

# Visuotactile Integration for Depth Perception in Augmented Reality

Nina Rosa, Wolfgang Hürst, Peter Werkhoven and Remco Veltkamp  
Utrecht University, the Netherlands  
{n.e.rosa, huerst, p.j.werkhoven, r.c.veltkamp}@uu.nl

## ABSTRACT

Augmented reality applications using stereo head-mounted displays are not capable of perfectly blending real and virtual objects. For example, depth in the real world is perceived through cues such as accommodation and vergence. However, in stereo head-mounted displays these cues are disconnected since the virtual is generally projected at a static distance, while vergence changes with depth. This conflict can result in biased depth estimation of virtual objects in a real environment. In this research, we examined whether redundant tactile feedback can reduce the bias in perceived depth in a reaching task. In particular, our experiments proved that a tactile mapping of distance to vibration intensity or vibration position on the skin can be used to determine a virtual object's depth. Depth estimation when using only tactile feedback was more accurate than when using only visual feedback, and when using visuotactile feedback it was more precise and occurred faster than when using unimodal feedback. Our work demonstrates the value of multimodal feedback in augmented reality applications that require correct depth perception, and provides insights on various possible visuotactile implementations.

## CCS Concepts

•Human-centered computing → Mixed / augmented reality; Haptic devices;

## Keywords

Augmented reality; multisensory perception; haptics; sensory redundancy; sensory substitution; depth perception

## 1. INTRODUCTION

When using an augmented reality (AR) head-mounted display (HMD), virtual visual cues differ from real visual cues. For example, humans can determine an object's depth from cues such as accommodation and vergence. However for virtual objects, accommodation for different distances is

often not possible since all objects are generally projected at a static distance from the eyes. Such a conflict can cause a bias in depth estimation [9, 18, 19]. As AR applications are increasingly employing gestural interaction, this bias will likely negatively affect reaching tasks such as touching and grabbing. In many real world scenarios, humans deal with similar errors by integrating redundant information through a different modality, for example the tactile sense [8].

The aim of our study is to determine the feasibility of using the tactile sense as an additional information channel for depth information. In particular, we investigate the depth estimation bias of virtual objects in an active task in peripersonal space in visual-only, visuotactile, and tactile-only settings. We expect that the users are able to translate tactile feedback into depth information and that the bias caused by visuals will decrease with the use of tactile feedback compared to visual feedback. Our study shows that this is indeed the case for two different implementations, mapping distance to vibration intensity and to vibration position on the skin, and that using tactile feedback without visuals can be more correct than using only visual feedback.

## 2. BACKGROUND

### 2.1 Depth Perception in Augmented Reality

One of the earliest works measuring depth estimation in AR is [12]. In this article, the authors discuss important parameters that need to be evaluated for accurate HMD calibration. Based on this discussion, an experiment for depth perception assessment is performed. The results initially showed that virtual objects were seen farther away than real objects when both objects were presented at the same depth in visual space, however in [13] it was shown that these errors no longer occurred when using an improved HMD. In [6], examiners investigated the influence of various aspects such as accommodation, vergence and display type (monocular, binocular, stereo) on depth perception and found that occluding objects resulted in depth judgments being biased towards the observer. When the occluder was in front of and close to the virtual object, the object seemed a bit closer to the user than it actually was.

More recently, [18] examined the difference between perceptual matching and blind reaching to study depth judgments, and the difference in depth accuracy between virtual and real objects. Matching was shown to be more accurate than reaching, and matching real objects was more accurate than matching virtual objects. However in contrast to expectation, reaching real objects was not more accurate than

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ICMI'16, November 12–16, 2016, Tokyo, Japan  
© 2016 ACM. 978-1-4503-4556-9/16/11...\$15.00  
<http://dx.doi.org/10.1145/2993148.2993156>



Figure 1: A participant reaching for a virtual cube.

reaching virtual objects. Accuracy for reaching both real and virtual objects improved using a different experimental setup, but accuracy for matching real objects worsened. Accuracy for real objects was then improved after an intervention phase using corrective feedback with matching. Lastly, [4] investigates depth perception in AR in a *vision-for-perception* versus a *vision-for-action* scenario. Results showed that for matching there was a bias towards the user, and no bias in the active scenario, however it is questionable whether this was due to the activeness of the task rather than the method used to collect the data.

It is clear that in a passive task, a bias occurs in virtual object depth perception towards the user. It is however unclear whether a bias occurs in an active task, a scenario that is very likely in any interactive application. A goal of our research is therefore to investigate depth estimation in an active reaching task.

## 2.2 Haptic Depth Perception

The application of haptics for depth perception was initially intended for the blind, and has been studied for over a century. Many studies have applied tactile arrays as sensory substitution for vision. Generally, these arrays make a 2D image of the world through a camera, map  $(x, y)$  coordinates of a (group of) pixel(s) to  $(\tilde{x}, \tilde{y})$  array coordinates, and light intensity to tactile intensity. See [3] for an overview.

In AR, the goal would not necessarily be to substitute vision for haptics, but to aid vision through haptics. An example of this is given in [5], where the user is asked to place his/her hand at the perceived position of a random dot pattern, ‘raised’ off the surface by various disparities. It was shown that coupling force feedback with stereopsis lead to a decrease in variance of depth perception. Similarly, [1] examined depth perception in virtual reality with real haptic feedback, and found that visual and haptic calibration increased accuracy of distance estimates. Consequently, a goal of our research is investigate the influence of various tactile feedback configurations on the depth estimation bias of virtual objects.

## 3. EXPERIMENT

### 3.1 Research Question and Objective

The goals of this study were formalized in the following research question: *Can tactile feedback correct the depth es-*

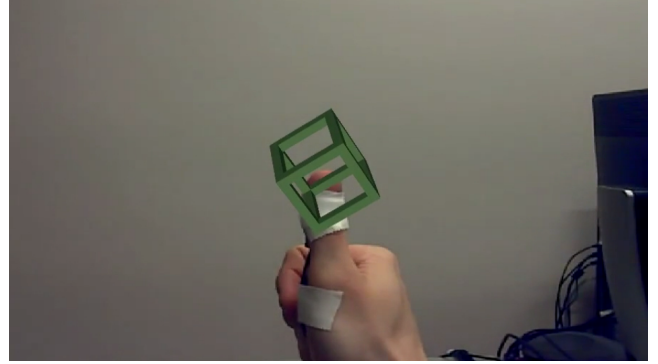


Figure 2: A participant’s view during the task.

*timization bias in an active reaching task in AR?* While [5] applied force feedback at a single position in space, our approach was to assist the reaching process by providing a gradient of vibrotactile feedback. The objective of our study was therefore to measure the perceived depth of virtual objects by placement of the real hand, which is guided by this gradient.

### 3.2 Material and Task

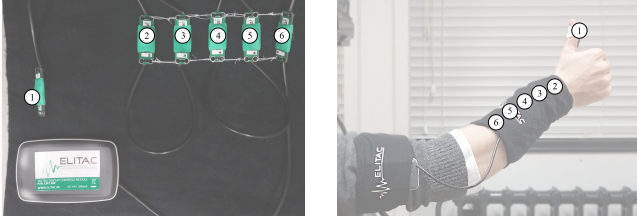
The Meta DK1 was used as HMD for this study. These glasses provide stereoscopic vision, have a resolution of  $480 \times 540$  per eye, a  $35^\circ$  FOV (using the expander lens), a SoftKinetic DepthSense time-of-flight depth camera with  $320 \times 240$  depth resolution (10cm to 2m), and  $360^\circ$  head tracking. An Elitac tactile display was used to provide vibrotactile feedback to the users. This display has 16 intensity levels on a logarithmic vibration power scale (where for PWM in % and intensity level  $I$ :  $PWM = 20 * 10^{\frac{I * \log_{10} 5}{15}}$ ) corresponding to a linearly perceived intensity scale with fundamental frequency  $158 \pm 2.4$  Hz at maximum vibration strength. The root mean square acceleration at maximum vibration strength is  $55.5 \pm 9.5$  m/s<sup>2</sup>. Five vibration units were placed on the skin of the forearm using elastic bands with Velcro, and one unit was taped to the tip of the right thumb, see Figures 1 and 3. Philips SHD8600 wireless headphones were used to provide masking pink noise and feedback beeps during the experiment. The experiment environment was created using Unity 5.3.2 and Meta SDK 1.3.3.

During the experiment, the participant was asked to locate a virtual object, a hollow cube, by placing the tip of their right thumb in the center of this cube (see Figure 2), after which the participant would press spacebar to log their answer. The cube was presented visually and/or tactually. In the cases including visual feedback, the virtual cube was not corrected for occlusion by the thumb. In the cases including tactile feedback, the Euclidean distance between the tip of the thumb and the center of the cube was mapped to a vibrotactile sensation.

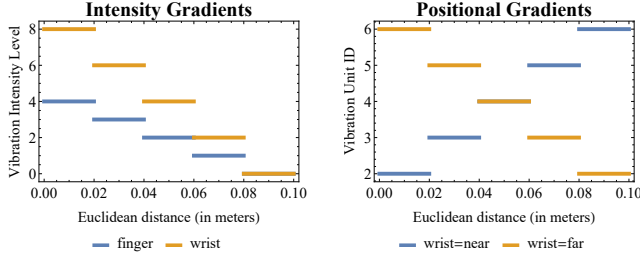
### 3.3 Experimental Design

#### 3.3.1 Participants

27 participants took part in the experiment. Ages of the included participants ranged between 21 and 26, with a average age of 23.1; 24 were male, 3 female; 22 right-handed, 3 left-handed and 2 mixed-handed; 4 had previous experience



**Figure 3: Tactile display setup with 6 vibration units: #1 for the thumb, #2-5 for the forearm.**



**Figure 4: The tactile mapping functions: (L) intensity gradient to vibration units #1, 2; (R) positional gradient, always with vibration intensity level 4.**

with AR apps and demos, 0 had experience with tactile displays, all but 2 had normal or corrected vision (these 2 had only very slightly impaired vision), and 0 had an affliction to their touch sense. Because of the basic nature of the task we do not expect these factors to be of influence, except possibly age [20] (see Section 4). Each participant signed a consent form prior to the experiment, and did not receive any form of monetary compensation.

### 3.3.2 Design

There were three modality conditions: visual-only (*V*), visuotactile (*VT*) and tactile-only (*T*). *VT* and *T* in turn had two types of tactile feedback: intensity gradient (*I*; the Euclidean distance was mapped to an intensity) and positional gradient (*P*; the Euclidean distance was mapped to a position on the arm); see Section 3.3.3 for the implementation. This tactile feedback type was tested between subjects, and for each tactile feedback type there were 2 subconditions: for *I* the feedback was provided to the *thumb* (intensities 0, 1, 2, 3, 4) or the *wrist* (intensities 0, 2, 4, 5, 8), for *P* the feedback was mapped using *wrist=near* or *wrist=far* (always intensity 4). This means each participant took part in 5 conditions. 3 depths were tested: 0.32m, 0.36m, and 0.40m. Within each condition, each depth was repeated 11 times, giving a total of 165 test trials per participant. The order of conditions was evenly spread over participants, where modality gave the first level of ordering, and tactile position the second.

### 3.3.3 Tactile Mapping

When designing the tactile mapping functions, a few aspects needed to be taken into account. Firstly, the continuous distance had to be mapped to a discrete intensity or position. Continuous intensity and position change were not attempted since correct implementation of apparent motion and phantom sensations on the skin are in themselves very difficult; see [10] for an example. Secondly, the number of

used intensity levels had to be equal to the number of used positions. Thirdly, the sensitivity of the skin on the thumb is much higher than that of the skin on the forearm, so the applied intensity on the finger had to be relatively lower than on the wrist. Lastly, constant intense vibrations can be irritating after long use, so for *P* a lower vibration intensity was used than for *I-wrist*. See Figure 4 for the used mapping functions.

## 3.4 Procedure

A participant was seated on a chair at a table facing a blank wall. Before using the HMD, participants were tested for stereoscopic vision using the TNO test for stereoscopic vision (8th edition, 1972), and for vibration detection using intensities 15, 10 and 5. If the participants passed these tests, we continued with the experiment. Interpupilar distance was measured and the lenses of the HMD were adjusted accordingly. After putting on the headset, calibration was performed using the built-in calibration application. This calibration was repeated every time the participant took off the headset during the experiment. To verify that calibration was successful, the hand point cloud was shown to the participants, and they were explicitly asked if the depth of the point cloud matched the depth of their own hand. If this was not the case, the HMD was recalibrated, otherwise the experiment started.

A trial had the following steps. A beep would occur and a cube would appear in front of the participant at eye-level at a certain depth with random edge size between 2cm and 4cm (the participant was told beforehand that the size was random). The user, keeping his right hand in a thumbs-up gesture, would stretch out his arm in order to position the tip of his thumb in the center of the cube; see Figure 1. Once the participant was satisfied with his/her thumb placement, they would press the spacebar with their left hand to log their answer. Once the spacebar was pressed, a different beep would occur, the cube would disappear and the trial was completed. If the log was unsuccessful for some reason, an auditory buzz would occur and the cube stayed in place. In this case the participant would simply try it again and continue. The next trial started 1 second after completion of the previous trial. For each trial, the trial ID, cube position and cube size were logged, followed by the timestamp and position of the tip of the thumb for each frame in which the right hand was visible. The participant was allowed to take short breaks during a trial to aid/avoid arm fatigue by putting their hand down and out of view. Auditory beeps/buzzes were used rather than textual information since the latter could bias the depth estimation.

Each condition was preceded by 5 training trials, during which the participant was able to experience the tactile (if present) and auditory beeps/buzzes. The participant was allowed to take off the headphones and HMD in between conditions, during which the experimenter verbally asked a few questions regarding the participant's experience, see Table 1. After a short break, the experiment continued with the next condition.

## 3.5 Results

To analyze the data, the logged thumb-tip positions were first checked for continuous motion: in the case of a fatigue-break, only the data after the last break during a trial was used. From the remaining data, the final depth estima-

**Table 1: Questions asked after completion of a condition, answered using a 7-point Likert scale.**

Mod.	Question
V	1. How <i>confident</i> were you at placing your finger at the target object location? (not at all - very)
VT	1. How <i>confident</i> were you at placing your finger at the target object location? (not at all - very) 2. How <i>easy</i> was it to use the tactile feedback? (not at all - very) 3. How did the <i>visual</i> feedback influence you at placing your finger? (negatively - positively) 4. How did the <i>tactile</i> feedback influence you at placing your finger? (negatively - positively)
T	1. How <i>confident</i> were you at placing your finger at the target object location? (not at all - very) 2. How <i>easy</i> was it to use the tactile feedback? (not at all - very)

tion bias (mean and standard deviation) and task duration (mean) were calculated for each participant and further analyzed. The data was then grouped by tactile feedback type, modality & tactile position, cube depth, and cube size. The latter was accomplished by splitting the cube sizes into two groups: *S* ( $2\text{cm} \leq \text{cube edge} < 3\text{cm}$ ) and *L* ( $3\text{cm} \leq \text{cube edge} < 4\text{cm}$ ). This factor was added to make sure participants indeed only relied on vergence cues to determine depth, and not cube size. Only two size groups were used in order to assure that no category was empty.

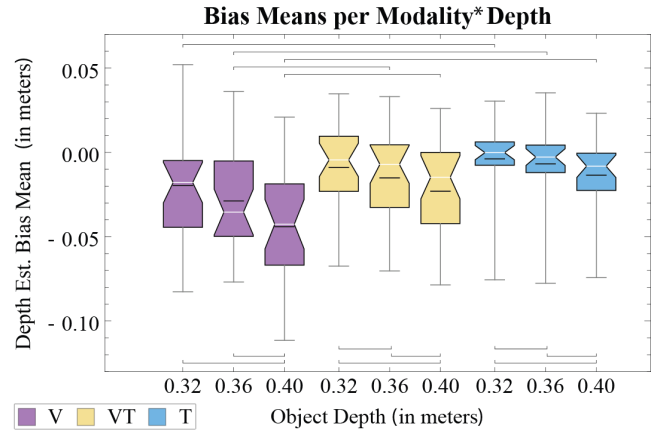
### 3.5.1 Mean of Depth Estimation Bias

The data was first checked for outliers. There were two *I* participants that caused outliers in modality *T*, but because they were particularly consistent they were preserved in the data. 48 paired samples t-tests were run for both feedback type groups individually to check for differences in bias over tactile positions. This showed no significant differences, so the data was combined over tactile positions for each participant. Next, 18 independent samples t-tests were run between both tactile feedback type groups. This also gave no significant differences, so the data for both groups was combined.

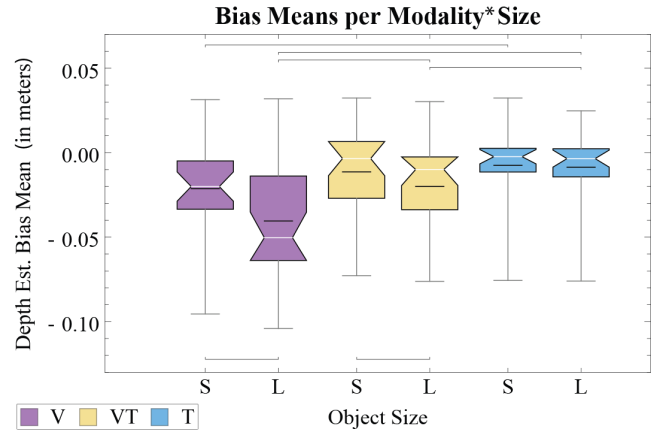
The data was further analyzed using a three-way ANOVA with within-factors modality (3 levels), depth (3 levels) and size (2 levels). This showed that all main effects were significant, and also interaction effects modality×depth and modality×size, see Table 2. To understand the significant interactions, the simple main effects are further investigated, see Table 3 and Figures 5, 6.

### 3.5.2 Standard Deviation of Depth Estimation Bias

The standard deviation values were checked for outliers. This gave a few sporadic cases in *VT* that were not consistent over any participants, so all data was considered for further analysis. The 48 paired samples t-tests over tactile positions showed no significant differences, so again the data was combined for each participant. The 18 independent samples t-tests between tactile feedback type groups gave 6 significant differences: all standard deviations in *T* were significantly smaller for *I* than for *P* (all  $p < 0.05$ ). This suggested a difference in behavior between the groups, indicating a between-subjects effect for *T*, so the data was



**Figure 5: Boxplots of the depth estimation biases in the modality×depth interaction, with median notches and mean markers. Purple represents *V*, yellow *VT*, and blue *T*. Horizontal brackets indicate a significant difference at  $p < 0.05$ .**



**Figure 6: Boxplots of the depth estimation biases in the modality×size interaction, with median notches and mean markers. Purple represents *V*, yellow *VT*, and blue *T*. Horizontal brackets indicate a significant difference at  $p < 0.05$ .**

first analyzed for the two groups separately.

For *I*, the three-way ANOVA showed only a significant main effect over modality, see Tables 2, 4 and Figure 7. Pairwise comparisons showed that *V* differed significantly from *VT* ( $p = 0.021$ ), and *VT* from *T* ( $p = 0.011$ ). For *P*, the three-way ANOVA also showed only significant main effect over modality. Pairwise comparisons showed that *V* differed significantly from *T* ( $p = 0.016$ ), and *VT* from *T* ( $p = 0.001$ ). A three-way ANOVA with within-factors depth and size and between-factor tactile feedback type on the *T* data confirms the between subjects effect, see Table 5.

### 3.5.3 Mean of Task Duration

Checking for outliers showed only two consistent cases, so all data was considered for further analysis. As before, 48 paired t-tests were run to compare task duration over tactile positions. This showed 3 significant differences in the

**Table 2: Results of three-way ANOVA with within-factors modality (3 levels), depth (3 levels), and size (2 levels), for mean and standard deviation of depth estimation bias and task duration mean.** <sup>G</sup> indicates a Greenhouse-Geisser correction and <sup>H</sup> a Huynh-Feldt correction.

Effect	Depth Mean	Depth St. Dev. - I	Depth St.Dev - P	Duration Mean
mod	$F(2,52)=8.918; p=0.0005$	$F(2,26)=4.133; p=0.028$	$F(1.4,16.7)=11.180; p=0.002^G$	$F(2,52)=20.862; p<0.000001$
depth	$F(1.4,35.3)=46.751; p<0.000001^G$	$F(2,26)=0.068; p=0.068$	$F(2,24)=0.365; p=0.698$	$F(2,52)=0.048; p=0.953$
size	$F(1,26)=31.873; p<0.00001$	$F(1,13)=0.053; p=0.821$	$F(1,12)=1.384; p=0.262$	$F(1,25)=1.574; p=0.221$
m×d	$F(3.3,86.7)=5.520; p=0.001^H$	$F(4,52)=1.193; p=0.325$	$F(4,48)=1.847; p=0.135$	$F(2.9,74.7)=0.738; p=0.543^G$
m×s	$F(1.6,42.1)=10.954; p=0.0004^H$	$F(1.2,15.0)=0.006; p=0.958^G$	$F(2,24)=1.142; p=0.336$	$F(2,52)=7.879; p=0.001$
d×s	$F(2,52)=1.127; p=0.332$	$F(2,26)=0.589; p=0.562$	$F(1.4,16.9)=0.782; p=0.431^G$	$F(2,52)=2.204; p=0.121$
m×d×s	$F(3.4,88.9)=0.294; p=0.954^H$	$F(2.1,27.0)=1.311; p=0.287^G$	$F(2.0,24.3)=0.336; p=0.721^G$	$F(2.7,70.7)=0.781; p=0.497^G$

**Table 3: Grand means of the depth estimation bias (in meters) in the modality×depth and modality×size interactions. Negative values indicate a bias towards the user.**

	V	VT	T
<b>0.32</b>	-0.020	-0.009	-0.004
<b>0.36</b>	-0.029	-0.015	-0.007
<b>0.40</b>	-0.044	-0.023	-0.014
<b>S</b>	-0.021	-0.011	-0.007
<b>L</b>	-0.040	-0.020	-0.009

**Table 4: Grand means of the depth estimation bias standard deviations for each modality and tactile feedback type (in meters).**

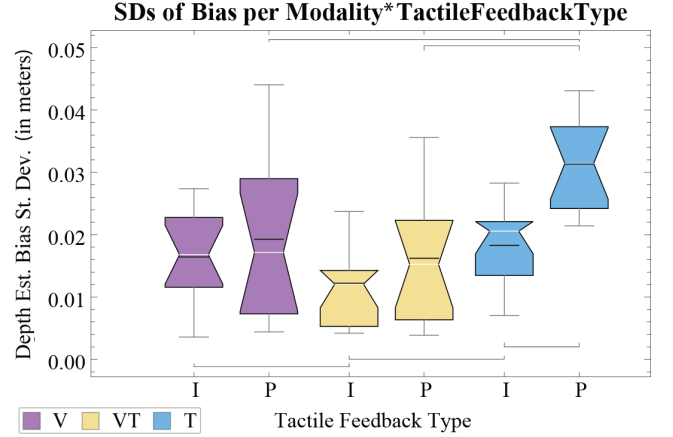
	V	VT	T
<b>I</b>	0.016	0.012	0.018
<b>P</b>	0.019	0.016	0.031

*I* group:  $I\text{-}VT\text{-}finger\text{-}32\text{-}S < I\text{-}VT\text{-}wrist\text{-}32\text{-}S$ ,  $I\text{-}VT\text{-}finger\text{-}32\text{-}L < I\text{-}VT\text{-}wrist\text{-}32\text{-}L$ , and  $I\text{-}T\text{-}finger\text{-}32\text{-}S > I\text{-}T\text{-}wrist\text{-}32\text{-}S$  (all  $0.01 < p < 0.05$ ). It was unlikely that the very small amount of differences would affect further analysis, so the data was combined for each participant. The 18 independent samples t-test between tactile feedback type gave two significant differences:  $I\text{-}T\text{-}32\text{-}L < P\text{-}T\text{-}32\text{-}L$  and  $I\text{-}T\text{-}36\text{-}L < P\text{-}T\text{-}36\text{-}L$  (both  $0.01 < p < 0.05$ ). As before, we continued by combining the data of both groups.

The three-way ANOVA showed a significant main effect over modality and a significant interaction effect over modality×size, see Table 2. To investigate the interaction, the simple main effects were analyzed, see Table 6 and Figure 8.

**Table 5: Results of three-way ANOVA with within-factors depth (3 levels) and size (2 levels) and between-factor feedbacktype (2 levels), for standard deviation of depth estimation bias on *T* data.**

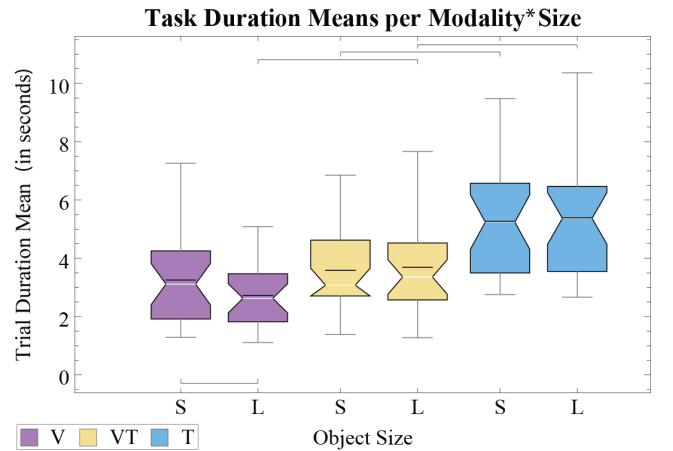
Effect	Depth St.Dev.
depth	$F(2,50)=0.036; p=0.964$
size	$F(1,25)=0.072; p=0.790$
feedbacktype	$F(1,25)=23.790; p=0.00005$
d×f	$F(2,50)=0.623; p=0.540$
s×f	$F(1,25)=0.017; p=0.898$
d×s	$F(2,50)=0.131; p=0.877$
d×s×f	$F(2,50)=1.055; p=0.356$



**Figure 7: Boxplots of the depth estimation bias standard deviations over modality and tactile feedback type, with median notches and mean markers. Purple represents *V*, yellow *VT*, and blue *T*. Horizontal brackets indicate a significant difference at  $p < 0.05$ .**

**Table 6: Grand means of the task duration in the modality×size interaction (in seconds).**

	V	VT	T
<b>S</b>	3.255	3.583	5.265
<b>L</b>	2.727	3.692	5.388



**Figure 8: Boxplots of the average task duration, with median notches and mean markers. Purple represents *V*, yellow *VT*, and blue *T*. Horizontal brackets indicate a significant difference at  $p < 0.05$ .**



**Table 7: Questionnaire results (7-point Likert) for groups *I* and *P* combined.**

Mod.	Question	Median	Mean	St.Dev.
V	1 - confidence	4	3.96	0.280
VT	1 - confidence	6	5.91	0.177
	2 - ease	6	5.61	0.216
	3 - visual	6	5.43	0.206
	4 - tactile	6	5.76	0.154
T	1 - confidence	4.5	4.50	0.243
	2 - ease	5	5.11	0.202

### 3.5.4 Questionnaire

Analogous to before, 12 Wilcoxon paired signed rank tests were used to check for differences between tactile position for both groups. There were no significant differences in either group, so the data was combined for each participant over tactile position. 7 Mann-Whitney tests were used to check for differences between groups. This showed only one significant difference between *I-T-ease* and *P-T-ease*, where medians were 6 and 5 respectively ( $U = 49.5, n_I = 14, n_P = 13, p = 0.041$ ). See Table 7 and Figure 9 for the data combined over groups.

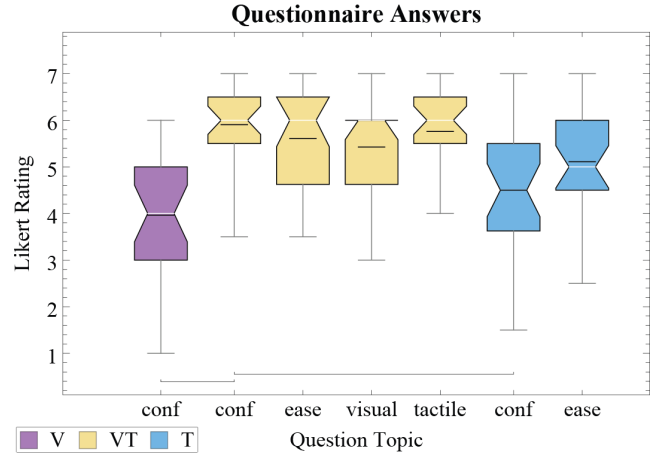
To compare confidence (Q1 for all modalities), a Friedman test was run over factor modality over the combined data. This showed a significant effect ( $X^2(2) = 29.152; p < 0.000001$ ), and according to Wilcoxon Signed Rank tests, *V-conf* vs. *VT-conf* and *VT-conf* vs. *T-conf* were significant ( $Z = -4.307; p = 0.00002$  and  $Z = -3.785; p = 0.0002$ , respectively). Next, for ease we observed a significant difference between *I* and *P* in *T* (see above), but this effect did not occur for *VT* ( $U = 64.0, n_I = 14, n_P = 13, p = 0.180$ ). Ease was also compared over modality for both groups separately using Wilcoxon Signed Rank tests. There were no significant differences in *I* ( $Z = -1.029; p = 0.303$ ) or in *P* ( $Z = -1.545; p = 0.122$ ).

Lastly, a Wilcoxon Signed Rank test was used to compare the influence of the visual and tactile feedback (Q3 and Q4 for *VT*, respectively) for both groups combined, and showed no significant difference ( $Z = -1.211; p = 0.226$ ).

## 4. DISCUSSION

The aim of this study was to determine the feasibility of using the tactile sense as an information channel for object information that is not usually perceived through haptics. The presented research question was *Can tactile feedback correct the depth estimation bias in an active reaching task in AR?* We expected that participants would not only be able to translate tactile information to meaningful depth information, but also that it would correct the bias caused by the (conflicting and/or unreliable) visuals. We showed that this was indeed the case, and that the rate at which this works does not depend on type of tactile mapping, nor does it depend on placement on the arm.

As in previous studies [18, 4], a bias occurred towards the user in the visual-only case. Note that for [4] the bias only occurred in the matching task and not in the active task; one must speculate whether the experimental setup was appropriate for the intended measurements. It has been shown that the bias is caused by, amongst other factors, the separation of accommodation and vergence cues [19, 9]. It must



**Figure 9: Boxplots of the questionnaire results, with median notches and mean markers. Purple represents *V*, yellow *VT*, and blue *T*. Horizontal brackets indicate a significant difference at  $p < 0.05$ .**

be noted, however, that all participants were young adults, and this conflict may be less of a problem for older adults [20]. In particular, tolerance of stereo depth perception to vergence-accommodation conflict was related to accommodation ability, an ability that declines with age. However, we saw that for the multimodal condition, not only did the bias decrease, but also that the variability was low, a factor that does not necessarily depend on accommodation ability. This suggests that visuotactile depth estimation can be valuable for all ages.

We also observed that, opposite to previous research, the bias actually increased when virtual object depth increased. A possible explanation is that there was a conflict due to the absence of occlusion of virtual objects by real objects. It has been shown that when a virtual object is positioned in front of an occluder (with constant depth), the bias is consistently towards the user, however there is a point where the participants are sure the virtual object is behind the occluder and the object ‘falls back’ [6, 15]. Our study has an even more complex setup, because the occluder (the thumb) is constantly moving in space. Once it is in the vicinity of the virtual object, depth ranking can take place, and participants may tend to move their finger forward, to place it in front of the initial placement. Our results then suggest that these ‘final adjustments’ occur to a greater extent when the virtual object is farther away.

However, the bias towards the user also occurred in the visuotactile and tactile conditions. This is especially surprising for the latter case, because there is no longer a visual conflict present. In the following, a few aspects are discussed that may have influenced tactile depth estimation. Firstly, the order of the conditions could have played a role: if participants experienced a condition with visual feedback before the tactile-only condition, they could have been biased towards the depth they perceived in those visually inclusive conditions, i.e. an overall learning effect could have occurred. This was checked for the results of the tactile-only condition (combined over size), where participants were split into two groups: *T-first*, *T-not-first*. However, a one-way ANOVA with within-factor depth and between-factor

$T$ -first did not show a significant between-subjects effect ( $F(1, 25) = 0.016; p = 0.901$ ). Another explanation is that the calibration was accurate for changes in hand depth, but not absolute hand depth. In an attempt to prevent this, participants were asked to compare the hand point cloud and real hand for depth differences, however there may have already been a bias present. Again, this does not explain the tactile-only results, since the tactile feedback is generated by the depth obtained through the depth sensor; if this measurement was always incorrect, the tactile feedback would be given at this distance and therefore counted as correct. A final possible explanation then is that the effort required to stretch out the arm was of influence. Firstly, this effort experienced in the arm muscles may cause the vibrotactile feedback to be interpreted differently. Such an effect was demonstrated in [17], where high muscle pain induced by a saline solution (5%) reduced the cutaneous mechanosensitivity. Secondly, although the virtual object was always straight ahead at eye-level, the participants may have had trouble staying on this line in space. This would mean that if they did not go exactly through the center of the virtual object, they would never reach the vibration limit (highest intensity or extreme position), so they would feel less changes in tactile feedback, possibly causing more error. It is more likely that the participants then settle for a distance that is closer to them than farther away.

Curiously, our results showed that participants (subconsciously) used virtual object size, even though instructed not to rely on it since it was random, causing a natural size bias. The size-bias was absent in the tactile-only condition since no indication of size was available, and for the larger cubes the size-bias was smaller in the multimodal condition, indicating that the integrated tactile feedback partially corrected this bias as well. According to [16], users rely more on retinal image size when the change in virtual object size is small than when the change is large, which was indeed the case in our experiment. This suggests that this interaction may not have occurred had the cube sizes been varied more. A remark must be made concerning previous AR depth estimation studies. The concern of using size as a depth cue was mentioned in [15], but not further analyzed. Also, the replication and extension of this study in [18] does not mention size randomization at all. Our results together with those from [16] give indication that this is in fact a necessary analysis, thus we recommend future studies to take this ‘instinctive’ use of size into account.

The decrease in both biases when moving from visual to visuotactile to tactile suggests that tactile feedback can correct various types of biases caused by visuals. Unfortunately, the bias correction coincides with an increase in variability of estimation. This increase was more severe for the positional gradient mapping than for intensity gradient mapping, which may have been caused by the incongruence in movement direction between the forearm and the tactile feedback on the arm itself. Based on this, and the fact that ease was higher for the intensity gradient mapping in the tactile-only condition, one could recommend intensity gradient over positional for future applications. However, since there was no difference in bias between these mappings, we do not exclude positional gradient as a feasible type of tactile feedback. Although tactile-only led to the highest variability, variability for the visual-tactile condition was lowest overall. This was expected through previous research in the

field of multimodal interaction [8]. For example, according to the theory of Maximum Likelihood Estimation in multimodal integration, the purpose of multimodal integration is the increase in reliability of perception, where reliability is defined as the inverse variance [7]. This is supported by the questionnaire results regarding confidence: confidence of finger placement was higher for the multimodal condition than for the unimodal conditions. An interesting note is that participants experienced the visual feedback as equally of influence as the tactile feedback. Many participants found this surprising, with remarks such as “I did not expect the tactile feedback to be so important.” We emphasize furthermore that the relatively small bias with relatively larger, but certainly acceptable, variability in the tactile-only case shows that the answers are not random, which reflects the intuitiveness and usefulness of the tactile mappings.

Concerning task duration, there was an unexpected interaction effect between modality and size. However, closer investigation of the simple main effects showed that there was only a significant difference in task duration between  $V$ - $S$  and  $V$ - $L$  and no other cases, so the difference was likely due to chance. All other significant differences were regarding different modalities, where overall tactile-only trials took longer than the visuotactile trials, and visuo-tactile trials longer than visual-only trials. This complements the above discussion on variability. For visual-only, the participant already has an idea what the position is and receives no feedback concerning whether it must be corrected, so a decision is made fast with little to no corrections. In the tactile-only condition, the user had to go over a straight line first to determine where the object approximately was, and then determine it more precisely. It is likely that a combined strategy was used in the visuotactile case: an approximate position was determined immediately through visuals, and then corrected using tactile feedback. This increase in duration in the multimodal condition is in contrast to what was shown in [5], and is likely due to the implementation: they applied tactile feedback at a static point in space, where in our setup the tactile feedback was a gradient over a larger space around the object. This requires more time, but aids the entire reaching process rather than a single point. The integration of tactile feedback in an AR application presents a trade-off situation between accuracy and duration, and thus its value will depend on the purpose of the application. An example application where precision would outweigh duration is (remote) AR surgery, a research field that has been struggling with such precision problems for years [14].

Lastly, there were no large differences between tactile positions. This suggests that both intensity and positional gradient can be used on different body parts. This is supported by previous research using haptics for vision substitution on, amongst other locations, the tongue [2] and forehead [11]. This is interesting since in a real version of the task used in our experiment, the thumb would touch the object and only receive tactile feedback there. However, participants performed equally well on the forearm, which demonstrates the simplicity of possible tactile setups for (near) future applications, for example using a smartwatch.

A possible limitation of this study is the use of the built-in calibration rather than a self-implemented calibration. The accuracy of the measured results are directly related to the decisions made in this calibration, which due its design cannot be adjusted or changed at real-time. As stated earlier,

the precision of this calibration was tested by participants themselves, however it may be that a certain bias already played a role at this stage.

## 5. CONCLUSION

This work was motivated by the known advantages of multimodal feedback, and consequently we investigated the feasibility of tactile feedback to aid object depth estimation in AR. We have shown that the bias that occurs in a visual-only setting is partially corrected when integrating tactile feedback, but also in a tactile-only setting, although a trade-off must then be made between time and consistent precision. The tactile feedback worked in two different mapping implementations, one mapping distance to vibration intensity, and the other to vibration position on the skin. It was also shown that virtual object size plays a larger role in depth estimation than expected, and must be taken into account in future research. Tactile feedback has not yet thoroughly been studied in multimodal AR settings, but can be beneficial in scenarios that require accurate depth estimation such as AR surgery. Other applications with purposes other than performance can benefit as well. For example, games may have the goal to offer imaginary scenarios that do not mimic the real world and can then thrive on non-natural feedback. We expect our results to drive future research in tactile depth estimation and other applications of the tactile information channel. Firstly, this study examined tactile feedback that was centered around the virtual object, but it is unclear whether this feedback is equally intuitive and interpretable in an egocentric world setting, i.e. when the feedback is centered around the real body. Secondly, we showed that multiple tactile mappings and corresponding positions were feasible for depth estimation. The limits of this malleability are unknown and must be further explored. We expect beneficial use cases to not only use tactile feedback for sensory redundancy, but also sensory substitution in its most abstract form, e.g. receiving meta-information.

## 6. REFERENCES

- [1] B. M. Altenhoff, P. E. Napieralski, L. O. Long, J. W. Bertrand, C. C. Pagano, S. V. Babu, and T. A. Davis. Effects of calibration to visual and haptic feedback on near-field depth perception in an immersive virtual environment. In *Proceedings of the ACM Symposium on applied perception*, pages 71–78. ACM, 2012.
- [2] P. Bach-y Rita, K. A. Kaczmarek, M. E. Tyler, and J. Garcia-Lara. Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *Journal of rehabilitation research and development*, 35(4):427, 1998.
- [3] P. Bach-y Rita and S. W. Kercel. Sensory substitution and the human-machine interface. *Trends in cognitive sciences*, 7(12):541–546, 2003.
- [4] S. Baldassi. Depth perception and action in wearable augmented reality: A pilot study. In *2015 IEEE International Symposium on Mixed and Augmented Reality Workshops (ISMARW)*, pages 12–14. IEEE, 2015.
- [5] L. Bouguila, M. Ishii, and M. Sato. What impact does the haptic-stereo integration have on depth perception in stereographic virtual environment? a preliminary study. In *Haptic Human-Computer Interaction*, pages 135–150. Springer, 2001.
- [6] S. R. Ellis and B. M. Menges. Localization of virtual objects in the near visual field. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(3):415–431, 1998.
- [7] M. O. Ernst and M. S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.
- [8] M. O. Ernst and H. H. Bühlhoff. Merging the senses into a robust percept. *Trends in cognitive sciences*, 8(4):162–169, 2004.
- [9] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):33, 2008.
- [10] A. Israr and I. Poupyrev. Tactile brush: drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2019–2028. ACM, 2011.
- [11] H. Kajimoto, Y. Kanno, and S. Tachi. Forehead electro-tactile display for vision substitution. In *Proc. EuroHaptics*, 2006.
- [12] J. P. Rolland, W. Gibson, and D. Ariely. Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators & Virtual Environments*, 4(1):24–49, 1995.
- [13] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence: Teleoperators and Virtual Environments*, 11(6):610–625, 2002.
- [14] J. H. Shuhaiber. Augmented reality in surgery. *Archives of surgery*, 139(2):170–174, 2004.
- [15] G. Singh, J. E. Swan II, J. A. Jones, and S. R. Ellis. Depth judgment measures and occluding surfaces in near-field augmented reality. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 149–156. ACM, 2010.
- [16] R. Sousa, E. Brenner, and J. B. Smeets. Judging an unfamiliar object’s distance from its retinal image size. *Journal of Vision*, 11(9):10–10, 2011.
- [17] C. S. Stohler, C. J. Kowalski, and J. P. Lund. Muscle pain inhibits cutaneous touch perception. *Pain*, 92(3):327–333, 2001.
- [18] J. E. Swan, G. Singh, and S. R. Ellis. Matching and reaching depth judgments with real and augmented reality targets. *Visualization and Computer Graphics, IEEE Transactions on*, 21(11):1289–1298, 2015.
- [19] J. P. Wann, S. Rushton, and M. Mon-Williams. Natural problems for stereoscopic depth perception in virtual environments. *Vision research*, 35(19):2731–2736, 1995.
- [20] S. Watt and L. Ryan. Age-related changes in accommodation predict perceptual tolerance to vergence-accommodation conflicts in stereo displays. *Journal of vision*, 15(12):267–267, 2015.