Influence of visuomotor action on visual-haptic simultaneous perception: A psychophysical study

Zhuanghua Shi *   Sandra Hirche †   Werner X. Schneider ‡   Hermann Müller §

*‡§ Allgemeine und Experimentelle Psychologie, Ludwig-Maximilians-Universität München, D-80802, München, Germany
† Institute of Automatic Control Engineering, Technische Universität München, D-80290, München, Germany

ABSTRACT

Perceiving temporal simultaneity of cross-modal (e.g. visuo-haptic) events is very important for coherent perception and high performance in multi-modal interaction (e.g. telepresence) systems. The perception of simultaneity, however, depends on many factors, including stimulus intensity, redundant information and selective attention. In this study, we examined the effects of two factors on perceived simultaneity in a visuo-haptic task scenario: active (self-generated) motor control and additional visual motion feedback. We used a psychophysical method to measure the differences in points of subjective simultaneity (PSSs) and just noticeable differences (JNDs) in visuo-haptic events. The results showed that the PSS decreased significantly under conditions of active motor control with concurrent visual feedback, and the JNDs were narrowed with either active motor control or additional visual motion feedback.

Keywords: Temporal integration, Simultaneity, PSS, JND

Index Terms: H.1.2 [User/Machine Systems]: tele-presence—cross-modal temporal integration

1 INTRODUCTION

Multi-modal human system interfaces (HSIs) represent one of the main components of telepresence and networked virtual reality systems. In this context multi-modal refers to sensory modalities, e.g. visuo-haptic. By means of the HSI, human operator commands a remote robot (avatar in virtual environments). In order to make the human operator experience the situation as if he/she were directly present and acting in the remote environment, multiple sources of sensory information must be presented so as to provide him/her with the perception of coherence. One key feature of coherent perception is cross-modal temporal simultaneity [8,13,26]. Simultaneous feedback is particularly important in daily life, since it provides one of the most salient cues to bind information from different sensory modalities together into a single perceptual object or event. Asynchronous feedbacks from the same event can give rise to perceptual misbindings and illusions [28].

In bilateral telepresence systems multi-modal signals are exchanged between the HSI and the remote robot (accordingly, between the HSI and the avatar in virtual environments) through a communication network, e.g. the Internet [9,18,20,22]. The signals processing and transmission times are modality-dependent. This is due to the different requirements of communication bandwidth, as well as compression algorithms, and communication protocols for the different modalities. As a result, the multi-modal information is actually asynchronously presented to the human operator, thus inevitably challenging the perceptual temporal integration. To what degree temporal asynchrony can be considered to be acceptable is of great interest to the designers of HSI, as well as of multi-modal communication protocols and multi-modal augmentation measures.

It is well-known that humans can tolerate a certain degree of temporal asynchrony between two signals from different modalities. This range of tolerance is referred to as temporal window of integration (TWI), within which cross-modal events can be easily bound together. One important measure of the TWI is provided by the just noticeable difference (JND) indicating the thresholds between synchrony and asynchrony. Past investigations of cross-modal temporal integration have mainly focused on the audio-visual modalities [3, 13, 14]. Recently, several studies on the visuo-tactile modalities have reported the JND to be asymmetrically distributed: on average, the visual stimuli must occur before the tactile stimuli for them to be perceived as simultaneous; Furthermore, the JND can be varied by selective attention, stimulus intensity, spatial location, and adaptation [16, 24–26].

However, note that most of the aforementioned studies were concerned with very simple stimulated events, for example, temporal discrimination between a very short time of tactile vibration and a briefly illuminated LED. To our knowledge, in almost all previous studies [1, 19, 23–27, 37, 38], participants were passively stimulated with cross-modal stimuli and then asked to make a temporal order judgement or a simultaneity judgement. By contrast, in bilateral telepresence and virtual reality systems, the human operator actively manipulates the HSI and therewith the remote robot or virtual object. Furthermore, the multi-modal sensor information is permanently fed back to the human operator, which contains not only the collision but also the motion information. These factors may influence the perception of cross-modal temporal integration.

On this background, the primary aim of the present study was to investigate the influence of two common factors on cross-modal temporal integration: first, the human operator's active manipulation; and, second, additional visual feedback of motion information. Both factors are very common in multi-modal interactive systems, yet little is known to date on how they influence the perceptual temporal integration.

2 BACKGROUND

The perception of temporal simultaneity is an old, yet 'current' issue in cognitive science. It is fundamentally important for object and event recognition, because synchrony is one of the most salient cues to form a coherent perception of an object or event, whereas asynchrony helps to separate the object from background [17].

Typical results of visual-tactile cross-modal studies are that the visual stimuli have to be presented prior to the tactile stimuli, by 20 to 50 ms, in order to reach the point of subjective simultaneity (PSS), and synchrony-asynchrony thresholds are asymmetrically distributed [5, 10, 23]. However, the perception of temporal simultaneity has also been revealed to depend on many aspects, ranging from low-level neural activity to high-level cognition. For example, Dixon and Spitz found a difference in thresholds between speech

*e-mail:strongway@psy.uni-muenchen.de

†e-mail:hirche@lsr.ei.tum.de

‡e-mail: wxs@mail.paed.uni-muenchen.de

§e-mail:mueller@mail.paed.uni-muenchen.de

and non-speech stimuli [3]. Fujisaki et al showed that training and adaptation can alter the point of subjective simultaneity across different modalities [6]. And Spence, Shore, and Klein demonstrated that attention can also shift the perception of simultaneity: When attention was directed to the tactile modality, visual stimuli had to lead the tactile stimuli by 155 ms in oder to be perceived as simultaneous with the latter; in contrast, when attention was directed to the visual modality, visual stimuli had to lead by only 22 ms [26] (for further detail, see review [23, 31]).

One factor that plays an important role in multi-modal telepresence is redundant information. For example, a collision event is accompanied by multiple sources of the information, such as force feedback, predictable motion trajectory and sound etc. There have been a few studies recently concerned with the role of spatial location as one of the redundant factors. When multi-sensory signals come from different locations, e.g. the left and the right, participants show better performance on temporal order judgments (TOJs) compared to signals from the same location [12, 24] - though the facilitation was manifest only when the separate signals were presented across the body midline [40]. Apart from studies on spatial location, to our knowledge, there are hardly any studies that have examined the effects of redundant information on cross-modal integration, especially the temporal simultaneity judgments, which are ubiquitous in multi-modal systems. In this paper we consider the visual motion information generated by the moving object as redundant information accompanying 'collision' events. Motion not only provides a cue to movement direction, but also renders future events, such as collisions, predictable. However, it is as yet unknown whether or not such additional visual motion signals affect the perception of simultaneity. One plausible prediction is that 'redundant' motion information facilitates the processing of visual events and the prediction of imminent events. As a result, they will increase the sensitivity for subjective simultaneity. Alternatively, additional motion signals may simply increase the mental work load and so act as distractors, i.e. they would not be benefical (but may rather be harmful) to the subjective simultaneity.

While little is known about the effects of redundant information, the role of active manipulation for perceptual temporal integration also remains unclear. Visuo-motor behavior and perception have been found to differ profoundly in their intrinsic processing capabilities. These differences are not only reflected in the existence of distinct neuronal circuits [7], but also related to the fact that the visuo-motor loop requires much faster updating of environmental information than merely visual perception. Several studies have shown that the subjective timing of sensory events changes when they are causally linked to subjects' action, as compared to when they are passively sensed [15, 34]. Recently Vogels compared the temporal accuracy between active touch and passive touch by using joystick and showed that the active touch degrades the perception of asychrony [31]. However, in her paradigm the touch space and visual space were separated. Participants required additional resource for mapping visuo-haptics information together. In the same visuo-motor coordinated space it has been demonstrated that additional movement-related information in the active movement often makes predictions more anticipatory [32]. Therefore, it is still unclear whether the accuracy of the simultaneity will be degraded or improved in that same visuo-motor coordinated space.

To examine the effects of visual motion information and self-generated action on perceptual temporal integration, the present study adopted a classic event-based temporal order judgment approach. The visual event was the breaching of (i.e. collision with) a vertical-line barrier by a moving object (visible object conditions), or simply the disappearance of the bar's middle section (invisible object conditions), and the haptic event was the impulse force feedback associated with the collision. The breaching event has a distinct onset and offset timing which could give the participant a clear

perception of the collision and the results can be comparable with previous passive studies. In addition, the availability (versus the non-availability) of visual motion feedback and of manual movement control were examined in a 2 x 2 full factorial design. For the hand-controlled movement, participants used their right-hand index finger to move an object, which was either visible or invisible to cause a collision with the barrier. This condition will hitherto be referred to as 'movement condition', in contrast to the 'no movement condition'. The points of subjective simultaneity (PSSs) and the just noticeable differences (JNDs) were estimated for each of the resulting 2 x 2 experimental condtions.

3 METHOD

3.1 General Psychophysical Methods

Two classical psychophysical methods are most often used to measure multi-sensory temporal processing: temporal order judgments (TOJ) and synchrony-asynchrony judgments (SAS) [4, 5, 10, 26, 29, 39]. Typically, participants are presented with two stimuli from different modalities at varying stimulus onset asynchronies (SOAs). With the TOJ method, participants are asked to judge which stimulus (or in which modality a stimulus) was presented first. The point of subjective simultaneity (PSS) is then estimated in terms of that SOA at which the two stimuli appear to be synchronous, i.e. the point at which the responses 'stimulus A first' and 'stimulus B first' are equally frequent. The 75% and 25% frequencies of 'stimulus A first' responses are referred to as upper and lower thresholds, respectively. Note that the upper and lower thresholds do not always have the same values. The just noticeable difference (JND) is usually defined by the difference between the upper and lower thresholds. Similar to the TOJ method, the synchrony-asynchrony (SAS) judgment method also aims to obtain PSS and JND estimates, but in a somewhat different way. In an SAS experiment, participants are asked to judge whether two stimuli were presented simultaneously or successively. The PSS is then estimated as the SOA at which the frequency of 'simultaneous' responses reaches its maximum, and the JND is estimated by the lower and upper synchrony-asynchrony thresholds at 50%. Some researchers have argued that the SAS method may not provide reliable estimates of participants' sensitivity to asynchrony, given that the criterion for synchrony judgments adopted by participants is likely to be affected by the particular range of SOAs presented [26, 39]. For this reason, the TOJ method was used in the present study.

The calculation of PSS and JND is more accurate if a psychometric curve is fitted to each participant's data first, and then the parameters are estimated from the curve. A logistic function is often used to estimate the psychometric curve [2]

$$P(x) = \frac{1}{1 + \exp(\frac{\alpha - x}{\beta})} \quad (1)$$

where α is a parameter indicating the location of the threshold at 50% and β is a parameter indicating the slope of psychometric function.

With the above logistic function the PSS is estimated by the following function

$$\hat{PSS} = \hat{\alpha}; \quad (2)$$

and the JND is estimated by:

$$\hat{JND} = (X_{P_{.75}} - X_{P_{.25}})/2 = \hat{\beta} \log 3, \quad (3)$$

where $\hat{\alpha}$, $\hat{\beta}$ are estimated parameters of the logistic function.

3.2 Subjects

Nine participants (2 females and 7 males, mean age of 21.9) took part in the experiment for payment. All had normal or corrected-to-normal vision and no history of somato-sensory disorders. All

were dominantly right-handed. And all were naive as to the purpose of the experiment except for two participants (Z.S. and H.Z.). Informed consent was obtained before the start of the experiment.

3.3 Apparatus

The tactile stimuli were generated and delivered by means of a haptic display (SenAble Technologies, PHANToM PREMIUM 1.5A). The haptic display was connected to a PC (Dell Precision 690 Xeon Dual-Core) that controlled the whole experimental procedure. The visual stimuli were generated by using OpenGL and displayed with an NVIDIA Quadro FX 3450 graphics card on a Philips 202P70 CRT monitor with a refresh rate of 100 Hz and a resolution of 1024x768 pixels. The computer program for controlling of the experimental procedure was developed with OpenHaptics (SensSable Technologies, Academic version 2.0) and Visual C++ 2005.

The experimental setup is illustrated in Figure 1. The participant looked down at a mirror, on which he/she saw a visual scene reflected from the above CRT monitor. Below the mirror, the participants's right-hand index finger was attached to the stylus of Phantom limb (SenAble Technologies, PHANToM PREMIUM 1.5A) device. The position of the index finger was adjusted, before the start of the experiment, such that the participant experienced it to be on the same plane as the visual 2D surface.

Figure 1: Experimental set-up. Visual stimuli were presented on a CRT monitor and mirrored to appear aligned with haptic stimuli. Haptic stimuli presentation was controlled by a PHANToM device, and the visual stimuli were generated by OpenGL with the GLUT toolkit. The participant's head position was fixed by the aid of a chin rest, ensuring a stable viewing distance of 57 cm to the mirror display surface.

3.4 Stimuli and Procedure

Participants were seated in front of the haptic display and placed their right-hand index finger into the stylus of PHANToM device, as illustrated in Figure 1. Participants were asked to rest their heads on the chinrest, so that the viewing distance was fixed at 57 cm.

Visual and haptic stimuli were presented in the experiment. The visual stimuli consisted of a white vertical bar (0.1 cm x 6.0 cm) and a white square box (0.3 cm x 0.3 cm), both 52.7 cd/m^2 in luminance. The tactile stimulus was set to an impulse force of 2N in a direction opposite to the moving direction of the white square; the force lasted 15ms. Auditory perception was 'masked' by white noise presented via earphones, in order to eliminate auditory artifacts produced by the onset of force feedback in PHANToM device.

There were four experimental conditions presented in four separate sessions.

Session 1: No movement with additional visual motion feedback. On each trial, a fixation cross was shown in the display center for 1 second. Next, a continuous vertical bar was presented in the center of the monitor. Then a square box moved from its starting position, 11.5 cm either to the left or to the right, through the

middle of the display to the opposite side, at a constant velocity of 11.5 cm/s. When the moving square collided with the vertical bar, the middle section of the bar disappeared from the screen, leaving behind two separated, upper and lower, sections (0.1 cm x 1.6 cm each) visible. The onset of the force feedback event was determined by the stimulus onset asynchronies (SOAs) relative to the visual collision. The time between the break-up of the bar and the haptic force feedback were set at one of seven SOA levels: -90, -60, -30, 0, 30, 60 and 90ms, respectively (for three participants who displayed relatively poor performance in the initial practice blocks, the SOAs range was set from -120 to 120 ms). Positive SOA values mean that the visual collision was presented first, while negative values stipulate that the haptic force feedback was presented first. The moving square continued to move and finally disappeared at its end position (11.5 cm away from the center). Then a probe text was presented on the monitor prompting the participant to respond:'vision first or touch first'. Participants were asked to make a non-speeded two-alternative forced-choice (2AFC) response by pressing either the UP or the DOWN arrow key.

Participants were instructed to keep their index finger aligned with the horizontal location of the vertical bar during the presentation of the visual display.
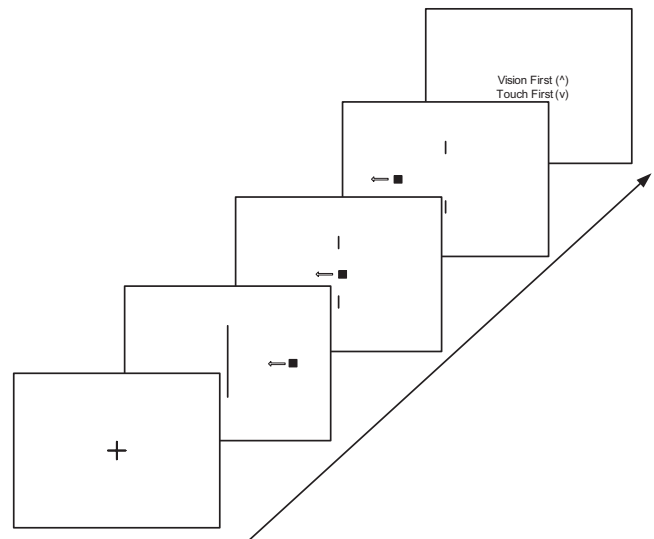

Figure 2: Schematic illustration of a visual stimulus display in session 1. The displays in sessions 2, 3, and 4 were similar. The haptic stimulus presentation is described in the text.

Session 2: No movement and no additional visual motion feedback. The stimuli and procedure were identical to session 1, except that the moving (square) object was invisible. In the absence of the 'redundant' motion information of the square, the 2AFC judgment was a pure comparison between two events: the breaking-up of the visual bar and the impulse force feedback. This condition can be considered as a baseline condition.

Session 3: Hand movement with additional visual motion feedback. The visual stimuli were identical to session 1. In addition, a red square was shown at the beginning of each trial to indicate the starting position of the movement. The horizontal position of moving square was determined by the horizontal position of the PHANToM stylus, which was actively controlled by the participant's index finger movement. The tactile stimuli were identical to session 1. The same seven levels of SOA as in session 1 were employed in the current session. However, due to the movement of the square being manually controlled by participants with instant (zero-delay) visual feedback, the negative SOAs (haptic force feedback prior to

visual collision) were dynamically estimated by the instant velocity and distance of the moving square from the vertical bar. The actual SOAs were then measured after the visual collision and haptic force feedback events had occurred. With positive SOAs, the onset of the force feedback occurred at a pre-selected time delay following the visual collision.

At the begin of each trial, the white fixation cross was shown in the display center, with the red square presented on the left or the right side (11.5 cm away from the center) to indicate the starting position of the moving object. A white square was also presented, whose horizontal position was controlled by participants' index finger position. Participants were asked to move this white square to the starting position indicated by the red square. After reaching the starting position, participants had to press any key with their left hand to start the trial. Immediately after this key-press, the red square disappeared and participants had to move the white square with their right-hand index finger to the opposite side at a constant velocity. The object moving speed (11.5 cm/s) was shown in demo block. When the white square collided with the vertical bar, the vertical bar broke into two segments (as in session 1), and a 2N impulse force feedback was provided for 15 ms in the opposite direction of movement at a given SOA relative to the visual collision. Participants were instructed to continue to move the white square until it disappeared, being 11.5 cm away from the center. Finally, the probe text 'vision first or touch first' prompted participants to respond as described above for session 1. It is known that movement velocity will influence the crossmodal simultaneity [31], the movement velocity was examined in each trial. If participants made a slow movement (mean velocity < 8 cm/s) or a fast movement (mean velocity > 15 cm/s), a feedback screen would present at the end of the trial informing participants that their movement was too slow or too fast.

Session 4: Hand movement without visual motion feedback. The stimuli and procedure were identical to session 3, except the moving (square) object was invisible.

All participants took part in the four sessions. The session order is randomized for each participant. There were six blocks of 42 trials in each session, with the first block devoted to practice. The experiment used within-subject full-factor design, with 2 (movement/no movement) x 2 (with/without visual motion feedback) x 7 (SOAs) conditions. There were 30 trials for each condition (the practice trials were excluded).

4 RESULTS

The data from the practice blocks were discarded prior to analysis. Moreover, for the active hand movement sessions, trials with a mean velocity slower than 8 cm/s and faster than 15 cm/s (7.7% of total trials) were also discarded from analysis. On the remaining trials, the average velocities were 12.08 cm/s and 12.01 cm/s for session 3 and 4, respectively. These velocities did not differ significantly from the constant object velocity of 11.5 cm/s in sessions 1 and 2 (t-tests, $p = 0.163$ and $p = 0.138$ for sessions 3 and 4, respectively). Furthermore, for the hand movement session, the actual SOA values used for psychometric function (PF) fitting were computed post-hoc (instead of the estimated) SOAs. A logistic function was used to estimate the PFs for each participant's data set. Figure 3 presents the four different PFs for each participant and Figure 4 shows the average PFs.

The mean PSSs are presented in the left panel of Figure 5. The mean PSS was the smallest in the condition 'active hand movement with visual motion feedback' (Mean±SE: 4.4 ± 5.3 ms). The other three mean PSSs were around 20ms. The data from all participants were examined by a repeated-measures analysis of variance (ANOVA) test with two factors: hand movement (absent/present) and visual motion feedback (absent/present). This ANOVA revealed a significant main effect of hand movement,
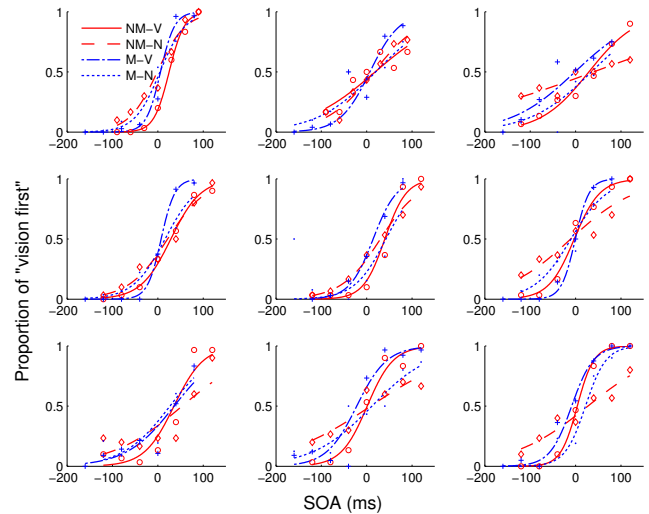
Figure 3: TOJ data and psychometric functions for each subject. The data points indicate the proportion of 'vision first' responses. The curves were estimated using the Logit model (see text for details). NM-V = No movement with visual feedback; NM-N = No movement, no visual feedback; M-V = Movement with visual movement feedback; M-N= Movement without visual feedback.

$F(1,8) = 20.804, p < 0.005$, but no significant main effect of visual motion feedback, $F(1,8) = 2.552, p = 0.149$. The hand movement effect indicates that the PSS is overall smaller under conditions in which participants actively controlled the movement. Furthermore, the two-way interaction was marginally significant, $F(1,8) = 4.653, p = 0.063$, mainly due to the smallest PSS being evident under the condition of active control with visual motion feedback. This interaction seems to indicate that coherent visuo-motor feedback can improve the (estimated) subjective simultaneity (PSS).

The mean JNDs are presented in the right panel of Figure 5. Interestingly, the pattern of the JNDs is different from that of the PSSs. The mean JNDs were smaller under conditions of visual motion feedback (whether or not there was active control), whereas the mean PSSs were smaller only in the condition of active control with visual motion feedback. A repeated-measures ANOVA of the JND data revealed the main effect of hand movement (absent/present) to be significant, $F(1,8) = 12.33, p < 0.01$. The estimated marginal means of the hand movement and no hand movement conditions were 43 and 61 ms, respectively. This indicates that active control of the moving object helps to increase the sensitivity of the temporal discrimination required. The ANOVA also revealed the main effect of visual motion feedback to be significant, $F(1,8) = 18.91, p < 0.01$. The estimated marginal means for the conditions with and without feedback were 37 and 67 ms, respectively. This means that additional information from visual feedback signals can also increase the sensitivity of temporal integration.

Finally, the interaction between hand movement and visual motion feedback was non-significant, $F(1,8) = 2.519, p = 0.151$ - suggesting largely additive effects on perceptual temporal discrimination.

5 DISCUSSION

In this paper we investigated whether or not active hand movements and additional visual motion feedback can influence the crossmodal temporal integration. We quantitatively measured the effects of visual motion feedback and active hand movements on crossmodal temporal integration by estimating the points of subjective
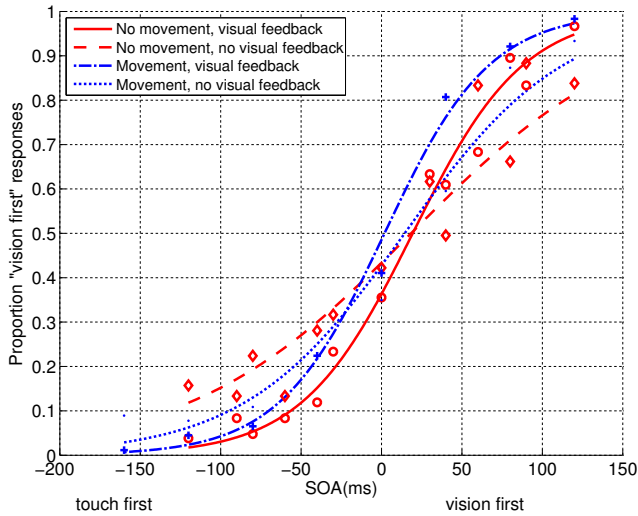
Figure 4: Average psychometric functions across all subjects for the four different conditions.
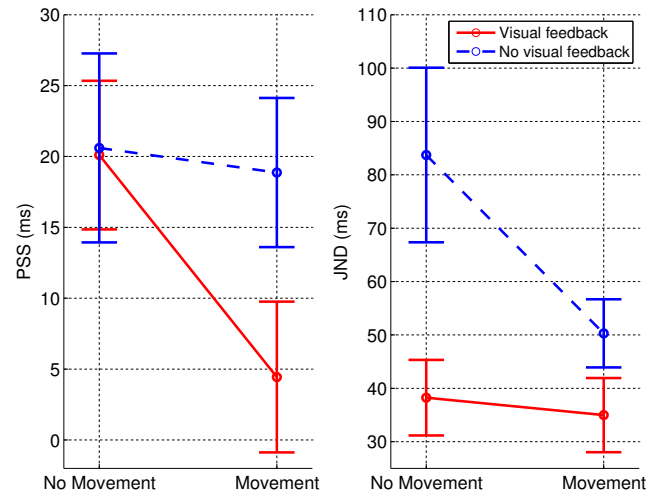

Figure 5: Left Panel: Mean values of the points subjective simultaneity (PSSs) for the four conditions (plus/minus the standard errors). Right Panel: Mean values of just noticeable differences (JNDs) for the four conditions (plus/minus the standard errors).

simultaneity (PSSs) and the just noticeable differences (JNDs). The first finding was that both factors, i.e. redundant visual signals (visual motion feedback) and hand movements, narrowed the synchrony-asynchrony thresholds. Redundant visual signals reduced the mean JND from 67 to 37 ms (a gain of 30 ms), while active hand movements reduced the mean JND from 61 to 43 ms (a gain of 18 ms). These effects are likely to be additive (as suggested by the non-significant interaction). Previous studies showed that participants may use redundant cues such as spatial locations to facilitate their performance on the multi-sensory TOJ task [10, 24]. Our study suggests that the additional visual motion cues can likewise improve the temporal synchrony-asynchrony discrimination. Some vision scientists have proposed that motion signals are usually processed faster than abrupt events, such as flashes and stimulus break-ups [21, 33]. Shorter processing times allow the visual system to compensate for part of the neural delay and to predict the position of moving object more precisely. As a consequence, it narrows the possible time range of forthcoming events. Furthermore, the distance between the moving object and the barrier can also provide a cue for judging whether or not the collision event has happened. In contrast to previous study by Vogels [31], we found the simultaneity sensitivity in active movement was improved. We argued that the main differences were due to the different space configuration. In her experiments, the space of the hand movement was separated from the visual space and the final positions of the joystick in the condition of the active movement were randomly shifted. Thus, as she argued that "moving the arm did not provide the participant with extra information about the moment at which the object would collide" [31]. In the present study, however, the hand movement and visual representation were set in the same space. It simulated natural visuo-motor loop feedback. We demonstrated the reduction of JNDs by active hand movements, which indicates that active motor control also contributes to improving the temporal discrimination. Previous studies have also demonstrated that the central nervous system (CNS) can use the motor command, in conjunction with internal models of both hand and visual feedback, to anticipate the resulting load force and the position of the object [30, 35, 36].

In our study, the PSS for pure visual-tactile events was around 20 ms, i.e. the visual stimulus has to appear 20 ms before the tactile stimulus in order to be perceived as simultaneous. This finding is consistent with previous studies of simple cross-modality com-

parisons [10, 24, 26]. Interestingly, although additional visual motion information reduces the JNDs, it alone does not shift the PSS; rather, the PSS remains constant under 'no movement' conditions regardless of the additional visual motion feedback. This suggests that the passive force feedback or non-motor-related visual motion signals may not be able to influence the PSS. However, the PSS dramatically decreases in the condition with coherent visuo-motor feedback condition (21 ms vs. 4 ms). Moreover, hand movement alone without visually feedback does not influence the PSS (21 ms vs. 19ms). Taken together, these findings suggest that self-generated movement combined with coherent visual feedback is the key to shift the PSS. Previous studies on motor control have suggested that closed-loop visuo-motor processes allow for fine control of the movement [11]. On-line visual feedback lets the visuo-motor system rapidly develop good predictors to finely predict forthcoming events [35]. As a result, the visual-haptic PSS shifts towards zero. In the other two conditions, i.e. 'no movement' with visual motion feedback and active control without visual motion feedback, the closed-loop visuo-motor is not readily formed. Therefore, such a strategy may not be employed.

The magnitude of the mean PSSs obtained in the current study were overall smaller than the values reported by Spence for the divided-attention condition (53ms), but similar to the value for the attend-to-vision condition (22ms). This may be due to the current experimental setting, where participants implicitly focused on the visual modality all the time and their index finger was invisible to them. Since all conditions were run under the same setting, the attentional bias to visual information would have been similar.

The current experimental setting with active manipulation and additional visual motion feedback is similar to the closed visuo-motor loop in typical multi-modal telepresence systems. The present findings indicate a higher sensitivity to temporal synchrony with a closed visuo-motor loop than suggested by earlier studies without a closed visuo-motor loop.

These results have implications for the design of multi-modal communication protocols as well as the design of augmenting measures. The tighter timing requirements represent a challenge especially for the synchronization in multi-modal communication over the communication network where large differences exist in the bandwidth requirements and the associated time consumption for data compression. The PSS and JND values can be considered to

provide design guidelines for such multi-modal data processing approaches.

6 CONCLUSION

In this study we investigated the perception of simultaneity for visuo-haptic events, motivated by the challenge of synchronizing multi-modal information in telepresence systems. Specifically, we examined the influence of active (self-generated) motor control and additional visual motion information on perceived simultaneity, a condition that is similar to the typical multi-modal telepresence scenario. We used a psychophysical method to measure the differences in points of subjective simultaneity (PSS) and just noticeable differences (JND) in visuo-haptic events. The result showed that the PSS decreased significantly under conditions of active motor control with concurrent visual feedback, and the JNDs were narrowed by either active motor control or visual feedback. The values provided may serve as a guideline for multi-modal telepresence system design. Ongoing research investigates the temporal integration window in the presence of a time delay between motor action and sensory perception, as is typical in most telepresence scenarios.

REFERENCES

[1] B. D. Adelstein, D. R. Begault, M. R. Anderson, and E. M. Wenzel. Sensitivity to haptic-audio asynchrony. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 73–76, Vancouver, British Columbia, Canada, 2003.

[2] D. Collett. *Modelling binary data*. CRC texts in statistical science series. Chapman & Hall, London, 2 edition, 2003.

[3] N. F. Dixon and L. Spitz. The detection of auditory visual desynchrony. *Perception*, 9:719–721, 1980.

[4] M. A. Elliott, Z. Shi, and F. Surer. The effects of subthreshold synchrony on the perception of simultaneity. *Psychological Research*, 4 2006.

[5] S. Exner. Experimentelle untersuchung der einfachsten psychischen processe. *Pflüger's Archiv für die geschichte Physiologie*, 11:403–432, 1875.

[6] W. Fujisaki, S. Shimojo, M. Kashino, and S. Nishida. Recalibration of audiovisual simultaneity. *Nat Neurosci*, 7(7):773–8, 2004.

[7] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends Neurosci*, 15(1):20–5, 1992.

[8] A. J. Hamlin. On the least observable interval between stimuli addressed to disparate sense and to different organs of the same sense. *American Journal of Psychology*, 6:564–575, 1895.

[9] S. Hirche, B. Stanczyk, and M. Buss. Haptic Tele-Assembly over the Internet. In *Proceedings of the EuroHaptics'2004 Conference*, pages 417–421, Munich, Germany, 2004.

[10] I. J. Hirsh and J. Sherrick, C. E. Perceived order in different sense modalities. *J Exp Psychol*, 62:423–32, 1961.

[11] S. W. Keele. Motor control. In K. R. Boff, L. Kaufman, and J. P. Thomas, editors, *Handbook of perception and human performance*, volume II. Cognitive processes and performance, pages 30–1 to 30–60. Wiley, Chichester, 1986.

[12] M. Keetels and J. Vroomen. The role of spatial disparity and hemifields in audio-visual temporal order judgments. *Exp Brain Res*, 167(4):635–40, 2005.

[13] D. J. Levitin, K. MacLean, M. Mathews, and L. Chu. The perception of cross-modal simultaneity. *International Journal of Computing and Anticipatory Systems*, pages 323–329, 2000.

[14] J. Lewald and R. Guski. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain Res Cogn Brain Res*, 16:468–478, 2003.

[15] B. Libet. Timing of conscious experience: reply to the 2002 commentaries on libet's findings. *Conscious Cogn*, 12(3):321–31, 2003.

[16] M. Miyazaki, S. Yamamoto, S. Uchida, and S. Kitazawa. Bayesian calibration of simultaneity in tactile temporal order judgment. *Nature Neuroscience*, 9:875–877, July 2006.

[17] K. Moutoussis and S. Zeki. Functional segregation and temporal hierarchy of the visual perceptive systems. *Proceedings of the Royal Society of London Series B*, 264:1407–1414, 1997.

[18] S. Munir. *Internet-Based Teleoperation*. PhD thesis, Georgia Institute of Technology, George W. Woodruff School of Engineering, 2001.

[19] J. Navarra, S. Soto-Faraco, and C. Spence. Adaptation to audiotactile asynchrony. *Neurosci Lett*, 413(1):72–6, 2007.

[20] G. Niemeyer and J. E. Slotine. Towards Force-Reflecting Teleoperation Over the Internet. In *Proceedings on Robotics and Automation, 1998.*, pages 1909–1915, Leuven, Belgium, 1998.

[21] R. Nijhawan. Motion extrapolation in catching. *Nature*, 370(6487):256–7, 1994.

[22] R. Oboe and P. Fiorini. Internet-Based Telerobotics: Problems and Approaches. In *Proceedings of the International Conference on Advanced Robotics ICAR 97*, pages 765–770, Monterey (CA), US, 1997.

[23] C. Spence. Multisensory attention and tactile information-processing. *Behav Brain Res*, 135(1-2):57–64, 2002.

[24] C. Spence, R. Baddeley, M. Zampini, R. James, and D. I. Shore. Multisensory temporal order judgments: when two locations are better than one. *Percept Psychophys*, 65:318–328, Feb. 2003.

[25] C. Spence, M. E. Nicholls, and J. Driver. The cost of expecting events in the wrong sensory modality. *Percept Psychophys*, 63:330–336, Feb. 2001.

[26] C. Spence, D. I. Shore, and R. M. Klein. Multisensory prior entry. *J Exp Psychol Gen*, 130:799–832, Dec. 2001.

[27] C. Spence and S. Squire. Multisensory integration: maintaining the perception of synchrony. *Curr Biol*, 13(13):R519–21, 2003.

[28] B. E. Stein and M. A. Meredith. *The merging of the senses*. MIT Press, Cambridge (MA), US, 1993.

[29] J. V. Stone, N. M. Hunkin, J. Porrill, R. Wood, V. Keeler, M. Beanland, M. Port, and N. R. Porter. When is now? perception of simultaneity. *Proc Biol Sci*, 268(1462):31–8, 2001.

[30] R. J. van Beers, A. C. Sittig, and J. J. Gon. Integration of proprioceptive and visual position-information: An experimentally supported model. *J Neurophysiol*, 81(3):1355–64, 1999.

[31] I. M. Vogels. Detection of temporal delays in visual-haptic interfaces. *Human factors: The journal of the Human Factors and Ergonomics society*, 46:118–134, 2004.

[32] M. Wexler and F. Klam. Movement prediction and movement production. *Journal of Experimental Psychology: Human Perception and Performance*, 27:48–64, 2001.

[33] D. Whitney. The influence of visual motion on perceived position. *TRENDS in Cognitive Sciences*, 6(5):211–216, 2002.

[34] A. G. Witney, S. J. Goodbody, and D. M. Wolpert. Predictive motor learning of temporal delays. *J Neurophysiol*, 82(5):2039–48, 1999.

[35] D. M. Wolpert and Z. Ghahramani. Computational principles of movement neuroscience. *Nat Neurosci*, 3 Suppl:1212–7, 2000.

[36] D. M. Wolpert, Z. Ghahramani, and M. I. Jordan. An internal model for sensorimotor integration. *Science*, 269(5232):1880–2, 1995.

[37] M. Zampini, T. Brown, D. I. Shore, A. Maravita, B. Roder, and C. Spence. Audiotactile temporal order judgments. *Acta Psychol (Amst)*, 118(3):277–91, 2005.

[38] M. Zampini, S. Guest, D. I. Shore, and C. Spence. Audio-visual simultaneity judgments. *Percept Psychophys*, 67(3):531–44, 2005.

[39] M. Zampini, D. I. Shore, and C. Spence. Audiovisual temporal order judgments. *Exp Brain Res*, 152(2):198–210, 2003.

[40] M. Zampini, D. I. Shore, and C. Spence. Multisensory temporal order judgments: the role of hemispheric redundancy. *Int J Psychophysiol*, 50(1-2):165–80, 2003.