Effects of packet loss and latency on the temporal discrimination of visual-haptic events

Zhuanghua Shi, Heng Zou, Markus Rank, Lihan Chen, Sandra Hirche, and Hermann J. Müller

Abstract—Temporal discontinuities and delay caused by packet loss or communication latency often occur in multimodal telepresence systems. It is known that such artifacts can influence the feeling of presence [1]. However, it is largely unknown how the packet loss and communication latency affect the temporal perception of multisensory events. In this article, we simulated random packet dropouts and communication latency in the visual modality and investigated the effects on the temporal discrimination of visual-haptic collisions. Our results demonstrated that the synchronous perception of crossmodal events was very sensitive to the packet loss rate. The packet loss caused the impression of time delay and influenced the perception of the subsequent events. The perceived time of the visual event increased linearly, and the temporal discrimination deteriorated, with increasing packet loss rate. The perceived time was also influenced by the communication delay, which caused time to be slightly overestimated.

Index Terms—visual-haptic temporal perception, packet loss, psychophysics, perception

1 Introduction

Multimodal telepresence systems have been adopted in a variety of applications, such as telesurgery, space and underwater teleoperation. In a typical multimodal telepresence system, the human operator manipulates the remote robot (teleoperator) through the local human system interface (HSI). Information from the remote environment, such as visual, auditory and haptic signals, is then fed back to the human operator. Multisensory feedback provides the operator with an enhanced immersive experience, permitting him/her to competently and efficiently explore and operate in the remote environment [2], [3], [4].

However, 'high-fidelity' experiences in telepresence systems are often compromised by many factors. For example, it is hard to maintain the synchronous feedback from multiple modalities within very short time windows. Yet, synchronicity of feedback is one of the most critical factors for ensuring that the virtual environment is consistent and remote events are transmitted with their causal relationships maintained [2], [5]. Due to data compression and bandwidth requirements for transmission, communication latencies differ significantly among various modalities. In addition, telepresence systems operating over large geographical distances suffer packet loss and network communication delays, as the data are transmitted via the Internet. As a result, 'synchronous' events may be turned into 'asynchronous' incidents. It is well known that such packet losses and communication delays deteriorate users' performance of the ongoing task [6], [7] and may even, at times, lead to dangerous situations, especially when the causality of the remote events is distorted. However, to our knowledge, it is still largely unknown precisely (in quantitative terms) how packet loss and communication delays affect the temporal perception of crossmodal events.

1.1 Packet loss and related models

In telepresence systems, the Internet is an attractive medium for transmitting information between the human system interface (HSI) and the remote teleoperator (TO). A number of studies have shown that packet loss is a key factor determining the quality-of-service (QoS) in delay-sensitive multimedia applications [6], [8], [9], [10], [11], [12]. For example, perceptual quality was dramatically reduced in frame-based coding schemes with packet loss rates equal to or greater than about 8% [6]. Other studies suggested that packet loss rate is tolerable within a somewhat wider range. Beigbeder et al. showed that, with standard network games using the DHCP service, users rarely notice packet losses as high as 5% during the game [11]. Another study examining the perception of information loss where participants watched continuous video stream at 30 frames per second indicated that with aggregate losses less than 17%, the loss is imperceptible; with losses between 17%-23%, it is tolerable; while above 23% it is unacceptable [10]. Besides the packet loss rate, the size of consecutive packet loss (i.e., the burst length) can also be noticed by the user and influence his/her performance [8]. In video streams, losses of two consecutive video frames (∼60ms) would be noticed by most users [10].

Several models have been proposed for describing the characteristics of packet losses in the Internet. The

- *Z. Shi, L. Chen and H. J. Müller are with the General and Experimental Psychology, Ludwig-Maximilians-Universität München, Munich, Germany, 80802. E-mail: shi@lmu.de*
- *H. Zou is with Graduate School of Systemic Neurosciences; the General and Experimental Psychology, Ludwig-Maximilians-Universität München, Munich, Germany, 80802.*
- *M. Rank and S. Hirche are with the Institute of Automatic Control Engineering, Technische Universität München, Munich, Germany, 80290.*

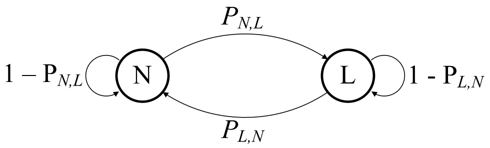


Fig. 1: The Gilbert-Elliot model is a 2-state Markov process and approximates the network characteristics in packet-based data transmission networks. 'N' and 'L' denote the states 'No packet loss' and 'Packet loss' state, respectively.

simplest one is the Bernoulli model, which assumes an uncorrelated loss probability over time. In so-called packet-switched networks, however, bursts of packet losses are often observed. In contrast to a Bernoulli process, the Gilbert-Elliot model [13], [14] is able to reflect this behavior. The Gilbert-Elliot model is described by a two-state Markov process. Indeed, this model can approximate the characteristics of packet loss in the Internet [15]. Essentially it states that the network can be sufficiently described by the probability $P_{n,l}$ for a transition from state N (No packet loss) to state L (Packet loss) and the probability $P_{l,n}$ for a transition from state L (Packet loss) to state N (No packet loss, see Figure 1). The mean loss rate P_l^0 can be computed as follows:

$$P_l^0 = \frac{P_{n,l}}{P_{n,l} + P_{l,n}}. \quad (1)$$

The mean burst length of consecutive packet losses $E(L)$ can then be calculated as

$$E(L) = \frac{1}{P_{l,n}}. \quad (2)$$

Given a desired loss rate and mean burst length, the packet loss process is fully defined and the transition between the mutually exclusive states (N, L) can be derived from the above formulas. Based on the superior performance of the Gilbert-Elliot model, we used it for the simulation of packet dropout.

1.2 Time delay and temporal perception

Time delay is ubiquitous in telepresence systems. It is generally known that time delays degrade user performance. For instance, delays in visual or haptic feedback increase both task completion times and error rates [16], [17], [18], [19]. Based on his/her own motor commands and proprioception, a delay of 250 ms in the visual feedback is easily recognized by the human operator [20]. In virtual environments, even the small system latency (of 33 ms) between input action and visual display change can be perceived by the observer based on the 'image slip' [21]. MacKenzie and Ware [1] examined the quantitative effect of time delays on users' completion time. Using a Fitts's law target acquisition task, they found a linear relationship between the magnitude of the time delay (ranging between 25 and 225 ms) and the completion time. The completion time has also been found to depend on the task difficulty: the harder the task, the greater the detrimental effect caused by time delay. In a recent study, Jay et al. [22] examined the effect of time delay between haptic and visual feedback in a collaborative virtual environment. In their study, the error rate rose steeply from delays of 0 to 25 ms, even though participants failed to notice the delay. This suggests that time delay can dramatically degrade performance even in a very short range of latencies.

Varying time delays in communication can be characterized by two values: the mean end-to-end delay (latency) and its variation (jitter). Both latency and jitter are produced mainly by the infrastructure of the communication system. In multimodal telepresence systems, the delay problem does not solely arise from the latency and jitter in the communication, but also from the temporal inconsistency between the different sensory modalities. And the latter is compounded by the fact that the human brain processes different types of sensory information with different latencies [23], [24], [25], [26]. For example, humans can detect the audiovisual asynchrony more easily when the sound precedes the picture information [27]. Similarly, asynchrony in visual-haptic events can be detected easily if the haptic feedback is delivered 50 ms before the visual event. But when the visual event is presented in advance of the haptic event, the two events are likely to be perceived as synchronous with differences in onset times less than 150 ms [26]. Recent studies have also shown that the crossmodal temporal synchrony-asynchrony threshold for visual-haptic events can be influenced by the visuo-motor control loop [28], [29].

In studies of crossmodal temporal perception, two types of thresholds are usually measured: the point of subjective simultaneity (PSS) and the just noticeable difference (JND) [26]. The PSS is the time interval between the onsets of two sensory stimuli at which the two stimuli are reported to be most synchronous. The JND, on the other hand, indicates the resolution of the temporal discrimination. In addition, two types of tasks are frequently used for determining the thresholds: the temporal-order judgment (TOJ) and the synchrony/asynchrony judgment (SJ). In the TOJ task, the JND is calculated as half the difference between the lower (25%) and upper bound (75%) of the threshold, whereas in a SJ task, the JND is defined by 50% of 'synchronous' responses. The PSS is defined by the 50%-threshold in the TOJ task and the maximum of the distribution of 'synchronous' responses in the SJ task.

1.3 Aim of the study

The present study was designed to systematically and quantitatively investigate how packet loss, with a constant latency of visual feedback, influences the synchrony perception of a visual-haptic collision. In particular, we examined the effects of the packet loss and latency in the visual modality since the visual information often occupies most of the communication bandwidth. We used the Gilbert-Elliot model to simulate packet loss in the end-to-end transmission of video frames. We hypothesized that the rate of packet loss would influence the quality of the signal, thus affecting the variance of temporal discrimination (indexed by the JND). In addition, packet loss may induce the general impression of a time delay and thus bias the perceived onset of the visual event. Higher packet loss rates give rise to the perception of longer time delays. This would be observable in terms of a modulation of the PSS by the packet loss rate. Experiment 1 was designed to test these two hypotheses.

Communication latencies have been shown to have a similar influence as packet losses on the perceived quality of videos [7]. However, it is as yet unknown how they influence crossmodal temporal perception in conjunction with packet losses. One possible outcome is that communication latency and packet loss influence crossmodal temporal perception independently. Alternatively, they may have an interactive effect on the temporal discrimination. Experiment 2 was designed to examine this issue.

2 General Methods

2.1 Participants

Ten healthy participants took part in Experiment 1 (5 females, mean age of 25.1 years) and eight in Experiment 2 (5 females, mean age of 25.0 years); They were paid at a rate of 8 Euros per hour. All participants had normal or corrected-to-normal vision and were right-handed; none of them reported any history of somato-sensory disorders. And all were naive as to the purpose of the experiments, except for one of the participants (H.Z.,one of the authors). The number of the participants reported above do not include two further observers who either failed to reach the criterion of finger movement speed or displayed a strong responses bias ($\geq 80\%$) towards one direction (see below for a more detailed explanation).

2.2 Apparatus

The haptic feedback force was generated via a PHANToM Premium 1.5A haptic device (SensAble Technologies, Inc.). The visual 3D environment was presented on a Philips 202p70 CRT monitor (screen resolution: 1024×768 pixels; refresh rate: 120 Hz), which was fixed above the haptic device and tilted 80 degrees towards the observer. The visual space was collocated with (i.e., projected into) the haptic space by means of a mirror, and participants viewed the mirrored image through a pair of shutter glasses for stereo-image presentation (StereoGraphics CrystalEYE3 with E2 emitter) (see Fig. 2). Ear masks were used to block out the auditory noise generated by the haptic device. To ensure accurate timing of visual and haptic events, we developed a calibration system with a luminance sensor and an acceleration sensor, designed to measure the onsets of visual stimuli and the pulses of the force feedback. After calibration, the physical asynchrony of visual-haptic ('synchronous') events was guaranteed to be no greater than 1 visual frame (8 ms).

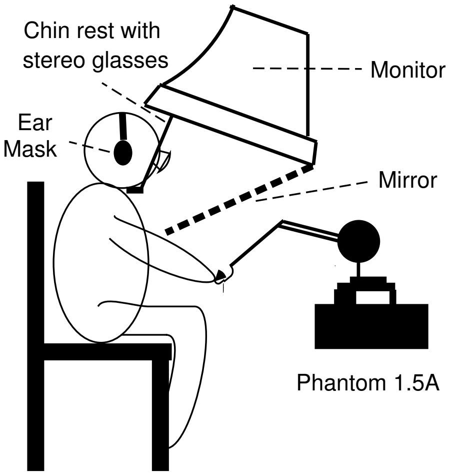


Fig. 2: Schematic illustration of the visual-haptic collocated setup: Participants viewed the mirror-reflected visual stimuli through the shutter glasses. The right-hand index finger was attached to the thimble of the PHANToM. The hand beneath the mirror could not be seen (see text for further details).

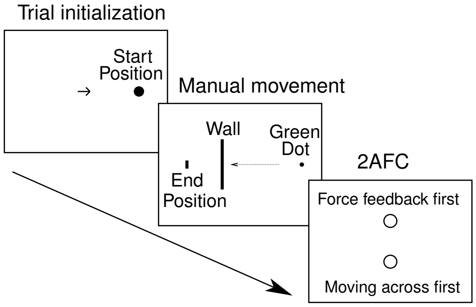


Fig. 3: Schematic illustration of a trial sequence. See Section 3.1 for a detailed description of the procedure.

3 Experiment 1: Packet loss and event discrimination

3.1 Method

To simplify the simulation, we used a frame-based encoding scheme and defined one video frame as one packet. Four different mean packet loss rates (0. 0.1, 0.2 and 0.3, respectively) were used in Experiment 1 in order to examine the relationship between the loss rate and the visual-haptic event judgment. To make the packet loss perceivable, the mean length of burst losses was fixed closely to the visual flicker fusion threshold (4 frames at the refresh rate 120 Hz, about 33 ms). During the burst of lost frames (packets), the last received frame remained visible until a new frame arrived. Phenomenally, packet loss in visual feedback interrupts video continuity.

Before the experiment, the participant's right index finger was attached to the thimble of the haptic device (See Fig. 2). Corresponding with the position of the finger tip, a green dot (subtending $0.05^\circ \times 0.05^\circ$ visual angle) was presented through the 3D display. At the beginning of each trial, an arrow was presented in display center, indicating the start side of the to-be-performed movement, as well as a wire-frame sphere ball (diameter of 0.4° visual angle) indicating the starting position (either on the left or on the right side of the visual space, 15.0 cm away from the center). The participant had to move the finger tip (i.e., the corresponding green dot) into the wire sphere to initialize a trial. Upon this, a short line ($0.05^\circ \times 0.2^\circ$ visual angle) appeared on the opposite horizontal side indicating the end position of the movement. After that, the participant had to move his/her finger tip horizontally at a constant speed from the start position towards the end position. Meanwhile, an inward attraction force field was generated to an invisible line along the movement direction. The force field helped the participant move the finger in the horizontal direction without deviating from the invisible line. There was no attracting force along the movement direction. In addition, a vertical line ($0.05^\circ \times 1^\circ$ visual angle) was displayed in-between the center and the end position (placed randomly 2.19 to 4.74 cm away from the center), which represented a 'wall'. When the green dot moved across the 'wall', we defined this 'crossing' event as visual collision.

With or without time delay, this event would be accompanied by a short pulse force feedback along the horizontal axis opposite to the movement direction. This force feedback was defined as haptic collision. The magnitude of the pulse force corresponded to a spring force with a stiffness of 100 N/m, and the maximal pop-through magnitude was 3 N. The visual-haptic stimulus onset asynchrony (SOA), measured from the onset of visual collision to the onset of the force feedback, was systematically varied from 0 to 120 ms with a step size of 20 ms. When the finger reached the end position, two alternative response options: 'dot moving across the wall first' and 'force feedback first', were presented on the screen. The participant had to make a two-alternative forced-choice (2AFC) by pointing with the finger to the corresponding button(see Fig. 3).

Before starting the formal experiment, participants received a training session of about 15 minutes. During this session, participants became familiar with the task and the appropriate speed of finger movement. In order to avoid (potential) positional discrepancies between vision and proprioception arising from the slow visual update rate of the monitor (120 Hz), the speed of finger movement was monitored by the program. When the participant moved too rapidly or too slowly (grand mean velocity greater than 20 cm/sec or, respectively, less than 12 cm/sec), the trial was discarded and repeated at the end of the experiment.

The experiment used a full factorial within-subject design, with 4 (loss rate) × 7 (visual-haptic SOA) conditions. For each condition, there were 20 'valid' (i.e., on the movement speed criterion acceptable) trials. Depending on the number of discarded trials that a participant produced, the experiment consisted of 10 or more blocks with 56 trials per block.

3.2 Analysis

Psychometric functions, such as logistic function, are often used to model the binomial response data [30]. A logistic function,

$$P(x) = \frac{1}{1 + e^{\frac{\alpha - x}{\beta}}}, \tag{3}$$

was used in current study for estimating PSS and JND. With above function, the parameters α and β can be easily estimated from the data. The PSS can then be obtained as:

$$P\hat{S}S = \hat{\alpha}, \tag{4}$$

and the JND can be calculated as:

$$J\hat{N}D = (x_{P.75} - x_{P.25})/2 = \hat{\beta} \log 3. \tag{5}$$

3.3 Results

Only the 560 valid trials were included in the following analyses. To rule out an influence of finger movement speed on performance, the average velocities of the finger movement during the 500-ms interval prior to the collision were calculated. These were 24.8, 24.4, 24.2 and 24.5 cm/s for the 0, 0.1, 0.2 and 0.3 loss rate conditions, respectively. A repeated-measures analysis of variance (ANOVA) showed that these velocities did not differ significantly between the loss rate conditions, $F(3, 27) = 0.503$.

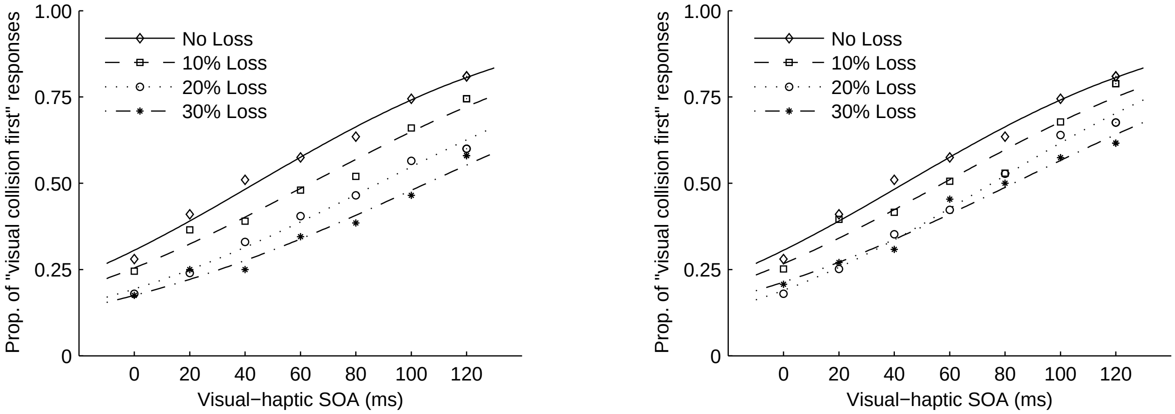
PSS and JND were estimated for each condition, individually for each participant. The group mean PSSs and JNDs are presented in Table 1. With all data combined, the overall mean psychometric functions for the four different loss rate conditions are shown in Fig. 4a.

Note that the visual 'continuous-movement' event could become a 'jump-movement' event when a burst of packet loss happened during the visual collision, that is, the moving dot stopped and then suddenly jumped over the wall, continuing the movement. The proportions of 'jump-movement' trials were close to the packet loss rates, that is, 0, 0.1, 0.2 and 0.3, respectively. In order to take this fact into account, we excluded 'jump-movement' trials and recalculated the psychometric curves and corresponding PSS's and JND's. These are presented in Fig. 4b and in Table 1. Furthermore, we calculated the mean image stagnation time around the visual collision event for each condition of packet loss rate (See Table 1).

The mean PSSs, PSS's, JNDs and JND's are shown in Fig. 5. The PSSs from all participants were examined by a

TABLE 1: Means and associated standard errors [in ms] of PSS, JND, PSS′ (without ‘jump-movement’ trials), JND′ (without ‘jump-movement’ trials) and ST (mean image stagnation time) from Experiment 1. A positive PSS means that the visual collision has to precede the haptic collision in order to be perceived as synchronous with the latter.

Loss rate	PSS(±SE)	JND(±SE)	PSS′(±SE)	JND′(±SE)	ST
0	43.5 (±17.2)	48.7 (±6.8)	43.5 (±17.2)	48.7 (±6.8)	0.
0.1	64.2 (±14.0)	59.8 (±7.1)	58.1 (±15.0)	56.1 (±6.5)	4.99
0.2	96.9 (±19.8)	70.8 (±12.1)	83.5 (±17.6)	55.4 (±8.1)	11.66
0.3	115.7 (±18.3)	81.3 (±12.7)	90.6 (±17.9)	72.3 (±10.3)	17.31



(a) With all trials

(b) Without ‘jump-movement’ trials

Fig. 4: TOJ data and psychometric functions for the four different loss rates. The curves were estimated using the logistic model. (a) psychometric curves with all trials. (b) psychometric curves without ‘jump-movement’ trials.

repeated-measures ANOVA, which revealed a significant main effect of packet loss, $F(3, 27) = 10.65, p < 0.001$. A further (within-subjects) linear-contrast test indicated that the PSSs are a linear function of the packet loss rate, $F(1, 9) = 18.28, p < 0.005$. The estimated linear equation is,

$$\text{PSS} = 42.7 + 249.3 \times \text{LossRate}, (r^2 = 0.989) \qquad (6)$$

A repeated-measures ANOVA of the PSS′s showed that, without the ‘jump-movement’ trials, the main effect of packet loss was still significant, $F(3, 27) = 7.34, p < 0.001$, and a follow-on linear-contrast test showed the PSS′s to increase linearly with the packet loss rate, $F(1, 9) = 13.37, p < 0.005$. Accordingly, the estimated linear equation is,

$$\text{PSS}' = 43.9 + 166.5 \times \text{LossRate}, (r^2 = 0.961) \qquad (7)$$

Interestingly, for each packet loss rate condition, the magnitude of PSS was close to the sum of the PSS′ and the corresponding stagnation time.

A further ANOVA test conducted on the JNDs revealed a significant main effect of packet loss, $F(3, 27) = 4.97, p < 0.01$, with the JNDs increasing linearly as a function of the loss rate, as confirmed by a linear-contrast test, $F(1, 9) = 8.26, p < 0.05$. The estimated linear equation is,

$$\text{JND} = 48.8 + 108.6 \times \text{LossRate}, (r^2 = 0.99) \qquad (8)$$

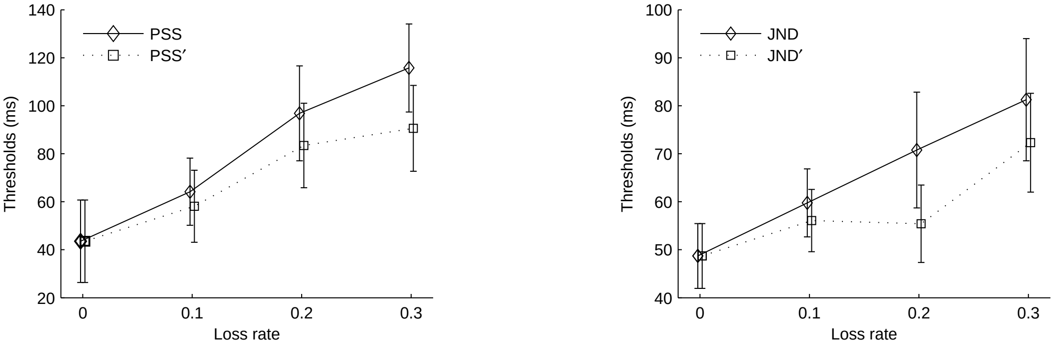
Similarly, without ‘jump-movement’ trials, the main effect of the packet loss rate on the JND′s was again significant, $F(3, 27) = 3.27, p < 0.05$. Further contrast tests showed that the JND′s had a linear trend with packet loss rate, $F(1, 9) = 5.96, p < 0.05$, but no quadric or cubic trends, F(1,9)=0.94, and F(1,9)=1.2, respectively. The further linear regression suggests,

$$\text{JND}' = 47.6 + 70 \times \text{LossRate}, (r^2 = 0.816) \qquad (9)$$

3.4 Discussion

Regardless of the packet loss rate, the task was accomplished with similar movement velocities. This indicates that, in the current experiment, packet losses had little influence on user action in this simple goal-directed movement task.

The positive PSS value in the baseline condition without packet loss reveals that the visual event has to be presented some 50 ms before the tactile event to reliably achieve perceptual simultaneity. This result is consistent with previous studies [26], [29], indicating that the processing of visual signals requires more time than that of haptic signals. Together with the JND, one can infer that the visual-haptic simultaneity window ([PSS-JND, PSS+JND]) is in the range of -5 to 92 ms. By contrast, in the packet loss rates conditions, the visual-haptic simultaneity windows are in the positive range, [4.4, 124 ms],[26.1, 167.7 ms] and [34.4, 207 ms] for the



(a) PSSs and PSS′s

(b) JNDs and JND′s

Fig. 5: (a) PSS and PSS′ as a function of the packet loss rate. (b) JND and JND′ as a function of the packet loss rate. PSS′s and JND′s were calculated from trials without 'jump-movement'. Error bar indicates the standard error (with $n = 10$).

loss rates 0.1, 0.2 and 0.3 respectively. This means that physically simultaneous visual-haptic events are more likely to be perceived as haptic event first and visual event second.

As shown in Table 1, packet losses can cause a 'jump-movement' at the visual collision, which leads to a delayed 'movement-across-the-wall' event. This delay, as measured by the mean stagnation time, contributes to the shifts of the PSSs. However, on average, the mean stagnation times are smaller than the shifts of the PSSs. Excluding the 'jump-movement' trials, the PSS′ is still shifted positively and increases linearly with the packet loss rate (see Fig. 5a). This indicates the packet loss before the collision may bias the time judgment of the forthcoming event. Packet loss causes stagnant images from time to time, which gives the user the impression of a signal delay. It is reasonable to assume that, in working memory, the user may use such delay information to predict the forthcoming events. As a consequence of this, the temporal perception of intact ('continuous-movement') events is also affected by the preceding packet loss.

It is known that packet loss can drastically reduce the perceptual quality [6], [7]. This is also reflected in our results. The linear increase of JNDs with packet loss suggests that the quality of the signal determines the temporal resolution of the event discrimination.

In summary, the major finding in Experiment 1 was that the packet loss rate linearly affects both the PSS and the JND. The general impression of the delay caused by the packet loss before the collision may influence the perceived timing of the forthcoming event.

4 Experiment 2: Packet loss with delay and event discrimination

Experiment 2 was designed to further examine the effects of the packet loss, but this time in combination with a communication delay on visual-haptic temporal order judgments.

4.1 Method

The procedure and the stimuli were the same as in Experiment 1, except for the following changes: A communication latency was introduced for the visual modality, that is, there was a time delay between the participant's actual finger movement and the visible movement of the green dot which represented the finger. Two levels of visual signal delay were compared: 0 ms versus 50 ms. Furthermore, there were two levels of packet loss rates: 0 and 0.2. In order to cover the visual-haptic simultaneity window, the range of visual-haptic SOAs was extended from 0 to 180 ms, varied in steps of 30 ms. Thus, Experiment 2 implemented a full factorial (within-subject) design, with 2 (loss rate) × 2 (visual latency) × 7 (visual-haptic SOA) conditions.

4.2 Results

The mean image stagnation time caused by 'jump-movement' events was 4.13 ms for the packet loss rate of 0.2, which was comparable to Experiment 1 (see Table 1).

The mean movement velocity in the 500-ms period before the haptic collision was 32.8, 33.5, 33.2 and 33.7 cm/s in four conditions 'no packet loss/no visual delay', 'no packet loss/50-ms visual delay', 'packet loss/no visual delay', and, respectively, 'packet loss/50-ms visual delay'. A repeated-measures ANOVA revealed these velocities to be statistically equivalent, $F(3, 21) = 1.3$. This indicates that the movements before the collision were comparable.

The PSSs and JNDs were estimated separately for each of the four conditions, individually for each participant. Fig. 6 presents the average proportion of 'visual collision first'-responses as a function of the visual-haptic SOA, separately for the four conditions, as well as the corresponding psychometric curves. A repeated-measures ANOVA of the PSS estimates with the factors packet loss rate and visual latency revealed both main effects to be

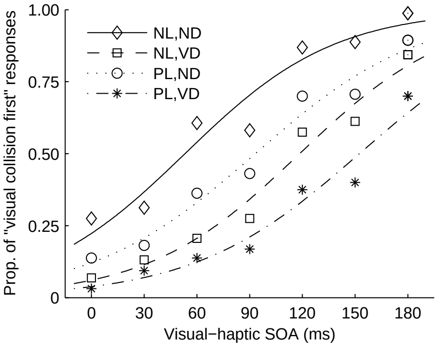
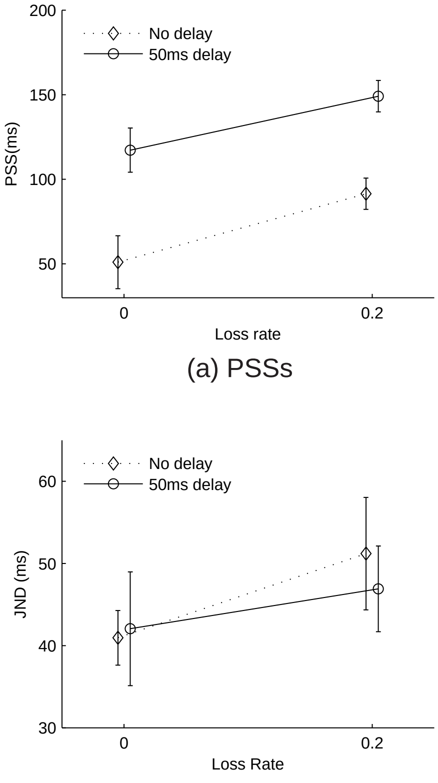


Fig. 6: TOJs as a function of the visual-haptic SOA (Experiment 2).The psychometric curves were estimated using the Logit model. 'NL' denotes no packet loss, 'PL' packet loss, 'ND' without visual delay and 'VD' with visual delay.



(a) PSSs

(b) JNDs

Fig. 7: (a) PSSs and (b) JNDs as a function of the packet loss rate. Error bars indicate the standard error of the means ($n = 8$).

significant: packet loss rate, $F(1,7) = 15.66, p < 0.01$, and visual latency, $F(1,7) = 284.83, p < 0.001$. However, the interaction was not significant, $F(1,7) = 1.79$. The loss rate of 0.2 caused a shift of the PSS by 36.2 ms, and the visual 50-ms delay caused a further (additive) shift of the PSS by 61.9 ms (see Fig. 7a). A one sample t-test revealed the shift of the PSS to be significantly larger than 50 ms, $t(15) = 3.4, p < 0.01$.

A repeated-measures ANOVA of the JND estimates showed the main effect of packet loss to be significant, $F(1,7) = 10.16, p < 0.05$, consistent with Experiment 1. The condition with packet loss exhibited a larger variability in temporal order judgments. However, the main effect of visual-signal delay was non-significant, $F(1,7) = 0.19, p = 0.67$, and there was no interaction effect, $F(1,7) = 0.72, p = 0.42$ (see Fig. 7b).

4.3 Discussion

Although the movement velocity was somewhat faster compared to Experiment 1 (which is likely due to the different groups of participants), the movement velocities did not differ significantly among the four experimental conditions. Again, this suggests that the incidence of packet loss and the existence of a visual-signal delay do not as such influence the user's action in a simple goal-directed movement task.

The PSS result obtained in Experiment 2 suggest that both factors influence the perception of crossmodal events and they do so in an additive way. Consistent with Experiment 1, the packet loss caused the visual (collision) event to be perceived as being delayed. More interestingly, the visual-signal latency induced a shift of 62 ms in the PSS, which is significantly longer than 50-ms latency. This indicates that the delay time was overestimated.

Similar to Experiment 1, packet loss affected the perceptual quality of the signal, which made the temporal-order judgment more difficult. This was evidenced by an increase in the JND. Although packet loss greatly influenced the JND, the delay of visual feedback did not affect the variability of the crossmodal event discrimination. However, this (non-finding) might be due to the fact that the visual-signal delay was relatively short and invariant (50ms). Indeed, it has been reported that, when the time delay is variable, it can affect the perceptual quality of the signal [6]. Therefore, it remains to be seen in future studies whether or not varying delay time influences the precision of temporal-order judgments.

5 General Discussion

Packet loss and communication latency are ubiquitous in multimodal telepresence systems. Several previous studies have focused on how these factors influence the perceptual quality of the service (QoS) and task performance [1], [6], [11], [12], [16], [17], [22]. The present study mainly examined how they influence the temporal perception of visual-haptic events. We consider the findings to make an important contribution to the issue of time delay in the field, since temporal-order judgments of multisensory events reflect a very fundamental process serving higher cognitive tasks, for example, determining causal relationships based on the temporal order of the events involved.

The first finding was that both PSS and JND increased as a function of packet loss rate. The PSS increased much faster than the increase in the mean stagnation time during the visual collision. Excluding such ‘jump-movement’ trials, the PSS′ still shifted linearly with the packet loss rate. Recently, Vatakis and Spence examined the influence of the frame rate on audiovisual temporal-order judgments; they found lower frame rate (6 fps) speech video clips to require larger visual-speech leads for the PSS to be achieved compared to higher frame rate video clips [31]. The low frame rate in their study and packet loss in our study both gave rise to a delayed temporal percept, that is, visual information was perceived as stagnant or delayed compared to the feedback from the other modality. This suggests that the perception of the stagnant(delayed) image may bias judgments of the temporal order of visual-haptic event. Another interesting finding in Experiment 1 was that with a packet loss rate greater than 0.1, physically simultaneous visual-haptic events were perceived as asynchronous. Consequently, as a guideline in visual-haptic telepresence system, the visual packet loss should be kept below 10% with frame-based encoding schemes.

In Experiment 2, the PSS was also found to be shifted by the communication (visual-signal) latency, consistent with a previous study of system latency in virtual environments [21]. Yet, interestingly, the magnitude of the shift was 12 ms longer than the visual-signal delay. One possible mechanism underlying this is the asymmetric time estimation between visual and tactile modalities. Short (empty) tactile intervals have been found to be perceived as longer than the same (empty) visual intervals [32]. In our study, the visual stimuli was delayed compared to the haptic event, the perceived onset time of the visual event might be influenced by the earlier haptic event. Although both packet loss and communication delay shifted the PSS, there was no interaction between two factors. This may suggest that the two factors influence the PSS independently and quite possibly on different levels of temporal processing. While the communication latency would mainly affect the delay on early, sensory-coding levels, the perception of a delay caused by the packet loss may arise on later stages of neural processing concerned with the coding of visual motion information.

In both experiments, the JNDs were positively correlated with the packet loss rate, indicating that temporal sensitivity deteriorates with packet loss. This is consistent with previous findings that packet losses can produce substantial reductions of perceptual quality [6], [7], [9], [12]. Although JNDs have been found to be influenced by packet loss (Experiment 1 and 2), they were unaffected by a constant communication latency (Experiment 2). One reason for this might be that a constant latency does not affect the quality of the visual signal. Consistent with this, the JND, that is, the measure of sensitivity of the temporal discrimination task, was also unchanged. However, one should note that, in the current study, the communication latency was relatively short, only 50 ms. It remains an open issue whether the JND would be affected with longer (and variable) latencies.

6 Conclusion

In this paper, we quantitatively examined the effects of packet loss and time delay on the temporal order judgments of the visual-haptic events. We found that the points of subjective simultaneity (PSSs) and just noticeable differences (JNDs) of visual-haptic events to increase linearly with the packet loss rate. When this rate is greater than 0.1, physically synchronous visual-haptic events would be perceived as asynchronous events. Both packet loss and communication latency result in an increase in the PSS, and a communication delay of 50 ms causes a further shift of about 12ms in the PSS. These results may provide some guidelines for the design of telepresence systems, such as for the choice of time window for presenting assistive functions and of the dynamic upper and lower limits of the simultaneity window for visual-haptic events as a function of the packet loss rate.

Acknowledgements

This work was supported in part by a grant from the German Research Foundation (DFG) within the Collaborative Research Centre SFB 453 on “High-Fidelity Telepresence and Teleaction”. The authors are indebted to Prof. Hong Z. Tan for stimulating discussions.

References

[1] S. I. Mackenzie and C. Ware, “Lag as a determinant of human performance in interactive systems,” in *Proceedings of the INTERACT ’93 and CHI ’93 conference on Human factors in computing systems*. New York, NY, USA: ACM, 1993, pp. 488–493.

[2] J. Steuer, “Defining virtual reality: Dimensions determining telepresence,” *Journal of Communication*, vol. 42, no. 4, pp. 73–93, 1992.

[3] G. Hirzinger, B. Brunner, J. Dietrich, and J. Heindl, “Sensor-based space robotics-rotex and its telerobotic features,” *IEEE Transactions on Robotics and Automation*, vol. 9, no. 5, pp. 649–663, 1993.

[4] S. Hirche and M. Buss, “Human perceived transparency with time delay,” in *Advances in Telerobotics*, M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer, Eds. Berlin: Springer STAR series, 2007, pp. 191–209.

[5] R. Held, “Telepresence, time delay and adaptation,” in *Pictorial communication in virtual and real environments*, S. R. Ellis, M. K. Kaiser, and A. J. Grunwald, Eds. Taylor and Francis, 1993, pp. 232–246.

[6] M. Claypool and J. Tanner, “The effects of jitter on the peceptual quality of video,” in *Proceedings of the seventh ACM international conference on Multimedia (Part 2)*. New York, NY, USA: ACM Press, 1999, pp. 115–118.

[7] Z. Wang, *Internet QoS: Architectures and Mechanisms for Quality of Service*, 1st ed. Morgan Kaufmann, 2001.

[8] D. Hands and M. Wilkins, “A study of the impact of network loss and burst size on video streaming quality and acceptability,” in *Proceedings of the 6th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services*. London, UK: Springer-Verlag, 1999, pp. 45–57.

[9] M. Yajnik, S. Moon, J. Kurose, and D. Towsley, “Measurement and modelling of the temporal dependence in packet loss,” in *Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1, 1999, pp. 345–352.

[10] D. Wijesekera, J. Srivastava, A. Nerode, and M. Foresti, “Experimental evaluation of loss perception in continuous media,” *Multimedia Systems*, vol. 7, no. 6, pp. 486–499, 1999.

[11] T. Beigbeder, R. Coughlan, C. Lusher, J. Plunkett, E. Agu, and M. Claypool, "The effects of loss and latency on user performance in unreal tournament 2003," in *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games*. New York, NY, USA: ACM Press, 2004, pp. 144–151.

[12] M. Dick, O. Wellnitz, and L. Wolf, "Analysis of factors affecting players' performance and perception in multiplayer games," in *Proceedings of 4th ACM SIGCOMM workshop on Network and system support for games*. New York, NY, USA: ACM, 2005, pp. 1–7.

[13] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, pp. 1253–1265, 1960.

[14] E. O. Elliot, "A model of the switched telephone network for data communications," *Bell System Technical Journal*, vol. 44, pp. 89–109, 1963.

[15] H. Sanneck, "Packet loss recovery and control for voice transmission over the internet," Ph.D. dissertation, Technischen University of Berlin, 2000.

[16] H. Kalmus, D. B. Fry, and P. Denes, "Effects of delayed visual control on writing, drawing, and tracing," *Language and Speech*, vol. 3, pp. 96–108, 1960.

[17] W. M. Smith, J. W. Mccrary, and K. U. Smith, "Delayed visual feedback and behavior," *Science*, vol. 132, no. 3433, pp. 1013–1014, 1960.

[18] T. B. Sheridan and W. R. Ferrell, "Remote manipulative control with transmission delay," *IEEE Transactions on Human Factors in Electronics*, vol. HFE-4, no. 1, pp. 25–29, 1963.

[19] W. R. Ferrell, "Delayed force feedback," *Human Factors*, vol. 8, pp. 449–455, 1966.

[20] M. Barth, T. Burkert, C. Eberst, N. O. Stoffler, and G. Farber, "Photo-realistic scene prediction of partially unknown environments for the compensation of time delays in telepresence applications," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 4, 2000, pp. 3132–3137.

[21] B. D. Adelstein, T. G. Lee, and S. R. Ellis, "Head tracking latency in virtual environments: Psychophysics and a model," *Human Factors and Ergonomics Society Annual Meeting Proceedings*, pp. 2083–2087, 2003.

[22] C. Jay, M. Glencross, and R. Hubbold, "Modeling the effects of delayed haptic and visual feedback in a collaborative virtual environment," *ACM Transactions on Computer-Human Interaction*, vol. 14, no. 2, pp. Article 8 / 1–31, 2007.

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

[24] 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," *Proceedings of the Royal Society B: Biological Sciences*, vol. 268, no. 1462, pp. 31–8, 2001.

[25] C. Spence, F. Pavani, and J. Driver, "Spatial constraints on visual-tactile cross-modal distractor congruency effects," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 4, no. 2, pp. 148–69, 2004.

[26] C. Spence, D. I. Shore, and R. M. Klein, "Multisensory prior entry." *Journal of Experimental Psychology: General*, vol. 130, no. 4, pp. 799–832, 2001.

[27] N. F. Dixon and L. Spitz, "The detection of auditory visual desynchrony," *Perception*, vol. 9, no. 6, pp. 719–721, 1980.

[28] K. P. Körding and D. M. Wolpert, "Bayesian decision theory in sensorimotor control." *Trends in Cognitive Sciences*, vol. 10, no. 7, pp. 319–326, 2006.

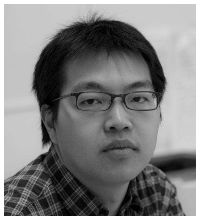
[29] Z. Shi, S. Hirche, W. X. Schneider, and J. H. Müller, "Influence of visuomotor action on visual-haptic simultaneous perception: A psychophysical study," in *2008 Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2008, pp. 65–70.

[30] D. Collett, *Modelling Binary Data, Second Edition (Texts in Statistical Science Series)*. Chapman & Hall/CRC, 2002.

[31] A. Vatakis and C. Spence, "Evaluating the influence of frame rate on the temporal aspects of audiovisual speech perception," *Neuroscience Letters*, vol. 405, no. 1-2, pp. 132–136, 2006.

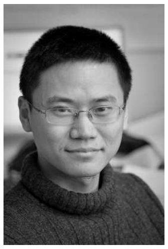
[32] J. B. F. van Erp and P. J. Werkhoven, "Vibro-tactile and visual asynchronies: Sensitivity and consistency," *Perception*, vol. 33, no. 1, pp. 103–111, 2004.

Zhuanghua Shi received his PhD degree in Psychology from Zhejiang University, P. R. China in 2002. He is currently a senior researcher at the Department of Psychology, Ludwig-Maximilians-Universität München, Munich, Germany. He is the principle investigator of the Multisensory Temporal Integration Lab at LMU. His research interests include multisensory temporal perception and integration, crossmodal attention, visual illusion and time perception.



Heng Zou received his MSc in Neuro-Cognitive Psychology in 2008 from Ludwig-Maximilians-Universität München, Munich, Germany. He is currently a PhD student at the Graduate School of Systemic Neurosciences, and a research assistant at the Department of Psychology, LMU. His research interest lies in the general mechanism of information processing in the brain, while his current project focuses on the temporal integration of multisensory information.

Markus Rank received his Dipl.-Ing. degree in electrical engineering from Technische Universität München, Munich, Germany in 2008. He is currently a PhD student at the Institute of Automatic Control Engineering at TUM, working in the Information-oriented Control group. His research interests include interaction with time-delayed multimodal environments, perception of delayed haptic information and human haptic performance modeling.



Lihan Chen received his PhD from the department of Psychology, Ludiwig-Maximilians-Universität München, Munich, Germany in 2009. He works in the Collaborative Research Centre SFB 453 on "High-Fidelity Telepresence and Teleaction" since October, 2007. His research interests are time perception and crossmodal temporal capture effects among auditory, visual and tactile modalities.

Sandra Hirche received her PhD of Engineering in Electrical Engineering and Computer Science in 2005 from the Technische Universität München, Munich, Germany. From 2005-2007 she has been a PostDoc at the Tokyo Institute of Technology, Tokyo, Japan. Since 2008 she holds an associate professor position at the Institute of Automatic Control Engineering, Technische Universität München. Her research interests include control over communication networks, networked control systems, cooperative control, human-machine interaction, mechatronics, multimodal telepresence systems and perception-oriented control.

Hermann J. Müller received his Psychology degree from the University of Würzburg, Germany, and his PhD from the University of Durham, UK. Following a post-doctoral fellowship award by the German Research Foundation, he worked at the School of Psychology, Birkbeck College, University of London, UK. In 1997, he was appointed Chair of Experimental Psychology at the University of Leipzig. In 2000, he became Chair of General and Experimental Psychology at the Ludwig Maximilian University (LMU) Munich. He has a broad range of research interests including: visuo-spatial attention, adaptive weighting dynamics in visual search, cross-modal processing and motor action, and adaptive control and plasticity of cognitive functions. He uses a combination of behavioral, neuroscientific, and computational-modelling approaches. In 2007, he was awarded a special LMU Research Professorship, and in 2008 he was made a member of the LMU Center for Advanced Studies.