



Composite Face Effect Predicts Configural Encoding in Visual Short-Term Memory

Lilian Azer and Weiwei Zhang*

Department of Psychology, University of California, Riverside, Riverside, CA, United States

In natural vision, visual scenes consist of individual items (e.g., trees) and global properties of items as a whole (e.g., forest). These different levels of representations can all contribute to perception, natural scene understanding, sensory memory, working memory, and long-term memory. Despite these various hierarchical representations across perception and cognition, the nature of the global representations has received considerably less attention in empirical research on working memory than item representations. The present study aimed to understand the perceptual root of the configural information retained in Visual Short-term Memory (VSTM). Specifically, we assessed whether configural VSTM was related to holistic face processing across participants using an individual differences approach. Configural versus item encoding in VSTM was assessed using Xie and Zhang's (2017) dual-trace Signal Detection Theory model in a change detection task for orientation. Configural face processing was assessed using Le Grand composite face effect (CFE). In addition, overall face recognition was assessed using Glasgow Face Matching Test (GFMT). Across participants, holistic face encoding, but not face recognition accuracy, predicted configural information, but not item information, retained in VSTM. Together, these findings suggest that configural encoding in VSTM may have a perceptual root.

Keywords: visual short-term memory, Gestalt, holistic face processing, receiver operating characteristic, individual difference

In natural vision, visual scenes often consist of individual items (e.g., trees) and global emergent properties of items as a whole (e.g., forest). These different levels of representations can all contribute to perception (Navon and Norman, 1983; Kimchi, 1992), natural scene understanding (e.g., Greene and Oliva, 2009), sensory memory (Cappiello and Zhang, 2016), visual short-term memory (Brady et al., 2011; Tanaka et al., 2012; Orhan et al., 2013; Nie et al., 2017), and long-term memory (Hunt and Einstein, 1981; Yonelinas, 2002). In addition, there could be significant interactions between these hierarchical representations, for example, enhanced item processing by the global context (Fine and Minnery, 2009; Santangelo and Macaluso, 2013). Despite these various hierarchical representations across perception and cognition, global representations receive considerably less attention in memory research than item representation (e.g., Brady et al., 2011). The present study has thus assessed whether configural information, one kind of global representations, retained in VSTM is related to overall holistic processing in vision.

The representations of global information in VSTM have gained some support in recent years (e.g., Brady and Alvarez, 2015; Nie et al., 2017). For example, experimental manipulation of configural information at retrieval could either impair or facilitate VSTM for item information

OPEN ACCESS

Edited by:

Tatiana Aloï Emmanouil,
Baruch College (CUNY),
United States

Reviewed by:

Qi-Yang Nie,
Sun Yat-sen University, China
Timothy Michael Ellmore,
City College of New York (CUNY),
United States

*Correspondence:

Weiwei Zhang
weiwei.zhang@ucr.edu;
wwzhang@ucr.edu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 29 July 2019

Accepted: 22 November 2019

Published: 11 December 2019

Citation:

Azer L and Zhang W (2019)
Composite Face Effect Predicts
Configural Encoding in Visual
Short-Term Memory.
Front. Psychol. 10:2753.
doi: 10.3389/fpsyg.2019.02753

(Jiang et al., 2000, 2004; Treisman and Zhang, 2006). Specifically, changes in configural context (e.g., by changing features of non-probed items) at test can impair VSTM for spatial locations (Jiang et al., 2004) and non-spatial features (Vidal et al., 2005; Jaswal and Logie, 2011). In addition, manipulation of configural encoding upon formation of VSTM representations can also affect the later access to stored VSTM contents (Delvenne et al., 2002; Xu, 2006; Gao and Bentin, 2011; Gao et al., 2013; Peterson and Berryhill, 2013). For example, surrounding circles on orientation bars can considerably reduce VSTM storage capacity for orientation information (Delvenne et al., 2002; Alvarez and Cavanagh, 2008), presumably because the surrounding circles severely interrupted configural encoding (Xie and Zhang, 2017). It is highly likely that these effects of perceptual organization on VSTM are a natural extension of configural encoding in perceptual processing. Consistent with this hypothesis, Gestalt cues such as connectedness (Woodman et al., 2003; Xu, 2006), similarity (Peterson and Berryhill, 2013), and closure (Gao et al., 2016) could facilitate grouping of individual items during VSTM encoding, leading to increased storage capacity. In addition, the configural superiority effect (CSE) has demonstrated that individuals' ability to detect a target among distractors is significantly faster in the presence of contextual cues and closure (Nie et al., 2016). In other words, closure of stimuli allows individuals to form Gestalts resulting in rapid detection of the target stimulus and successful inhibition of distractor stimulus. Nonetheless, it is unclear whether holistic encoding, for example configural and holistic encoding as opposed to first-order processing of isolated feature in object and face recognition (Kimchi, 1994; Maurer et al., 2002; Piepers and Robbins, 2012), is related to VSTM for configural information. The present study has thus assessed whether holistic face processing (e.g., Tanaka and Farah, 1993) can predict configural VSTM across participants, using an individual differences approach.

Configural encoding in VSTM was estimated using a change detection task for orientation and a recently developed Receiver Operating Characteristic (ROC) model (Xie and Zhang, 2017). In this task **Figure 1A**, participants memorized five briefly presented orientation bars in a memory array. Following a 1,000-ms delay interval, participants reported whether an orientation bar in a test array contained a new orientation ("new" response) or the old orientation ("old" response) on a 6-point confidence scale, as compared to the corresponding bar presented at the same location in the memory array. These responses were used to construct ROC curves, the function relating the probability of "old" responses on *old* trials (hit rate) to the probability of "old" responses on *new* trials (false alarm rate) using Signal Detection Theory (Macmillan and Creelman, 2005). The resulting ROCs were then fitted with a Dual-Trace Signal Detection ROC model (DTSD, Xie and Zhang, 2017) to quantitatively assess contributions of item-based encoding (i.e., individual orientation) and configural encoding (i.e., the overall shape of all orientation bars) in VSTM (for details of the model, see Xie and Zhang, 2017).

Holistic encoding in face processing was estimated using the composite face effect (CFE) of the Le Grand face task (Le Grand et al., 2001, 2004; Mondloch et al., 2002). Unlike the face inversion task that taps on second-order relational

information in face processing, the CFE is a more robust measure of holistic face processing (Maurer et al., 2002). In this task, two brief displays of composite faces were presented sequentially (**Figure 1B**). Each face consisted of a top half and a bottom half that form a complete face when combined together. Participants reported whether the top halves of the two composite faces were the same (e.g., faces in the right column in **Figure 1C**) or different (e.g., faces in the left column in **Figure 1C**) across the two displays while ignoring the two bottom halves which were always the same across the two displays. Although the two bottom halves were completely task irrelevant, they can empirically and phenomenologically interfere with the same/different judgments of the two top halves (e.g., Le Grand et al., 2004), due to holistic face processing that encodes the top and bottom face halves as an integrated face instead of separate face segments. Orthogonal to the manipulation of the same versus different top face halves, the top and bottom face halves were misaligned (e.g., faces in the top row in **Figure 1C**) on half of the trials and properly aligned (e.g., faces in the bottom row in **Figure 1C**) on the remaining trials. The interference from the irrelevant bottom face halves tends to be reduced for the misaligned condition, as compared to the aligned condition, because the top and bottom halves from the misaligned condition are less likely to be perceived as an integrated face (Le Grand et al., 2001, 2004). This difference in performance between the aligned and misaligned conditions, the CFE, is thus an operational definition of the interference caused by task-irrelevant bottom face halves.

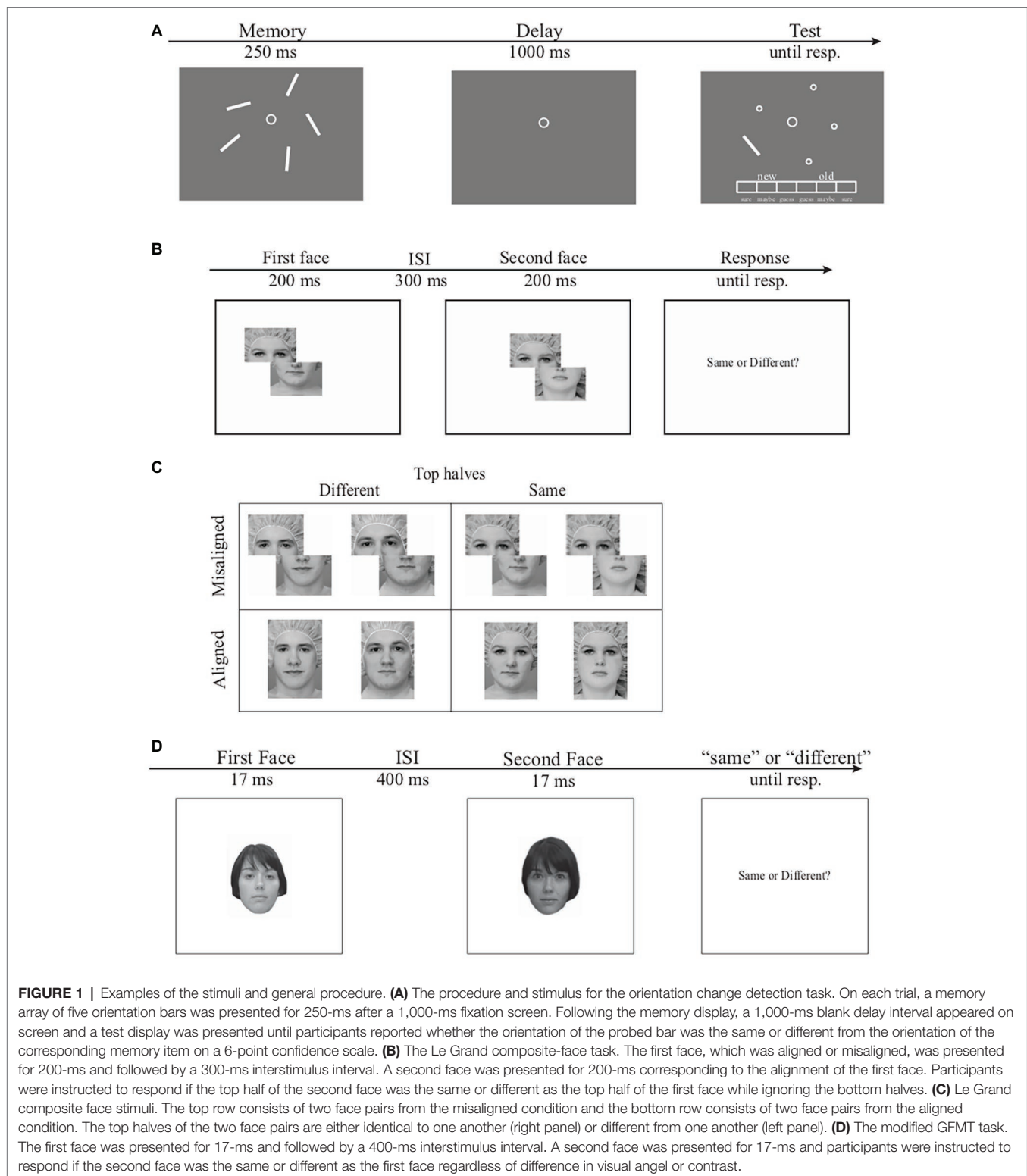
Overall face discrimination was also assessed, using a two-interval forced choice task (**Figure 1D**) with face stimuli from the Glasgow Face Matching Test (GFMT, for details see Burton et al., 2010). Participants in this task reported whether the two sequentially presented faces had the same or different identity. This modified GFMT was chosen over other face identification tasks (Bruce et al., 1999) to minimize potential involvement of VSTM (Xie and Zhang, 2017). Specifically, in the Bruce et al. face identification task for example, a target face is matched to one of the 10 simultaneously presented faces. The matching process in this task could involve several eye movements across stimuli, leading to significant involvements of VSTM (e.g., Irwin, 1991).

We hypothesized that holistic encoding in face processing assessed as CFE, but not the overall face matching ability assessed as accuracy in the modified GFMT discrimination task, would predict configural encoding, but not item encoding, in VSTM across participants.

METHODS

Participants

Forty-six UC Riverside students (31 females) between the ages of 18 and 30 with normal color vision and normal or corrected-to-normal visual acuity participated in this study for course credit. Four additional participants were excluded because they did not complete all three tasks within a 1-h experimental



session. The experimental procedure was approved by the Institutional Review Board of University of California, Riverside. All participants were provided written informed consent. *A priori* power analysis (Faul et al., 2009) for *r*-based effect

size at a medium level (0.35) suggested that a total sample size of 50 participants would provide 80% statistic power. *Post hoc* power analysis for 46 subjects for a *r*-based effect size of 0.38 yielded 84% statistical power.

Stimuli and Procedure

All stimuli were presented, using PsychToolbox-3 (Brainard, 1997) for Matlab (The MathWorks, Cambridge, MA), on a LCD monitor with a homogeneous gray background (6.7 cd/m^2) on a macOS operating system with a refresh rate of 60 Hz at a viewing distance of 57 cm.

The stimuli and procedure were the same as the uncircled bar condition in Experiment 1 of Xie and Zhang (2017). In this VSTM change detection task (Figure 1A), the memory array consisted of five white orientation bars (3° in length and 0.15° in width) in different orientations quasi-randomly selected from 180° circular space. The angular differences between any two orientations were more than 12° . The orientation bars were presented at five locations randomly chosen from eight equally spaced locations on an imaginary circle 4.5° in radius. Each trial began with a 1,000-ms fixation at the center of the screen, followed by a 250-ms memory array of five orientation bars. Participants were required to memorize and retain as many orientation bars as possible over a 1,000-ms blank delay interval. At test, one bar randomly selected from the memory set reappeared at its original location, whereas other memory items were replaced with circles (0.3°) as placeholders. Participants reported whether this bar had the “old” or a “new” orientation as compared to the corresponding item at the same location of memory array. The “old”/“new” decision and the confidence for this decision (e.g., sure new, maybe new, or guess new, sure old, maybe old, or guess old) were reported on a 6-point confidence scale (16.2 by 0.8° in visual angle) presented at the bottom of the screen using a computer mouse by the participants. The test orientations were equally likely to be the same as and different from the corresponding memory items. On “new” trials, the orientation was always perpendicular to the original orientation of the memory item (90° apart). Note, this manipulation rendered mnemonic precision of retained VSTM representations largely irrelevant. That is, either coarse-grain or fine-grained VSTM representation for the test orientation is sufficient for detecting the change between memory and test (for extended discussion, see Xie and Zhang, 2017). Each participant completed 120 trials that were split into three experimental blocks. Responses from this task were fit with the DTSD model, yielding separate estimates of item and configural VSTM encoding for each participant. The details of the DTSD model (e.g., the equations and the theoretical interpretation of the model parameters) and the model fitting procedure were provided in Xie and Zhang (2017), and are thus omitted here.

In the Le Grand face task (Figure 1B), each trial began with an 800-ms fixation, followed by sequential presentations of two composite faces with a 300-ms interstimulus interval. Each face was presented for 200-ms. Participants reported whether the top halves of the two sequentially presented faces were the same or different (same and different trials were equally likely), while ignoring the bottom halves, which were always different. Note, to fit the entire experiment within a 1-h session, a partial design was used here, as compared to the complete design that also includes the condition in which the bottom halves were the same (e.g., Richler et al., 2011).

On half of the trials, either the top or bottom halves of the two faces were shifted horizontally to the left by half a face width (the misaligned condition, the top row in Figure 1C), whereas on the remaining half of the trials, the top and bottom halves of each face were properly aligned (the aligned condition, the bottom row in Figure 1C). The faces in the aligned condition and misaligned condition were presented within a $4.8^\circ \times 7.2^\circ$ and $7.2^\circ \times 7.2^\circ$ rectangular area, respectively. The same and different trials were randomly mixed within experimental blocks, whereas misaligned and aligned conditions were blocked with the order counterbalanced across participants. Participants were instructed to make a *Same-or-Different* decision specifically to the two top halves, while ignoring the bottom halves, by pressing button “s” for *Same* or button “d” for *Different* on a computer keyboard as quickly and accurately as possible once the second face appeared. Both response time (RT) and accuracy were recorded. Twenty-five *same* trials and 25 *different* trials were presented for each of the aligned and misaligned conditions, yielding 100 trials in total.

CFE of reaction time (RT) was calculated by subtracting the median RT for misaligned trials from the median RT for aligned trials (Le Grand et al., 2004; Konar et al., 2010). Note, only RTs from trials with correct responses were used in this analysis. CFE was also calculated on mean d' , a signal detection theory measure (Macmillan and Creelman, 2005), in the same way as CFE on RT (Konar et al., 2010; Richler et al., 2011).

The modified GFMT task is a two-interval forced choice task using face stimuli adopted from Burton and colleagues (Burton et al., 2010). In this task, 150 pairs of gray-scale front-view Caucasian faces, subtending $5^\circ \times 7^\circ$ of visual angle, were randomly selected from the GFMT set (see Burton et al., 2010 for details). Half of these pairs had matching identities and the other half had different identities. On each trial, two brief displays of faces (17-ms each) were presented sequentially with a 400-ms interstimulus interval. Note, this was different from the original GFMT in which the two faces were presented side by side simultaneously. The sequential presentation in the present study was to match the sequential presentations of face stimuli in the Le Grand task. Participants judged whether the two faces had the same identity or different identities while ignoring visual features that were irrelevant for identities. For example, the two faces with the matching identity in Figure 1D had subtle differences in contrasts, hairstyles, face contours, viewing angles, etc. This variance in identity-independent visual features is to ensure participants encode face identities across different views, which mimics faces recognition in natural vision (Burton et al., 2010).

Grubbs' test (Grubbs, 1969) was conducted to detect potential outliers in all measures, although no outlier was identified in the present data, leading to zero outlier rejection.

RESULTS AND DISCUSSION

All three tasks were performed with reasonable accuracy. The change detection task yielded an average accuracy of 75% (72%–77%) [mean (95% confidence interval)] and average capacity

(assessed as Cowan's K , Cowan et al., 2005) of 2.48 (2.23–2.72). For the Le Grand face task, accuracy was averaged at 85% (83%–87%). More importantly, RTs on correct trials were significantly faster on misaligned trials than aligned trials [$t(45) = 4.23, p < 0.001$, Cohen's $d = 0.63$, Bayes factor = 211.96], indicating more interference on aligned trials (and hence significant CFE on RT). Although CFE on d' was significant [$t(45) = 4.69, p < 0.001$, Cohen's $d = 0.70$, Bayes factor = 818.14], it was not significantly correlated with any measures in the present study (p 's > 0.30) and was thus not discussed further. For the modified GFMT discrimination task, accuracy was averaged at 78% (76%–80%).

Of central importance, participants with more holistic face processing in the Le Grand face task had larger configural encoding in VSTM (**Figure 2A**), but comparable item encoding (**Figure 2B**), as compared to participants with less holistic face processing. That is, holistic face processing assessed as CFE significantly correlated with configural encoding [Pearson correlation: $r = 0.38$ (0.10–0.60), $p = 0.009$; Spearman's rank-order correlation for non-Gaussian distribution in RT: $r = 0.37$ (0.08–0.60), $p = 0.011$], but not with item encoding [Pearson correlation: $r = -0.10$ (–0.38–0.20), $p = 0.512$; Spearman's rank-order correlation: $r = -0.10$ (–0.39–0.20), $p = 0.495$]. Additionally, a multivariate regression analysis suggested that CFE significantly predicted configural VWM [$\beta = 0.95$ (0.25–1.64), $p = 0.009$] but not item VWM [$\beta = -0.090$ (–0.37–0.19), $p = 0.512$]. Critically, the correlation between CFE and configural encoding was significantly greater than the correlation between CFE and item encoding ($z = 2.09, p = 0.018$, one-tailed), based on a Fisher's r to z transformation one-tailed test of correlated correlation (Meng et al., 1992). The relationship between configural VSTM and holistic face processing seems to be specific in that VSTM configural encoding did not significantly correlate with overall face processing assessed as the accuracy of the Le Grand face task (**Figure 3A**) [Pearson correlation: $r = -0.16$ (–0.43–0.13), $p = 0.27$; Spearman's rank-order correlation: $r = -0.25$ (–0.51–0.06), $p = 0.10$] or the accuracy of the modified GFMT (**Figure 3B**) [Pearson correlation: $r = -0.01$ (–0.30–0.28), $p = 0.94$; Spearman's rank-order correlation: $r = -0.08$ (–0.37–0.23), $p = 0.61$].

The lack of significant correlation between CFE and the overall face matching ability assessed as the accuracy of 2IFC discrimination task using GFMT face stimuli, although consistent with some previous findings (e.g., Konar et al., 2010), could simply result from the partial design (Richler et al., 2011). As a result, the present study was not optimal for assessing the relationship between configural face encoding and face recognition, which is beyond the scope of the present study.

Interestingly, CFE was significantly correlated with the overall performance of the VSTM change detection task [Accuracy, Pearson correlation: $r = 0.30$ (0.01–0.54), $p = 0.04$; Spearman's rank-order correlation: $r = 0.31$ (0.01–0.56), $p = 0.04$; Capacity, Pearson correlation: $r = 0.30$ (0.01–0.55), $p = 0.04$; Spearman's rank-order correlation: $r = 0.31$ (0.02–0.56), $p = 0.03$], potentially driven by the significant correlation between the CFE and VSTM configural encoding. The CFE could be considered as the magnitude of distractor processing in a way similar to the flanker compatibility effect (Zhang and Luck, 2014) given that both effects reflect how much distracting information (flanker distractor letters in the flanker task or bottom face halves in the Le Grand composite-face task) is processed. However, the positive correlation between VSTM capacity and the CFE seems to be inconsistent with the load theory of attention which predicts that higher WM capacity (equivalent to low cognitive load) would reduce distractor processing (for a recent review, see de Fockert, 2013; but see, Konstantinou and Lavie, 2013; Zhang and Luck, 2014). In other words, individuals with lower working memory capacities process distractors in an equivalent manner as higher working memory load conditions in tasks investigating the load theory of attention (de Fockert, 2013). In the present study, higher VSTM accuracy and capacity predicted a higher CFE, which is indicative of greater distractibility in the aligned face condition compared to the misaligned face condition. This inconsistency may result from the holistic nature of face processing in the Le Grand task. Further research is needed to directly compare effects of cognitive load on CFE and flanker compatibility effects.

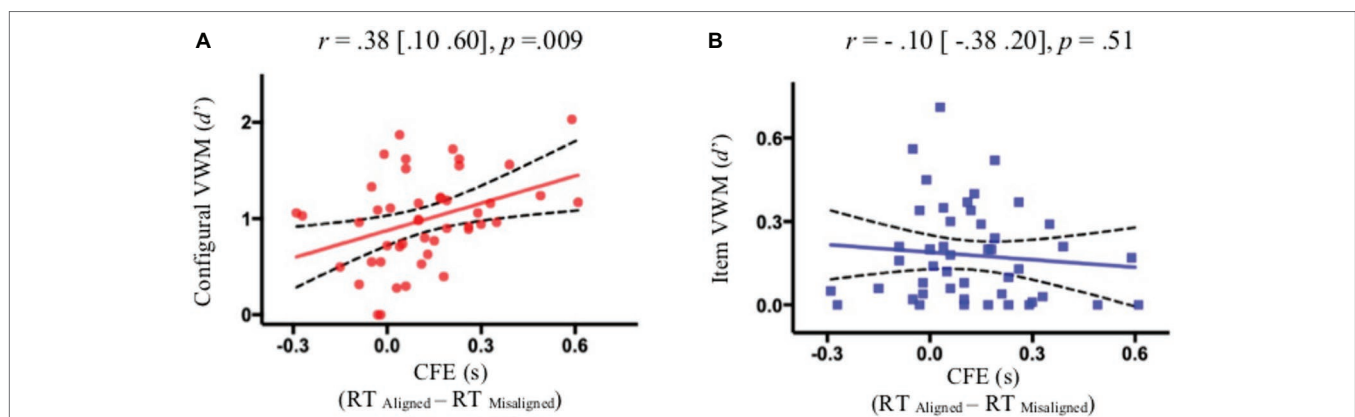
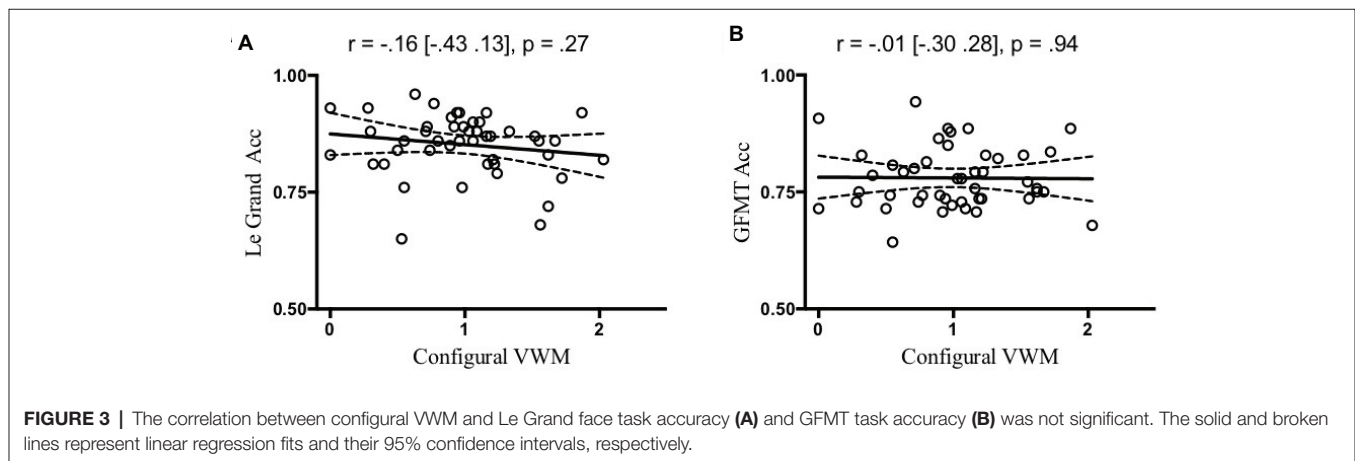


FIGURE 2 | The significant correlation between holistic face encoding (CFE) and configural information retained in VSTM (**A**), but not between holistic face encoding (CFE) and item information retained in VSTM (**B**). The solid and broken lines represent linear regression fits and their 95% confidence intervals, respectively.



GENERAL DISCUSSION

Given the hierarchical nature of visual representations in natural vision, it is essential for VSTM to retain hierarchical representations such as item and configural information. The present study assessed configural encoding and item encoding in VSTM for orientation using ROC modeling of change detection performance and assessed holistic face processing using the Le Grand CFE. We found that configural encoding, but not item encoding, for orientations in VSTM significantly correlated with holistic processing in face discrimination in that participants with more configural VSTM also showed larger holistic face encoding. These findings add to the growing literature on the effects of perceptual organization on VSTM.

In addition to its selective correlation with holistic face processing, configural information retained in VSTM could also be experimentally dissociated from item information retained in VSTM using selective experimental manipulations (Xie and Zhang, 2017). Together these findings are essential for scaling up models for VSTM to account for both item information and global stimulus structures (Kroll et al., 1996; Reinitz et al., 1996), especially for natural stimuli (Greene and Oliva, 2009; Brady et al., 2011; Brady and Tenenbaum, 2013).

The Dual Trace Signal Detection model for VSTM is mathematically equivalent to the Dual-Process Signal Detection (DPSD) of Recognition memory (Yonelinas et al., 1999). According to DPSD model, recognition is based on a high-threshold discrete recollection and a continuous familiarity (modeled as d'). The d' component shared across the two models seems to suggest some relationship between configural encoding and familiarity. For instance, configural information may be more important for supporting familiarity than item information. Consequently, face inversion that profoundly impairs holistic face processing (for review, see Maurer et al., 2002) can significantly reduce familiarity (Yonelinas et al., 1999). In addition, processing of higher level representations (e.g., summary statistics and configural information) is more automatic than processing of discrete item representations (e.g., Alvarez and Oliva, 2009), consistent

with the proposal that familiarity is less controlled than recollection (Yonelinas and Jacoby, 1996; Yonelinas, 2002).

In the present study, configural VSTM correlates with holistic face encoding (CFE), but not with overall face discrimination (the 2IFC using GFMT stimuli). The latter finding was not necessarily inconsistent with a previously reported correlation between VSTM and face identification (Megreya and Burton, 2006). The face identification task in Megreya and Burton's (2006) study used a procedure in which the participants match one target face to one of 10 possible faces. This procedure could exert a high demand on VSTM. When demand on VSTM was significantly reduced using a two-alternative force choice task with two simultaneously presented faces in the GFMT, no significant correlation was found between face matching and VSTM (Burton et al., 2010). It is thus possible that a robust relationship between VSTM and face processing could manifest to face recognition tasks with high working memory load.

The present study has focused on how configural processing is shared between face perception and VSTM for orientation. This issue is orthogonal to the role of holistic processing in face processing (Konar et al., 2010; Richler et al., 2011; Gold et al., 2012; DeGutis et al., 2013), which is beyond the scope of the present study. While Konar et al. (2010) used a full factorial design in which the bottom face halves were equally likely to be the same or different the current study used a partial design. The partial design may be suboptimal in that it could induce systematic biases of reporting "same," which may account for some conflicting findings on the effects of holistic encoding in face identification (Konar et al., 2010; Richler et al., 2011). However, it is less of an issue in the present study given a significant relationship was found between holistic face processing and configural VSTM. Nonetheless, it is important for future research to replicate this finding using a full factorial design in the Le Grand CFE task. It will also be interesting to link item and configural encoding in VSTM to part-based encoding and holistic encoding in processing of facial expressions (Tanaka et al., 2012). Furthermore, a stronger test of the relationship between holistic face processing and configural encoding in VSTM is to assess whether holistic face processing can predict increased VSTM performance in VSTM

from a condition where configural encoding is minimized (Xie and Zhang, 2017) to conditions where configural encoding is prominent (Xie and Zhang, 2017).

In summary, the present study, using an individual differences approach, has demonstrated that VSTM for configural information can be accounted for by holistic face perception, providing preliminary evidence that configural encoding in VSTM may result from configural encoding in perception.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of California, Riverside IRB. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

WZ designed the study, collected data, and conducted data analysis. Both authors contributed to manuscript preparation.

FUNDING

This study was made possible by funding support from the National Institute of Mental Health (R01MH117132 and 1F32MH090719 to WZ).

ACKNOWLEDGMENTS

We would like to thank Mike Burton for sharing the GFMT face stimuli.

REFERENCES

- Alvarez, G. A., and Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Percept. Psychophys.* 70, 346–364. doi: 10.3758/PP.70.2.346
- Alvarez, G. A., and Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proc. Natl. Acad. Sci. USA* 106, 7345–7350. doi: 10.1073/pnas.0808981106
- Brady, T. F., and Alvarez, G. A. (2015). Contextual effects in visual working memory reveal hierarchically structured memory representations. *J. Vis.* 15, 6–24. doi: 10.1167/15.15.6
- Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* 11, 4–4. doi: 10.1167/11.5.4
- Brady, T., and Tenenbaum, J. (2013). A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* 120, 85–109. doi: 10.1037/a0030779
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436.
- Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., and Miller, P. (1999). Verification of face identities from images captured on video. *J. Exp. Psychol. Appl.* 5, 339–360. doi: 10.1037/1076-898X.5.4.339
- Burton, A. M., White, D., and McNeill, A. (2010). The glasgow face matching test. *Behav. Res. Methods* 42, 286–291. doi: 10.3758/BRM.42.1.286
- Cappiello, M., and Zhang, W. (2016). A dual-trace model for visual sensory memory. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1903–1922. doi: 10.1037/xhp0000274
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi: 10.1016/j.cogpsych.2004.12.001
- de Fockert, J. W. (2013). Beyond perceptual load and dilution: a review of the role of working memory in selective attention. *Front. Psychol.* 4, 1–12. doi: 10.3389/fpsyg.2013.00287
- DeGutis, J., Wilmer, J., Mercado, R. J., and Cohan, S. (2013). Using regression to measure holistic face processing reveals a strong link with face recognition ability. *Cognition* 126, 87–100. doi: 10.1016/j.cognition.2012.09.004
- Delvenne, J. F., Braithwaite, J. J., Riddoch, M. J., and Humphreys, G. W. (2002). Capacity limits in visual short-term memory for local orientations. *Curr. Psychol. Cogn.* 21, 681–690.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Fine, M. S., and Minnery, B. S. (2009). Visual salience affects performance in a working memory task. *J. Neurosci.* 29, 8016–8021. doi: 10.1523/JNEUROSCI.5503-08.2009
- Gao, Z., and Bentin, S. (2011). Coarse-to-fine encoding of spatial frequency information into visual short-term memory for faces but impartial decay. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1051–1064. doi: 10.1037/a0023091
- Gao, Z., Ding, X., Yang, T., Liang, J., and Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PLoS One* 8:e57913. doi: 10.1371/journal.pone.0057913
- Gao, Z., Gao, Q., Tang, N., Shui, R., and Shen, M. (2016). Organization principles in visual working memory: evidence from sequential stimulus display. *Cognition* 146, 277–288. doi: 10.1016/j.cognition.2015.10.005
- Gold, J. M., Mundy, P. J., and Tjan, B. S. (2012). The perception of a face is no more than the sum of its parts. *Psychol. Sci.* 23, 427–434. doi: 10.1177/0956797611427407
- Greene, M. R., and Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58, 137–176. doi: 10.1016/j.cogpsych.2008.06.001
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21. doi: 10.1080/00401706.1969.10490657
- Hunt, R. R., and Einstein, G. O. (1981). Relational and item-specific information in memory. *J. Verbal Learn. Verbal Behav.* 20, 497–514. doi: 10.1016/S0022-5371(81)90138-9
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cogn. Psychol.* 23, 420–456. doi: 10.1016/0010-0285(91)90015-G
- Jaswal, S., and Logie, R. (2011). Configural encoding in visual feature binding. *J. Cogn. Psychol.* 23, 586–603. doi: 10.1080/20445911.2011.570256
- Jiang, Y., Chun, M. M., and Olson, I. R. (2004). Perceptual grouping in change detection. *Percept. Psychophys.* 66, 446–453. doi: 10.3758/BF03194892
- Jiang, Y., Olson, I. R., and Chun, M. M. (2000). Organization of visual short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 683–702. doi: 10.1037/0278-7393.26.3.683
- Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: a critical review. *Psychol. Bull.* 112, 24–38. doi: 10.1037/0033-2909.112.1.24
- Kimchi, R. (1994). The role of wholistic/configural properties versus global properties in visual form perception. *Perception* 23, 489–504.
- Konar, Y., Bennett, P. J., and Sekuler, A. B. (2010). Holistic processing is not correlated with face-identification accuracy. *Psychol. Sci.* 21, 38–43. doi: 10.1177/0956797609356508

- Konstantinou, N., and Lavie, N. (2013). Dissociable roles of different types of working memory load in visual detection. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 919–924. doi: 10.1037/a0033037
- Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., and Tulving, E. (1996). Cohesion failure as a source of memory illusions. *J. Mem. Lang.* 35, 176–196. doi: 10.1006/jmla.1996.0010
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2001). Neuroperception. Early visual experience and face processing. *Nature* 410:890. doi: 10.1038/35073749
- Le Grand, R., Mondloch, C. J., Maurer, D., and Brent, H. P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychol. Sci.* 15, 762–768. doi: 10.1111/j.0956-7976.2004.00753.x
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection theory. 2nd Edn.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Maurer, D., Grand, R. L., and Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260. doi: 10.1016/S1364-6613(02)01903-4
- Megreya, A. M., and Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Mem. Cogn.* 34, 865–876. doi: 10.3758/BF03193433
- Meng, X.-L., Rosenthal, R., and Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychol. Bull.* 111, 172–175. doi: 10.1037/0033-2909.111.1.172
- Mondloch, C. J., Le Grand, R., and Maurer, D. (2002). Configural face processing develops more slowly than featural face processing. *Perception* 31, 553–566. doi: 10.1068/p3339
- Navon, D., and Norman, J. (1983). Does global precedence really depend on visual angle? *J. Exp. Psychol. Hum. Percept. Perform.* 9, 955–965.
- Nie, Q.-Y., Maurer, M., Müller, H. J., and Conci, M. (2016). Inhibition drives configural superiority of illusory Gestalt: combined behavioral and drift-diffusion model evidence. *Cognition* 150, 150–162. doi: 10.1016/j.cognition.2016.02.007
- Nie, Q.-Y., Müller, H. J., and Conci, M. (2017). Hierarchical organization in visual working memory: from global ensemble to individual object structure. *Cognition* 159, 85–96. doi: 10.1016/j.cognition.2016.11.009
- Orhan, A., Jacobs, R., and Jacobs, R. (2013). A probabilistic clustering theory of the organization of visual short-term memory. *Psychol. Rev.* 120, 297–328. doi: 10.1037/a0031541
- Peterson, D. J., and Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychon. Bull. Rev.* 20, 1282–1289. doi: 10.3758/s13423-013-0460-x
- Piepers, D., and Robbins, R. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. *Front. Psychol.* 3, 1–11. doi: 10.3389/fpsyg.2012.00559
- Reinitz, M. T., Verfaellie, M., and Milberg, W. P. (1996). Memory conjunction errors in normal and amnesic subjects. *J. Mem. Lang.* 35, 286–299. doi: 10.1006/jmla.1996.0016
- Richler, J. J., Cheung, O. S., and Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychol. Sci.* 22, 464–471. doi: 10.1177/0956797611401753
- Santangelo, V., and Macaluso, E. (2013). Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity. *J. Neurosci.* 33, 4110–4117. doi: 10.1523/JNEUROSCI.4138-12.2013
- Tanaka, J. W., and Farah, M. J. (1993). Parts and wholes in face recognition. *Q. J. Exp. Psychol. A* 46, 225–245.
- Tanaka, J. W., Kaiser, M. D., Butler, S., and Le Grand, R. (2012). Mixed emotions: holistic and analytic perception of facial expressions. *Cognit. Emot.* 26, 961–977. doi: 10.1080/02699931.2011.630933
- Treisman, A., and Zhang, W. (2006). Location and binding in visual working memory. *Mem. Cogn.* 34, 1704–1719. doi: 10.3758/BF03195932
- Vidal, J. R., Gauchou, H. L., Tallon-Baudry, C., and O'Regan, J. K. (2005). Relational information in visual short-term memory: the structural gist. *J. Vis.* 5, 244–256. doi: 10.1167/5.3.8
- Woodman, G. F., Vecera, S. P., and Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychon. Bull. Rev.* 10, 80–87. doi: 10.3758/BF03196470
- Xie, W., and Zhang, W. (2017). Discrete item-based and continuous configural representations in visual short-term memory. *Vis. Cogn.* 25, 21–33. doi: 10.1080/13506285.2017.1339157
- Xu, Y. (2006). Understanding the object benefit in visual short-term memory: the roles of feature proximity and connectedness. *Percept. Psychophys.* 68, 815–828. doi: 10.3758/bf03193704
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46, 441–517. doi: 10.1006/jmla.2002.2864
- Yonelinas, A. P., and Jacoby, L. (1996). Noncritical recollection: familiarity as automatic, irrelevant recollection. *Conscious. Cogn.* 5, 131–141. doi: 10.1006/ccog.1996.0008
- Yonelinas, A. P., Kroll, N. E., Dobbins, I. G., and Soltani, M. (1999). Recognition memory for faces: when familiarity supports associative recognition judgments. *Psychon. Bull. Rev.* 6, 654–661. doi: 10.3758/BF03212975
- Zhang, W., and Luck, S. J. (2014). Opposite effects of capacity load and resolution load on distractor processing. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 22–27. doi: 10.1037/xhp0000013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Azer and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.