BRILL | *Multisensory Research 30 (2017) 363–385* | brill.com/msr

Cognitive Styles Differentiate Crossmodal Correspondences Between Pitch Glide and Visual Apparent Motion

Lu Guo [1 2], **Ming Bao** [1 2 *], **Luyang Guan** [1 2] **and Lihan Chen** [3 4 *]

[1] Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China

[4] Key Laboratory of Machine Perception, Peking University, Beijing 100871, China

Received 4 September 2016; accepted 13 February 2017

Abstract

Crossmodal correspondences are the automatic associations that most people have between different basic sensory stimulus attributes, dimensions, or features. For instance, people often show a systematic tendency to associate moving objects with changing pitches. Cognitive styles are defined as an individual's consistent approach to think, perceive, and remember information, and they reflect qualitative rather than quantitative differences between individuals in their thinking processes. Here we asked whether cognitive styles played a role in modulating the crossmodal interaction. We used the visual Ternus display in our study, since it elicits two distinct apparent motion percepts: element motion (with a shorter interval between the two Ternus frames) and group motion (with a longer interval between the two frames). We examined the audiovisual correspondences between the visual Ternus movement directions (upward or downward) and the changes of pitches of concurrent glides (ascending frequency or descending frequency). Moreover, we measured the cognitive styles (with the Embedded Figure Test) for each participant. The results showed that congruent correspondence between pitch-ascending (decreasing) glides and moving upward (downward) visual directions led to a more dominant percept of 'element motion', and such an effect was typically observed in the field-independent group. Importantly, field-independent participants demonstrated a high efficiency for identifying the properties of audiovisual events and applying the crossmodal correspondence in crossmodal interaction. The results suggest cognitive styles could differentiate crossmodal correspondences in crossmodal interaction.

* To whom correspondence should be addressed. E-mail: baoming@mail.ioa.ac.cn, clh20000@gmail.com

DOI:10.1163/22134808-00002556

Keywords
Crossmodal correspondence, cognitive style, Ternus display, field-dependent/independent, auditory glide

1. Introduction

In everyday life, we are bombarded with a deluge of information arriving via several sensory modalities. To maintain a coherent presentation of the events in the environment, our brain needs to integrate these discrete unisensory signals. For many years, research has targeted the spatial and temporal factors that modulate or constrain multisensory integration (Chen and Vroomen, 2013; Frens *et al.*, 1995; Jones and Jarick, 2006; Spence and Driver, 2004). In addition to those factors, it has been shown that multisensory integration is affected by crossmodal correspondences between different sensory events. These correspondences include the natural (synaesthetic) mapping between basic properties such as visual luminance and auditory pitch (Evans and Treisman, 2010; Parise and Spence, 2008; Spence, 2011), and high-level mappings such as the semantic congruency (e.g., matching meaning *vs* mismatching meaning) between crossmodal events (Molholm *et al.*, 2004; Van Atteveldt *et al.*, 2004). The associations between different sensory properties should interact with spatio-temporal factors to generate the observed result of multisensory integration.

Crossmodal correspondence research has received widespread attention in recent years (Gallace and Spence, 2006; Makovac and Gerbino, 2010). Crossmodal correspondences describe the automatic associations people tend to make between different basic physical attributes, dimensions, or features (e.g., pitch, lightness, brightness, size) in different sensory modalities (Spence, 2011). For example, auditory loudness has been shown to map onto multiple visual properties, including direction of visual motion (Clark and Brownell, 1976), brightness (Marks, 1987), and even spatial frequency (Evans and Treisman, 2010). One property typical of the many types of crossmodal correspondence is the association that exists between auditory pitch changes and visual movement direction. Maeda *et al.* (2004) demonstrated that gratings with ambiguous motion accompanied by ascending pitch were more likely to be perceived as an upward motion, while those accompanied by descending pitch were more likely to be perceived as a downward motion. Mossbridge *et al.* (2011) used an exogenous spatial cuing paradigm to demonstrate that sounds with ascending frequency could guide visual spatial attention upwards, whereas sounds with descending frequency direct attention downwards. Using speeded classification, Evans and Treisman (2010) found spontaneous mappings between the auditory feature of pitch and the features of vertical location, size, and spatial frequency (but not contrast). They also proved these interactions took place in an automatic way (i.e., at the perceptual level), during which the strength of the interactions between pitch and (visual) spatial position was largest among many dimensions (such as the interaction between pitch and spatial frequency) (Evans and Treisman, 2010). In contrast, Parise and Spence (2012) utilized a modified version of the implicit association test (IAT) to measure crossmodal congruency effects, and found that crossmodal correspondences that involve elementary stimulus features, such as pitch and size, as well as more complex stimuli, such as nonsense words and line drawings, all had very similar effect sizes (Parise and Spence, 2012). Taken together, these results demonstrated that crossmodal correspondences operate on a wide range of sensory properties and are largely on a perceptual level (Parise and Spence, 2012).

Multisensory processing has recently been shown to be subject to individual differences (Cecere *et al.*, 2015). Studies suggest that the Bouba/Kiki effect, a well-known shape–sound symbolism effect, is partly tuned by differing perceptual styles in Eastern and Western culture, and therefore indicates that human perception, perhaps surprisingly, may be affected by cultural background (Bremner *et al.*, 2013; Chen *et al.*, 2016a). Empirically, crossmodal correspondence as a form of multisensory processing may differ among individuals who exhibit various perceptual styles (Rader and Tellegen, 1987). Therefore, in the present study, we aimed to examine how cognitive style could affect crossmodal correspondence and hence subsequent crossmodal interaction.

For more than half a century, a number of studies have supported that cognitive style plays an important role in affecting people's intellectual activities across a range of domains. Cognitive style initially reflects qualitative rather than quantitative differences between individuals in their thinking processes (Rayner and Riding, 2002). Among the cognitive styles identified to date, Witkin's concept of the field-dependent (FD)/field-independent (FI) dimension has been the most extensively studied. The demarcation of the two types has been widely measured by the Embedded Figure Test (EFT) (Kozhevnikov, 2007; Witkin *et al.*, 1975). The EFT requires participants to locate a simple shape embedded in a complex figure. Respondents' scores on the EFT can be used to classify them as either field-independent, if their scores are high, or as field-dependent, if their scores are low.

FD describes the "degree to which a learner's perception or comprehension of information is affected by the surrounding perceptual or contextual field" (Jonassen and Grabowski, 1993). The FD group usually takes a passive approach, exhibiting more dependency on the surrounding field, and cannot easily perceive the embedded part (Goodenough, 1976; Kozhevnikov, 2007; Witkin *et al.*, 1975). On the contrary, the FI group tends to adopt an analytical approach to problem solving and sees objects or details as discrete from their backgrounds, so that they can perform better in the EFT. Using the FI and FD categorization, many studies have revealed evidence of self-consistency extending across the embedded-figure tasks involving different senses. For example, if a person takes a long time to find the simple figures embedded in the 'global' pictures, then he or she also finds it difficult to identify a simple tune from a complex melody, or fails to identify a simple tactile pattern (composed by raised contours) embedded in a sophisticated tactile figure (Axelrod and Cohen, 1961; White, 1954; Witkin *et al.*, 1968).

However, the distinctions between FI and FD indicated in the studies cited were mostly confined to a single modality. The current study aimed to explore the role of cognitive style in modulating the association between two properties of crossmodal events: auditory pitch and visual movement direction. Specifically, we used the vertical Ternus display (containing two visual frames) as a visual probe and examined how auditory glide with changing/fixed pitches affected discrimination of Ternus apparent motion. Moreover, we investigated how the individual differences would modulate the outcome of crossmodal interaction. We expected that the FI group would be less affected by the auditory cues than the FD group, and have less crossmodal bias as compared with their FD peers. Alternatively, if the degree of 'dependency' in the FD and FI groups mainly hinges on the ability to identify 'individual' sensory events from a relatively complex crossmodal scenario, then the FI group would show a better ability to identify the elementary components of audiovisual events, and FI individuals could further exploit the audiovisual correspondence and show the typical effect (bias) of crossmodal interaction. However, for the FD group, their potential lower efficiency in separating and identifying individual events/properties may weaken the binding of different sensory properties and lead to less crossmodal bias. Thereforerson2-3226ding ofiThere032
iio,for visual movement direction- [illegible]
the of Ternus frames.or P(F)17.2(a).r(rtipa2(nts219 [(were)-206.requipared)-318.1(t).1(o)-655.8(d).

2.1. Participants

Forty right-handed undergraduate and graduate students were paid for participation (29 women, 11 men, $_{age} = 22$ 2 years, SD = 2.07, age range: 18–25 years). All subjects had normal (or corrected-to-normal) vision and normal hearing, and were naïve to the purpose of the study. Prior to participating in the study, all participants provided written informed consent. The experiments were performed in compliance with all institutional guidelines set by the Academic Affairs Committee, School of Psychological and Cognitive Sciences at Peking University.

2.2. Apparatus and Stimuli

Visual stimuli were presented on a 17-inch CRT monitor (ViewSonic; 100 Hz refresh rate; 1024 × 768 pixel resolution), controlled by a normal PC (HP AMD Athlon 64 Dual-Core Processor). Auditory stimuli (65 dB) were monaurally presented via two mini-speakers (DK-601, diameter 3.6 cm; vertical center-to-center distance 45 cm), which were placed at the central points of both edges above and below the monitor. The computer programs for controlling the experiments were developed with Matlab (Mathworks Inc.) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The test cabin was semi-anechoic and dimly lit throughout the experiment. The viewing distance was set at 60 cm, maintained by using a chin-rest.

2.2.1. Visual Stimuli

The area of the monitor was divided equally into two parts: the upper visual field (UVF) and the lower visual field (LVF), respective to a central fixed point. On each trial, the visual stimuli were presented randomly (but with equal probabilities) in the upper or low visual field. A variant of the visual Ternus display was used. The display contained two consecutively presented vertical frames, each lasting 30 ms. Each frame contained two vertical black dots and was presented on a gray background (10.6 cd/m^2 in luminance). The dots were 1.3° of visual angle in diameter, 0.24 cd/m^2 in luminance and had a 2° separation between them. When overlaid, the two frames shared one dot location at the center of the current visual field, with the other two dots vertically located on opposite sides relative to the center (Fig. 1). For a given trial, the inter-frame-interval (IFI) between the two visual frames was randomly selected from one of the following five durations: 80, 110, 140, 170, or 200 ms.

2.2.2. Auditory Stimuli

For the sound presentation conditions, three 500-ms auditory configurations (ascending-pitch, descending-pitch, and fixed-pitch) were used. For the ascending pitch, the glide had frequencies linearly modulated from 300 Hz to 4300 Hz, while for the descending pitch, the glide changed frequencies from

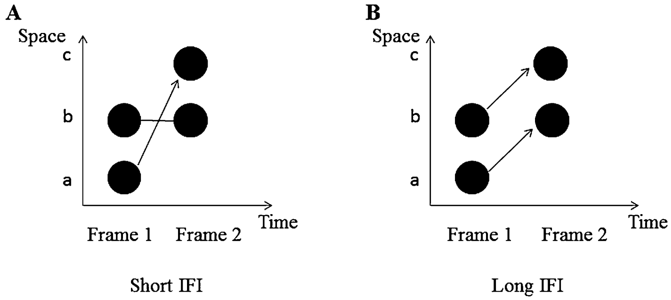

Figure 1. Vertical Ternus apparent motion, consisting of two partially overlapping frames of elements. In the first frame, the two dots are presented in vertical locations *a* and *b,* and in the second frame, they appear in locations *b* and *c*, so that they share a common position of 'b'. The Ternus display induces two possible motion percepts: (A) Element motion (EM) for short inter-frame intervals (IFIs) with the middle disk being perceived as static and the outer disk being perceived as moving from one side to the other. (B) Group motion (GM) for long IFIs with the two disks being perceived as moving together as a group.

4300 Hz to 300 Hz instead. For the fixed-pitch sound, the frequency was maintained at 500 Hz. For all the above auditory stimuli, the durations were 500 ms, with 50 ms linear ramps at the onset and offset. All the auditory stimuli were presented monaurally, in a random order, from two mini-speakers with an intensity of 65 dB SPL (with reference to the participant's head). The temporal middle point of the glide (i.e., 250 ms from the onset of the glide) was coincident with the middle point of the IFIs between the two visual frames.

2.3. Design and Procedures

A 4 (audiovisual stimulus configurations: congruent, incongruent, fixed, and visual only) $\times$ 5 (IFIs: 80, 110, 140, 170, or 200 ms) factorial and block-wise design was implemented. As shown in Fig. 2, in the 'congruent' condition, an upward Ternus apparent motion was paired with an ascending-pitched glide, or a downward Ternus apparent motion was paired with a descending-pitched glide. In the 'incongruent' audiovisual condition, pairings of upward *vs* descending and downward *vs* ascending were used.

Each audiovisual configuration had four blocks, with each block having 80 trials. The orders for visual fields (upper or lower area) in which the apparent motion appeared and the sound locations (above or below the monitor) were counterbalanced. In total, there were 16 blocks and 1280 trials. Considering that it would take a long time (around three hours) to complete all the trials, the first test of the experiment was divided into two parts, with 40 min of rest in between. Between blocks, participants took a short rest for 2–3 min.

To accurately render the timing of the auditory and visual stimuli, the duration of the visual stimuli and the synchronization of the auditory and visual stimuli were controlled by the monitor's vertical synchronization pulse.

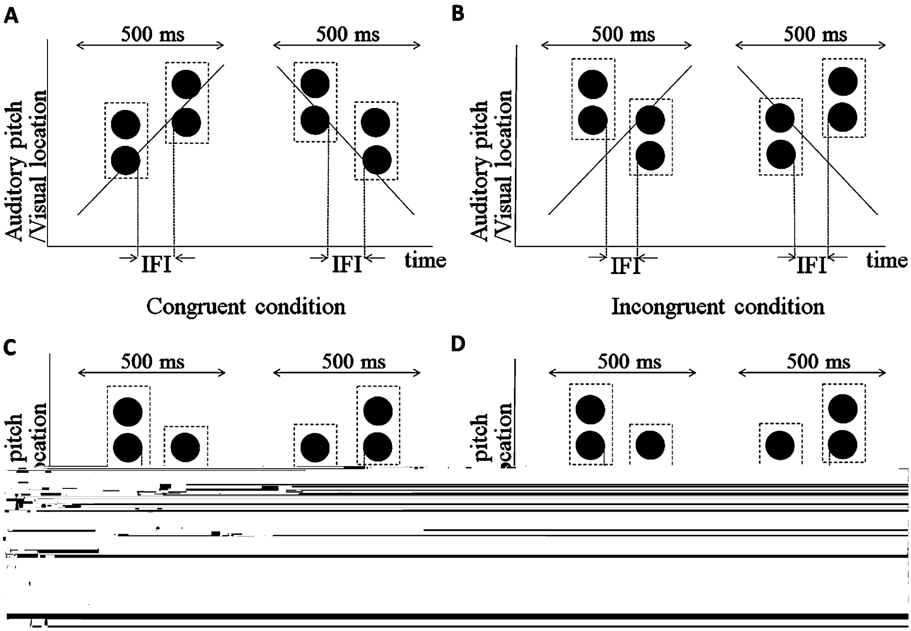

Figure 2. Illustrations of the audiovisual configurations. In ‘congruent’, ‘incongruent’, and ‘fixed-pitch’ conditions, the stimuli consisted of both auditory and visual events. In each sub-figure (A–D), the straight solid line indicates the glide with the ascending, descending, or invariant pitch, and the two dotted frames represent the upward or downward Ternus display. ‘Visual-only’ indicates sound-absent stimuli. The duration of each stimulus configuration was 500 ms. In the ‘congruent’ condition, upward/downward Ternus apparent motions were always accompanied with ascending-pitched/descending-pitched glides. In the ‘incongruent’ condition, upward (downward) visual displays and glides with descending (ascending) pitch were presented simultaneously. In the ‘fixed-pitch’ condition, the auditory stimuli consisted of a pure tone with a fixed frequency of 500 Hz. For the ‘visual-only’ condition, the visual Ternus display was shown without auditory stimuli.

Prior to the experiment, participants were shown demonstrations of the ‘EM’ and ‘GM’ variants of the typical Ternus display. They then practiced for 80 trials. All participants reported having a good understanding of the task and achieved a correct response rate of about 95% for perceiving ‘EM’ (IFI = 50 ms) and ‘GM’ (IFI = 230 ms). During the first test, participants were instructed to gaze upon the fixation cross and make a two-alternative forced choice (2-AFC) to show the perceptual state of visual Ternus apparent motion (EM *vs* GM). Throughout the experiment, they were required to focus on the visual task and ignore the auditory stimuli (for the sound-present conditions). As shown in Fig. 3A, a typical trial began with a central filled black circle on the monitor lasting for 600 ms (0.24 cd/m^2 in luminance). Then a fixation cross appeared at the center of the upper or lower field, randomly, for 600 ms.

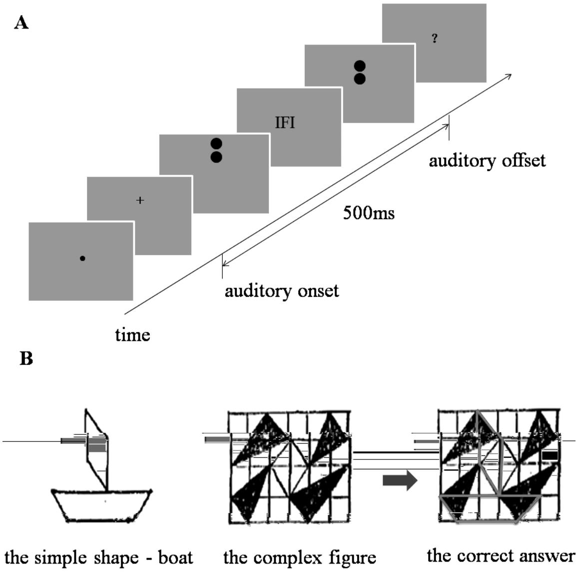

Figure 3. Schematic illustration of the stimuli presented on one trial in the first test and a sample test of the Embedded Figure Test (EFT) in the experiment. (A) In this sound-present trial, the glide started before the first frame and ended after the second frame. The two visual frames of the downward motion were presented vertically in the upper visual field. The IFI between the two frames was selected from five durations: 80, 110, 140, 170, or 200 ms. After presentation of the audiovisual stimuli, a question mark appeared to prompt participants to choose between two options by pressing either the left or right key. (B) In this EFT sample test, the simple shape was a boat-like figure embedded in a complex figure. The task for participants was to try to trace the outline (labelled as red lines) of this 'boat'.

After a blank interval of 500 ms, participants were required to fixate upon the cross. Upon offset of the fixation cross, there was another blank of 500 ms and the (audio)visual stimuli appeared. In the sound-present trials, participants heard the glide first and then saw the first visual frame lasting for 30 ms. After a given IFI (five levels from 80 to 200 ms), the second visual frame appeared. The temporal midpoint of the 500-ms glide was coincident with that of the IFI. For the visual-only trials, participants waited for the same time interval as in the sound-present trials before seeing the first visual frame. After all stimuli were presented, with a random delay of 300–450 ms, participants were presented with a question mark. They then pressed the left arrow for 'EM' and the right arrow for 'GM' (Fig. 3A).

After the experiment, participants were required to take the EFT (Oltman *et al.*, 1971; Witkin and Goodenough, 1977). We used a Chinese version of the EFT (Xie and Zhang, 1988), which comprises three sections: section A (9 figures) for practice, section B (10 figures) and section C (10 figures) for formal tests. The task was to draw the outline of the simple shape in each complex figure as quickly and accurately as possible. Each section could last up to four minutes. A sample test from the EFT is given in Fig. 3B. Participants used pencil to depict the target figure and could use an eraser to correct it if necessary. When the time (12 min in total) was up, they had to stop immediately and hand in the sheets of the EFT. Afterwards, the experimenter evaluated participants' performance by awarding one point for each correct answer. The highest score possible was 20.

3. Results

The proportion of group motion reports was plotted as a function of IFI and fitted by a logistic regression for each participant. The points of subjective equality (PSE) and just noticeable difference (JND) were calculated across each audiovisual condition. PSE refers to the transitional temporal point at which percepts of 'EM' and 'GM' were perceived with equal probabilities. This can be calculated by estimating the point of 50% of the percentages for reporting 'GM' on a fitted logistic function. JND represents the point at which the difference between the two motion perceptions becomes apparent, which is obtained by estimating the IFI difference of one half between 25% and 75% of the 'GM' responses from the psychometric curves (Treutwein and Strasburger, 1999). Figure 4 shows the average psychometric curves in the first session of

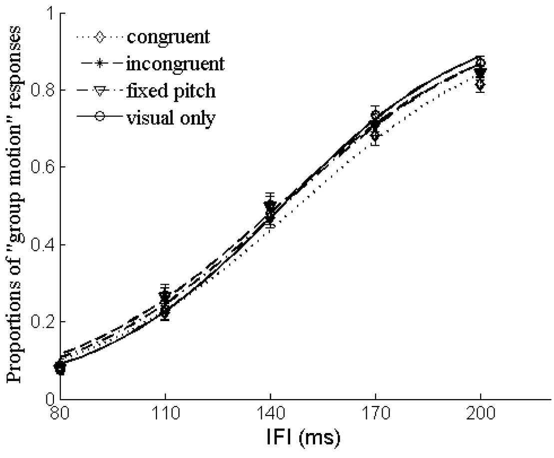

Figure 4. The average psychometric curves for all participants under the four audiovisual stimulus conditions in the first session of the experiment: the curves represent the proportions of group motion responses as a function of the IFI between the two visual frames. The solid line with circles shows the proportions of group motion for the 'visual only' condition; the dotted line with diamonds illustrates the 'congruent' condition; the dashed line with stars indicates the 'incongruent' condition; the dot-dashed line with triangles indicates the 'fixed-pitch' condition. The error bars represent the standard errors of the means.

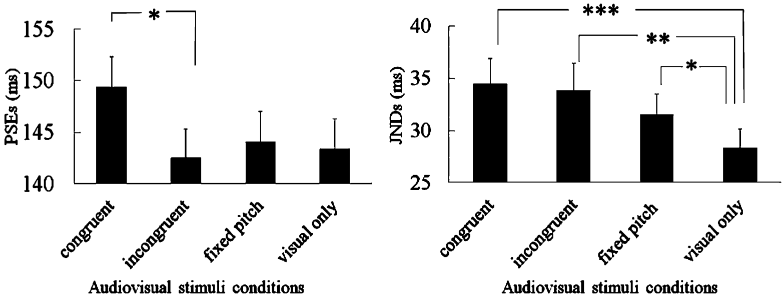

Figure 5. Points of subjective equality (PSEs) and just noticeable differences (JNDs) for Ternus motion classification in the first session of the experiment. The error bars represent standard errors (* 0 05; ** 0 01; *** 0 001).

the experiment for all participants. Figure 5 shows the mean PSEs and JNDs with associated standard errors.

A repeated measures ANOVA was conducted on both the obtained PSEs and JNDs, with audiovisual stimulus condition (‘congruent’, ‘incongruent’, ‘fixed pitch’, and ‘visual-only’) as a within-subjects factor. For the PSEs, the main effect of stimulus conditions was significant, (3 117) = 3 765, 0 05, ^2_p = 0 088. Bonferroni correction was used for the full set of six possible cohorts. Pairwise comparison revealed the PSE in the congruent condition (149.4 ms) was statistically larger than that in the incongruent condition (142.6 ms), s 0 05. However, no differences were found among the remaining conditions: ‘incongruent’ (142.6 ms), ‘fixed pitch’ (144.1 ms) and ‘visual only’ (143.4 ms), all s 0 1. Therefore, the results indicate a ‘congruency’ between pitch change and visual moving direction, in which the upward/downward direction of Ternus motion with ascending/descending glide strengthened the subjective connectedness of the two Ternus frames and hence gave rise to more reports of EM (i.e., with larger PSEs) (Kramer and Yantis, 1997).

The ANOVA performed on the JNDs revealed a significant main effect of audiovisual stimulus conditions, (3 117) = 11 048, 0 001, ^2_p = 0 221. No significant differences were found between sound-present conditions: congruent condition (34.6 ms), incongruent condition (33.8 ms) and fixed-pitch condition (31.5 ms), all 0 05, while they were all larger than the JND in the visual-only condition (28.3 ms), with s 0 001, s 0 01, and s 0 05 respectively. The results indicate that task-irrelevant glides indeed reduced the sensitivity for discriminating Ternus apparent motion, independent of whether the correspondences between auditory glides and visual stimuli were congruent or incongruent, as shown in Fig. 5.

We also performed a two-way repeated-measures ANOVA on the PSEs and JNDs, with audiovisual stimulus conditions (‘congruent’ pattern, ‘incongruent' pattern, 'fixed pitch', and 'visual-only') and visual motion direction (upward and downward) as within-subjects factors. For PSEs, the analyses revealed significant main effects of stimulus conditions [(3 117) = 3 567, 0 05, $^{2}_{p}$ = 0 084] and visual motion directions [(1 39) = 16 323, 0 001, $^{2}_{p}$ = 0 295], but no significant interaction effect was found, (3 117) = 0 797, = 0 498. The mean PSE for an upward direction was smaller (140.3 ± 2.6 ms) than the one for a downward direction (149.2 ± 2.9 ms). The downward direction gave rise to the perception of 'falling' and the metaphor of gravity has led to illusory fast movement of the Ternus frames (i.e., short IFI). This gave rise to a dominant percept of 'EM' (with larger PSE).

Two-way ANOVAs on JNDs showed the main effect of audiovisual stimulus conditions was significant, (3 117) = 9 378, 0 001, $^{2}_{p}$ = 0 194. However, the main effect of the visual motion direction was not significant, (1 39) = 1 747, = 0 194. There was a significant interaction between the stimulus conditions and visual motion directions, (3 117) = 3 172, 0 05, $^{2}_{p}$ = 0 075. Further, simple main effects suggested that the visual motion direction affected sensitivity for discriminating the Ternus motion in solely the visual-only condition, with low sensitivity for upward motion (30.1 ± 2.3 ms) and relatively higher sensitivity for downward motion (26.1 ± 1.5 ms), (1 39) = 8 35, 0 01, but there were no differences between upward and downward trials within 'congruent', 'incongruent', or 'fixed pitch' conditions, all s 0 1. Therefore, people seemed more adaptive to the downward motion, due to that the daily experience of 'gravity'. However, the concurrent sound signals might have interfered with this visual experience, as shown in Fig. 6.

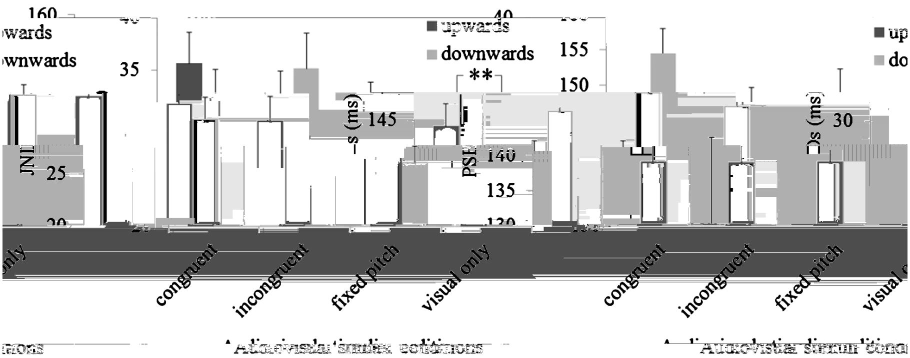

Figure 6. Points of subjective equality (PSEs) and just noticeable differences (JNDs) for perception of Ternus motion, with audiovisual conditions (congruent, incongruent, fixed-pitch and visual-only) and motion direction (upward and downward) as within-participant factors. The error bars represent standard errors of mean (** 0 01).

Taking visual field into consideration, a further look into the JND pattern for the LVF found that two participants had larger JNDs, out of the range of three standard deviations, even though the collapsed mean for lower and upper field was within the range. We excluded these two participants and carried out two-way ANOVAs on PSEs and JNDs. We used stimulus conditions (congruent, incongruent, fixed, and visual-only) and visual field (upper *vs* lower) as within-participant independent factors. For PSEs, the analyses still revealed significant main effects of stimulus conditions [(3 111) = 3 753, 0 05, $^2_{\mathrm{p}}$ = 0 092]. The main effect of visual field was not significant, with the upper-field PSE at 145.7 ms and the lower-field one at 145.9 ms, (1 37) = 0 014, 0 05. However, the interaction between stimulus conditions and visual fields was significant, (3 111) = 3 170, 0 05, $^2_{\mathrm{p}}$ = 0 079. Simple-effects analysis indicated a trend in the ‘congruent’ condition, in which the PSE for the lower field (154.3 ms) was marginally larger (i.e., more reports of element motion) than the one in the upper field (146.9 ms), (1 37) = 3 95, = 0 054. This difference indicates a potent influence of ‘gravity’, in which speed of visual apparent motion seems to be faster, hence the dominant percept of ‘element motion’ in the LVF.

Two-way ANOVAs on JNDs showed that the main effect of audiovisual stimulus conditions was significant, (3 111) = 9 704, 0 001, $^2_{\mathrm{p}}$ = 0 208. Again, the JND in the congruent condition (32.8 ms) was larger than the JND in the incongruent (31.5 ms), fixed (30.3 ms) and visual-only (26.7 ms) conditions, s 0 05. However, the main effect of visual field was not significant, with the upper field JND at 29.6 ms and the lower field JND at 31.1 ms, (1 37) = 2 368, = 0 132. There was borderline interaction between the stimulus conditions and visual field directions, (3 111) = 2 244, = 0 087. Further simple-effects analysis also indicated that only for the ‘congruent’ condition, the JND for the LVF was larger (35.4 ms) than the one in the UVF (30.4 ms), (1 37) = 4 33, 0 05 (Fig. 7).

In the second session of the experiment, we used the EFT to examine participants’ cognitive styles. Participants were divided into either the FI group (= 19) or FD group (= 19) based on the ‘mean-split’ of their scores. Here, the FI group had a mean score of 14.53 (±2.29) out of 20, and the FD group scored 8.68 (±2.14) on average. We ran four-way ANOVAs on PSEs and JNDs, with audiovisual conditions, movement direction, and visual field as within-participant factors, and group (field-dependent, field-independent) as between-participants factor.

Here we focus on the effects with only the factor of ‘group’ included. For PSEs, the main effect of the group factor was not significant, (1 36) = 0 695, = 0 410. The interaction between audiovisual conditions and group was significant, (3 108) = 2 884, 0 05, $^2_{\mathrm{p}}$ = 0 074. Further simple-effects analysis showed that for the field-dependent group, the main effect

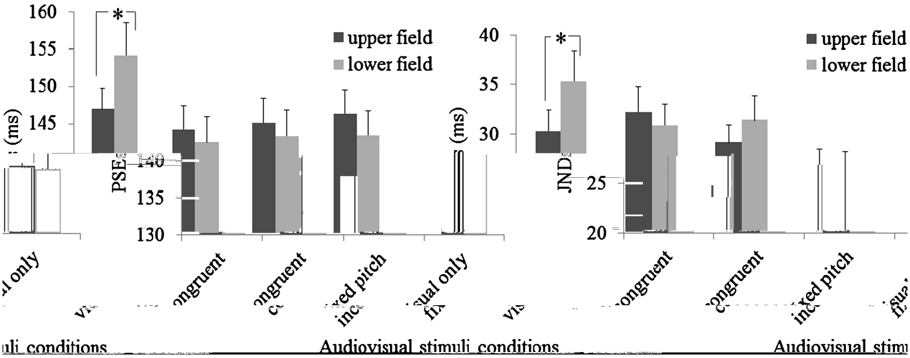

Figure 7. Points of subjective equality (PSEs) and just noticeable differences (JNDs) for Ternus motion classification in the first session of the experiment, with the stimulus conditions and visual fields as within-participant factors. The error bars represent standard errors of mean (* 0 05).

of audiovisual conditions was not significant, (3 108) = 0 48, = 0 698. For the field-independent group, however, the main effect of audiovisual conditions was significant, (3 108) = 6 11, 0 01. The mean PSE for the congruent condition (151.7 ms) was larger than the mean PSEs for the incongruent condition (140.0 ms) and the visual-only condition (140.2 ms), s 0 05. No remaining potential interactions were found when the factor of ‘group’ was included. The result pattern indicated that the FI group, rather than the FD group, showed the bias effect of crossmodal integration by audiovisual correspondence.

For JNDs, the main effect of ‘group’ was significant, (1 36) = 5 080, 0 05, ^2_p = 0 128. The JND for the independent group (34.0 ms) was larger than the one for the dependent group (25.9 ms). The interaction between audiovisual conditions and groups was significant, (3 108) = 3 421, 0 05, ^2_p = 0 087. We probed this interaction with a simple-effects analysis, finding that for the congruent condition, the JND in the independent group (37.9 ms) was larger than the one in the dependent group (27.7 ms), (1 36) = 5 21, 0 05. For the incongruent condition, the JND in the independent group (35.9 ms) was larger than the one in the dependent group (26.3 ms), (1 36) = 5 33, 0 05. Also, in the fixed pitch condition, the JND in the independent group (34.4 ms) was larger than the one in the dependent group (25.0 ms), (1 36) = 7 77, 0 01. However, for the visual-only condition, the JNDs in both groups were statistically equal, (1 36) = 1 36, = 0 252. Of note, simple-effects tests on the independent group showed a significant main effect of audiovisual condition, (3 108) = 12 47, 0 001. The JNDs in all the sound-present conditions were larger than the one in visual-only condition (27.7 ms), s 0 05. As shown in Fig. 8, however,

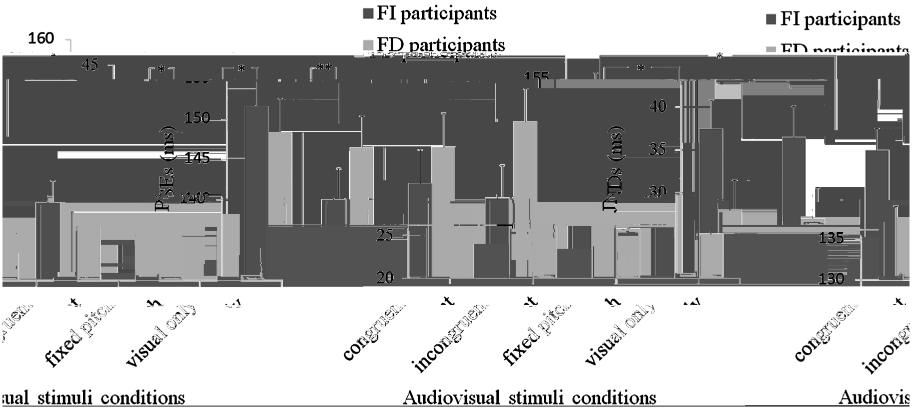

Figure 8. Points of subjective equality (PSEs) and just noticeable differences (JNDs) for discriminating ‘EM’ and ‘GM’ in the first session of the experiment from the two groups of participants. The error bars represent standard errors of the mean (* 0 05; ** 0 01; *** 0 001).

Table 1.
Summary of the critical findings. The left column describes the critical experimental conditions. The checks in the right two columns indicate which type of apparent motion (group motion

motion'. However, this effect was typically observed in the FI group. Moreover, more reports of 'group motion' were found in the upward direction and UVF conditions than in the downward and LVF conditions.

4. Discussion

The interaction between audiovisual timing and perceptual grouping between crossmodal events has recently been studied with the paradigm of the Ternus display (Chen *et al.*, 2010; Shi *et al.*, 2010; Wang *et al.*, 2014, 2015). A typical (horizontal) Ternus display is composed of two sequential visual frames, each containing two horizontally arranged dots. One dot is 'central' and one is 'lateral'. The central dot is repeatedly shown in the same position in successive frames whereas the lateral dot changes positions (e.g., from the left to right by crossing the stationary central dot). Such displays elicit two distinct motion percepts: element motion (only the lateral dot is perceived to move from one side to the other) or group motion (both dots are perceived to move as a whole), with the form of apparent motion mainly determined by the inter-frame interval (IFI) between the two frames (Alais and Lorenceau, 2002; Harrar and Harris, 2007; He and Ooi, 1999; Petersik and Rice, 2008; Shi *et al.*, 2010; Ternus, 1926). When the IFI is shorter, dominant element motion is perceived, but when the IFI is longer, dominant group motion is perceived. Thus, the Ternus display provides a promising tool for manipulating temporal perceptual grouping to generate a clear-cut state of apparent visual motion.

In this study, we examined how individual differences, typified by cognitive styles, affect audiovisual interaction by using the vertical type of Ternus apparent motion. We compared the combinations of pitch of auditory glides (ascending and descending) and visual motion direction (upward and downward) and explored how the 'congruent' (such as glide with ascending pitch paired with upward Ternus motion) and 'incongruent' audiovisual correspondence (i.e., glide with ascending pitch paired with downward Ternus motion) influenced the perceptual classification of Ternus apparent motion (element motion *vs* group motion).

4.1. Audiovisual Correspondence Precedes Crossmodal Integration

In general, we found that congruent audiovisual correspondence facilitated the subjective connectedness of the two Ternus frames and gave rise to the dominant percept of 'element motion' (with increased PSEs). In the task-irrelevant sound-present conditions, the auditory glides reduced sensitivity for judging Ternus motion, making the JNDs larger than those in the visual-only condition. For the field-independent group, the congruent audiovisual stimulus led to the dominant percept of element motion (compared with 'incongruent' and 'visual-only' conditions), with reduced sensitivities for judging Ternus motion in sound-present conditions. However, those featured differences were absent in the field-dependent group. This counter-intuitive finding suggests that crossmodal correspondence and crossmodal integration might be two separate processes. The separation of attentional abilities for FD and FI individuals might provide a means to examine this hypothesis.

A person's attention influences various stages of the multisensory integration of crossmodal events, including the identification of unisensory properties (Gallace and Spence, 2006; Harrar *et al.*, 2014; Parise and Spence, 2009; Sweeny *et al.*, 2012; Talsma *et al.*, 2010). The FI group had the competitive ability (relevant to attention) to separate and identify the components signals, and they further exploited the crossmodal correspondence of audiovisual events (pitches of glides and directions of visual motion) to facilitate the discrimination of Ternus motion. On the contrary, for the FD group, the corresponding dimensions for individual sensory events were less 'salient' to them, and the 'correspondence' would be deterred by their low efficiency to separate the individual events/properties from the complex audiovisual scene. For the FD group, this low efficiency thus made the effects of audiovisual interaction less distinguishable in the given four conditions ('congruent', 'incongruent', 'fixed' and 'visual-only'). Our experiment further indicates that the crossmodal correspondence process is determined (in part) by the context of the experiment and the detailed sensory stimuli we adopted (Chiou and Rich, 2012a, b; Krugliak and Noppeney, 2016). Nevertheless, future studies should adopt simple configurations of audiovisual stimuli to corroborate the dissociation of crossmodal correspondence and crossmodal integration.

The correspondence between auditory pitch and direction of visual motion, in our current experimental setting, is different from previous studies, which used static and short stimuli. For simple and short stimuli, the information processing is largely automatic. For short stimuli, evidence has shown that attention has no obvious effect on the measured temporal window of integration (Donohue *et al.*, 2015), hence attention has little engagement in the integration of short stimuli (crossmodal) events in our case. However, in our study, the relative complex stimuli and dynamic scenario required relative longer time to establish (note that the duration of audiovisual stimuli is 500 ms) the crossmodal correspondence, especially by tracking the changes of pitches. Therefore, it might largely prevent the automatic binding of different sensory properties but requests attentional engagement (Chiou and Rich, 2012a, b; Hackley, 1993; Klapetek *et al.*, 2012). Indeed, Bien *et al.* (2012) concluded that the first neural signs associated with a distinction between congruent and incongruent stimulus pairings started around 250 ms after stimulus onset in the right intraparietal sulcus (identified with the parietal P2), so that initial processing of crossmodal correspondences with attentional inputs might precede, and then determine, the subsequent outcome of crossmodal interaction.

The JND results also suggest the potential role of attentional engagement. Previous studies have shown that congruent ‘synaesthetic’ correspondence between static auditory pitch and static visual size enhances the sensitivity of time perception as shown in the temporal order judgment task (Carnevale and Harris, 2016; Chen *et al.*, 2016b; Parise and Spence, 2008). According to our hypothesis, when the motion direction of the visual Ternus displays was congruent with the changing pitch, for example, when the upward (downward) apparent motion as well as a glide with ascending (descending) pitch were presented simultaneously, we would expect to observe higher sensitivities to judge the Ternus apparent motion. Against our expectation, we observed decreased sensitivities (larger JNDs) for discriminating Ternus motion in all sound-present conditions. It indicated that overall, the sound distracted from rather than boosted the performance of classifying the visual Ternus motion. We attributed this to be the result of a potential ‘attention’ mechanism. In our case, the two successive visual frames were temporally separated by stimulus onset asynchronies from 110 to 230 ms (also with the IFI range changing from 80 to 200 ms), and the concurrent auditory stimutu were ctng co.2(lotion-]TJ T* tgong icotionuus sung glldewartentuted The eiscriminatinnse-225.7(sor)-238.9(3v.2(irua 'element motion' in the lower versus upper visual hemifields. Since the observed JNDs were larger in the LVF, the attentional resolution did not show a lower field preference (He *et al.*, 1996; Levine and McAnany, 2005; Rubin *et al.*, 1996).

However, in terms of motion direction, LVF was more sensitive to the downward motion, and the downward Ternus apparent motion turned out to have larger PSEs, with a dominant percept of element motion. For the visual-only condition, participants had higher sensitivities for the 'downward' motion than for the 'upward' motion (but the distinction was generally blurred in the presence of auditory glides). The evidence from magnetic response has shown that the downward apparent motion elicited great amplitudes than the upward motion did. The neurons in the extrastriate cortex of humans are sensitive to the downward motion. Evidence has shown that humans perform various visual tasks better when the stimulus is presented in the lower than in the upper visual field, with some neurophysiological evidence that in V5 the representation of the LVF is larger than that of the UVF (Maunsell and Van Essen, 1987). Overall, the findings above suggest that performance capacities in the upper and lower visual hemifields are task-specific. Generally, LVF has been shown to be more sensitive to downward motion (Amenedo *et al.*, 2007; He *et al.*, 1996; Levine and McAnany, 2005; Raymond, 1994; Rubin *et al.*, 1996). However there has been evidence supporting the opposite effect (Danckert and Goodale, 2003; Previc, 1990; Previc and Blume, 1993).

4.3. The Role of Cognitive Styles and Perceptual Grouping

Research into cognitive styles has mainly focused on individual differences in carrying out perceptual (decision making) and cognitive tasks. Previous study has shown that in a serial choice response task, field-dependent individuals were slower and less accurate than field-independent and neutral individuals. Moreover, they were more sensitive to changes in stimulus modality than the other two groups (Yan, 2010). The field-independent individuals could focus on the task goal with controlled attention and response selection that was not distracted by the stimulus setup and the changes in the stimulus modality (Amador and Kirchner, 1999; Liu, 2003; Riding and Al-Salih, 2000; Sternberg and Grigorenko, 1997). The choice responses of the FDs could be subject to the influence of the changing stimuli and targets. Indeed, the FI group was more internal referencing while the FD group was more external referencing.

Analogously, in the current experiment, FI individuals tended to rely more on internal criteria so that they were keener on the fine differentiations of the audiovisual conditions. That is to say, FI participants 'corresponded' well from the single glide frame (without a gap) to the visual frames (with a gap in between). Despite the complex audiovisual stream, FI participants could form differential representations. Therefore, we observed that in the 'congruent' condition, the potent audiovisual correspondence led more to the dominant percept of 'element motion' (with increased PSE). Meanwhile, attentional demands (associated with the separating of audiovisual events/properties) somehow decreased the 'sensitivity' of discrimination and generally led to increased JNDs for the FI individuals.

Crossmodal attention has been shown to decrease cortical responses to distracting or competing stimuli. Attentional processing for separation of audiovisual events is more demanding than the fusing of different sensory properties. For separation, identifying properties of events from a single modality increases brain activity of the corresponding sensory cortices and suppresses activity in non-corresponding sensory cortices. This interaction was achieved by selective attention (Johnson and Zatorre, 2005). However, in present experiment, for the integration (fusing) of bimodal stimuli, the suppression of activity in given sensory cortices was less. Moreover, the properties associated with auditory and visual stimuli prompted the observers to form a representation of a coherent event. For the field-dependent group, since they were more external-referencing, in general they had some difficulty in separating/corresponding the element auditory glides and visual events, thus rendering the physically distinctive audiovisual pattern less perceptually distinguishable. Therefore, we observed no statistical difference in (PSEs for) discriminating Ternus motion under the given audiovisual conditions. In addition, since the FD group could not invest much attentional input in the task, they seemed ready to make quick perceptual decisions for Ternus apparent motion (with reduced JNDs).

In conclusion, we propose that field-dependent and field-independent individuals have different 'abilities' for identifying the element properties of stimuli from complex crossmodal scenarios. These individual differences in cognitive style have been revealed to differentiate the potentially initial step of crossmodal correspondences and could affect subsequent crossmodal integration. We are currently not able to tease apart the potential cognitive bias from the pure perceptual processing for the FD and FI groups in dealing with crossmodal interaction, nor are we able to pinpoint the time course of potential different processes for crossmodal correspondence and crossmodal integration in the present study. Yet to our best knowledge, this is the first time that the role of cognitive style in shaping crossmodal interaction has been addressed. The present research has important implications for future studies in the field of multisensory integration. For instance, our findings may inform further investigation into how internal (FI) and external (FD) referencing constrains or otherwise influences the effectiveness of perceptual training, and may be applied to developmental perspectives as well.

Acknowledgements

This study was supported in part by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA06020200) to Bao, National Natural Science Foundation of China (NSFC 11304345, NSFC 61621136008, NSFC 61527804).

References

Alais, D. and Lorenceau, J. (2002). Perceptual grouping in the Ternus display: evidence for an 'association field' in apparent motion, *Vis. Res.* **42**, 1005–1016.

Amador, C. J. A. and Kirchner, N. T. (1999). Correlations among scores on measures of field dependence and independence cognitive style, cognitive ability, and sustained attention, *Percept. Mot. Skills* **88**, 236–239.

Amenedo, E., Pazo-Alvarez, P. and Cadaveira, F. (2007). Vertical asymmetries in pre-attentive detection of changes in motion direction, *Int. J. Psychophys.* **64**, 184–189.

Axelrod, S. and Cohen, L. D. (1961). Senescence and embedded-figure performance in vision and touch, *Percept. Mot. Skills* **12**, 283–288.

Bien, N., Ten Oever, S., Goebel, R. and Sack, A. T. (2012). The sound of size: crossmodal binding in pitch-size synesthesia: a combined TMS, EEG, and psychophysics study, *NeuroImage* **59**, 663–672.

Brainard, D. H. (1997). The psychophysics toolbox, *Spat. Vis.* **10**, 433–436.

Bremner, A., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. and Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners, *Cognition* **126**, 165–172.

Carnevale, M. J. and Harris, L. R. (2016). Which direction is up for a high pitch? *Multisens. Res.* **29**, 113–132.

Cecere, R., Rees, G. and Romei, V. (2015). Individual differences in alpha frequency drive crossmodal illusory perception, *Curr. Biol.* **25**, 231–235.

Chen, L. and Vroomen, J. (2013). Intersensory binding across space and time: a tutorial review, *Atten. Percept. Psychophys.* **75**, 790–811.

Chen, L., Shi, Z. and Muller, H. J. (2010). Influences of intra- and crossmodal grouping on visual and tactile Ternus apparent motion, *Brain Res.* **1354**, 152–162.

Chen, Y. C., Huang, P. C., Woods, A. and Spence, C. (2016a). When "Bouba" equals "Kiki": cultural commonalities and cultural differences in sound–shape correspondences, *Sci. Rep.* **6**, 26681. DOI:10.1038/srep26681.

Chen, L., Zhang, M., Ai, F., Xie, W. and Meng, X. (2016b). Crossmodal synesthetic congruency improves visual timing in dyslexic children, *Res. Dev. Disabil.* **55**, 14–26.

Chiou, R. and Rich, A. N. (2012a). Cross-modality correspondence between pitch and spatial location modulates attentional orienting, *Perception* **41**, 339–353.

Chiou, R. and Rich, A. N. (2012b). Perceptual difficulty and speed pressure reveal different behavioural effects of voluntary and involuntary attention, in: *39th Australasian Experimental Psychology Conference*, Sydney, Australia.

Clark, H. H. and Brownell, H. H. (1976). Position, direction, and their perceptual integrality, *Percept. Psychophys.* **19**, 328–334.

Maeda, F., Kanai, R. and Shimojo, S. (2004). Changing pitch induced visual motion illusion, *Curr. Biol.* **14**, R990–R991.

Makovac, E. and Gerbino, W. (2010). Sound–shape congruency affects the multisensory response enhancement, *Vis. Cogn.* **18**, 133–137.

Marks, L. E. (1987). On cross-modal similarity: auditory-visual interactions in speeded discrimination, *J. Exp. Psychol. Hum. Percept. Perform.* **13**, 384–394.

Maunsell, J. H. and Van Essen, D. C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries, *J. Comp. Neurol.* **266**, 535–555.

Molholm, S., Walter, R., Daniel, C. J. and John, J. F. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study, *Cereb. Cortex* **14**, 452–465.

Mossbridge, J. A., Grabowecky, M. and Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention, *Cognition* **121**, 133–139.

Naito, T., Kaneoke, Y., Osaka, N. and Kakigi, R. (2000). Asymmetry of the human visual field in magnetic response to apparent motion, *Brain Res.* **865**, 221–226.

Oltman, P. K., Raskin, E. and Witkin, H. A. (1971). *Group Embedded Figures Test*. Consulting Psychologists Press, Palo Alto, CA, USA.

Parise, C. V. and Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect, *Neurosci. Lett.* **442**, 257–261.

Parise, C. V. and Spence, C. (2009). 'When birds of a feather flock together': synesthetic correspondences modulate audiovisual integration in non-synesthetes, *PLoS One* **4**, e5664. DOI:10.1371/journal.pone.0005664.

Parise, C. V. and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: an IAT study, *Exp. Brain Res.* **220**, 319–333.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies, *Spat. Vis.* **10**, 437–442.

Petersik, J. T. and Rice, C. M. (2008). Spatial correspondence and relation correspondence: grouping factors that influence perception of the Ternus display, *Perception* **37**, 725–739.

Previc, F. H. (1990). Functional specialization in the lower and upper visual fields in humans: its ecological origins and neurophysiological implications, *Behav. Brain Sci.* **13**, 519–575.

Previc, F. H. and Blume, J. L. (1993). Visual search asymmetries in three-dimensional space, *Vis. Res.* **33**, 2697–2704.

Previc, F. H., Breitmeyer, B. G. and Weinstein, L. F. (1995). Discriminability of random-dot stereograms in three-dimensional space, *Int. J. Neurosci.* **80**, 247–253.

Rader, C. M. and Tellegen, A. (1987). An investigation of synesthesia, *J. Pers. Soc. Psychol.* **52**, 981–987.

Raymond, J. E. (1994). Directional anisotropy of motion sensitivity across the visual field, *Vis. Res.* **34**, 1029–1037.

Rayner, S. and Riding, R. (2002). *Cognitive Styles and Learning Strategies*. David Fulton, London, UK.

Riding, R. J. and Al-Salih, N. (2000). Cognitive style and motor skills and sports performance, *Educ. Stud.* **26**, 19–32.

Rubin, N., Nakayama, K. and Shapley, R. (1996). Enhanced perception of illusory contours in the lower versus upper visual hemifields, *Science* **271**, 651–653.

Shi, Z., Chen, L. and Muller, H. J. (2010). Auditory temporal modulation of the visual Ternus effect: the influence of time interval, *Exp. Brain Res.* **203**, 723–735.

Spence, C. (2011). Crossmodal correspondences: a tutorial review, *Atten. Percept. Psychophys.* **73**, 971–995.

Spence, C. and Driver, J. (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxforsd, UK.

Sternberg, R. J. and Grigorenko, E. L. (1997). Are cognitive styles still in style, *Am. Psychol.* **52**, 700–712.

Sweeny, T. D., Guzman-Martinez, E., Ortega, L., Grabowecky, M. and Suzuki, S. (2012). Sounds exaggerate visual shape, *Cognition* **124**, 194–200.

Talsma, D., Senkowski, D., Soto-Faraco, S. and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration, *Trends Cogn. Sci.* **14**, 400–410.

Ternus, J. (1926). Experimentelle Untersuchungen über phänomenale Identität, *Psychol.e Forsch.* **7**, 81–136. (English translation as “The problem of phenomenal identity”, in: *A Source Book of Gestalt Psychology*, W. D. Ellis (Ed.), pp. 149–160, 1938, Routledge and Kegan Paul, London).

Treutwein, B. and Strasburger, H. (1999). Fitting the psychometric function, *Percept. Psychophys.* **61**, 87–106.

Van Atteveldt, N., Formisano, E., Goebel, R. and Blomert, L. (2004). Integration of letters and speech sounds in the human brain, *Neuron* **43**, 271–282.

Wang, Q., Bao, M. and Chen, L. (2014). The role of spatiotemporal and spectral cues in segregating short sound events: evidence from auditory Ternus display, *Exp. Brain Res.* **232**, 273–282.

Wang, Q., Guo, L., Bao, M. and Chen, L. (2015). Perception of visual apparent motion is modulated by a gap within concurrent auditory glides, even when it is illusory, *Front. Psychol.* **6**, 564. DOI:10.3389/fpsyg.2015.00564.

White, B. W. (1954). Visual and auditory closure, *J. Exp. Psychol.* **48**, 234–240.

Witkin, H. A. and Goodenough, D. R. (1977). Field dependence and interpersonal behavior, *Psychol. Bull.* **84**, 661–689.

Witkin, H. A., Birnbaum, J., Lomonaco, S., Lehr, S. and Herman, J. L. (1968). Cognitive patterning in congenitally totally blind children, *Child Dev.* **39**, 767–786.

Witkin, H. A., Moore, C. A., Goodenough, D. R. and Cox, P. W. (1975). Field-dependent and field-independent cognitive styles and their educational implications, *ETS Res. Bull. Series* **1975**(2), 1–64.

Xie, J. and Zhang, H. (1988). *Cognitive Style: Experimental Study on a Personality Dimension*. Beijing Normal University Press, Beijing, P.R. China.

Yan, J. H. (2010). Cognitive styles affect choice response time and accuracy, *Pers. Indiv. Diff.* **48**, 747–751.