, M. Paquette, J. Bergeron, A. Delorme, Perception of relative distance in a driving simulator1,2, Japanese Psychological Research 47 (3) (2005) 230–237. doi:10.1111/j.1468-5884.2005.00292.x.
URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5884.2005.00292.x

[73] F. Panerai, J. Droulez, J. M. Kelada, A. Kemeny, E. Balligand, B. Favre, Speed and safety distance control in truck driving: comparison of simulation and real-world environment, in: Proceedings of driving simulation conference, Citeseer, 2001, pp. 91–107.

[74] E. Langbehn, T. Raupp, G. Bruder, F. Steinicke, B. Bolte, M. Lappe, Visual blur in immersive virtual environments, in: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology - VRST '16, ACM Press, New York, New York, USA, 2016, pp. 241–250. doi:10.1145/2993369.2993379.
URL http://delivery.acm.org.ezproxy1.bath.ac.uk/10.1145/3000000/2993379/p241-langbehn.pdf?ip=138.38.44.95&id=2993379&acc=ACTIVESERVICE&key=BF07A2EE685417C5.85B475708465C551.4D4702B0C3E38B35.4D4702B0C3E38B35&__acm__=1519249102_08dd2ee352e035966593a4d3bb58ed44

[75] T. Rousset, C. Bourdin, C. Goulon, J. Monnoyer, J.-L. Vercher, Misperception of egocentric distances in virtual environments: More a question of training than a technological issue?, Displays 52 (2018) 8–20. doi:10.1016/j.displa.2018.02.004.
URL http://www.sciencedirect.com/science/article/pii/S0141938217300732

[76] A. Rungta, N. Rewkowski, R. Klatzky, D. Manocha, P-Reverb: Perceptual Characterization of Early and Late Reflections for Auditory Displays.
URL http://arxiv.org/abs/1902.06880

[77] S. Persky, A. P. Dolwick, Olfactory Perception and Presence in a Virtual Reality Food Environment, Frontiers in Virtual Reality 0, publisher: Frontiers. doi:10.3389/frvir.2020.571812.
URL https://www.frontiersin.org/articles/10.3389/frvir.2020.571812/full

[78] J. L. Campos, J. S. Butler, H. H. Bülthoff, Multisensory integration in the estimation of walked distances, Experimental Brain Research 218 (4) (2012) 551–565. doi:10.1007/s00221-012-3048-1.