International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Working Memory Capacity for Gesture-Command Associations in Gestural Interaction

Qi Gao, Zheng Ma, Quan Gu, Jiaofeng Li & Zaifeng Gao

To cite this article: Qi Gao, Zheng Ma, Quan Gu, Jiaofeng Li & Zaifeng Gao (2023) Working Memory Capacity for Gesture-Command Associations in Gestural Interaction, International Journal of Human–Computer Interaction, 39:15, 3045-3056, DOI: 10.1080/10447318.2022.2091213

To link to this article: https://doi.org/10.1080/10447318.2022.2091213

Published online: 17 Jul 2022.

INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
2023, VOL. 39, NO. 15, 3045–3056
https://doi.org/10.1080/10447318.2022.2091213

Working Memory Capacity for Gesture-Command Associations in Gestural Interaction

Qi Gao[a*], Zheng Ma[a*], Quan Gu[a,b], Jiaofeng Li[a], and Zaifeng Gao[a]

[a]Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China; [b]Advanced Display and Human Percpetion Research Center, Yongjiang Laboratory (Y-LAB), Ningbo, China

ABSTRACT
The study addressed the working memory capacity (WMC) of gesture-command associations in gestural interaction and investigated the impact of compatibility between users' mental models and the predefined gesture-command associations on WMC. Gestural interaction is a popular representative of natural interactions. Although gestural interaction intends to be natural, it has been criticized for not being so. One of the critical problems lies in learning and memorizing. WMC is a pivotal bottleneck that underlies learning and memorizing gesture-command associations, yet it remains unknown. Two standardized paradigms were used to estimate the WMC of gesture-command associations: change-detection task in Experiment 1, and span task in Experiment 2. Besides, we further examined the impact of compatibility on WMC. We found that two gesture-command associations can be retained in working memory under low-compatibility conditions, while three to five associations can be retained under high-compatibility conditions. The result implies that WMC of gesture-command associations is highly limited, while this cognitive limitation could be reduced by promoting the compatibility between users' mental model and the predefined gesture-command associations. Designers for gestural interactions may require users to memorize two to five gesture-command associations, considering different application scenarios.

1. Introduction

Gesture language application has been burgeoning in recent years. Traditional gesture communication is universally adopted in situations where auditory information is hard to reach, like sign language for the deaf, traffic control gestures, and gestures in military operations, etc. In addition to these special populations, ordinary people are also involved in gestural interactions, for example, thumbs up for approval or encouragement. Moreover, the proliferation of recognition technologies boosts gesture interaction as one of the most natural and popular human-machine interaction methods (GrandViewResearch, 2019; Tractica, 2018; Vuletic et al., 2019). It enables users to interact with a computer system through postures and movements (e.g., mid-air gestures by hands) (Koutsabasis & Vogiatzidakis, 2019). Gestural interaction, not only improves users' engagement (Bianchiberthouze et al., 2007) and emotional experience (van Beurden et al., 2012), but also avoids distractions (Graichen et al., 2019) and reduces users' cognitive load (Pantic et al., 2007).

However, the current so-called "natural user interface" via gestural interaction is not as natural as it intends to be (e.g., Norman & Nielsen, 2010). One of the critical problems lies in learning and memorizing. The learning of the gesture language encompasses the association process between the gestures and commands. The ease of technology implementation, instead of the naturalness and appropriateness of gesture-command associations, takes precedence in gesture design (Vuletic et al., 2019). Moreover, designers tend to present excessive gesture-command associations, leading to formidable challenges for users to memorize (Liang, 2013; Pang et al., 2014), which violates the "reduce memory load" golden rule of interface design (Cohen et al., 2018). Therefore, engineers and designers must understand the underlying cognitive bottleneck in learning and memorizing gesture-command associations, which remains largely unclear. Delving into this bottleneck will not only help to fathom human limitations on interacting with gestural interfaces, but also benefit sign language acquisition in specific domains, such as military training and deaf education.

2. Literature review

Working memory, a buffer temporarily maintaining and manipulating a limited set of information from perception and long-term memory (Baddeley & Hitch, 1974), is a fundamental cognitive component underlying learning and memorizing gesture-command associations. Specifically, users need to learn and memorize gesture-command associations presented in perception during the study phase of

CONTACT Zaifeng Gao zaifengg@zju.edu.cn Department of Psychology and Behavioral Science, Zhejiang University, Xixi Campus, Hangzhou, China
*Equal contribution.

gestural interaction (Anderson & Bischof, 2013), and later have to retrieve them from long-term memory after a period without interacting with the system (Harrison et al., 2013). This process innately requires the involvement of working memory storage, or short-term memory (see Baddeley, 2012 for a review). In the multi-component working memory model, the storage component is referred to as the traditional short-term memory term, while working memory includes extra active processing components, episodic buffer, and central executive (see Baddeley, 2012; Cowan, 2016 for reviews). Corroborating this view, many studies have revealed that working memory plays a pivotal role in a variety of high-level cognitive activities, including learning, reasoning, and comprehension (e.g., Just & Carpenter, 1992; Kyllonen & Christal, 1990; Maxwell et al., 2003).

A key characteristic of working memory is that its capacity is limited. Miller (1956) pioneered in revealing the working memory capacity (WMC), by suggesting that about seven chunks could be simultaneously retained in working memory (magic number seven). Although the seven-item notion had been popular in cognitive psychology and is still widely acknowledged in many human-machine interaction domains, the progress in cognitive science revealed that the WMC was overestimated (see Cowan, 2001 for a review), as participants used supplementary mnemonic strategies (such as rehearsal and chunking) during the memorization in previous studies of magic number seven. Researchers have now revealed converging evidence that working memory can retain at most three to four chunks (magic number four; Cowan, 2001; Luck & Vogel, 1997; Oberauer et al., 2016). Moreover, the WMC is modulated by the complexity of stimuli (e.g., Gao et al., 2009; Luria et al., 2010; Schurgin & Brady, 2019; van den Berg et al., 2014; but see Awh et al., 2007 for a different view). For instance, the information load contained in a chunk dramatically affects the WMC, by exhibiting a negative correlation between information load and WMC (e.g., Alvarez & Cavanagh, 2004; Gao et al., 2009; Luria et al., 2010). This fact motivates the explorations of WMC for different types of stimuli to better understand the cognitive bottleneck in our brain (e.g., Eng et al., 2005; He et al., 2019; Jiang et al., 2008; Lorenc et al., 2014; Meconi et al., 2014).

Studies have explored the WMC of human actions and sign languages, but that of gesture-command associations for gestural interaction, to the best of our knowledge, has not been investigated. Early studies on sign languages revealed similar results as the long-recognized seven-item notion by Miller. In this line of studies (e.g., Boutla et al., 2004; Wilson & Emmorey, 2006), participants watched gesture videos for the corresponding digits or letters, and recalled all the gestures from the memory array in the same order as they presented. Though the result showed that five to seven gestures can be retained in working memory, these findings could not be generalized to the gesture-command associations. This was because participants did not need to retain the associations between gesture and verbal stimuli, which were more complex than those simple gestures. In line with the "magic number four" view (Cowan, 2001), it has been revealed that three to four distinct human actions (Shen et al., 2014; Wood, 2007) or two associations between action and clothing color of actors (Ding et al., 2015) can be retained in working memory recently. Moreover, although some studies have manipulated the set size of gesture associations in a memorization task of gestural interactions (Nacenta et al., 2013; Pereira et al., 2015), these studies did not directly examine the WMC. It is worth noting that Jégo et al. (2013) investigated the memorization of gesture-related associations in a virtual environment and found that it was difficult for users to recall more than two abstract gesture-related associations. The procedure used by Jégo et al. (2013), however, was not a standardized paradigm for estimating the WMC, which may contaminate the estimation of WMC.

Meanwhile, we postulate that compatibility between users' mental model and the real gesture-command association may leverage the WMC of gesture-command associations. Mental models are users' representations of the physical world, explaining human behavior and internal mechanisms that enable them to understand, interpret, and predict the states of systems (Revell & Stanton, 2017). Previous studies on gestural interaction suggest that compatibility improves memorability (Dong et al., 2015; Nacenta et al., 2013) and reduces learning time (Pereira et al., 2015). As mental model combines knowledge stored in long-term memory, the associations in working memory can benefit from a better compatibility with the help of information from long-term memory. Therefore, we predicted that mental model compatibility could alleviate the limited WMC for gesture-command associations. Janczyk et al. (2019) provided preliminary evidence that compatibility influenced mental load in gestural interactions via using the NASA-TLX questionnaire. However, so far, no empirical research directly explored the influence of compatibility on WMC of gesture-command associations.

Determining the WMC of gesture-command associations is crucial for long-term gesture-command association learning. WMC predicts the efficiency of long-term memory encoding (Fukuda & Vogel, 2019), and items can be better encoded into long-term memory when working memory is not overloaded (Forsberg et al., 2021). A meta-analysis study also implied that WMC is positively associated with second language processing and proficiency outcomes (Linck et al., 2014). These studies propelled us into the investigation of WMC of gesture-command associations, in hope of providing designing guidelines for gesture interactions.

3. Methodology overview

In the current study, we estimated the WMC of gesture-command associations by focusing on a natural interaction scenario, where participants interact with a visual display using mid-air hand gestures. We adopted well-accepted paradigms in the working memory field to estimate the WMC during gestural interaction when memorizing a set of information simultaneously: change-detection task and simple span task (cf. Cowan, 2016; Luck & Vogel, 2013).

In both tasks, participants are presented a set of to-be-remembered items (i.e., memory array), then after a blank interval (usually at least 900 ms), they have to make a response. Critically, the number of to-be-memorized items is manipulated to estimate the WMC. In the change detection task (e.g., Luck & Vogel, 1997; Shen et al., 2014; Wood, 2007; see Rouder et al., 2011 for a review), an item is typically displayed as the probe, and participants decide whether the probe belongs to one of the items in the memory array. The change detection task has been broadly adopted in the recent two decades and is considered as an ideal paradigm in estimating WMC (Luck & Vogel, 2013), as the estimation is not affected by many other factors, such as proactive and retroactive interference, unbalanced decay, etc. (Cowan, 2001; Rouder et al., 2011). As to the simple span task (Hitch et al., 2001; Leather & Henry, 1994; Towse et al., 1998, 2000), participants have to recall the memory array one by one as much as possible. The span task can simulate a situation where a memory recall rather than simple recognition is required. It has been documented that both change-detection task and simple span task can provide a direct measure of the capacity for maintaining or refreshing (Bunting et al., 2006; Dahlin et al., 2008) the representations stored in WM, and produce reasonably similar variations in WMC (Broadway & Engle, 2010; Cowan, 2001). To have an accurate estimation of the WMC of gesture-command associations, we decided to first use the change detection task to have a strict examination of WMC, then adopt the simple span task which allows a real-application simulation with higher external ecological validity.

The second aim of the current study is to examine the impact of compatibility between users' mental model and the tested gesture-command associations on WMC, by testing the WMC under low and high compatibility conditions, separately. The compatibility (low versus high) was manipulated by the degree of match between gesture and commands, rated in a pilot study by a group of veterans on gesture-based interfaces (see Section 4.1.2). However, as the degree of compatibility is highly related to long-term memory which directly affects the estimation of WMC (Brady et al., 2016), we had to rule out the contribution of long-term memory (but see Cowan, 2001 for a different view) when considering a pure WMC. Therefore, we focused only on the low-compatibility condition (i.e., low match degree between users' mental model and the target gesture-command associations) for estimating the WMC of gesture-command associations, where the involvement of long-term memory is avoided as much as possible.

4. Experiment 1

Experiment 1 measured WMC using a change detection task. To mimic the fact that gesture-command associations are presented sequentially when learning and memorizing them in real practice, we presented the stimuli sequentially. Moreover, considering that working memory could hold two action-clothing color associations (Ding et al., 2015) and abstract gesture-associations (Jégo et al., 2013), and an accurate estimation of WMC requires a memory load condition being higher than the real WMC (cf. Rouder et al., 2011), we required the participants to memorize two or four gesture-command associations. In this case, the four-items setting avoids the underestimation of the WMC under the low-compatibility condition, and also avoids a floor effect of the memory performance due to an unreasonably high memory load.

4.1. Method

4.1.1. Participants

We used a Sequential Bayes Factor design, where Bayes factors (BFs) are computed repeatedly during data collection until the BF exceeds an a priori defined grade of evidence following the instruction from Schönbrodt et al. (2017). Accordingly, the minimum sample size was set to 20; the critical BF_{10}, which indicates the Bayes factor in favor of H1 over H0, for stopping the sequential sampling was set to 5 (resp. 1/5). We calculated the BFs via JASP (Version 0.14.1). As we have no strong prior evidence for any favorable prior, the scale parameter of r for the effect size prior was set to 1. This finally resulted in 20 participants (M_{age} = 20.7 ± 2.1 years; 11 males). All participants were undergraduate and graduate students from Zhejiang University. These participants had no mid-air gestural interface interaction experience. They were all right-handed, with normal color vision and normal or corrected-to-normal visual acuity, and were all native Chinese speakers. Payment or course credit was rewarded for participation. This study was exempted by the Research Ethics Board of Zhejiang University.

4.1.2. Apparatus and stimuli

We selected 24 common gestures and corresponding commands from gesture-interaction-related published literatures (Chen et al., 2018; Choi et al., 2012; Nacenta et al., 2013; Pereira et al., 2015; Wu & Wang, 2012). The 24 commands were "accept" (接受), "switch" (切换), "help" (帮助), "reject" (拒绝), "rotate" (旋转), "cut" (剪切), "move" (移动), "amplify" (放大), "wake up" (唤醒), "yes" (是的), "pause" (暂停), "next page" (下页), "delete" (删除), "minimum"(最小), "open" (打开), "mute" (静音), "select" (选中), "improve" (提高), "close" (关闭), "volume" (音量), "cancel" (取消), "box" (框选), "change channels" (换台), and "no" (不是). All the gestures were re-generated using Poser animation-creation software (Version 11; Smith Micro Software, Aliso Viejo, CA) for physical features homogeneity.

We conducted a pilot study to evaluate the degree of compatibility between the gesture and the command. We randomly combined 24 commands with 24 gestures into 576 command-gesture pairs. Each pair was scored on a seven-point scale by 20 veterans on gestural interaction, who had been researching gestural interaction for at least half a year. Twenty-four pairs of highest match scores ($M = 4.25$, $SD = 0.98$) were selected as the stimuli for the matched condition (high-compatibility); twenty-four pairs of lowest

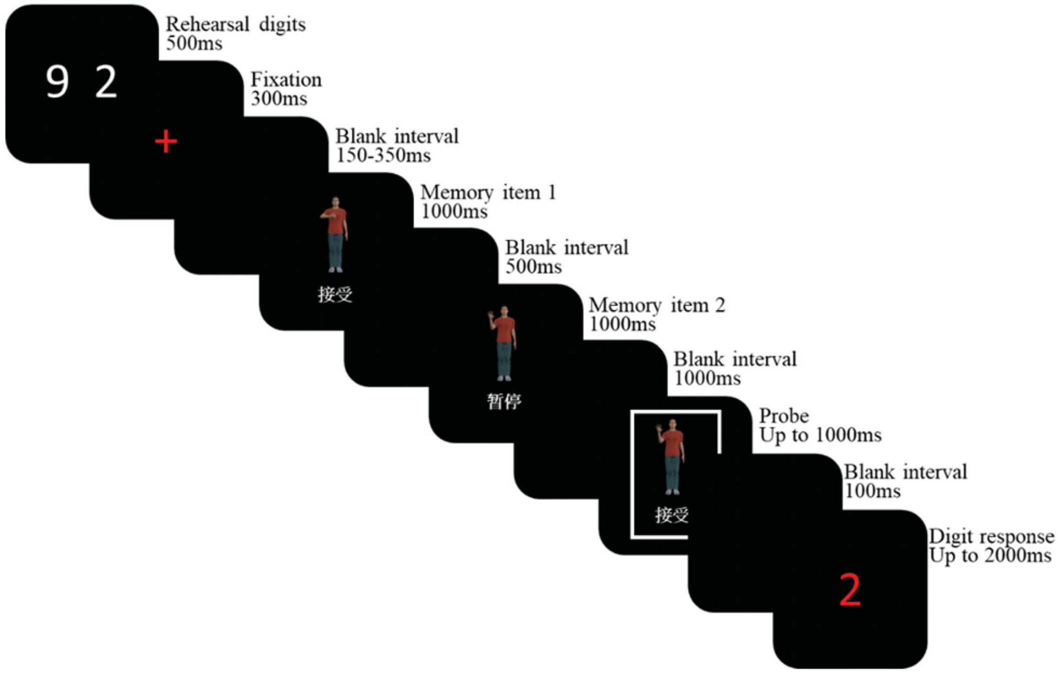

Figure 1. An example trial in Experiment 1 for set size two. In this trial, participants memorized two sequentially presented gesture-command associations, and then a probe pair was presented within a white box indicating a timely response. The commands were written in Chinese characters. The participant also conducted an articulatory suppression task to prevent verbal encoding of the gesture-command associations.

match scores (with all scores being 1) were selected as the stimuli for the unmatched condition (low-compatibility). See supplementary material for each gesture-command association.

The whole experiment was carried out in a dark and sound-shielded room from a viewing distance of 60 cm. The stimuli were presented with MATLAB and Psychtoolbox against a black (0, 0, 0, in RGB value) background on a l7-inch cathode ray tube (CRT) monitor with a resolution of 1024×768 pixels, at a refresh rate of 60-Hz. Each gesture consisted of 30 distinct frames (refresh rate of 30 Hz), leading to a one-second display. The gesture subtended approximately at a 4.11°(height) × 3.10°(width) visual angle at the screen center, and the command was conveyed by Chinese characters in white color, which was presented below the gesture and subtended at a 1.59° (level) × 0.80°(width) visual angle.

4.1.3. Procedure

Each trial began with the presentation of two digits at the screen center (see Figure 1) for 500 ms. Participants were required to repeat the two digits (e.g., by stating "nine" and "two," not "ninety-two") out loud throughout the trial at their own pace. This concurrent articulatory suppression task was used to prevent participants from verbally rehearsing the gesture (Curby & Gauthier, 2007; Shen et al., 2014). Next, a red fixation appeared for 300 ms to inform the participants of the upcoming memory task. After a blank interval of 150–350 ms, the command-gesture pairs were presented, with each gesture accompanied by a command. Each command-gesture association lasted for 1000 ms, followed by a 500-ms blank interval. After the memory array, a 1 s blank interval was presented, followed by a probe in a white box (8.10° × 7.30° visual angle), asking the participant to make a judgment about whether the probed pair had appeared in the memory array within 3 s. Participants pressed "*J*" on the keyboard for yes, and "*K*" for no. After the response, or if no response was made within 3 s, a digit judgment task was presented after a 100 ms delay to ensure the validity of the articulatory suppression task. Participants had 2 s to determine whether the red digit was one of the previously rehearsed digits. The probed gesture-command pair and digit were changed independently, both with the appearance in the memory array for half of the trials. When the gesture-command pairs changed, only a gesture or command changed into another one from the memory array. There was a 2000 ms blank interval between trials.

Experiment 1 adopted a 2 (compatibility: matched, unmatched) × 2 (set size: two, four) within-subjects design. The two set sizes were randomly displayed. The two compatibility conditions were tested in different blocks, the order of which was counterbalanced between participants. There were 16 trials within each set size condition, resulting in 32 trials in each block and 64 trials in total. There was a 3-min break between the blocks. The whole experiment lasted for approximately 15 min.

4.1.4. Analysis

Cowan's formula was adopted to estimate the WMC (Cowan, 2001): $K = N \times (H - F)$ where K was the WMC, N was the number of to-be-memorized stimuli, H was the hit rate, *and* F was the false alarm rate. We calculated K estimates for each set size for each participant. A two-way repeated-measures analysis of variance (ANOVA) was conducted on K estimates by taking both compatibility and set size as within-subject factors via JASP.

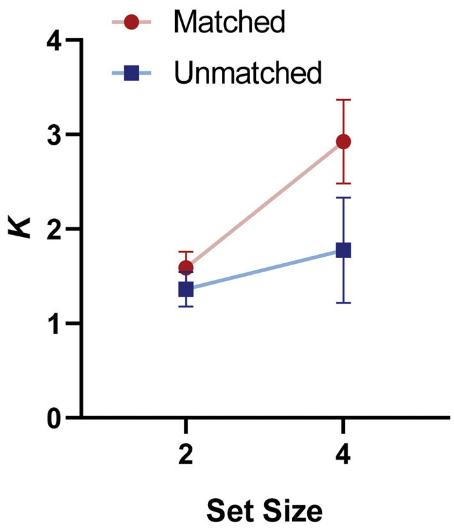

Figure 2. The working memory capacities under different conditions. Error bars indicate 95% confidence interval (CI).

To estimate the WMC of gesture-command associations, we took the maximum K (K_{max}) between the two set size conditions as one's WMC (e.g., Curby & Gauthier, 2007; Shen et al., 2014), and the averaged value of all participants' K_{max} as the final WMC of gesture-command associations.

We also calculated the BF (Rouder et al., 2009, 2012) to examine the ratio of the alternative hypothesis (H1) relative to the null hypothesis (H0), which was done via JASP 0.14.1 (JASP Team, 2020). As to the computation of the BFs for the interaction, we compared a linear model containing only one main effect with a null model (including subject) to compute the BFs for the main effects, and compared a full model (numerator; including the two main effects and the interaction) with a reduced model (denominator), wherein the interaction effect of interest was not included. A BF_{10} of 3 indicates substantial evidence for the selection of H1 over H0, whereas a BF_{10} of 10 is considered to provide strong evidence for the selection of H1 over H0 (cf. Jeffreys, 1998).

4.2. Result

The two-way ANOVA (see Figure 2) revealed a significant main effect of compatibility on K, F (1, 19) = 13.70, $p < 0.01$, $\eta_p^2 = 0.42$, $BF_{10} = 155.79$, indicating that WMC under the matched condition was higher than that under the unmatched condition. The main effect of set size was significant, F (1,19) = 22.82, $p < 0.001$, $\eta_p^2 = 0.55$, $BF_{10} = 5649.58$, suggesting that memory load modulated the number of associations in working memory. The interaction between set size and compatibility was significant, F (1, 19) = 11.16, $p < 0.01$, $\eta_p^2 = 0.37$, $BF_{10} = 6.65$. A simple effect analysis revealed that K was both significantly higher under the matched condition than under the unmatched condition when memorizing two associations, t (19) = 2.44, $p < 0.05$, Cohen's $d = 0.27$ and four associations, t (19) = 3.66, $p < 0.01$, Cohen's $d = 1.38$.

We took K_{max} as the estimated WMC of each participant. According to Figure 3, participants could memorize about 2 gesture-command associations in the unmatched condition

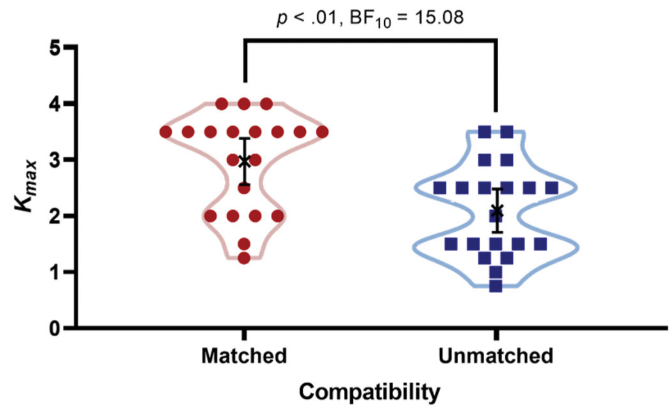

Figure 3. The WMC for each participant under the two compatibility conditions. The crossing demonstrates the mean value under each condition, with error bars indicating 95% CI.

($M = 2.09$, $SE = 0.18$), while around 3 gesture-command associations in the matched condition ($M = 2.96$, $SE = 0.20$). As the data met the normality assumption (Shapiro-Wilk, $W = 0.93$, $p = 0.13$), a further paired t-test was conducted. It revealed that the average K_{max} was significantly higher under matched condition than under unmatched condition, t (19) = 3.43, $p < 0.01$, Cohen's $d = 0.77$, $BF_{10} = 15.08$.

4.3. Discussion

The averaged WMC indicated that at most two gesture-command associations could be retained in working memory under the unmatched condition. Meanwhile, three gesture-command associations can be retained under the matched condition, which, to some extent, was underestimated as three participants' performance was at ceiling level even at the high load condition. More associations were memorized under the match condition than the unmatched condition, suggesting that compatibility modulates the memory of gesture-command associations.

Although the change detection task has been popularized in estimating the WMC in the working memory field (Logie et al., 2020), it is quite different from the real-world common gestural interaction situation. In reality, a common situation is that users imitate each gesture when the to-be-memorized gesture-command associations are presented; moreover, they have to recall and perform the gestures to fulfill a command. To reach a conclusion with high ecological validity, we further used the simple span task to measure WMC, in which participants have to imitate the gestures as the memory array was presented, then recall and reproduce all the memorized associations in the test phase.

5. Experiment 2

In Experiment 2, the simple span task was adopted for a higher ecological validity. As WMC is a stable measurement for the short-term storage ability, and both the change detection task and the simple span task bear a resemblance for the WMC estimation in the previous studies (Cowan, 2016), we expected that a similar WMC result between Experiments 1 and 2.

In the simple span task, participants started the memory task from a specific memory load (i.e., memorizing four items). The memory load is then adaptively adjusted according to the memory performance on the very memory load, by increasing or decreasing one memory load (Towse et al., 1998, 2000). Considering that some participants can even memorize more than three associations in the unmatched condition in Experiment 1 (see Figure 3), and that conducting gestures benefits memory for the gestures (Zhao et al., 2016), we hence set the beginning memory load as four gesture-command associations in Experiment 2.

5.1. Method

5.1.1. Participants, apparatus, and stimuli

Twenty new students (M_{age} = 22.1 ± 2.1 years; 11 males) participated in Experiment 2, and Sequential Bayesian Factor design was adopted. The stimuli were presented on a 34-inch LCD monitor with a resolution of 1366 × 768 pixels. Subjects sat in a luminance-controlled room with a view distance of 120 cm to gain a proper visual angle. A camera (Kinect for Xbox One, Microsoft) was used to record all the gestures performed by the participants. The other aspects were the same as Experiment 1.

5.1.2. Procedure

Figure 4 presents an example trial and the procedure for Experiment 2. Each trial started with a 1000-ms fixation, followed by a 450–1050 ms blank interval. Then, the participants started to learn and memorize gesture-command associations. To give participants enough time to learn and reproduce the gestures in the learning phase, the exposure time of each memorized association was increased to 3000 ms. When a gesture-command association was presented in the learning phase, participants were asked to repeat the same gesture twice within 3000 ms. There was a 1500-ms blank interval between each item as well as between the last memorized item and the testing phase. In the test phase, all commands that had been presented in the memory array were presented in random order with each command lasting 6 s. Participants had to perform the corresponding gesture of the command twice. The inter-trial interval was randomized within 4500–6000 ms. The articulatory suppression task was removed because we intended to simulate the real-life scenario to the greatest extent.

Participants started the task by memorizing four gesture-command associations. There were three trials for each memory load condition. If participants successfully reproduced all the associations in one trial, they succeeded in that trial. When participants succeeded in two or three trials of the same memory load, they were considered to pass the tested memory load and entered a higher memory load condition. Otherwise, a lower memory load was tested. A trained experimenter scored the performance in real-time. Our pilot study found that memorizing five actions was too difficult to fulfill under the unmatched condition; hence, we set the upper boundary of the testing as five gesture-command associations. This decision may underestimate the WMC under highly compatible conditions. However, since the exact estimation of WMC under highly compatible conditions has a direct relation with the compatible degree of mental model between the tested gestures and the users, it is difficult to reach a stable and accurate estimation like that under the low compatible condition.

Experiment 2 was divided into matched and unmatched blocks with a 10-min break in between, and the order was counterbalanced. Before the formal experiment, four practice trials were conducted, where participants were asked to memorize two gesture-command pairs. The entire experiment lasted for 30 min and was videotaped.

5.1.3. Analysis

The WMC is estimated using the span score (e.g., Towse et al., 1998). A participant's span is initially set as the highest memory load that they can pass. For example, if a participant passed set size four and failed set size five, the span score would be 4; if a participant failed set size four and passed set size three, the span score would be 3. The initial span score is further adjusted to reach a final estimation: participants earn 0.5 points for one correct trial for the set size they fail. For example, if a participant was correct in all three trials when memorizing four pairs but only one trial when memorizing five pairs then the final span score would be 4.5 (4 + 0.5). A paired *t*-test was conducted on the span scores under matched and unmatched conditions via JASP 0.9.0.1.

The BF was calculated to examine the ratio of the alternative hypothesis (H1) relative to the null hypothesis (H0). A BF_{10} of 3 indicates substantial evidence for the selection of H1 over H0, whereas a BF_{10} of 10 is considered to provide strong evidence for the selection of H1 over H0.

5.2. Result

According to Figure 5, participants could memorize about 2–3 gesture-command associations in the unmatched condition ($M = 2.23$, $SE = 0.07$), while 4–5 gesture-command associations in the matched condition ($M = 4.10$, $SE = 0.14$). Though the data did not meet the normality assumption (Shapiro–Wilk, $W = 0.81$, $p < 0.01$), a paired *t*-test revealed that the span score was significantly higher under matched condition than unmatched condition for both student test, $t\ (19) = 13.40$, $p < 0.01$, Cohen's $d = 3.00$, Wilcoxon signed-rank test, $T = 210.00$, $p < 0.01$, $rb = 1.00$. Bayesian test indicates similar results for student test, $BF_{10} = 2.21\text{e} + 8$, and Wilcoxon signed-rank test, $BF_{10} = 562.32$.

5.3. Discussion

Experiment 2 used a different paradigm to estimate the WMC, yet reached a similar conclusion as Experiment 1: about two gesture-command associations could be retained

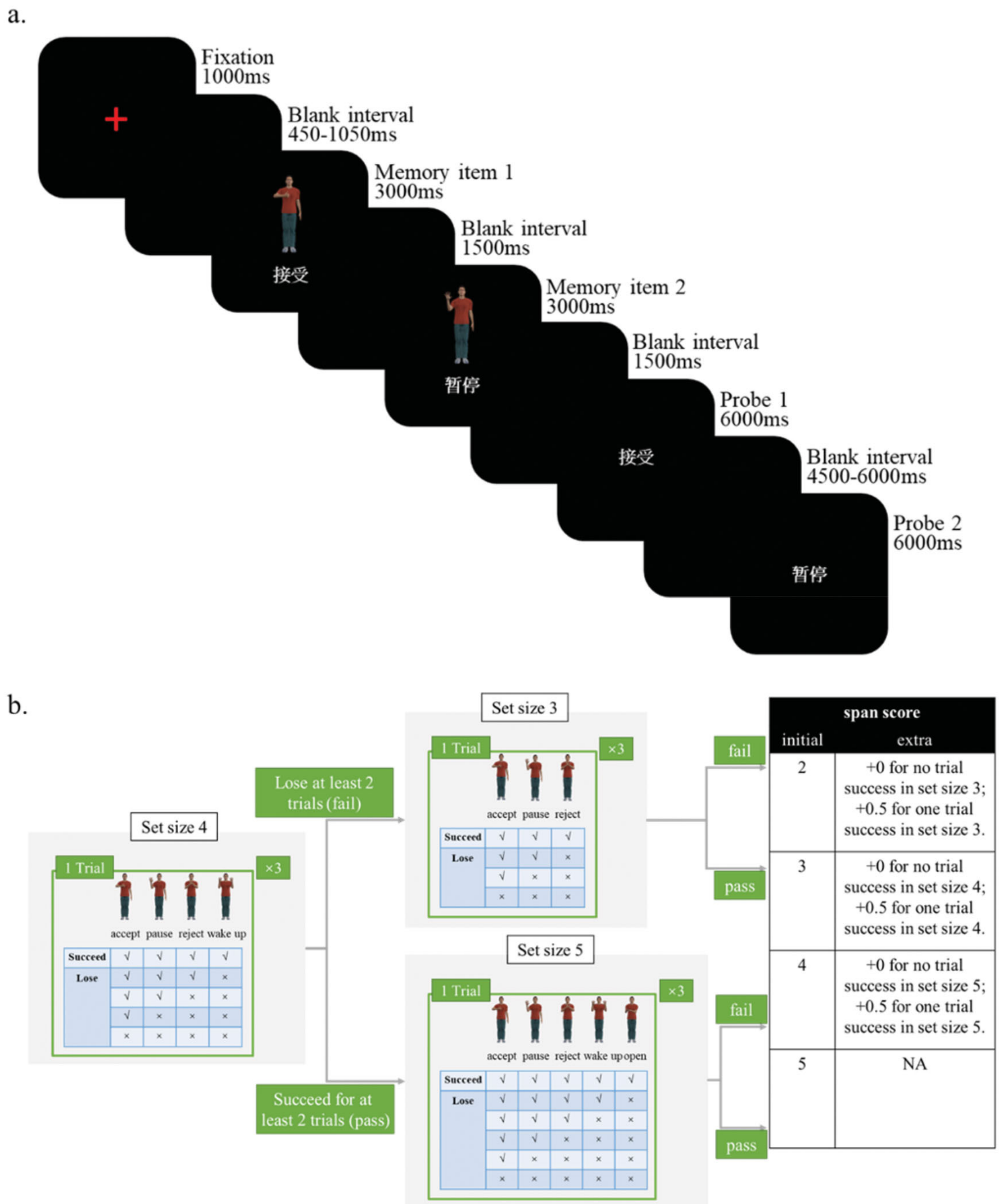

Figure 4. An example trial and procedure for Experiment 2 under set size two condition. (a) An example trial in Experiment 2 for set size two. In this trial, participants memorized two gesture-command associations and performed the gestures simultaneously as the memory array was displayed. When the probe command appeared, they had to reproduce the corresponding gesture; (b) The procedure of Experiment 2. Each participant started with set size four, which contained 3 trials. They were considered success in a trial only if they correctly reproduce all four gestures. They passed the certain set size to the next level only if they succeeded for at least 2 trials. The span score was the summation of the largest set size they passed and 0.5 × trial number they succeeded under the set size they failed.

in working memory under the unmatched condition, and compatibility significantly improved participants' memory performance.

The WMC under the matched condition in Experiment 2 was much higher than that in Experiment 1. At least two factors contributed to this phenomenon. First, Experiment 2 included a memory load of five associations. Although it remains open as to the caveat of underestimating WMC, this setting significantly reduced this caveat relative to Experiment 1. Second, participants performed gestures during the learning phase; this operation has been revealed to boost memory performance, which is called the enactment effect (Hornstein & Mulligan, 2004). However, it seems that the enactment effect does not dramatically affect the memory performance when the mental model between the designer and the user is incompatible. Though the lower bound was increased as no span score was below 2 in Experiment 2, the average WMC is largely comparable between Experiment 1 ($K = 2.09$) and Experiment 2 ($K = 2.23$) under the unmatched condition (Mann–Whitney test, $W = 178.00$, $p = 0.55$, $rb = -0.11$, Bayesian Mann–Whitney, $BF_{10} = 0.37$). This point further pinpoints the necessity of considering users' models while

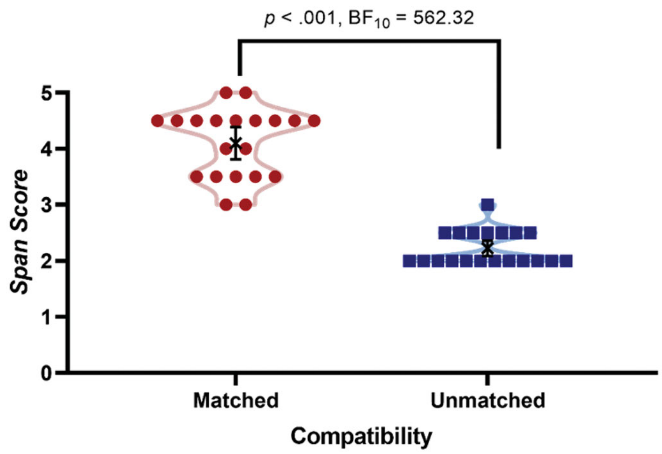

Figure 5. Span scores for each participant under the two compatibility conditions. The crossing demonstrates the mean value under each condition, with error bars indicating 95% CI.

designing gestural interaction for maximizing the enactment benefit.

6. General discussion

Multiple studies have shown that reduced workload is linked with increased gesture acceptance (e.g., Graichen et al., 2019) and system security (e.g., Gregoriades et al., 2010). At the current stage, one of the key problems that hinders the effectiveness and popularization of gestural interaction is the learnability and memorability of the gesture-command associations. However, the role of working memory limitation has not been well acknowledged. The current study investigated the WMC of gesture-command associations in two experiments. Participants exhibited clear limitations in WMC in both experiments: approximately a maximum of two gesture-command associations can be retained in working memory under a low-compatibility condition, while three to five associations under a high-compatibility condition.

The current study, for the first time, provided a pure estimation of the WMC of gesture-command associations underlying the gestural interaction. In two different paradigms estimating WMC, we consistently found that a pure WMC of gesture-command associations is around two gesture-command associations, both under a strictly-controlled experiment environment (i.e., the change detection task) and under a higher ecological validity environment (i.e., the simple span task). The WMC of two is even held when participants had to perform the gestures during the learning phase (Experiment 2). Although the WMC of the current study is apparently congruent with Jégo et al. (2013), which estimated a different situation of gesture-related associations, the current conclusion is not comparable to Jégo et al. (2013). Specifically, as the gestures in Jégo et al. (2013) were user-elicited and hence essentially belonged to a matched condition, the WMC in that study, theoretically, should be higher (see the matched condition in the current study). We argue that the low WMC in Jégo et al. (2013) was due to the high similarity between the gestures because participants were required to memorize a set of gestures corresponding to different boxes, and the same command "open" was obeyed. On the other hand, we had to point out that the current study focused on young adults. Since WMC is significantly lower for children and old adults (e.g., Wingfield et al., 1988), future studies should elucidate the WMC of gesture-command associations in other age groups.

The results under highly compatible conditions also contribute to our understanding of the cognitive bottleneck underlying gestural interaction. First, the current study closed a critical gap for the user elicitation method advantage, by examining the impact of compatibility of the mental model on the WMC of gesture-command associations in a strictly-controlled task. User elicitation method, a method which requests users to propose gestures for given commands (Wobbrock et al., 2009), is prevailing in gestural interaction design for improving user preference, memorability, and comfort (e.g., Chen et al., 2018; Morris et al., 2010; Nacenta et al., 2013; Vatavu & Zaiti, 2014). The benefit is believed to be driven by users' mental models (Koutsabasis & Vogiatzidakis, 2019; Wobbrock et al., 2009), and ample studies emphasized the importance of compatibility of the mental model of the gesture on the performance of users. However, researchers either did not directly verify the influence of mental model on user's performance (Nacenta et al., 2013; Pereira et al., 2015) or tested it indirectly (Janczyk et al., 2019). Our study showed clear-cut evidence that compatibility of mental load played a pivotal role in the user elicitation method, and demonstrated that WMC may be the bridge connecting the compatibility and the gesture interaction improvement (see Vogiatzidakis & Koutsabasis, 2018 for a review). Moreover, our study implied that the enactment benefit emerged only under a high compatibility condition, further pinpointing the importance of considering user's mental model in the gesture design.

Second, although we could not have a precise estimation of WMC under high compatible conditions as it has an intimate relationship with the degree of compatibility, our study found that most of the participants could not memorize beyond five gestures-command associations in working memory simultaneously (17 and 18 out of 20 participants, respectively, in Experiment 1 and 2). Therefore, it seems that although the mental model from long-term memory could alleviate the limitation of working memory, a potential upper limitation still exists, which conservatively disproves the unlimited-capacity view from Brady et al. (2016).

The current findings have direct implications for the gestural design. For the designers, we suggest that they should require users to learn and memorize no more than two gesture-command associations under new interaction circumstances for conservative consideration, for example, (1) gesture acquisition for novice users for a specific interface or gesture interaction vocabulary, and (2) gesture acquisition in the military, such as navy army or police, as the gestures barely encompass elements that can be linked with long-term memory. On the other hand, an approximate upper limitation of five gesture-commands could be retained by the users under familiar interaction vocabulary circumstances, for instance, (1) sign language learning as some signs can be ideographic or metaphoric, and (2) transfer learning within a device ecosystem as the gestures can be homogeneous and consecutive.

In broad terms, the WMC limitation in our study is consistent with previous cognitive research of binding problems in working memory. For bindings of static visual features (e.g., a red square involves binding between color and shape), memory deficit was observed across multiple studies compared with three to four capacity for single features (e.g., Hardman & Cowan, 2015). The same is true for bindings between human actions and other features: The WMC is only one to two for bindings between human actions and visual features (Ding et al., 2015) but three to four for actions alone (Shen et al., 2014). However, binding is feature integration within object, association in our study refers to constructing relationships across objects (i.e., across gestures and commands). Notwithstanding the difference, we gained similar results for association memory deficits. Five to seven single gestures could be retained simultaneously (Boutla et al., 2004; Wilson & Emmorey, 2006), in contrast with two gesture-command associations in our study. This memory deficit adds new evidence to support the view that (1) the category of memorized stimuli affects WMC, where mental complexity reduces WMC (e.g., Alvarez & Cavanagh, 2004; Gao et al., 2009; Luria et al., 2010); (2) maintaining bindings or associations is an attention-consuming process (e.g., Gao et al., 2017; Shen et al., 2015).

Though we provide concrete WMC under different compatibility conditions, a thorough application requires additional compatibility evaluations for specific gesture sets. The compatibility level was determined by the veteran's evaluation in our study, which is a rather coarse method. Many factors should be taken into consideration for evaluating compatibility (Proctor & Vu, 2016), for example, affective compatibility (Norman, 2004), ideomotor compatibility (Greenwald, 1972), affordance compatibility (Gibson, 1978), and cultural background (Wu et al., 2020). Therefore, additional tools for evaluation should be further developed and applied to the specific gesture-command sets when applying the WMC we provided here. Additionally, future studies need to uncover the underlying mechanisms of encoding and maintenance of gesture-command associations. There have been attempts to generate gestures through theoretical models of human performance (e.g., Cao & Zhai, 2007; Leiva et al., 2020), which would be facilitated by a comprehensive understanding of the cognitive processing of gesture-command associations. Debate continues about the form of mental representations of gestures: visuospatial (Emmorey et al., 2017) or propositional (Wagner et al., 2004). However, little attention has been paid to the association between gesture and commands.

7. Conclusion

In summary, our findings suggest that rather than bearing the traditional seven items in mind, designers should pay attention to the two to five chunks of WMC we obtained for gesture-command associations. This cognitive limitation might be extended to other gesture communication domains as well in addition to the current gesture interface, which needs to be verified in other gesture communication contexts.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by Science and Technology Innovation 2030 Program [2018AAA0101605], Fundamental Research Funds for the Central Universities National, Natural Science Foundation of China Grants [31771202], and Key Program of Natural Science Foundation of Zhejiang Province [LZ20C090001].

ORCID

Zaifeng Gao http://orcid.org/0000-0001-9727-8524

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Anderson, F., & Bischof, W. F. (2013). *Learning and performance with gesture guides* [Paper presentation]. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological science*, *18*(7), 622–628.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1974). *Working memory: Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.

Bianchiberthouze, N., Kim, W. W., & Patel, D. (2007). *Does body movement engage you more in digital game play? And why?* [Paper presentation]. Paper presented at the Affective Computing and Intelligent Interaction.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience*, *7*(9), 997–1002. https://doi.org/10.1038/nn1298

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–570. https://doi.org/10.3758/BRM.42.2.563

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

Cao, X., & Zhai, S. (2007). *Modeling human performance of pen stroke gestures* [Paper presentation]. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA. https://doi.org/10.1145/1240624.1240850

Chen, Z., Ma, X., Peng, Z., Zhou, Y., Yao, M., Ma, Z., Wang, C., Gao, Z., & Shen, M. (2018). User-defined gestures for gestural interaction: Extending from hands to other body parts. *International Journal of Human–Computer Interaction*, *34*(3), 238–250. https://doi.org/10.1080/10447318.2017.1342943

Choi, E., Kwon, S., Lee, D., Lee, H., Chung, M. K. (2012). Can user-derived gesture be considered as the best gesture for a command?: Focusing on the commands for smart home system. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1253–1257. https://doi.org/10.1177/1071181312561222

Cohen, M. S., Elmqvist, N., Plaisant, C., Jacobs, S. M., & Shneiderman, B. (2018). *Designing the user interface: Strategies for effective human-computer interaction*. Pearson.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/s0140525x01003922

Cowan, N. (2016). *Working memory capacity* (Classic edition). Routledge.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review*, *14*(4), 620–628. https://doi.org/10.3758/bf03196811

Dahlin, E., Neely, A. S., Larsson, A., Bäckman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510–1512. https://doi.org/10.1126/science.1155466

Ding, X., Zhao, Y., Wu, F., Lu, X., Gao, Z., & Shen, M. (2015). Binding biological motion and visual features in working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(3), 850–865. https://doi.org/10.1037/xhp0000061

Dong, H., Danesh, A., Figueroa, N., & Saddik, A. E. (2015). An elicitation study on gesture preferences and memorability toward a practical hand-gesture vocabulary for smart televisions. *IEEE Access*, *3*, 543–555. https://doi.org/10.1109/ACCESS.2015.2432679

Emmorey, K., Giezen, M. R., Petrich, J. A. F., Spurgeon, E., & O'Grady Farnady, L. (2017). The relation between working memory and language comprehension in signers and speakers. *Acta Psychologica*, *177*, 69–77. https://doi.org/10.1016/j.actpsy.2017.04.014

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*(6), 1127–1133. https://doi.org/10.3758/bf03206454

Forsberg, A., Guitard, D., & Cowan, N. (2021). Working memory limits severely constrain long-term retention. *Psychonomic Bulletin & Review*, *28*(2), 537–547. https://doi.org/10.3758/s13423-020-01847-z

Fukuda, K., & Vogel, E. K. (2019). Visual short-term memory capacity predicts the "bandwidth" of visual long-term memory encoding. *Memory & Cognition*, *47*(8), 1481–1497. https://doi.org/10.3758/s13421-019-00954-0

Gao, Z., Li, J., Liang, J., Chen, H., Yin, J., & Shen, M. (2009). Storing fine detailed information in visual working memory—Evidence from event-related potentials. *Journal of Vision*, *9*(7), 17–17. https://doi.org/10.1167/9.7.17

Gao, Z., Wu, F., Qiu, F., He, K., Yang, Y., & Shen, M. (2017). Bindings in working memory: The role of object-based attention. *Attention, Perception & Psychophysics*, *79*(2), 533–552. https://doi.org/10.3758/s13414-016-1227-z

Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, *11*(3), 227–235. https://doi.org/10.2307/1574154

Graichen, L., Graichen, M., & Krems, J. F. (2019). Evaluation of gesture-based in-vehicle interaction: User experience and the potential to reduce driver distraction. *Human Factors*, *61*(5), 774–792. https://doi.org/10.1177/0018720818824253

GrandViewResearch. (2019). *Gesture recognition market size, share & trends analysis report by technology (touch-based, touchless), by industry (automotive, consumer electronics, healthcare), and segment forecasts* (GVR-2-68038-019-4). https://www.grandviewresearch.com/industry-analysis/gesture-recognition-market.

Greenwald, A. G. (1972). On doing two things at once: Time sharing as a function of ideomotor compatibility. *Journal of Experimental Psychology*, *94*(1), 52–57. https://doi.org/10.1037/h0032762

Gregoriades, A., Sutcliffe, A., Papageorgiou, G., & Louvieris, P. (2010). Human-centered safety analysis of prospective road designs. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *40*(2), 236–250. https://doi.org/10.1109/TSMCA.2009.2037011

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(2), 325–347. https://doi.org/10.1037/xlm0000031

Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: Literature review and rationale for a new usability model. *Journal of Interaction Science*, *1*(1), 1. https://doi.org/10.1186/2194-0827-1-1

He, J., Guo, D., Zhai, S., Shen, M., & Gao, Z. (2019). Development of social working memory in preschoolers and its relation to theory of mind. *Child Development*, *90*(4), 1319–1332. https://doi.org/10.1111/cdev.13025

Hitch, G., Towse, J., & Hutton, U. (2001). What limits working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*(2), 184–198. https://doi.org/10.1037/0096-3445.130.2.184

Hornstein, S. L., & Mulligan, N. W. (2004). Memory for actions: Enactment and source memory. *Psychonomic Bulletin & Review*, *11*(2), 367–372. https://doi.org/10.3758/bf03196584

Janczyk, M., Xiong, A., & Proctor, R. W. (2019). Stimulus-response and response-effect compatibility with touchless gestures and moving action effects. *Human Factors*, *61*(8), 1297–1314. https://doi.org/10.1177/0018720819831814

JASP Team. (2020). *JASP (Version 0.14.1)* [Computer software]. https://jasp-stats.org/2020/12/17/jasp-0-14-1-minor-updates/

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Clarendon Press; Oxford University Press.

Jégo, J., Paljic, A., & Fuchs, P. (2013). *User-defined gestural interaction: A study on gesture memorization* [Paper presentation]. Paper presented at the 2013 IEEE Symposium on 3D User Interfaces (3DUI). https://doi.org/10.1109/3DUI.2013.6550189

Jiang, Y. V., Shim, W. M., & Makovski, T. (2008). Visual working memory for line orientations and face identities. *Perception & Psychophysics*, *70*(8), 1581–1591. https://doi.org/10.3758/PP.70.8.1581

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149. https://doi.org/10.1037/0033-295x.99.1.122

Koutsabasis, P., & Vogiatzidakis, P. (2019). Empirical research in mid-air interaction: A systematic review. *International Journal of Human–Computer Interaction*, *35*(18), 1747–1768. https://doi.org/10.1080/10447318.2019.1572352

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *! Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology*, *58*(1), 88–111. https://doi.org/10.1006/jecp.1994.1027

Leiva, L. A., Vatavu, R.-D., Martín-Albo, D., & Plamondon, R. (2020). Omnis Praedictio: Estimating the full spectrum of human performance with stroke gestures. *International Journal of Human-Computer Studies*, *142*, 102466. https://doi.org/10.1016/j.ijhcs.2020.102466

Liang, S.-F M. (2013). Control with hand gestures in home environment: A review. In Y.-K. Lin (Ed.), *Proceedings of the Institute of Industrial Engineers Asian Conference 2013* (pp. 837–843). Springer Singapore.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Logie, R., Camos, V., & Cowan, N. (2020). Working memory: The state of the science. Oxford University Press.

Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 1975–1984. https://doi.org/10.3758/s13414-014-0653-z

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Luria, R., Sessa, P., Gotler, A., Jolicoeur, P., & Dell'Acqua, R. (2010). Visual short-term memory capacity for simple and complex objects. *Journal of Cognitive Neuroscience*, *22*(3), 496–512. https://doi.org/10.1162/jocn.2009.21214

Maxwell, J. P., Masters, R. S. W., & Eves, F. F. (2003). The role of working memory in motor learning and performance. *Consciousness and Cognition*, *12*(3), 376–402. https://doi.org/10.1016/S1053-8100(03)00005-9

Meconi, F., Luria, R., & Sessa, P. (2014). Individual differences in anxiety predict neural measures of visual working memory for untrustworthy faces. *Social Cognitive and Affective Neuroscience*, *9*(12), 1872–1879. https://doi.org/10.1093/scan/nst189

Miller, G. A. (1956). The magic number seven, plus minus two. *Psychological Review*, *63*(2), 81–96. https://doi.org/10.1037/h0043158

Morris, M. R., Wobbrock, J. O., Wilson, A. D. (2010). *Understanding users' preferences for surface gestures* [Paper presentation]. Paper presented at the Proceedings of Graphics Interface 2010, Ottawa, Ontario, Canada.

Nacenta, M. A., Kamber, Y., Qiang, Y., & Kristensson, P. O. (2013). *Memorability of pre-designed and user-defined gesture sets* [Paper presentation]. Paper presented at the Human Factors in Computing Systems. https://doi.org/10.1145/2470654.2466142

Norman, D. A. (2004). *Emotional design: Why we love (or hate) everyday things*. Basic Books.

Norman, D. A., & Nielsen, J. (2010). Gestural interfaces: A step backward in usability. *Interactions*, *17*(5), 46–49. https://doi.org/10.1145/1836216.1836228

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10.1037/bul0000046

Pang, X., Guo, R., Yao, N., Yu, J., Yu, S., Wang, C., & Gao, Z. (2014). Human factor studies on gestural interaction: Past, present, and future. *Chinese Journal of Applied Psychology*, *20*(3), 243–251. https://doi.org/10.3969/j.issn.1006-6020.2014.03.007

Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). *Human computing and machine understanding of human behavior: A survey* [Paper presentation]. Paper presented at the International Joint Conference on Artificial Intelligence.

Pereira, A., Wachs, J. P., Park, K., & Rempel, D. (2015). A user-developed 3-D hand gesture set for human-computer interaction. *Human Factors*, *57*(4), 607–621. https://doi.org/10.1177/0018720814559307

Proctor, R. W., & Vu, K.-P L. (2016). Principles for designing interfaces compatible with human information processing. *International Journal of Human-Computer Interaction*, *32*(1), 2–22. https://doi.org/10.1080/10447318.2016.1105009

Revell, K. M., & Stanton, N. A. (2017). *Mental models: Design of user interaction and interfaces for domestic energy systems*. CRC Press.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–330. https://doi.org/10.3758/s13423-011-0055-3

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Schönbrodt, F. D., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*(2), 322–339. https://doi.org/10.1037/met0000061

Schurgin, M. W., & Brady, T. F. (2019). When "capacity" changes with set size: Ensemble representations support the detection of across-category changes in visual working memory. *Journal of Vision*, *19*(5), 3–3. https://doi.org/10.1167/19.5.3

Shen, M., Gao, Z., Ding, X., Zhou, B., & Huang, X. (2014). Holding biological motion information in working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(4), 1332–1345. https://doi.org/10.1037/a0036839

Shen, M., Huang, X., & Gao, Z. (2015). Object-based attention underlies the rehearsal of feature binding in visual working memory. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(2), 479–493. https://doi.org/10.1037/xhp0000018

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language*, *39*(2), 195–217. https://doi.org/10.1006/jmla.1998.2574

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, *28*(3), 341–348. https://doi.org/10.3758/bf03198549

Tractica. (2018). *Gesture control interfaces – Mobile devices, gaming, smart buildings and smart appliances, automotive, healthcare, augmented and virtual reality, sign language and computing: Global market analysis and forecasts*. Tractica.

van Beurden, M., Ijsselsteijn, W., & De Kort, Y. (2012). *User experience of gesture based interfaces: A comparison with traditional interaction methods on pragmatic and hedonic qualities* [Paper presentation]. Paper presented at the GW. Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, 2011.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149. https://doi.org/10.1037/a0035234

Vatavu, R.-D., & Zaiti, I.-A. (2014). *Leap gestures for TV:* Insights from an elicitation study [Paper presentation]. Paper presented at the Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video, Newcastle Upon Tyne, United Kingdom. https://doi.org/10.1145/2602299.2602316

Vogiatzidakis, P., & Koutsabasis, P. (2018). Gesture elicitation studies for mid-air interaction: A review. *Multimodal Technologies and Interaction*, *2*(4), 65. https://doi.org/10.3390/mti2040065

Vuletic, T., Duffy, A., Hay, L., McTeague, C., Campbell, G., & Grealy, M. (2019). Systematic literature review of hand gestures used in human computer interaction interfaces. *International Journal of Human-Computer Studies*, *129*, 74–94. https://doi.org/10.1016/j.ijhcs.2019.03.011

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaving spatial? *Journal of Memory and Language*, *50*(4), 395–407. https://doi.org/10.1016/j.jml.2004.01.002

Wilson, M., & Emmorey, K. (2006). Comparing sign language and speech reveals a universal limit on short-term memory capacity. *Psychological Science*, *17*(8), 682–683. https://doi.org/10.1111/j.1467-9280.2006.01766.x

Wingfield, A., Stine, E. A. L., Lahar, C. J., & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental Aging Research*, *14*(2–3), 103–107. https://doi.org/10.1080/03610738808259731

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. In *Chi2009: Proceedings of the 27th Annual CHI Conference on Human Factors in Computing Systems,* (Vols 1–4, pp. 1083–1092). ACM.

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology. General*, *136*(4), 639–652. https://doi.org/10.1037/0096-3445.136.4.639

Wu, H., Wang, J. (2012). *User-defined body gestures for TV-based applications* [Paper presentation]. Paper presented at the 2012 Fourth International Conference on Digital Home (pp. 415–420). https://doi.org/10.1109/icdh.2012.23

Wu, H., Gai, J., Wang, Y., Liu, J., Qiu, J., Wang, J., & Zhang, X. (2020). Influence of cultural factors on freehand gesture design. *International Journal of Human-Computer Studies*, *143*, 102502. https://doi.org/10.1016/j.ijhcs.2020.102502

Zhao, M.-F., Zimmer, H. D., Zhou, X., & Fu, X. (2016). Enactment supports unitisation of action components and enhances the contribution of familiarity to associative recognition. *Journal of Cognitive Psychology*, *28*(8), 932–947. https://doi.org/10.1080/20445911.2016.1229321

About the Authors

Qi Gao is a PhD candidate for psychology at Zhejiang University in Hangzhou. She received her Bachelor of Science in psychology from Nankai University and Bachelor of Management in business administration from Tianjin University in 2019.

Zheng Ma is a fourth year PhD candidate at the department of Industrial and Manufacturing Engineering in the Penn State University. He recieved a B.S. in Psychology from Zhejiang University in 2018.

Quan Gu obtained a PhD of cognitive psychology and engineering psychology from Zhejiang University in 2021.

Jiaofeng Li received a bachelor's degree in psychology at Sun Yat-sen University in 2018. At present, she is a PhD candidate in applied psychology at Zhejiang University.

Zaifeng Gao is a professor at Zhejiang University, Hangzhou, and obtained Ph.D from Zhejiang University, Hangzhou, China, 2009. His research focuses on engineering psychology and cognitive psychology.