
Phenomenal Regression to the Real Object in Physical and Virtual Worlds

Department of Computer Science, University of Hull, Hull HU6 7RX, UK
Tel.: +44-(0)1482-465247
Fax: +44-(0)1482-466666
E-mail: k.w.elner@hull.ac.uk
E-mail: h.wright@hull.ac.uk

Kevin W. Elner · **Helen Wright**

Accepted: 24 November 2014

Abstract In this paper, we investigate a new approach to comparing physical and virtual size and depth percepts that captures the involuntary responses of participants to different stimuli in their field of view, rather than relying on their skill at judging size, reaching, or directed walking. We show, via an effect first observed in the 1930s, that participants asked to equate the perspective projections of disc objects at different distances make a systematic error that is both individual in its extent and comparable in the particular physical and virtual setting we have tested. Prior work has shown that this systematic error is difficult to correct, even when participants are knowledgeable of its likelihood of occurring. In fact, in the real world the error only reduces as the available cues to depth are artificially reduced. This makes the effect we describe a potentially powerful, intrinsic measure of VE quality that ultimately may contribute to our understanding of VE depth compression phenomena.

Keywords Virtual environments · Depth and size perception · Depth compression · Index of phenomenal regression · Thouless ratio · Brunswik ratio

1 Introduction

The virtual environment (VE) is an established component of modern training systems. Some scenarios train for situational awareness but many train specific skills involving judgements of size and distance. Clearly for these it is preferable to have a veridical representation or at least to know if the accuracy of spatial perception falls short. Perception of space relies jointly on the perception of distances and sizes of objects, since retinal size is an ambiguous indicator of objective size until distance is understood, and vice versa. Ability to judge distance in VEs has therefore been extensively studied and generally it is found to be underestimated by participants. Judgement of size has been less intensively investigated but recent findings also show underestimation, possibly due to distance underestimation.

However, in spite of many studies, ensuring the spatial veracity of VEs is an ongoing challenge. Our contribution to this field has been to investigate the potential value in VEs of 'phenomenal regression', a systematic perceptual error made by test subjects when matching standard and response stimuli under certain experimental conditions. The effect was first reported by Thouless in the 1930s (Thouless, 1931a,b). He found that when a subject is asked to match the perspective sizes of two discs at different distances, they consistently overestimate the diameter of the nearer disc. The effect is perhaps most easily understood by comparing a subject's actual field of view (FOV) with how they perceive this FOV. Consider the classic 'trick shot' in Figure 1: when asked to open thumb and forefinger just wide enough to 'pick up' some feature of the scene, a person will make a larger estimate when observing with both eyes in person than is required when seen through the camera's lens. The effect is symmetrical—in the experimental setting, adjusting the nearer disc results in overestimation whereas adjusting the further disc causes underestimation. Moreover, it occurs not only for the sizes of objects but also their brightness and shapes. Thouless called this tendency 'phenomenal regression to the real object' (Thouless, 1931a,b):

phenomenal what his subjects perceived
regression the tendency to report a modified value for the response variable, compared to expectation
to the real object this modified value is always towards the standard, fixed stimulus.

Regression occurs innately, to different extents in different subjects under different conditions of depth cue reduction. This latter property is key to its potential to indicate the quality of a VE implementation, since it offers the opportunity to compare physical and virtual spatial percepts directly.

In this paper we first summarise in Section 2 the considerable body of prior work in VEs investigating distance underestimation, including its possible origins in the quality of the VE graphics and viewing conditions, hardware influences,

Fig. 1 Matching perspective size is remarkably easy in principle but very hard to accomplish in practice. The distant window appears to occupy a larger portion of the FOV when viewed with the subject's two eyes in person, than it does through the camera's lens. Setting up the shot by alternately focussing on the two objects results in a wider-than-needed grasp. In other words the hand's adjustment regresses towards the actually larger, but more distant, window frame.

user self-embodiment and cognition. The extent of regression for any individual is quantified using the Thouless ratio (TR; Thouless, 1931a,b) so Section 3 next explains the derivation of TR in the size-distance formulation that we use, and contrasts it with traditional methods of distance estimation in VEs. Our work in this field concentrates on the depth cue approximations necessary when implementing VEs in large-screen immersive displays (LSIDs), and their potential effect on spatial perception. Sections 4 and 5 describe our experiments and findings when comparing regression in the physical realm and virtual worlds of this type, including the precautions necessary for evoking a consistent response in participants. Finally in Sections 6 and 7 we discuss the contribution of our research to establishing a direct measure of VE visual fidelity, independent of participants' learned skill in distance and size estimation tasks.

2 Depth compression in VEs

A large body of work on depth perception in VEs has investigated medium-field distances (~2m to ~30m), i.e. what Cutting & Vishton (1995) categorised as action space. Its focus has been to understand the phenomenon that revealed itself in early VE studies (see review in Loomis & Knapp, 2003), whereby users with adequate real-world estimation skills nonetheless judge egocentric distances to be shorter than intended in VE scenes.

Studies that made improvements to the quality of the graphics found that depth compression was present whether rendering panoramic photographs or computer generated polygons (Willemsen & Gooch, 2002). Compression was evident even when employing realistic high-resolution photographic textures (Thompson *et al.*, 2004). Willemsen *et al.* (2008) investigated inaccuracies in stereoscopic viewing conditions—inter-pupil distance, binocular and monocular viewing—but concluded they were not likely causes of depth compression.

Other work has focused on aspects of head-mounted display (HMD) hardware. Willemsen *et al.* (2009) investigated mechanical aspects of HMDs. By comparing results between the real world, mock, and real HMDs, they concluded that mass, inertia and FOV of a HMD contributed some but not all of the depth compression, with restriction of the FOV being the largest identifiable source. However, Knapp & Loomis (2004) found that limiting FOV in the real world without the additional mass and inertia typical of a HMD did not result in depth compression. Creem-Regehr *et al.* (2005), in a similar study, found that limiting FOV did result in depth compression but only if there was also restricted head movement. Yet other work has concentrated on the optics of HMDs: Kuhl *et al.* (2009) demonstrated that minification and magnification in HMDs significantly affected distance judgements but that other miscalibrations, notably pitching of the VE and pincushion distortion, did not.

In contrast with studies which have shown depth compression in synthetic VE spaces, Interrante *et al.* (2006) did not observe significant compression when the VE space shown to participants corresponded with the physical space hosting their experiment. They concluded that the cause of depth compression in VEs may not be inherent in the technology itself, i.e. the HMD, but a higher-level cognitive issue with interpreting virtual stimuli. Further work by Interrante *et al.* (2008) investigated whether exposure to the real physical space provides users with a metric for calibrating size and distance judgements in a synthetic VE replication of that physical space; however, the results did not support this hypothesis.

Mohler *et al.* (2008, 2010) found that presence of a self- or displaced avatar in the VE reduced the depth compression, especially if its animation corresponded to the user's own body movement. They theorised that an awareness of one's body (self-embodiment) in a VE may serve as the reference frame required for scaling synthetic space correctly. A similar study by Ries *et al.* (2008) also found a reduction in depth compression and concluded that self-embodiment may facilitate a stronger sense of presence in a VE, suggesting that compression in a VE is a result of cognitive dissonance.

Much of the depth compression work in VEs has been conducted in medium-field distances and has employed the use of HMDs; however, studies have observed underestimation of distances in LSIDs. Murgia & Sharkey (2009) found distance underestimation in rich cue conditions (textured background surfaces) and even greater underestimation in poor cue conditions (limited background cues), though not to the extent found in other studies. They concluded that their experimental methodology, which provided information on the relative sizes of objects in the VE, may have increased users' accuracy in estimating egocentric distance. Piryankova *et al.* (2013) investigated 3 different types of LSID VE and found that distance underestimation was present in all. They concluded that inclusion of stereoscopic and motion parallax cues did not mitigate compression for medium-field distances.

Far less attention has been given to VE depth perception at near-field distances (within ~2m), i.e. what Cutting & Vishton (1995) categorised as personal space. Napieralski *et al.* (2011) found significant underestimation in the perception of near-field distances in a HMD VE compared to a real-world condition. When using LSIDs, Piryankova *et al.* (2013) found that, although distances are still underestimated, stereoscopic depth cues mitigate some of the depth compression effect in the near field.

Ellis & Menges (1998) examined users' perception of distance to augmented reality (AR) objects in the near field. It was found that as depth cues degrade there is an overestimation as virtual objects are judged to be at the distance of the background surface whereupon they are superimposed. Older participants particularly struggled to localise virtual objects placed at shorter distances from them. They concluded that this was due to their inability to accommodate to these focal lengths and, relying instead on the disparity cue, they converge to the surface behind the object. In contrast it was found that distances to AR objects are underestimated in the presence of a real occluding surface that is nearer to the viewer.

Replicating the apparatus of Ellis & Menges (1998), Singh *et al.* (2010) found depth judgements to AR objects in the near field to be underestimated. The presence of a salient occluding surface had a complex effect on depth judgements, with the virtual object appearing nearly at the depth of this surface if it could feasibly be associated with it.

The experiment described in the present paper investigates near-field perception in a LSID VE, neither of which has hithero been the subject of much attention. Moreover, we take a completely different approach to that taken so far. Whereas others in the field have used the accuracy of distance estimation as an indicator of VE spatial quality, we propose to use the regression effect, introduced in Section 1. The next section derives the TR equation needed to quantify the effect, preparatory to using it to compare some real and virtual stimuli in our target VE.

3 Derivation of the Thouless ratio in size-distance experiments

We begin the derivation of TR for size-distance, not with different diameter circles at different distances as our subsequent experiments in Section 4 will actually use, but with inclined circles. The aim is to aid understanding the concepts and equation we will eventually use in the size-distance experiments. In brief, when investigating the perception of shape, Thouless asked participants to look at circles inclined by different amounts to their line of sight. When asked to report the perspective shapes of these circles, either by drawing them or responding to different ellipses held up by the experimenter, the regression effect described in Section 1 caused participants to reproduce or choose a shape that was intermediate between the actual perspective shape and the objective, i.e. circular, shape. In this variant of phenomenal regression, 'phenomenal' refers to the shape as it appeared to his subject, whilst the 'real' object was the circular shape without inclination. Thouless (1931a, p.341) remarks that the real, objective character is known due to observing it "with both eyes fully open and focussed". In his description Thouless also used the term 'stimulus character' which in this experiment means the perspective shape of the inclined object. Figure 2 shows these real, phenomenal and stimulus characters (R, P and S), as a participant in this experiment might have experienced them.

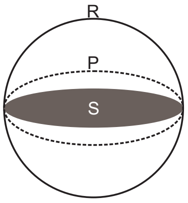


Fig. 2 R (solid line) shows the real, objective shape of the presented disc which, due to its inclination to the participant's line-of-sight, has a stimulus shape S (grey shading). The phenomenal or apparent character P (dotted line) was found always to lie somewhere between R and S. Redrawn from Thouless (1931a, Fig. 2, p.342).

Thouless noted individuals' different degrees of regression and developed a numerical measure, the index of phenomenal regression (IPR), linking R, P and S

$$IPR = \frac{\log P - \log S}{\log R - \log S} \tag{1}$$

In essence, the denominator of this equation measures the total distance separating the stimulus character from the real, objective character. Equation 1 therefore shows the degree to which the perceptual, or phenomenal, character differs from the stimulus, normalised to unity. When a subject makes a perfect perspective match the index is zero and, conversely, a perfect objective match yields an index of unity. IPR later became known as the Thouless ratio.

In the shape experiment R, P and S are the ratios of the minor and major axes of the ellipses in Figure 2; clearly in this case $R = 1$. In size-matching experiments, circular discs of different diameters D and d are presented without inclination at respective distances L and l, and the participant adjusts the size of one to phenomenal equality with the other. If, as in our experiment, the participant adjusts the further disc D, the comparable stimulus character is the size that d would have to assume, when placed at distance L, in order to appear as it does at l, i.e. dL/l. This gives IPR, or now TR, as

$$\frac{\log D - \log \frac{dL}{l}}{\log d - \log \frac{dL}{l}}$$

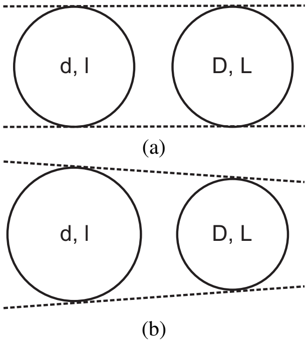
A simple rearrangement is to add and subtract $\log D$ from both the numerator and denominator and collect terms, giving

$$\frac{-\log \frac{dL}{Dl}}{\log \frac{d}{D} - \log \frac{dL}{Dl}} \tag{2}$$

In this formulation we can see parallels with the shape experiment that used the axes' ratios of the stimulus and response shapes. Whereas in the shape experiment $R = 1$ because the object was circular, in the size-distance experiment the real character is the ratio of the discs' diameters d/D. The stimulus character is the ratio of the discs' perspective sizes dL/Dl.

According to Eqn. 2, when $TR = 0$ in the size-distance experiment, participants are matching the perspective sizes of objects ($d/l = D/L$); conversely when $TR = 1$, they are estimating objective equivalence ($d = D$). Since the experiment by design involves objects at different distances, values of TR near to unity therefore demonstrate size-constancy, or the ability to estimate sizes correctly at different depths in a scene. Possibly for this reason, Thouless' work has frequently been cited in later studies of size-constancy, but in the original Thouless (1931a) quite clearly is interested principally in participants' ability to judge perspective size. Although Thouless (1931a,b, p.353, p.1) is consistent in describing his findings only as "a **tendency** [our emphasis] to constancy", this in itself could have contributed to subsequent confusion regarding the aims of his experiments.

The formulation in Eqn. 2 neatly emphasises two points. Firstly, (D, L) and (d, l) are interchangeable without affecting TR (Thouless, 1931a, p.353 footnote). This is not the case for TR's close relative the Brunswik ratio $(P - S)/(R - S)$ (Sedgwick, 1986, p.21–5). A practical consequence is that a Brunswik ratio from experiments adjusting the further disc would require scaling to compare with values from experiments adjusting the nearer disc. The logarithmic formulation of TR requires no such conversion. Secondly, by comparison with Eqn. 1 we see that $\log P = 0$ hence $P = 1$, termed 'phenomenal equality'. This condition is captured visually in Figure 3—participants perceive the discs as subtending the same angle (Figure 3(a)) but their actual, measured projections differ (Figure 3(b)). The greater the difference, the larger is that participant's TR value.



(a)

(b)

Fig. 3 When particpants report the apparent, or phenomenal, equality of near and far discs d and D (a), their typical perspective sizes are actually as at (b). The discs are perceived as if subtending the same angle at the eye but in fact their perspective projections differ.

Whilst it would be expected that estimating objective size is a skill acquired through practice that might therefore deviate from perfection, the finding that participants cannot match *perspective* sizes exactly under normal viewing conditions is somewhat unexpected. It seems the perception of the retinal image itself results from some form of processing involving depth. Thouless describes the response of participants that leads to values of $TR > 0$ as an involuntary reaction that cannot be defeated even when pointed out to them, with TR approaching zero only when various cues of depth are systematically removed (Thouless, 1931b).

We can contrast the proposed TR measure with the methods for VE distance estimation used in the studies summarised in Section 2. By far the commonest is ‘blind walking’ (also called direct or directed walking) whereby participants observe a target object and then walk to its perceived position blindfolded. Examples of this technique in use occur in the work of Willemsen & Gooch (2002); Knapp & Loomis (2004); Creem-Regehr *et al.* (2005); Interrante *et al.* (2006, 2008); Mohler *et al.* (2008, 2010); Ries *et al.* (2008); Kuhl *et al.* (2009) and Willemsen *et al.* (2009). Direct walking requires locomotion to the actual target position and where this is impractical due to space or other considerations, other variants have developed, notably triangulated walking used for example by Thompson *et al.* (2004) and Willemsen *et al.* (2008, 2009). Yet others have used verbal reporting of distance (Knapp & Loomis, 2004; Napieralski *et al.*, 2011; Piryankova *et al.*, 2013), visually guided reaching (Singh *et al.*, 2010; Napieralski *et al.*, 2011) and blind reaching (Ellis & Menges, 1998). Less frequently, studies of depth compression invoke size estimation skills. For example, Murgia & Sharkey (2009) asked participants to estimate distance supported by information on the relative sizes of real and virtual objects, whereas Loomis & Knapp (2003) used affordance to pass through an aperture.

Distance estimation accuracy is commonly used to infer the spatial quality of a VE, but it could also conceivably be used to achieve correct perception by artificially inflating geometry to offset the compression effect. Perceptual calibration is also the aim of Ponto *et al.* (2013), but their approach is conversely to allow users to adjust the VE’s projection parameters by making objects look level, square and stationary. Phenomenal regression falls between these two, being both potentially an indicator of VE visual quality and a target for calibration approaches. Our justification for this study is TR’s sheer simplicity compared with traditional methods; as captured in Figure 3, it depends only on participants attempting to match, all the while remaining seated, the perspective projections of two objects in their FOV.

4 Comparing TR in physical and virtual environments

An experiment was designed to compare individual participants’ Thouless ratios utilising physical and virtual stimuli. In brief, participants were asked to adjust the size of a physical far disc to match the perspective size of a near disc, the latter presented in both physical and virtual forms. The distances to the discs are nominally 1.22m and 2.72m so the experiment can be classified as taking place in personal space and action space (Cutting & Vishton, 1995), respectively the distances within and slightly beyond an arm’s length, and beyond this but within 30m. The following subsections provide more detail.

4.1 Apparatus

The Hull Immersive Visualization Environment (HIVE) is a rear-projected, single surface display, 5.33m wide by 2.44m high. A raised stage area in front is 5.33m wide by 3.22m deep. The display is driven by two active stereo projectors arranged horizontally with hardware edge blending. LCD shutter glasses are tracked optically and the resulting eye positions are used to calculate asymmetric parallel axis stereoscopic viewing frusta. The user’s head orientation is also taken into account, in total generating perspective, motion parallax and binocular disparity cues to depth (Figure 4). The depth at which an object is perceived is given by the vergence point as the stereo pairs are fused (Figure 4(b) and (e)). In the real world, vergence works in conjunction with accommodation of the eye’s lens, to focus the light rays coming from the object onto the retina. However, in a virtual world implemented on a projection screen, the light rays always emanate from the display surface. This creates an accommodation-vergence mismatch for the user which worsens progresssively as objects’ perceived distances from the screen increase.

Three discs are projected on three displays: the HIVE screen used as a large, physical monitor, a standard LED monitor and a virtual monitor implemented using the artificial depth cues in Figure 4. The displays are positioned so they do not overlap in the participant’s FOV and at such an angle that they face the participant (Figures 5 and 6).

The centre disc displayed in the plane of the HIVE screen has zero parallax and therefore no stereo separation. The focus and vergence cues are correct here so the disc has the perceptual properties of being physical. When the participant faces this disc it is 2.72m from them. This is the disc that participants adjust and displaying it on the large HIVE screen eliminates extraneous size cues that would be available from a monitor surround.

The right-hand disc on the LED monitor is positioned 1.22m from the participant and supported on a tripod. Like the centre disc, a disc on the LED monitor has the perceptual properties of being physical because the focus and vergence cues are correct. The physical LED monitor is tracked to ensure that the virtual monitor is reflected across the centre line

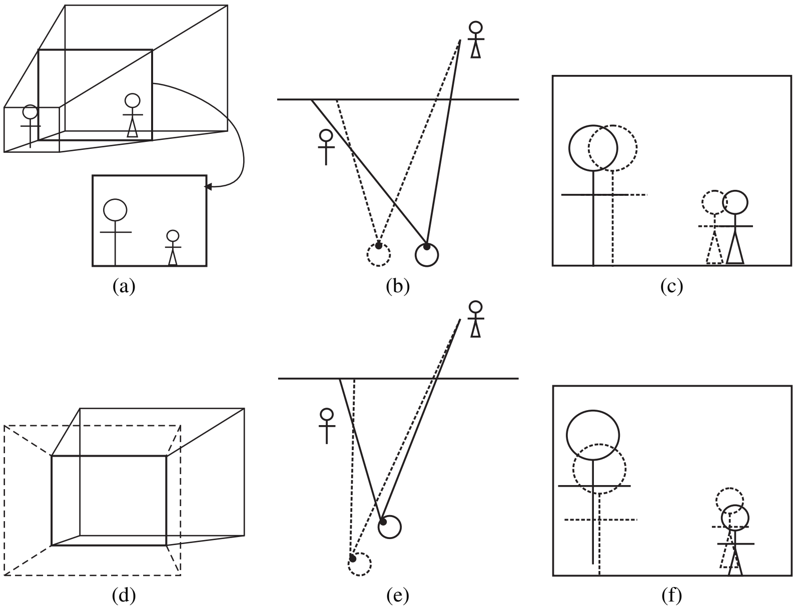


Fig. 4 Depth cues in our LSID VE. (a) shows the truncated pyramidal form of the viewing frustum that implements a perspective projection onto the image plane (bold) of objects at different distances. In (b), computing separate left- (dotted) and right-eye images renders near objects with negative parallax and far objects with positive parallax (c). When fused, these objects respectively appear to be out of and into the screen. (d) shows how the frustum is recalculated as the user moves, in this case towards the right (dashed lines; only the frustum portion beyond the fixed image plane is shown for clarity), thereby generating motion parallax. As well as tracking position information, the user's head orientation is also used to foreshorten the stereo separation during rotation around the vertical axis (e). Rotation around the horizontal axis perpendicular to the screen causes the images to separate vertically (f), so the stereo pairs remain fusible throughout a wide range of head motions. Taken together, the illusion of depth from these various measures is compelling, but correct focus cues cannot be simulated by this type of environment (see text).

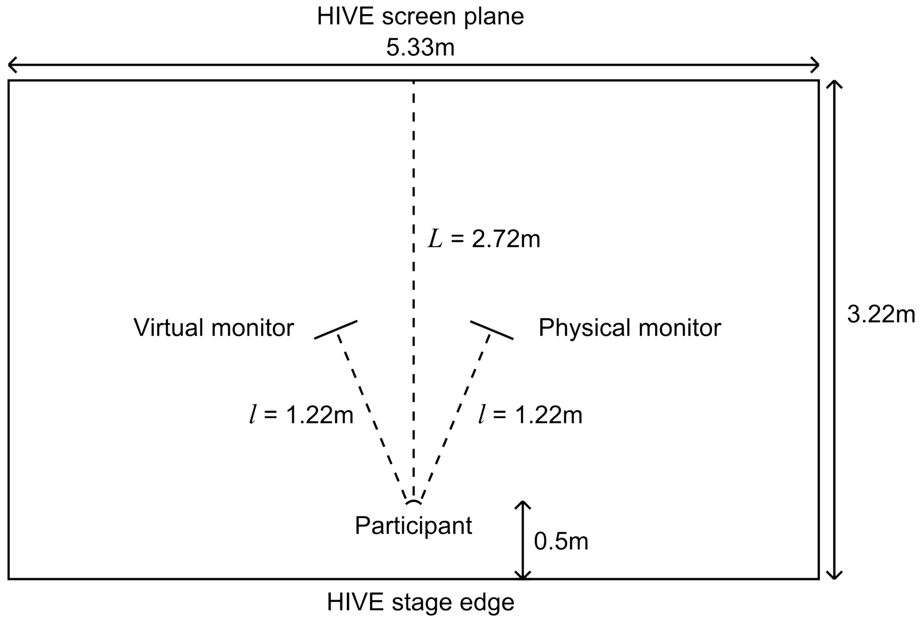


Fig. 5 Plan view of apparatus on HIVE stage.

of the stage. The monitor's full physical dimensions are 56.8cm by 33cm; however, the screen is masked by black felt to reveal only an area of 0.29cm by 0.29cm. This corresponds to the maximum size of virtual monitor that is reproducible without clipping by the HIVE screen edge.

The left-hand disc is projected on the virtual monitor, a black rectangle corresponding to the visible portion of the physical LED monitor but at exactly the mirror position. The disc on the virtual monitor and the monitor surround itself (simply a dark grey boundary) are displayed on the HIVE screen with negative parallax so the whole assembly appears to be in front of the screen, 1.22m from the participant. Note however that the accommodation distance when observing

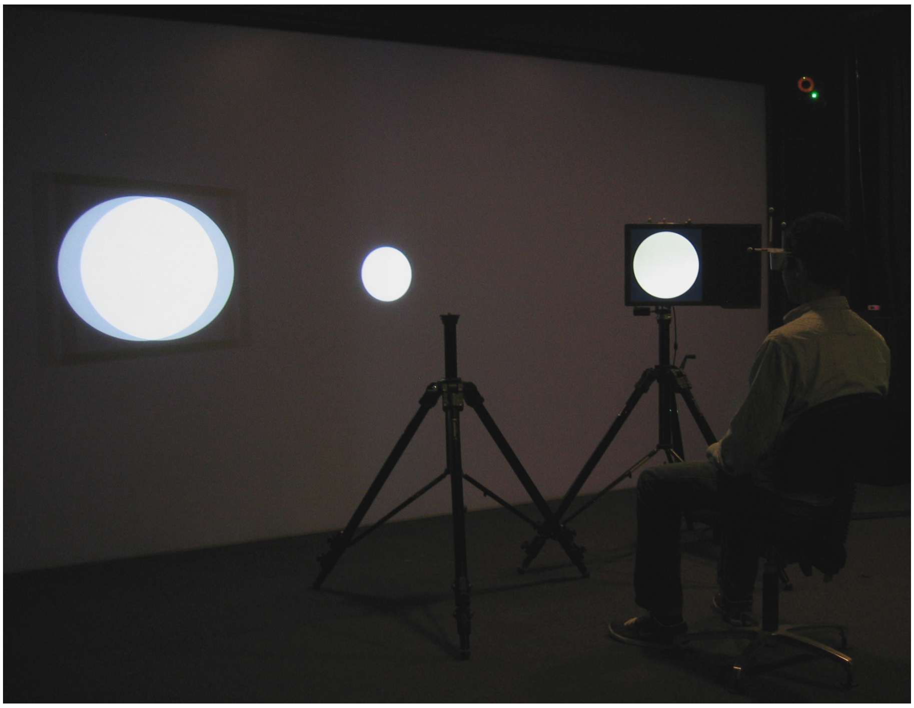


Fig. 6 Photograph of apparatus. Note that the scene is illuminated so the brightness and contrast of the displays are not representative of the actual experimental conditions experienced by a participant. Furthermore, the stereo separations seen on the left-hand, virtual disc are not evident whilst wearing the shutter glasses and the virtual disc appears above the physical tripod on the left when these separations are fused.

this disc is on the HIVE screen, some 2.97m away when the participant turns to face it. A physical tripod is placed below the virtual monitor.

All three discs are white with no texture and all are observed through the shutter glasses. Efforts were made to ensure uniform brightness across the three discs, measured using a light meter at each display, through the shutter glasses. Brightness was kept within a range of ±1 lux.

To perform the task correctly the participant must look at and focus on each disc with both eyes. Forming any kind of strategy such as placing both discs in the FOV but focusing on neither is undesirable because it eliminates the binocular cues of focus and vergence. The solution adopted was to render a disc only when the participant was facing it.

4.2 Procedure

A potential weakness when measuring TR lies in knowing whether participants are actually attempting to match perspective or objective size. Thus although the experimenter may have clear intentions it requires the participant to have understood these in order to respond as required. The work of Gilinsky (1955) sheds light, by describing the effect of different participant instructions on the results of size-matching experiments in the physical environment. Explicit instruction is needed to ensure participants do actually match perspective or objective size, depending on the aims of the particular study.

In the present experiments we need participants to attempt to match perspective size, that is, subjects are to report when they think the visual angles subtended by the discs are the same, not when the objective sizes of the discs match. This point is emphasised using two physical, white felt discs. The participant is invited to take a seat on the HIVE stage and swivel to face the left wall. A 27cm disc is fixed to a black curtain and with the participant's verbal assistance adjusted to their eye level. Holding a 19cm disc at the participant's eye level and next to the larger disc, the experimenter states "This is a large disc, this is a small disc, this one looks larger than the other, doesn't it?" The goal of this instruction is to make it clear that the smaller disc is really physically smaller than the large one and therefore when at the same distance it subtends a smaller angle. The participant acknowledges this and the smaller disc is then slowly carried by hand towards them at their eye level, being careful not to occlude the larger disc. As the experimenter does this they state "As I bring this disc closer to your eyes it looks bigger in your field of view until finally it looks bigger than the other disc".

The purpose of these instructions is to emphasise that the disc is not physically getting bigger, but it does fill a larger portion of the FOV as it moves closer. Once the participant acknowledges this the experimenter states "Now if I slowly move the disc away from you is there a distance where both discs look the same size, i.e. where their perspective sizes match?" Again, once the participant acknowledges a distance where this statement is true the experimenter states "During the experiment you will perform a similar task, but instead of moving a disc, both discs will be fixed at different distances. You will adjust the diameter of a far disc until its perspective size matches a near disc."

Before the experiment the participant's interpupillary distance (IPD) is measured using software and a web-cam, and the apparatus is calibrated to ensure the discs are at the correct distance, height and orientation. It is made clear to the participant that calibration is not part of the experiment. The distance between the participant's chair and the HIVE screen is adjusted so that, when they turn to face each disc, the distance from their eyes to the near displays is 1.22m and, to the HIVE screen, 2.72m. The participant sits, turns to look at each disc and if adjustment is necessary stands up while the chair is moved forwards or back. This is repeated until the desired distances are achieved and it takes 5-10 adjustments to get the desired position. The participant's head is not held in position following this adjustment but is subsequently tracked continuously and their actual distances from the discs are used in the calculations of TR. We found that no participant performed the experiment at more (and most considerably less) than 7cm from their initial calibration position and during measurements participants' positions remained stable, with $|l_{physical} - l_{virtual}| < 2$cm.

Next, the participant checks the height of their seat and if necessary adjusts it so their feet are on the floor and they feel comfortably supported. At this point the experimenter gives a command to the display software to match the height of the centre disc to the participant's tracked eye level. The tripod holding the LED monitor is then raised or lowered manually until the participant reports that the disc height matches that of the centre disc. Because the LED monitor is tracked the virtual monitor on the left also rises to the same position. The second tripod beneath the virtual disc is then adjusted to the same height as the first. As a final check the participant is asked "If you glance across all three discs, they should all appear to be at the same height and level with your eyes?" Finally the angle of the LED monitor is adjusted until the participant states they are facing it directly and the disc therefore appears circular. The angle of the virtual monitor and disc automatically adjusts to mirror the LED monitor. Figure 7 illustrates the apparatus set-up following the calibration procedure.

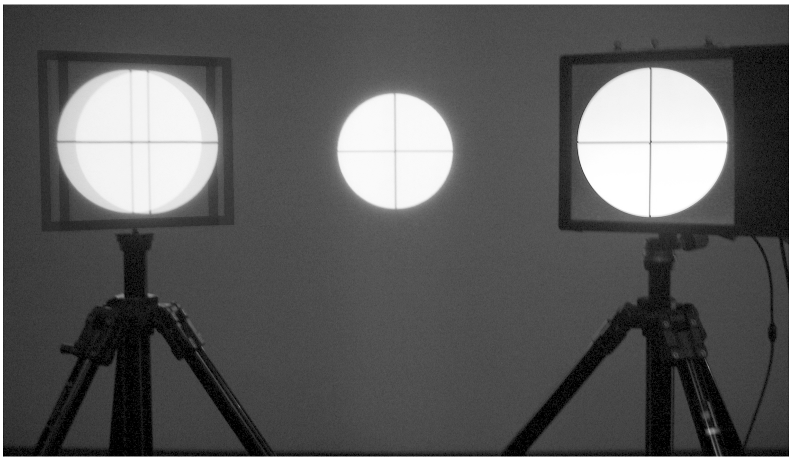


Fig. 7 Photograph of apparatus from the participant's viewpoint following the calibration procedure. All 3 discs are at the same height, facing the viewer and level with the viewer's eye. The black crosses drawn on the discs are to assist the calibration procedure. The stereo separation of the virtual disc and monitor (left) is not visible when the participant is wearing the LCD shutter glasses. This photograph illustrates how the virtual monitor appears to rest upon the physical tripod. The photograph has been manipulated for clarity of printing: the brightness and contrast do not match those of the experiment conditions; the image is desaturated; and the borders of the virtual monitor have been enhanced.

Once the apparatus has been calibrated the participant is instructed on the experimental procedure. They are given the specific instruction "Upon each experimental trial you will be presented with two discs, one on the centre screen and one on either the left or right monitor". No distinction is drawn between the virtual and physical monitors. It is explained that the discs are only visible when participants turn and face the display directly. They are handed a wireless mouse—"With the mouse wheel you are to adjust the size of the centre disc until its perspective size matches the standard disc." Paricipants are told that between each trial all the discs will disappear and a white box with a black cross will be displayed on the large screen in front for a period of three seconds, during which they are to look at the cross. It is explained that at the start of each trial the experimenter will say 'left' or 'right' to indicate which monitor the standard disc will appear on. They are reminded that for every trial the adjustable disc will be displayed on the centre screen.

Participants are told that they can turn back and forth between the two discs as many times as required and, to avoid any potential fatigue or strain from turning the head, are encouraged to do this by swivelling their chair. Once they are satisfied that both discs look the same they are asked to acknowledge this verbally and the next trial commences. Finally, participants are told that they should not narrow or close their eyes, and that they cannot use their fingers to aid them

Table 1 Slope, intercept, coefficient of determination (R^2) and p-value of virtual versus physical TR values, for each disc size

Disc size	Slope	Intercept	R^2	p
15cm	0.8237	0.0295	0.7277	$< .01$
19cm	1.3108	-0.1324	0.7800	$< .01$
23cm	0.9815	-0.0135	0.8359	$< .01$
27cm	0.9804	-0.0201	0.7906	$< .01$

with the task. The participant is given two practice trials, one using the left monitor, the other using the right, and during these they can ask any questions. After completing the two practice trials the experiment begins.

5 Results

Sixteen participants were chosen at random from the postgraduate students and lecturers in the Computer Science department at the University of Hull. All but three were male between the ages 23-58 with a mean age of 39. No participant was working within the field or knowledgeable of VEs or psychology, but all had had some exposure to head tracked stereo displays at university open day events. Participants had normal or corrected-to-normal vision but to maintain comparability with the Thouless (1931a) experiment we deliberately did not screen for stereo vision.

There are two experimental factors for the near disc—type and size. Disc type has two categorical levels—physical and virtual, whilst disc size has four interval levels—disc diameters were 15cm, 19cm, 23cm and 27cm. The chosen disc sizes are comparable with those of the Thouless (1931a) experiment, with 27cm representing the maximum disc size that could be displayed on the available apparatus. Each participant performed five trials for each condition, making a total of 40 trials per experiment. During each trial, participants' head positions were logged at each moment they faced a visible disc; these distances were used to calculate average values for l and L, in order to calculate the TR value for that trial. Participants were tested for all conditions within a single session with the order of conditions randomised.

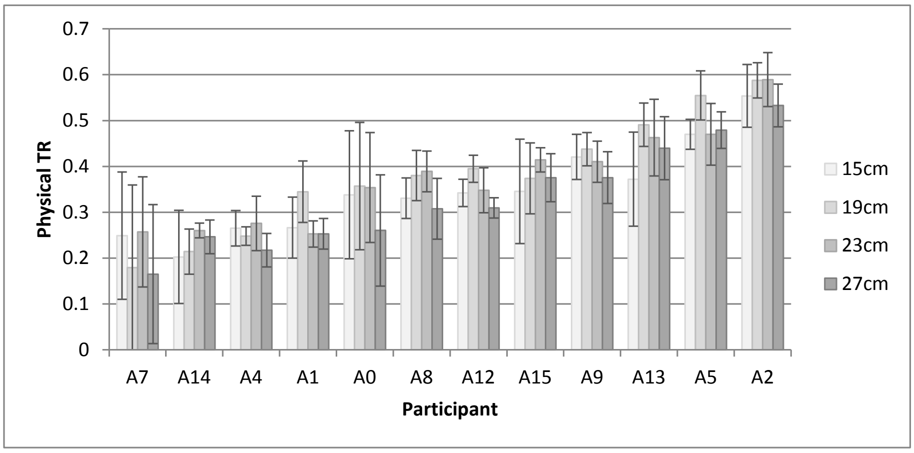
Four datasets were eliminated prior to analysis: one participant ended the experiment prematurely due to eye strain caused by not wearing their corrective lenses; a second stated they had engaged in a strategy to use retinal after-images; a third participant said they had forgotten what they were supposed to do and might have performed objective instead of perspective matches; a fourth was discovered to have performed the experiment at the wrong IPD due to an error in the procedure. Figure 8 shows the resulting means and standard deviations across five trials of physical and virtual TR values, per participant, per disc. Participants have been arranged in order of increasing overall mean physical TR value in order to demonstrate the range of TR values observed.

Mean TR values were subjected to a two-way repeated measures ANOVA. There were no outliers and the data was normally distributed for each group, as assessed by boxplot and Shapiro-Wilk test ($p > .05$), respectively. There was no significant interaction between disc type and disc size, $F(3, 33) = .281$, $p > .05$ and no significant main effect of disc type, $F(1, 11) = 2.852$, $p > .05$. However for the disc size factor, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(5) = 11.874$, $p = .037$, and therefore the Greenhouse-Geisser correction was applied ($\varepsilon = .669$), giving a significant main effect of disc size, $F(2.008, 22.086) = 9.758$, $p < .05$. For the disc size, *post-hoc* pairwise analysis of TR with a Bonferroni adjustment revealed that TR increased statistically significantly from 15cm to 19cm ($M = .026$, 95% CL[.009, .042], $p = .002$). TR decreased from size 19cm to 27cm ($M = .048$, 95% CL[.013, .084], $p = .007$), and from size 23cm to 27cm ($M = .046$, 95% CL[.085, .006], $p = .021$).

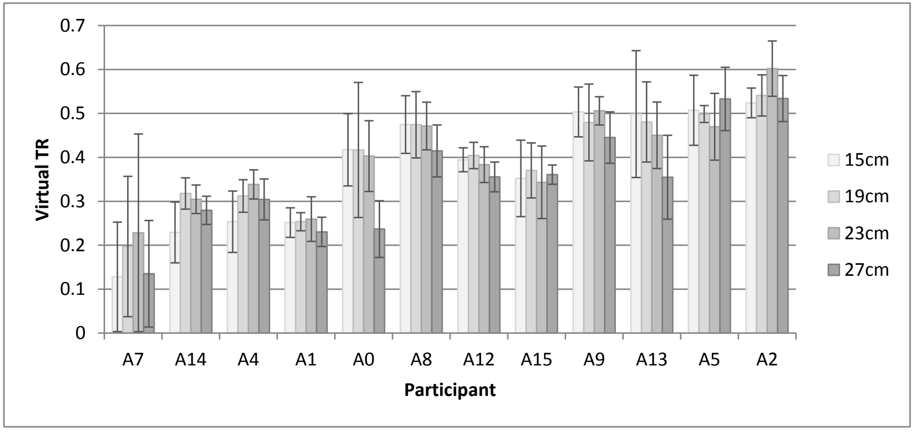
The degree of correlation of virtual and physical TR values was investigated by fitting a straight line, adopting the approach of Press *et al.* (1992, p.666–668) in order to take account of the presence of errors in both sets of measurements. Due to the significant main effect of disc size this was done for each disc separately, with the results shown in Table 1 and the corresponding plots in Figure 9.

6 Discussion

Figure 8(a) shows that participants matching perspective size in the physical world do indeed exhibit $TR > 0$ as Thouless (1931a,b) described and, furthermore, there is individual variation shown in this measure. In some individuals the departure from $TR = 0$ is very marked (two with mean TR around or above 0.5), bearing in mind that $TR = 1$ equates to size-constancy, or perfect objective matching. We are confident this is a genuine finding and not the result of inadvertent objective matching by participants, given the precautions we described in Section 4.2 and the disclosure of one participant who realised for themselves they had strayed from the required task. We also experienced one participant who asked to view their results after the experiment and could not believe they had not performed a perspective match. Thouless (1931b) mentioned that some of his participants made the same comment.



(a)

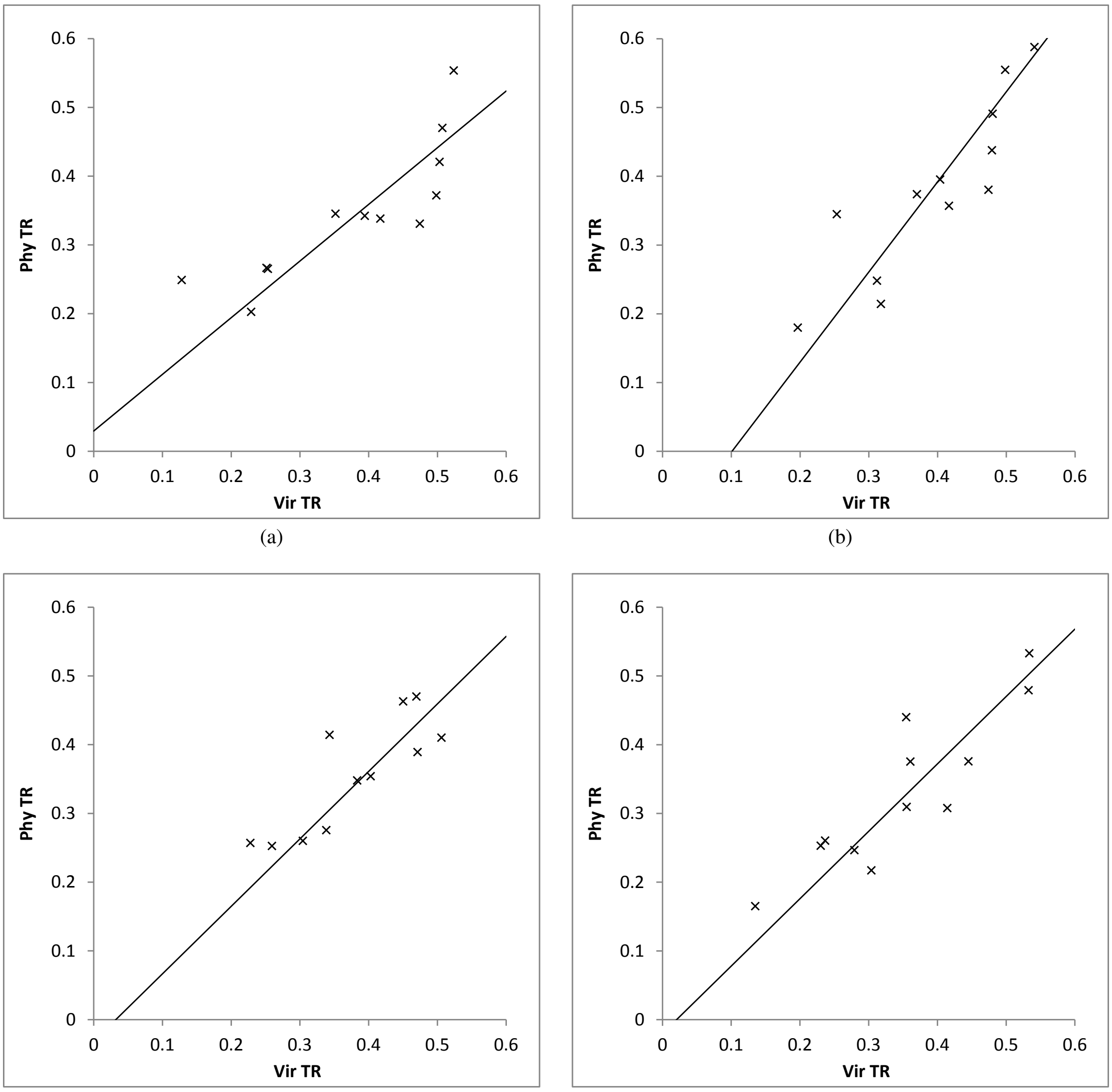


(b)

Fig. 8 Mean TR values per participant for each disc in the physical and virtual conditions, arranged in ascending order of overall mean value in the physical condition. Error bars denote $\pm\sigma_t$, the trial standard deviation for each condition tested, showing that the regression effect exhibited by participants is considerably more variable in size than the variability in its measurement.

The error bars in Figure 8(a) and (b) demonstrate that TR is generally reproducible for both the physical and virtual stimuli, at least for the duration of our experiment, except for two individuals A0 and A7 who exhibit rather large trial standard deviations. The reasons for this require more investigation but indications from the other participants are encouraging, since searching for VE perceptual effects using TR will require it to be stable under experimental conditions. Also encouraging is the correlation demonstrated in Figure 9 between physical and virtual TR values and the large coefficients of determination reported in Table 1. Although the slopes and intercepts in Figure 9 are not precisely unity and zero these plots show sufficient relationship to warrant further investigation—no correlation would surely have ruled out TR as a potential quality metric.

The ANOVA results show no main effect on TR of disc type but do show an effect of disc size. The former was unexpected but on reflection perhaps not surprising in the AR set-up we describe. The virtual disc is augmented by a physical tripod as our aim was to make the appearance of the virtual and physical components as similar as possible. We surmise that participants converge to the correct distance of the physical tripod, outweighing the incorrect accommodation cue presented by the virtual monitor. This is consistent with the findings of Ellis & Menges (1998) and Singh *et al.* (2010), even though in our case the real element is only adjacent to the virtual object in the FOV, not actually occluding it. Whatever the mechanism, most participants remarked how the virtual monitor appeared uncannily real to them. Some had not even realised it was virtual and expressed their shock when, upon completion of the experiment, they removed the shutter glasses and saw that there was actually no monitor on the left of the stage. We consider the augmentation of the environment further in Section 7 when we discuss our future experiments. The latter effect of disc size is surprising and requires more investigation. It might be due to weak control of the ambient light conditions, not from the room lights which were always set the same, but from variable light wash coming from the HIVE screen as participants adjusted the centre disc to phenomenal equality. Especially for participants with small TR, the far disc is physically large even though it subtends much the same angle at the eye as the small disc. We discuss how to mitigate this effect in Section 7.



(a) (b)

(c) (d)

Fig. 9 Correlation of physical and virtual TR values for disc sizes (a) 15cm, (b) 19cm, (c) 23cm and (d) 27cm. Model-fitting parameters, weighted R^2 and p values can be found in Table 1.

7 Conclusions and future work

For TR to be a truly useful tool we recognise that, as well as demonstrating it is measurable and comparable in the specific case described in this paper, we also have to establish whether it is sensitive to depth cue approximations in the VE. Thouless himself has proved this in the physical domain with his cue-reduction experiments but it remains to be seen if arguably more subtle effects, such as accommodation-vergence mismatch, have a measurable effect.

Our next experiment proposes to investigate precisely this. Using a variant of the apparatus described in Section 4.1 we will gradually mitigate the incorrect focus cue present in the virtual disc, by repeating the experiment at progressively shorter distances from the HIVE screen. We also intend to remove the tripod below the virtual disc so that the only true cue to its distance is the vergence point as the participant fuses the stereo pairs.

We will also pay due regard to improving our overall experimental procedure. Variable ambient light level is one issue which we believe can be adequately controlled by the simple expedient of using grey rather than white discs. Additionally we need to establish whether the current, mostly moderate variability in repeated trial values can be reduced still further. Several participants commented that perspective matching was a difficult task to perform and they started slowly; however, it became much easier as they relaxed. Therefore it is feasible that more practice trials would prove beneficial.

In summary, the work of Thouless has often been inappropriately cited when investigating size-constancy; it has been satisfying to reproduce his perspective-matching results many decades later with modern equipment and thereby re-read his accounts with fresh understanding. Our study is at a preliminary stage but initial findings are encouraging that TR

1. can be determined independently of participants' skill in spatial tasks, and
2. has potential as an intrinsic measure of VE spatial quality.

Acknowledgements We acknowledge useful discussions with Derek Wills and James Ward of the Department of Computer Science at the University of Hull. KWE was supported by a doctoral student scholarship funded by the Department of Computer Science at the University of Hull.

References

Creem-Regehr, S.H., Willemsen, P., Gooch, A.A. & Thompson, W.B. (2005). The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual environments. *Perception*, **34**, 191–204.

Cutting, J.E. & Vishton, P.M. (1995). *Handbook of perception and cognition; Perception of space and motion*, vol. 5, chap. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth, 69–117. San Diego, CA: Academic Press.

Ellis, S.R. & Menges, B.M. (1998). Localization of virtual objects in the near visual field. *Hum. Factors*, **40**, 415–431.

Gilinsky, A.S. (1955). The effect of attitude upon the perception of size. *The American Journal of Psychology*, **68**, pp. 173–192.

Interrante, V., Ries, B. & Anderson, L. (2006). Distance perception in immersive virtual environments, revisited. In *Proceedings of the IEEE conference on Virtual Reality*, VR '06, 3–10, IEEE Computer Society, Washington, DC, USA.

Interrante, V., Ries, B., Lindquist, J., Kaeding, M. & Anderson, L. (2008). Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence: Teleoper. Virtual Environ.*, **17**, 176–198.

Knapp, J.M. & Loomis, J.M. (2004). Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoper. Virtual Environ.*, **13**, 572–577.

Kuhl, S.A., Thompson, W.B. & Creem-Regehr, S.H. (2009). HMD calibration and its effects on distance judgments. *ACM Trans. Appl. Percept.*, **6**, 19:1–19:20.

Loomis, J.M. & Knapp, J.M. (2003). *Virtual and Adaptive Environments Applications, Implications, and Human Performance Issues*, chap. Visual Perception of Egocentric distance in Real and Virtual Environments, 21–46. CRC Press.

Mohler, B.J., Bülthoff, H.H., Thompson, W.B. & Creem-Regehr, S.H. (2008). A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, APGV '08, 194–, ACM, New York, NY, USA.

Mohler, B.J., Creem-Regehr, S.H., Thompson, W.B. & Bülthoff, H.H. (2010). The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoper. Virtual Environ.*, **19**, 230–242.

Murgia, A. & Sharkey, P.M. (2009). Estimation of distances in virtual environments using size constancy. *The International Journal of Virtual Reality*, **8**, 67–74.

Napieralski, P.E., Altenhoff, B.M., Bertrand, J.W., Long, L.O., Babu, S.V., Pagano, C.C., Kern, J. & Davis, T.A. (2011). Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Trans. Appl. Percept.*, **8**, 18:1–18:19.

Piryankova, I.V., de la Rosa, S., Kloos, U., Bulthoff, H.H. & Mohler, B.J. (2013). Egocentric distance perception in large screen immersive displays. *Displays* 34(2):153–164

Ponto, K., Gleicher, M., Radwin, R.G. & Shin, H.J. (2013). Perceptual calibration for immersive display environments. *IEEE Trans. Visual Comput. Graphics*, **19**, 691–700.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., eds. (1992). *Numerical Recipes in C: The Art of Scientific Computing*, chap. Modelling of Data, 666–668. Cambridge University Press, 2nd edn.

Ries, B., Interrante, V., Kaeding, M. & Anderson, L. (2008). 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, 167–170, ACM, New York, NY, USA.

Sedgwick, H.A. (1986). *Handbook of Perception and Human Performance*, vol. 1, chap. Space perception, 129–158. New York: John Wiley and Sons.

Singh, G., Swan, J.E., II, Jones, J.A. & Ellis, S.R. (2010). Depth judgment measures and occluding surfaces in near-field augmented reality. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, 149–156, ACM, New York, NY, USA.

Thompson, W.B., Willemsen, P., Gooch, A.A., Creem-Regehr, S.H., Loomis, J.M. & Beall, A.C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence: Teleoper. Virtual Environ.*, **13**, 560–571.

Thouless, R.H. (1931a). Phenomenal regression to the real object. I. *British Journal of Psychology. General Section*, **21**, 339–359.

Thouless, R.H. (1931b). Phenomenal regression to the real object. II. *British Journal of Psychology. General Section*, **22**, 1–30.

Willemsen, P. & Gooch, A.A. (2002). An experimental comparison of perceived egocentric distance in real, image-based, and traditional virtual environment using direct walking tasks. Tech. rep., University of Utah Computer Science.

Willemsen, P., Gooch, A.A., Thompson, W.B. & Creem-Regehr, S.H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoper. Virtual Environ.*, **17**, 91–101.

Willemsen, P., Colton, M.B., Creem-Regehr, S.H. & Thompson, W.B. (2009). The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Trans. Appl. Percept.*, **6**, 8:1–8:14.